Chapter 19: Developing an integrated model of marine bird distributions with environmental covariates using boat and digital video aerial survey data *This chapter is in draft form

Final Report to the Department of Energy Wind and Water Power Technologies Office, 2015

Nathan J. Hostetter[1], Beth Gardner[1], and Andrew T. Gilbert[2]

[1]North Carolina State University, Department of Forestry and Environmental Resources, Raleigh, NC
[2]Biodiversity Research Institute, Portland, ME

Project webpage: *www.briloon.org/mabs*

Suggested citation: Hostetter NJ, Gardner B, Gilbert AT. 2015. Developing an integrated model of marine bird distributions with environmental covariates using boat and digital video aerial survey data. In: Wildlife Densities and Habitat Use Across Temporal and Spatial Scales on the Mid-Atlantic Outer Continental Shelf: Final Report to the Department of Energy EERE Wind & Water Power Technologies Office. Williams KA, Connelly EE, Johnson SM, Stenhouse IJ (eds.) Award Number: DE-EE0005362. Report BRI 2015-11, Biodiversity Research Institute, Portland, Maine. 23 pp.

Acknowledgments: This material is based upon work supported by the Department of Energy under Award Number DE-EE0005362. Additional funding support came from the Maryland Energy Administration and Maryland Department of Natural Resources. HiDef Aerial Surveying, Inc., Richard Veit (College of Staten Island), Holly Goyert, Melissa Duron, Emily Connelly, Kathryn Williams, and Capt. Brian Patteson made significant contributions towards the completion of this study.

NC STATE UNIVERSITY bri

Chapter 19 Highlights

Developing an integrated model of marine bird distributions with environmental covariates using boat and digital video aerial survey data

Context[1]

A broad geographic and temporal scale of analysis is required to assess exposure to wildlife from proposed development projects. In this study, data were collected via traditional methods (boat-based distance sampling) and with newer technologies (high definition videography) with the intention of addressing similar questions related to marine wildlife abundance and distribution. Chapter 18 explored the two datasets to determine if similar patterns were detected by each sampling method. Based on those results, Chapter 19 aims to develop a method of integrating these two datasets into a combined approach that makes use of the strengths of each survey type, to produce a single prediction of marine bird abundance and distribution.

In this approach, predictions of marine bird abundance and distribution are jointly informed by aerial surveys, which encompassed a large geographic area, and boat surveys, which allowed for estimation of detection probability. We analyze data from the same four species groups as Chapter 18 (terns, alcids, loons, and Northern Gannets), and incorporate remotely collected environmental covariate data into the hierarchical modeling structure. This approach accounts for imperfect detection to estimate "true" abundance, and predicts marine bird distributions to help identify important habitat use areas and patterns.

Study goal/objectives

Evaluate potential exposure of the marine bird community to offshore development by: (1) developing a model to integrate data from the two survey platforms; and (2) producing a single prediction of abundance and distribution to identify ecological drivers of distribution, abundance, and local hotspots.

Highlights

- Distance to shore was generally the most common predictor of abundance, as was found in Chapters 12 and 18.
- Integrated models predicted species-specific hotspots that generally concurred with Chapters 12 and Chapter 17, with terns largely distributed along near shore habitats, alcids distributed across large areas of the study area, loon hotspots near the mouth of the Chesapeake Bay, and Northern Gannet hotspots consisting of large, localized aggregations.
- Integrated models improved the identification of abundance hotspots (and areas of lower than expected abundance) within a survey period, but further work is required to explore their predictive ability across surveys.

Implications

Developing new approaches to jointly model disparate datasets can improve identification of important habitat use areas, while also providing a framework to combine historical and new sources of data.

[1] For more detailed context for this chapter, please see the introduction to Part IV of this report.

Abstract

We investigated an approach to combine shipboard and digital video aerial survey data for marine birds to produce a single prediction of marine bird abundance and distribution. Modeling frameworks were similar to those in other chapters (for example, the boat survey data are modeled similarly to Chapters 11, 12, and 18, but with single species instead of a community), but we aimed to create a method for integrating the boat and aerial datasets. Our approach in this chapter creates a covariate based on the data from the digital video aerial surveys to be included as a predictor variable in the boat surveys. We compare models with and without the aerial covariate and evaluate the model performance. As in Chapter 18, we focused on terns (summer 2013), Northern Gannets (winter 2012), loons (winter 2012), and alcids (winter 2012). The preliminary results showed that integrated and boat-only models predicted similar total abundance across the study area, but distributions and hotspot locations often varied between approaches. Northern Gannets and loons showed strong associations between aerial and boat data, which led to concentrated hotspots for both species. The influence of integrated models were less evident for terns and alcids, but differences in predicted spatial patterns of abundance were still evident. Model evaluation indicated that integrated models outperformed models that only used boat data when predicting back to the same boat and aerial data used for the analysis, but boat-only models were better at predicting distributions from separate surveys (i.e., boat and aerial surveys conducted in the same season but during a different month than the data used in the analysis). The results of this chapter generally support conclusions of Chapters 12, 17, and 18, which found that the distribution of marine birds was often patchy, species- and survey-specific, and correlated with habitat covariates. Developing new joint modeling approaches can improve identification of important habitat use areas (particularly local dynamic hotspots) and provides a framework to compare historical and new sources of data. Further exploration of seasonal, annual, and species-specific results will be useful to evaluate the performance of integrated models to predict important habitat use areas.

Introduction

Shipboard and traditional aerial survey methodologies have been compared extensively in their performance at estimating marine bird species richness and abundance (see Camphuysen et al. 2004 for an overview). Advantages of shipboard and aerial survey methods vary by species of interest, logistical constraints, size of area surveyed, and specific research questions (Camphuysen et al. 2004). High resolution digital video aerial surveys (hereafter “digital video aerial surveys”) have become common practice for monitoring of marine wildlife in Europe in relation to offshore wind energy development, but the technology is still relatively new, and thus has seldom been compared to more traditional survey approaches. Digital video aerial surveys provide several benefits, including coverage of a larger geographic area in a faster timeframe than is possible with boat surveys (Buckland et al. 2012). Likewise, boat-based surveys provide advantages such as the ability to estimate detection probability based on distance sampling, and thereby adjust raw counts to develop estimates of abundance (Buckland et al. 1993, see Chapters 11-12 for details). Application of concurrent shipboard and digital video aerial surveys in this study provided a unique opportunity to integrate data collected from both these survey types, thus utilizing the strengths of each survey platform.

Herein, we investigate an approach to combine shipboard and digital aerial surveys for marine birds into a single model that uses information from both datasets (hereafter "integrated model"). We aim to compare integrated models and models that use strictly boat data (hereafter "boat-based"). This chapter includes the same suite of species as Chapter 18 (terns, Sternidae; alcids, Alcidae; loons, *Gavia* spp.; and Northern Gannets, *Morus bassanus*) to allow for improved comparisons and insights across multiple species groups.

Our objectives include:

1. Develop an integrated modeling approach that combines digital video aerial and boat-based data to produce a single prediction of marine bird abundance and distribution.

2. Compare the performance of integrated and boat-based models by evaluating their predictive ability to (1) the original boat and aerial datasets (Fitted surveys) and (2) independent boat and aerial surveys (i.e., independent surveys conducted in the same season but during a different month).

It is important to note that there are methodological differences in sampling from the boat versus digital videography. Some differences are inherent to the two survey methods, such as transect width; the boat surveys sample wider transect widths for most species, and use distance sampling to account for variation in detection, while aerial surveys sample a defined strip width for all species. Other differences are specific to the survey design utilized in this study (e.g., boat and aerial transects were located in slightly different geographic areas and occurred at different dates and times). To minimize the study-specific sources of variation, we accounted for the differences in area sampled, and only used data from boat and aerial surveys that occurred within a similar temporal period (i.e., within the same or consecutive months of each other).

In general, our approach first utilized the large geographic coverage of digital video aerial surveys to predict areas of higher or lower than average expected marine bird abundance across the study area, essentially mapping raw hot and cold spots for each species. Smoothed counts from the digital video aerial surveys (i.e., the degree to which grid cells were above or below expected values) were then integrated as a covariate in analysis of boat surveys. As in Chapters 11-12, 15, and 18, habitat covariates were also included in the analysis. This approach allowed predicted marine bird abundance from boat data to not only vary by habitat covariates (similar to Chapters 11-12, and 18), but also by information on expected abundance derived from digital video aerial surveys. This additional covariate should inform model predictions if similar trends in distribution and abundance were observed in both boat and digital video aerial surveys (Chapter 18). Specific details on our modeling approach are provided in the Methods section.

We caution the reader that the approach we use in this chapter is not a fully integrated model, as we use the aerial survey data as a covariate for analyzing the boat survey data. However, we implemented this approach to see if the aerial data would provide useful information for estimating abundance and local hotspots of abundance. After building the models, we compare and contrast those with and without the aerial data covariate to see how well the models do at (1) estimating overall abundance and

local hotspots and (2) predicting future seabird patterns. This chapter is an important step towards simultaneously modeling the two data types, which is the ultimate goal and will continue to be pursued in an addendum to the final report, to be completed in 2016.

Methods

Field methods for the aerial and boat surveys were explained elsewhere in this report (Chapters 3 and 7, respectively). Aerial identification protocols for video analysis were discussed in Chapter 4. For this comparison, we used boat survey observations that were sampled from the forward quadrant on one side of the vessel, extending up to 1 km from the trackline, and digital aerial observations that were collected from four cameras, which each recorded a 50 m band (totaling 0.2 km strip width). For both the boat and aerial surveys, we divided survey transects into 4 km segments; this resulted in some shorter segments at the transect ends, and segment area (the segment length by the abovementioned strip widths for each survey method) was included in our analysis as an offset. The number of individuals for each species was summed by segment and survey. We compared two modeling methods to estimate abundance and covariate relationships for the same species groups examined in Chapter 18, using data from boat and aerial surveys that were closely coincident in time.

Species

We investigated the same suite of species as Chapter 18 (terns, alcids, loons, and Northern Gannets).

Terns

Terns included Least Terns (*Sternula antillarum*), Caspian Terns (*Hydroprogne caspia*), Black Terns (*Chlidonias niger*), Common Terns (*Sterna hirundo*), Roseate Terns (*Sterna dougallii*), Royal Terns (*Thalasseus maximus*), and Sandwich Terns (*Thalasseus sandvicensis*), as well as those individuals classified as "unidentified terns." Vague identifications that could have included other species such as gulls (e.g., "large tern or small gull,") were excluded. Terns were primarily present in the study area during spring, summer and fall (Chapters 5, 8, 12), so we focused on boat and aerial surveys during the summer (Chapter 18), specifically August (boat) and September (aerial) 2013 (Table 19-1).

Northern Gannets

Northern Gannets are the only gannet species found in the study area. Because Northern Gannets were primarily present in the study area in late fall to early spring (Chapters 5, 8, 12), it made sense to focus on boat and aerial surveys during the winter season (Chapter 18), and for this analysis, we selected the survey from December 2012.

Loons

We considered loons as a group (all loons, which included Common Loons, *Gavia immer*, Red-throated Loons, *G. stellata,* and all unidentified loon observations). Similar to Northern Gannets, loons were primarily present in the study area from late fall to early spring (Chapters 5, 8, 12). We used the same survey, December 2012, for the analysis of loons.

Alcids

The alcid group included Razorbills (*Alca torda*), Dovekies (*Alle alle*), Atlantic Puffins (*Fratercula arctica*), Common Murres (*Uria aalge*), Thick-billed Murres (*U. lomvia*), and Black Guillemots (*Cepphus grille*), as

well as those individuals classified as "unidentified alcids." Alcids were primarily present in the study area during winter (Chapters 5, 8, 12), so again we focused on boat and aerial surveys during the winter season (Chapter 18), specifically December 2012.

Covariates

As in Chapter 18, we used five covariates in our analyses: three static (distance to shore, slope, and grain size), and two dynamic (sea surface temperature and salinity). We excluded chlorophyll-*a* in these analyses because it was co-linear with distance to shore in some of the surveys and we wanted to keep the covariates consistent across species. As in Chapter 18, remotely sensed covariate data corresponded to the values located at the midpoint of each transect segment. For the static covariates, we calculated distance to shore (m) within ArcGIS 10.2 (ESRI, Redlands, CA) and extracted slope (% rise, 370-m resolution) and grain size (ϕ = -log2[mean grain diameter in mm], 370-m resolution) from the data layer derived by NOAA/NOS National Centers for Coastal Ocean Science (Kinlan et al. 2013). For the dynamic covariates, we used Marine Geospatial Ecology Tools in ArcGIS (Roberts et al. 2010) to download remotely-sensed data at the highest resolution available for all segments. We compiled daily values for sea surface temperature (SST, °C, 1-km GHRSST L4) and salinity (Practical Salinity Units, 9-km HYCOM GLBa0.08 Equatorial 4D). In the boat survey analysis, we additionally included Beaufort sea state on the binary scale as a covariate to detection, which varied by segment (0 = calm seas, Beaufort state 0-2; 1 = rough seas, Beaufort state 3-6; see Chapters 12 and 18).

We overlaid a predictive grid (approximately 4x4 km grid cells) that encompassed the entire sampled area, including the three proposed wind energy areas (Figure 19-1 and Figure 19-2; note that this predictive grid is a restricted version of the one used in Chapter 12, to represent the area covered by digital aerial surveys). As in Chapter 12, we used data from the midpoint of each cell and the central date for each season to predict overall flock abundance to a representative day (summer [terns]: 25 July 2013, winter [Northern Gannets, loons, and alcids]: 25 December 2012).

Models

To facilitate comparisons, we used the same modeling approach across all species. We summarized the aerial data such that y_i is the count at segment i. For each species or group, we then modeled the aerial data using an overdispersed Poisson conditional autoregressive (CAR) model. This approach allowed us to capture excess heterogeneity in counts at the segment level (overdispersion), while also allowing spatial clustering of counts at a broader level with the CAR portion. To implement the model, we assigned each segment to the predictive grid cell that it fell within, thus the notation $i[j]$ indicates that segment i is within grid cell j.

The model for aerial surveys was:

$$y_i \sim \text{Poisson}(\lambda_i)$$

$$\log(\lambda_i) = \alpha_0 + offset\ (segment\ area_i) + \varepsilon_i + \theta_{i[j]}$$

where α_0 is the intercept, ε_i is a random effect at the segment level (i), and $\theta_{i[j]}$ is the spatially correlated random effect at the predictive grid cell level (j). Random segment effects (ε_i) were

distributed $\text{Normal}(0, \tau^2)$. Spatial autocorrelation was evaluated at the grid cell level, thus repeated segments (i) within grid cell j were assigned the same $\theta_{i[j]}$. Specifically,

$$\theta_j | \, \theta_k \neq \theta_j \sim \text{Normal}\left(\frac{1}{m_j}\sum_{k \in c_j}\theta_k, \frac{\sigma^2}{m_j}\right)$$

where $m_j$is the number of neighbors for predictive grid cell j and c_j is the specific set of neighbors for predictive grid cell j. The set of neighbors for each predictive grid cell (c_j) was all adjacent grid cells (i.e., Queen's neighborhood). The CAR model also allows inference to unsampled grid cells by utilizing the spatial correlation observed in counts across the sampled area. Grid cell specific random effects (θ_j), which indicate higher or lower expected grid cell abundance, were then used to inform abundance estimates from the boat data.

Next, for each species or group, we conducted preliminary diagnostics to evaluate boat-based data and select the best model for flock abundance, considering the Poisson and Negative Binomial distributions (see Chapters 11-12 and 18 for details). It should be noted that in this chapter, as well as Chapters 11 and 12, abundance of flocks was the sampling unit of analysis; however, Chapter 18 used individuals instead of flocks to make direct comparisons between boat and aerial surveys. Boat-based models for Northern Gannets, loons, and terns used a Negative Binomial distribution on abundance. The abundance component of the boat-based model was constructed such that the flock abundance at segment i, N_i, was modeled as:

$$N_\text{i} \sim \text{NegBin}(\lambda_i, \, r)$$

$$\log(\lambda_i) = \beta_0 + offset\,(segment\ area_i) + \beta_1 \text{Dst}_\text{i} + \beta_2 \text{Slp}_\text{i} + \beta_3 \text{Grn}_\text{i} + \beta_4 \text{Sst}_\text{i} + \beta_5 \text{Sal}_\text{i} + \, \beta_5 \theta_{i[j]}$$

where Dst = distance to shore, Slp = slope of the seafloor, Grn = sediment grain size, Sst = sea surface temperature, Sal = salinity, $\theta_{i[j]}$ = the estimated spatially correlated random effect defined above from the aerial data, and r is the overdispersion parameter. In this approach, a positive parameter estimate for the aerial covariate (β_5) indicates that aerial and boat surveys were observing similar trends, and data integration is informative for understanding the abundance patterns between the two surveys. Alcid boat-based data were adequately fit with a Poisson distribution, which was identical to the above model except that $N_\text{i} \sim \text{Pois}(\lambda_\text{i})$.

Raw count of flocks at segment i, n_i were linked to true abundance (N_i) through estimation of detection probability using a half-normal distribution (i.e., distance sampling, see Chapter 11-12, 18; Buckland et al. 1993). As in Chapters 12 and 18, we allowed the detection function to vary by a binary indicator of sea state.

Model evaluation

We evaluated integrated and boat-based models by predicting abundance to the original boat and aerial survey data (i.e., fitted surveys) and independent boat and aerial data (i.e., independent surveys in the same season). Specifically, we compared tern predictions to the original survey data, August 2013 (boat)

and September 2013 (aerial), as well as to independent surveys conducted during September 2013 (boat) and July 2013 (aerial). For Northern Gannets, loons, and alcids we compared predictions to the original survey data in December 2012 (boat and aerial) and independent surveys conducted during January 2013 (boat and aerial). Using the posterior means of each model parameter, we predicted the abundance of each species using survey-specific habitat covariates. Predictive ability was evaluated using root mean squared error (smaller indicates better model fit) with each segment considered as a replicate.

Implementation

We implemented all models in a Bayesian framework using the package "R2OpenBUGS" (Sturtz et al. 2010) to run the software OpenBUGS (Thomas et al. 2006) in program R version 3.2.0 (R Core Team 2014). We standardized the covariates for analysis to center them on a mean = 0, with a variance close to 1. We ran three parallel Markov chains for 40,000 iterations following a burn-in of 20,000 iterations, thinning by 5. We checked for chain convergence visually (posterior density and trace plots), and quantitatively using the Gelman-Rubin statistic (Gelman et al. 2014). This statistic (termed R-hat) is a measure of among-chain versus between-chain variance and values < 1.1 indicate convergence (Gelman et al. 2014). We also assessed goodness of fit by computing Bayesian p-values. We used Freeman-Tukey fit statistics to evaluate the model for abundance, and the Half Normal detection function (Gelman et al. 2014).

Results

The number of species/group-specific observations were > 100 for loons, Northern Gannets, and alcids in both boat (flocks) and digital aerial surveys (individuals), and > 60 for terns (Table 19-1). Mean flock size varied by species, but most observed flocks were $\leq$ 2 individuals (Table 19-1). Distributions of flock size were right-skewed for all species, with larger flocks ($\geq$ 18) only observed on rare occasions (Table 19-1). Bayesian p-values suggested that model fit was adequate for all of the abundance and detection model components (Table 19-2).

Integrated and boat-based models often resulted in similar abundance estimates (Table 19-3). For example, predicted tern abundance from the integrated and boat-based models was 3,367 and 3,727, respectively. Covariate relationships were also similar between integrated and boat-based models (Table 19-4 to Table 19-7). The noticeable exception to this trend was parameter estimates for distance to shore, especially for Northern Gannets. Integrated models for Northern Gannets predicted a significant positive relationship, while boat-based models predicted a negative relationship (Table 19-6). A reversal of the distance to shore parameter estimate was also observed for loons and alcids, but differences were generally smaller or non-significant (Table 19-6, Table 19-7, see next paragraph and Discussion for detailed explanations). Parameter estimates for the aerial covariate were positive and significant for both Northern Gannets and Loons (Table 19-5, Table 19-6). Positive parameter estimates indicated that variation in expected grid cell abundance in the aerial-based models were positively correlated with variation in expected grid cell abundance estimated from the boat data. Mean parameter estimates for terns and alcids were negative, but not significant (Table 19-4, Table 19-7).

Predicted distribution and abundance from both integrated and boat-based models showed similar relationships, with terns, loons, and Northern Gannets predicted to be closer to shore (Figure 19-2 to Figure 19-4). Integrated models, however, often identified hotspots that were not predicted in boat-based models (Figure 19-2 to Figure 19-5). Hotspots were particularly evident for Northern Gannets (Figure 19-3) and loons (Figure 19-4). Predicted distributions for terns and alicds were often more similar across models, but areas of higher and lower abundances were still noticeable (Figure 19-2, Figure 19-5). For instance, boat-based models for Northern Gannets predicted a rather uniform trend in abundance that decreased with distance to shore. Integrating information from aerial surveys, however, predicted a much more clustered distribution, even though total abundance was similar between the models (Figure 19-3, Table 19-1). Similarly, boat-based models for terns predicted a relatively ubiquitous near-shore abundance that quickly decreased with distance to shore (Figure 19-2). Integrated models predicted a similar trend with distance to shore for terns, but indicated several areas of higher and lower abundances in the nearshore environment (Figure 19-2). Over the entire study area, the range of predicted grid cell-specific flock abundances was much smaller for alcids (range = 0 - 40 flocks per grid cell) and terns (0 - 60 flocks per grid cell) relative to Northern Gannets (0 - 400 flocks per grid cell) and loons (0 - 200 flocks per grid cell; note species-specific scales for Figure 19-2 to Figure 19-5).

Root mean squared error for integrated models was generally lower, or at least equal to boat-based models when predicting back to the original boat and aerial survey data (Table 19-8). Boat-based predictions, however, often outperformed integrated models when predicting to independent boat and aerial survey data (i.e., predicting to a different survey in the same season, Table 19-8).

Discussion

Jointly modeling aerial and boat survey data can improve our understanding of several important ecological phenomena important to proposed wind energy development, especially (1) clustering of marine wildlife within the study area and (2) relationships between marine wildlife abundance and spatially varying habitat covariates. An integrated approach utilizes beneficial aspects of both survey methods, with study area predictions informed by both aerial surveys, which encompassed a large geographic area, and boat surveys that allowed for estimation of detection probability (see Chapters 11-12, and 18; Winiarski et al. 2014). The integrated model presented herein had noticeable improvements in predicting local hotspots and marine bird distribution relative to models that only included boat-based data. The integrated model, however, had relatively low predictive power to independent surveys (data collected from a survey different than the one used to fit the models), which was likely a consequence of interseasonal variation in local hotspots, changes in habitat covariates, and possibly changes in the relationships with those covariates (Winiarski et al. 2013, Winiarski et al. 2014, Chapter 18).

In general, habitat relationships were similar to those presented in Chapter 18. For instance, relationships with distance to shore were consistently negative for terns, loons, and Northern Gannets in boat-based models herein and in Chapter 18. Conversely, parameter estimates for distance to shore were sometimes reversed (positive) in the integrated model. Both boat-based and integrated models, however, still predicted increased nearshore distribution patterns (e.g., areas of higher predicted abundance for Northern Gannet were generally closer to shore). Distance to shore and the aerial covariate used in the

integrated model were collinear for all species ($\leq$ 0.60), but likely not strong enough to completely explain the differences in the distance to shore parameter estimate. Instead, integrating aerial data likely provided more information on areas of particularly high or low abundances, and was a more informative covariate than the less variable distance to shore covariate. These examples demonstrate the necessity of investigating both parameter estimates and predicted abundance maps to determine proper interpretations from boat-based and integrated models.

In line with our hypotheses, integrated models improved the identification of abundance hotspots and areas of lower than expected abundances. The greater spatial coverage of aerial surveys improved the detection of latitudinal gradients and hotspots, especially those occurring outside of areas surveyed by the boat. Covariate relationships identified during boat surveys, however, remained important predictors of marine bird abundance and distribution. Differences between modeling approaches were most evident for Northern Gannets and loons, where integrated models predicted increased clustering relative to boat-based model predictions. For loons, integrated models predicted both a distance to shore effect and latitudinal gradient, with abundances highest near the mouth of the Chesapeake Bay (southwest region of the study area). Similarly, integrated model predictions for Northern Gannets were clustered near the mouth of the Chesapeake Bay and a few nearshore areas off the coasts of Virginia.

Due to the high proportion of birds in the aerial dataset that were not identified to species, we could not model species-specific abundances (e.g., Common and Red-throated Loons; see Chapter 16 for details). However, one goal for creating an integrated model with the two datasets is to estimate species-specific abundances by accounting for birds that were not identified to species. This is not straightforward, as the work in Chapter 16 uses the information from the boat surveys to inform species identification in the aerial survey, while in the joint model, we want to use both data simultaneously to estimate abundance. After developing the fully integrated models, we hope to incorporate a model for species identification like the one presented in Chapter 16. Until that point, our integrated approach is limited to the best available data for each survey type, and we will continue to use species groups (e.g., loons).

Integrated models were an improvement over boat-only models at identifying hotspots and predicting to the original surveys. Boat-only models, however, were better at predicting patterns observed in independent surveys conducted in the same season, which are likely longer-term patterns of abundance. Short-term changes in local hotspot locations, possibly due to shifting prey distributions, affect the predictive ability of both modeling approaches. Predictions from integrated models may be particularly affected by shifting hotspot locations, as aerial data from one sampling period will not necessarily improve predictions for a different sampling period. Further exploration of seasonal, annual, and species-specific differences beyond those explored here will help evaluate the performance of integrated models. Overall, integrated models improved identification of important habitat use areas, but further work is required to explore their predictive ability across surveys.

Developing new approaches to jointly model disparate datasets can improve identification of important habitat use areas, while also providing a framework to compare and possibly combine historical (i.e., boat-based) and new sources of data (i.e., boat-based and high definition videography). Here, we have used the aerial data as a covariate for estimating abundance from the boat data, but have not formally

integrated the two data types. Identification of positive relationships between boat and aerial surveys suggests that these survey types often identify similar trends in abundance and distribution across the study area. Verifying the consistency of these results across multiple surveys, species, and geographic areas will be vital to developing fully integrated modeling approaches.

Literature cited

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., 1993. Distance sampling: Estimating abundance of biological populations. Chapman and Hall, London.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E., Woodward, R., 2012. Aerial surveys of seabirds: the advent of digital methods. Journal of Applied Ecology 49, 960-967.

Camphuysen, C.J., Fox, A.D., Leopold, M.F., Petersen, I.K., 2004. Towards standardised seabirds at sea census techniques in connection with environmental impact assessments for offshore wind farms in the U.K. Royal Netherlands Institute for Sea Research, Texel.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2014. Bayesian Data Analysis, Third Edition. Taylor & Francis, New York, NY, USA.

Kinlan, B.P., Poti, M., Drohan, A., Packer, D.B., Nizinski, M., Dorfman, D., Caldow, C., 2013. Digital data: Predictive models of deep-sea coral habitat suitability in the U.S. Northeast Atlantic and Mid-Atlantic regions. Downloadable digital data package. Department of Commerce (DOC), National Oceanic and Atmospheric Administration (NOAA), National Ocean Service (NOS), National Centers for Coastal Ocean Science (NCCOS), Center for Coastal Monitoring and Assessment (CCMA), Biogeography Branch. Released August 2013. Available at: http://coastalscience.noaa.gov/projects/detail?key=35.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Roberts, J., Best, B., Dunn, D., Treml, E., Halpin, P., 2010. Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. http://mgel.env.duke.edu/mget.

Sturtz S., Ligges U., Gelman A., 2010. R2OpenBUGS: a package for running OpenBUGS from R. http://openbugs.info/w/UserContributedCode.

Thomas, A., O'Hara, B., Ligges, U., and Sturtz, S., 2006. Making BUGS Open. R News 6, 12-17.

Winiarski, K.J., Miller D.L., Paton P. W.C., and McWilliams S.R., 2013. A spatially explicit model of wintering common loons: Conservation implications. Marine Ecology Progress Series 492:273–278.

Winiarski, K.J., Burt, M.L., Rexstad, E., Miller, D.L., Trocki, C.L., Paton, P.W.C., McWilliams, S.R., 2014. Integrating aerial and ship surveys of marine birds into a combined density surface model: A case study of wintering Common Loons. The Condor, 149-161.

Figures and tables

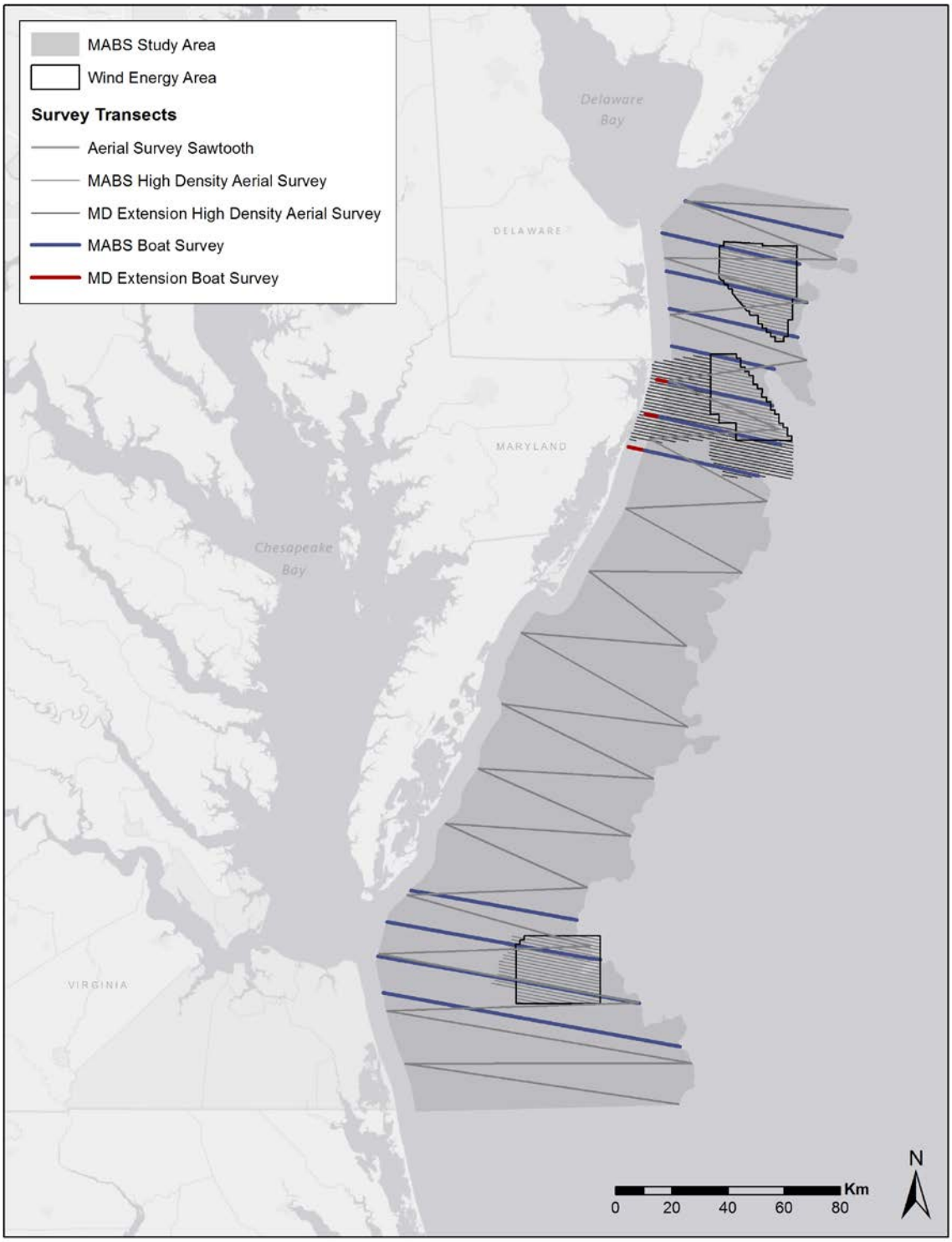

Figure 19-1. Study area. Boat transects are shown in blue and red, and aerial transects in gray; Maryland extension transects (funded by the state of Maryland and conducted only in the second year of surveys) are shown in red (boat) and dark gray (aerial). Department of Energy (DOE)-funded high density aerial surveys were located within federally designated wind energy areas (WEAs).

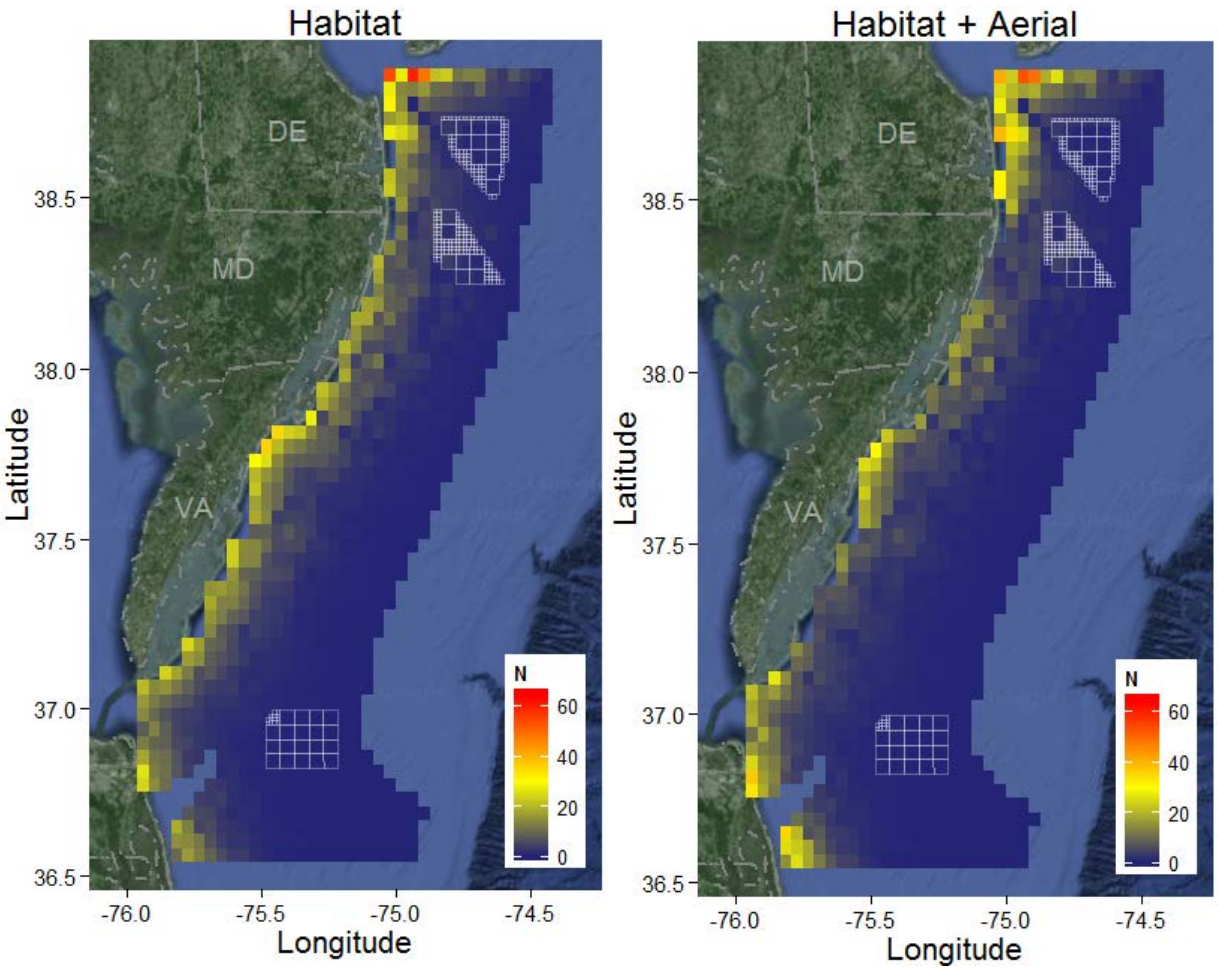

Figure 19-2. Predicted abundance of tern flocks using the boat-based model (Habitat, left) or integrated model (Habitat + Aerial, right). Note species-specific flock abundance scale. Median tern flock size was 1.0 individuals (Table 19-1). Covariate values were from the midpoint date for the summer season (25 July 2013). White grids represent proposed wind energy areas. See Table 19-1.

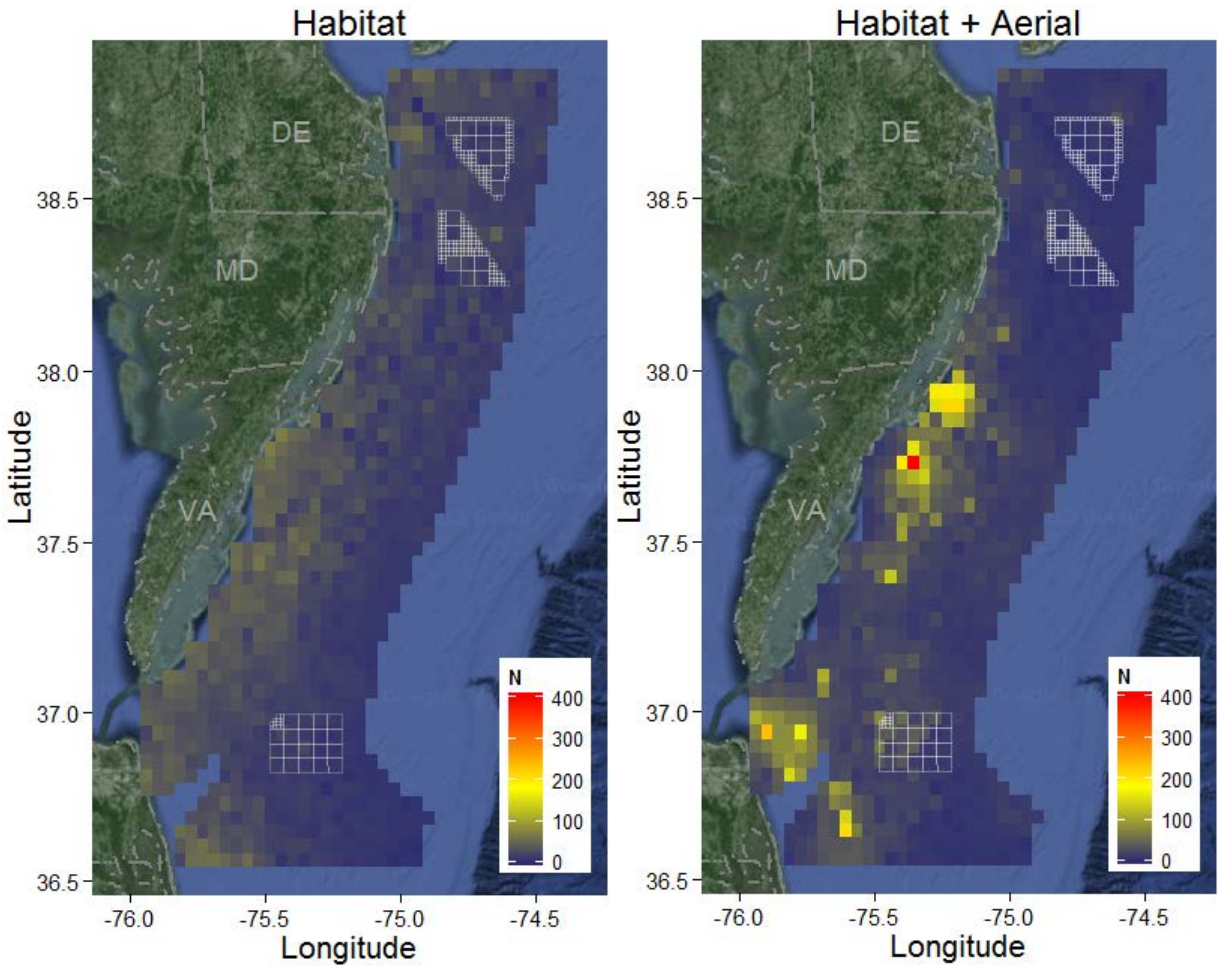

Figure 19-3. Predicted abundance of Northern Gannet flocks using the boat-based model (Habitat, left) or integrated model (Habitat + Aerial, right). Note species-specific flock abundance scale. Median Northern Gannet flock size was 1.0 individuals (Table 19-1). Covariate values were from the midpoint date for the winter season (25 December 2012). White grids represent proposed wind energy areas.

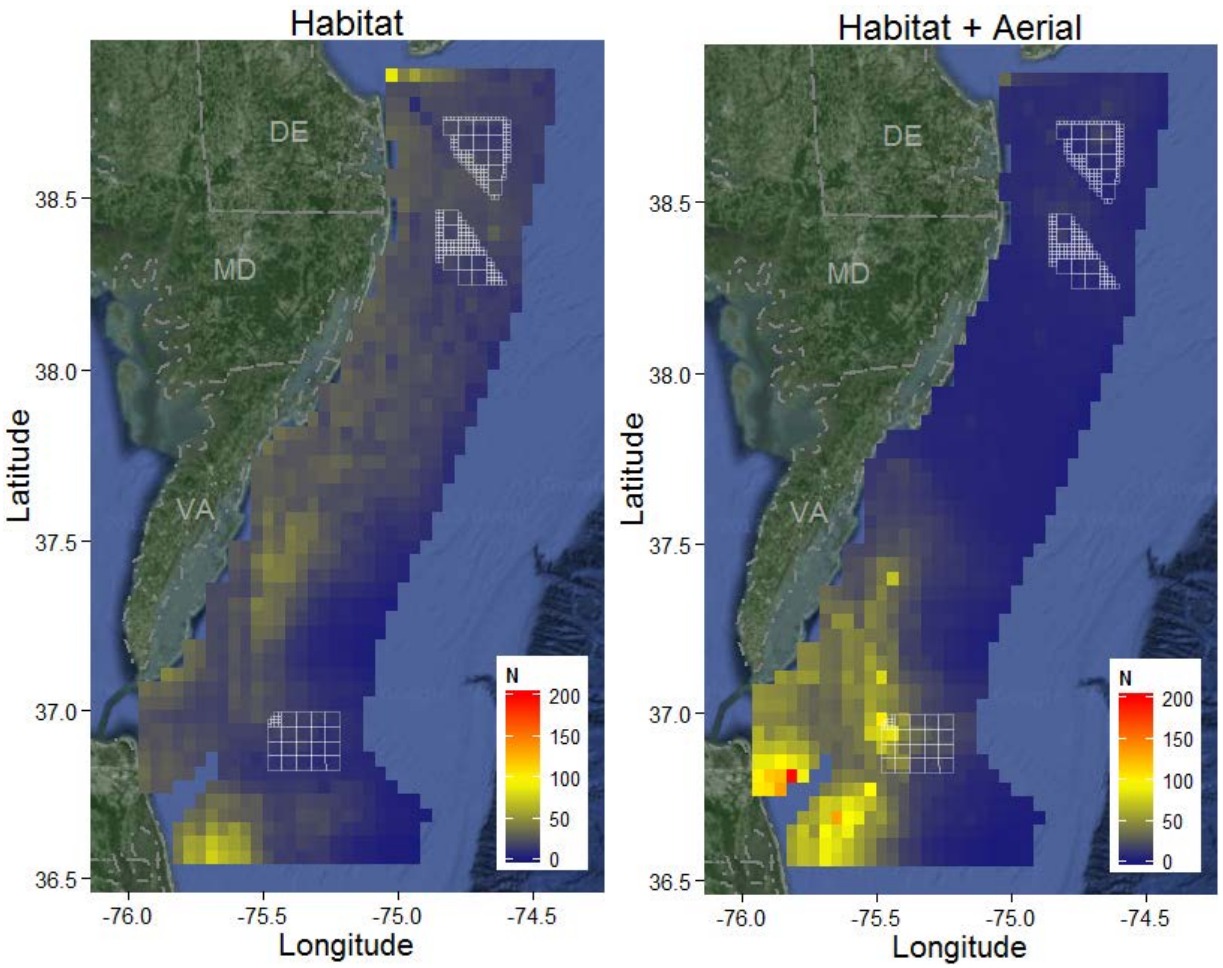

Figure 19-4. Predicted abundance of loon flocks using the boat-based model (Habitat, left) or integrated model (Habitat + Aerial, right). Note species-specific flock abundance scale. Median loon flock size was 1.0 individuals (Table 19-1). Covariate values were from the midpoint date for the winter season (25 December 2012). White grids represent proposed wind energy areas.

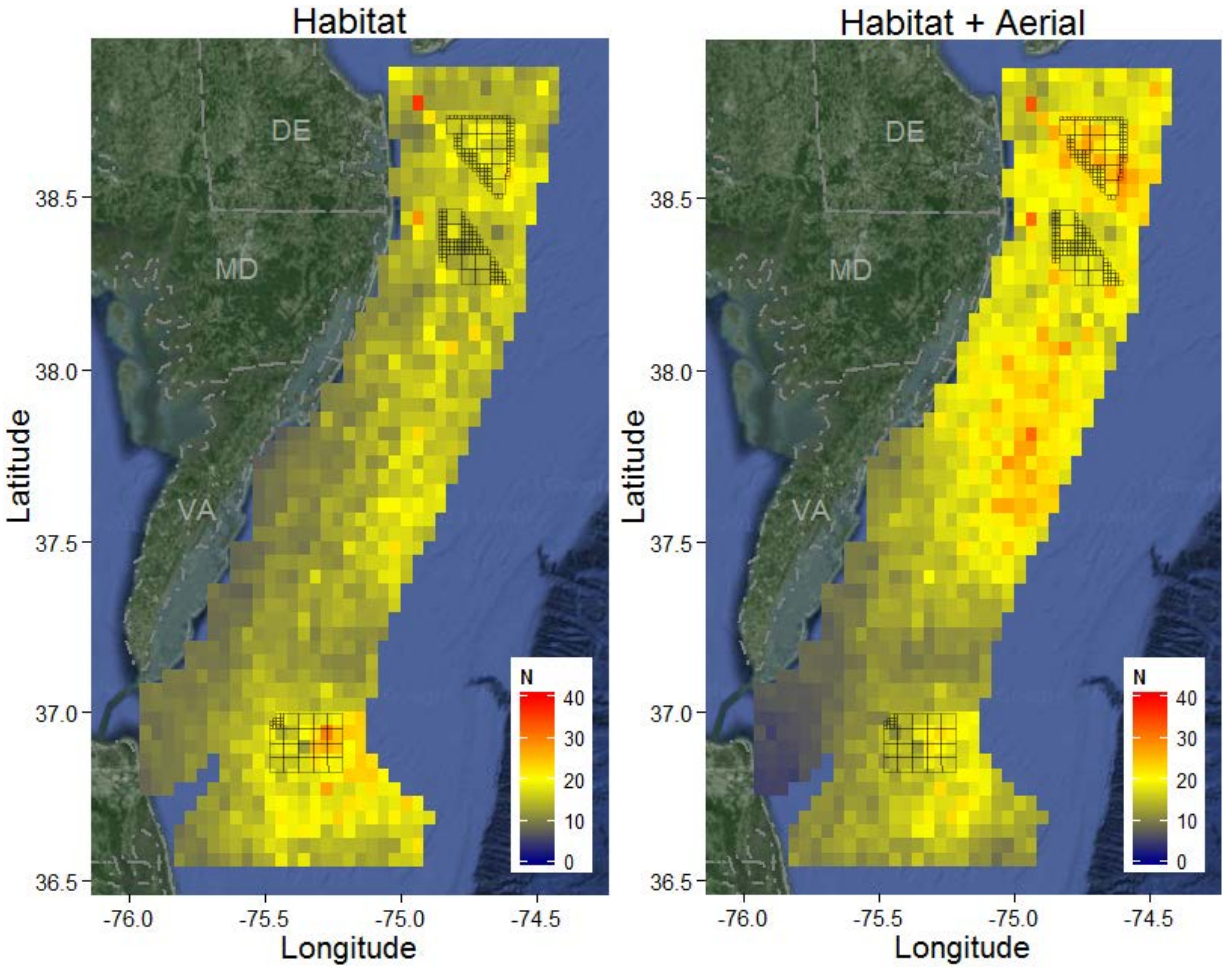

Figure 19-5. Predicted abundance of alcid flocks using the boat-based model (Habitat, left) or integrated model (Habitat + Aerial, right). Note species-specific flock abundance scale. Median alcid flock size was 2.0 individuals (Table 19-1). Covariate values were from the midpoint date for the winter season (25 December 2012). Black grids represent proposed wind energy areas.

Table 19-1. Surveys used in the analysis for each species/group, and the raw counts (observations) for each species/group.

Boat Survey	Aerial survey	Group	Boat observations					Aerial observations
			Flocks	Flock size				
				Mean	Median	Min	Max	Individuals
Aug 2013	Sep 2013	Terns	67	1.7	1.0	1	30	69
Dec 2012	Dec 2012	Northern Gannets	306	3.4	1.0	1	350	407
Dec 2012	Dec 2012	Loons	299	1.4	1.0	1	25	703
Dec 2012	Dec 2012	Alcids	122	2.3	2.0	1	18	148

Table 19-2. Bayesian p-values for the abundance and detection components of the models, using either the boat-based model (Habitat) or integrated model (Habitat +Aerial). Values close to 0.5 indicate good model fit.

Group	Model	Boat		Aerial
		Abundance	Detection	Abundance
Terns	Habitat	0.51	0.48	NA
	Habitat + Aerial	0.51	0.48	0.44
Northern Gannets	Habitat	0.44	0.23	NA
	Habitat + Aerial	0.48	0.23	0.57
Loons	Habitat	0.57	0.32	NA
	Habitat + Aerial	0.58	0.31	0.26
Alcids	Habitat	0.58	0.31	NA
	Habitat + Aerial	0.58	0.31	0.44

Table 19-3. Predicted flock abundance of each species/group to a representative summer day (25 Jul. 2013; terns) or winter day (25 Dec. 2012; Northern Gannets, loons, and alcids) using either the boat-based model (Habitat) or integrated model (Habitat +Aerial). Prediction area was constant across species and designed to represent the surveyed area (see Figure 19-2, Figure 19-3, Figure 19-4, and Figure 19-5).

Boat survey	Species/Group	Predicted abundance	
		Habitat	Habitat + Aerial
Aug 2013	Terns	3,726.8	3,366.6
Dec 2012	Northern Gannets	19,576.9	20,275.4
Dec 2012	Loons	18,503.6	17,371.6
Dec 2012	Alcids	12,969.0	14,653.6

Table 19-4. Parameter estimates for terns from the boat based-model (Habitat) or integrated model (Habitat + Aerial). Abundance was modeled using a Negative Binomial distribution. SD is the standard deviation, 2.5% and 97.5% are the respective quantiles, $\boldsymbol{r}$ is the overdispersion parameter, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters are on the log scale. Dst = distance to shore, Slp = slope of the seafloor, Grn = sediment grain size, Sst = sea surface temperature, Sal = salinity, Aerial = aerial covariate (i.e., smoothed aerial counts), and Beaufort sea state 3-6 are rough seas (as opposed to calm, 0-2). The posterior mean for covariates where the 95% credible interval does not overlap zero are in bold italics.

Terns									
Negative Binomial		**Habitat**				**Habitat + Aerial**			
Component	**Term**	**Mean**	**SD**	**2.5%**	**97.5%**	**Mean**	**SD**	**2.5%**	**97.5%**
Abundance	**Intercept**; α_0	***-2.12***	0.35	-2.85	-1.48	***-2.29***	0.39	-3.10	-1.58
	Dst; α_1	***-1.91***	0.37	-2.67	-1.23	***-2.12***	0.42	-2.99	-1.35
	Slp; α_2	-0.33	0.25	-0.86	0.15	-0.39	0.26	-0.94	0.10
	Grn; α_3	0.22	0.23	-0.22	0.68	0.18	0.23	-0.27	0.64
	Sst; α_4	-0.20	0.30	-0.79	0.41	-0.29	0.33	-0.96	0.33
	Sal; α_5	-0.31	0.23	-0.77	0.14	-0.07	0.29	-0.62	0.52
	Aerial; α_6	-	-	-	-	-0.26	0.18	-0.63	0.08
	Overdisp.; r	1.10	0.74	0.38	2.89	1.17	0.93	0.39	3.17
Detection	**Beaufort 0-2**;β_0	***5.28***	0.11	5.08	5.50	***5.28***	0.11	5.08	5.50
	Beaufort 3-6; β_1	***5.02***	0.14	4.76	5.32	***5.03***	0.14	4.77	5.32

Table 19-5. Parameter estimates for Northern Gannets from the boat based-model (Habitat) or integrated model (Habitat + Aerial). Abundance was modeled using a Negative Binomial distribution. SD is the standard deviation, 2.5% and 97.5% are the respective quantiles, $\boldsymbol{r}$ is the overdispersion parameter, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters are on the log scale. Dst = distance to shore, Slp = slope of the seafloor, Grn = sediment grain size, Sst = sea surface temperature, Sal = salinity, Aerial = aerial covariate (i.e., smoothed aerial counts), and Beaufort sea state 3-6 are rough seas (as opposed to calm, 0-2). The posterior mean for covariates where the 95% credible interval does not overlap zero are in bold italics.

Northern Gannets									
Negative Binomial		**Habitat**				**Habitat + Aerial**			
Component	**Term**	**Mean**	**SD**	**2.5%**	**97.5%**	**Mean**	**SD**	**2.5%**	**97.5%**
Abundance	**Intercept**; α_0	0.11	0.14	-0.17	0.41	0.11	0.13	-0.15	0.37
	Dst; α_1	***-0.48***	0.21	-0.88	-0.07	***0.80***	0.32	0.21	1.46
	Slp; α_2	-0.33	0.18	-0.68	0.02	-0.31	0.16	-0.64	0.00
	Grn; α_3	0.20	0.16	-0.13	0.52	0.00	0.15	-0.31	0.29
	Sst; α_4	0.06	0.18	-0.30	0.40	***-0.86***	0.24	-1.33	-0.39
	Sal; α_5	-0.14	0.19	-0.54	0.21	-0.24	0.19	-0.64	0.12
	Aerial; α_6	-	-	-	-	***0.86***	0.17	0.54	1.24
	Overdisp.; r	0.54	0.10	0.37	0.76	0.77	0.17	0.50	1.14
Detection	**Beaufort 0-2**;β_0	***5.67***	0.05	5.58	5.77	***5.66***	0.05	5.57	5.76
	Beaufort 3-6; β_1	***5.80***	0.08	5.65	5.97	***5.83***	0.08	5.68	6.01

Table 19-6. Parameter estimates for loons from the boat based-model (Habitat) or integrated model (Habitat + Aerial). Abundance was modeled using a Negative Binomial distribution. SD is the standard deviation, 2.5% and 97.5% are the respective quantiles, $\boldsymbol{r}$ is the overdispersion parameter, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters are on the log scale. Dst = distance to shore, Slp = slope of the seafloor, Grn = sediment grain size, Sst = sea surface temperature, Sal = salinity, Aerial = aerial covariate (i.e., smoothed aerial counts), and Beaufort sea state 3-6 are rough seas (as opposed to calm, 0-2). The posterior mean for covariates where the 95% credible interval does not overlap zero are in bold italics.

Loons									
Negative Binomial		**Habitat**				**Habitat + Aerial**			
Component	**Term**	**Mean**	**SD**	**2.5%**	**97.5%**	**Mean**	**SD**	**2.5%**	**97.5%**
Abundance	**Intercept**; α_0	***0.26***	0.12	0.03	0.49	0.02	0.12	-0.22	0.25
	Dst; α_1	***-0.86***	0.20	-1.25	-0.48	0.09	0.26	-0.45	0.60
	Slp; α_2	-0.08	0.11	-0.30	0.15	0.02	0.11	-0.20	0.24
	Grn; α_3	0.06	0.12	-0.17	0.30	0.00	0.11	-0.21	0.22
	Sst; α_4	***0.69***	0.15	0.40	0.98	-0.39	0.26	-0.89	0.15
	Sal; α_5	***-0.51***	0.14	-0.79	-0.25	***-0.44***	0.12	-0.68	-0.22
	Aerial; α_6	-	-	-	-	***1.26***	0.26	0.75	1.77
	Overdisp.; r	1.42	0.42	0.81	2.42	1.94	0.65	1.04	3.55
Detection	**Beaufort 0-2**; β_0	***5.42***	0.05	5.33	5.52	***5.41***	0.05	5.33	***5.50***
	Beaufort 3-6; β_1	***5.48***	0.09	5.33	5.66	***5.53***	0.09	5.36	***5.71***

Table 19-7. Parameter estimates for alcids from the boat based-model (Habitat) or integrated model (Habitat + Aerial). Abundance was modeled using a Poisson distribution. SD is the standard deviation, 2.5% and 97.5% are the respective quantiles, $\boldsymbol{r}$ is the overdispersion parameter, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters are on the log scale. Dst = distance to shore, Slp = slope of the seafloor, Grn = sediment grain size, Sst = sea surface temperature, Sal = salinity, Aerial = aerial covariate (i.e., smoothed aerial counts), and Beaufort sea state 3-6 are rough seas (as opposed to calm, 0-2). The posterior mean for covariates where the 95% credible interval does not overlap zero are in bold italics.

Alcids									
Poisson		**Habitat**				**Habitat + Aerial**			
Component	**Term**	**Mean**	**SD**	**2.5%**	**97.5%**	**Mean**	**SD**	**2.5%**	**97.5%**
Abundance	**Intercept**; α_0	-0.02	0.12	-0.26	0.21	0.03	0.12	-0.21	0.27
	Dst; α_1	0.09	0.15	-0.20	0.37	-0.20	0.25	-0.70	0.30
	Slp; α_2	0.08	0.09	-0.10	0.25	0.07	0.09	-0.12	0.23
	Grn; α_3	-0.12	0.10	-0.33	0.08	-0.10	0.10	-0.31	0.10
	Sst; α_4	-0.02	0.14	-0.28	0.25	0.30	0.26	-0.21	0.82
	Sal; α_5	-0.15	0.10	-0.33	0.06	-0.17	0.10	-0.36	0.04
	Aerial; α_6	-	-	-	-	-0.27	0.19	-0.65	0.11
Detection	**Beaufort 0-2**; β_0	***5.07***	0.07	4.93	5.22	***5.07***	0.07	4.94	5.22
	Beaufort 3-6; β_1	***4.79***	0.13	4.54	5.05	***4.77***	0.13	4.53	5.03

Table 19-8. Root mean squared error (RMSE) evaluating the ability of each model to predict abundance in (a) the original boat and aerial datasets (Original surveys) and (b) independent boat and aerial datasets (Independent surveys). Predictions were from the boat-based (Habitat) or integrated model (Habitat + Aerial). Dates for the original surveys and independent surveys are provided. Root mean squared error values closer to zero indicate better model fit (lower value for each comparison is in bold italics).

(a) Original surveys		Original survey date	RMSE	
Group	**Dataset**		**Habitat**	**Habitat + Aerial**
Terns	Boat	Aug-13	0.9	0.9
	Aerial	Sep-13	0.7	***0.6***
Northern Gannets	Boat	Dec-12	3.6	***3.1***
	Aerial	Dec-12	1.8	***1.0***
Loons	Boat	Dec-12	2.2	***2.0***
	Aerial	Dec-12	3.2	***2.7***
Alcids	Boat	Dec-12	1.0	1.0
	Aerial	Dec-12	***1.1***	1.4

(b) Independent surveys		Independent survey date	RMSE	
Group	**Dataset**		**Habitat**	**Habitat + Aerial**
Terns	Boat	Sep-13	1.0	1.0
	Aerial	Jul-13	***0.5***	0.6
Northern Gannets	Boat	Jan-13	***4.5***	14.4
	Aerial	Feb-13	***21.5***	22.1
Loons	Boat	Jan-13	***2.1***	3.9
	Aerial	Feb-13	***3.2***	3.4
Alcids	Boat	Jan-13	1.6	1.6
	Aerial	Feb-13	1.7	***1.6***