Nantucket Sound Sea Ducks

Adam D. Smith, Benjamin Hofner, Jason E. Osenkowski, Taber Allison,
Scott R. McWilliams, and Peter W. C. Paton

15 July, 2015

A basic tenet of ecology is that biophysical features affect the distribution and abundance of organisms at multiple spatial and temporal scales (MacArthur 1972, Andrewartha and Birch 1986, Rose and Leggett 1990, Maurer 1999). However, identifying the biophysical features preponderant to animal distribution and abundance is complicated because this influence may be direct via biological or physiological constraints (e.g., bathymetry constrains foraging depth) or indirect via the regulation of the distribution and abundance of important resources (**do we need citations for this sentence?**). These indirect influences are particularly important to the distribution and abundance of higher trophic-level predators in marine ecosystems, but the typically complex spatiotemporal dynamics of marine prey make them challenging to monitor and often relegates us to the use of biophysical features as speculative surrogates for the distribution and abundance of key resources (Kirk et al. 2008, Robinson et al. 2011, Boyd et al. 2015). Human activities that alter the distribution and availability of potential prey, either directly or via biophysical alterations of the environment, likewise can affect the spatial ecology of higher trophic-level marine organisms (Bertrand et al. 2012, Langston 2013). Characterizing the spatial ecology of marine predators, as well as detecting and quantifying the potential effects of human activities on distribution and abundance, requires a flexible, spatially-explicit modeling approach that arbitrates many potentially relevant biophysical features and accommodates the analytical challenges inherent in most animal survey data.

Model-based approaches are a useful framework to generate spatially-explicit estimates of animal abundance and changes in animal distribution independent of abundance (Borchers et al. 2002, see, e.g., Certain and Bretagnolle 2008, Nur et al. 2011, Kinlan et al. 2012, Winiarski et al. 2014). Nonetheless, the spatial ecology of marine organisms typically presents several analytical challenges: the spatiotemporal distribution of marine organisms can be highly irregular which, when surveyed, produces zero-inflated (Martin et al. 2005, Cunningham and Lindenmayer 2005), overdispersed (Richards 2008) count data that may vary with biophysical features in complex, non-linear ways (Austin 2007). Moreover, an adequate modeling approach must identify a typically small suite of important covariates among many potentially correlated covariates (i.e., high dimensionality; multicollinearity) while avoiding overfitting models (i.e., describing noise), a process complicated by data from repeated surveys that regularly exhibit spatial and temporal autocorrelation (Hoeting 2009). We demonstrate a model-based approach to predict animal distribution and abundance that incorporates these considerations using aerial transect survey data of sea ducks in Nantucket Sound, Massachusetts, USA. We implemented generalized additive models (GAMs; Hastie and Tibshirani 1990, Wood 2006) and their recent extension (GAMs for location, scale, and shape; GAMLSS; Rigby and Stasinopoulos 2005) in a gradient descent boosting framework

(Hothorn et al. 2010, Mayr et al. 2012, Hofner et al. 2015b). GAMLSS methods extend GAMs to allow all parameters of the conditional response distribution to be modeled as a function of relevant covariates (Rigby and Stasinopoulos 2005); this approach provides spatially-explicit insights into the covariates associated with sea duck abundance and its variability. The boosting algorithm accommodates the inclusion of various effect types (e.g., linear, smooth, spatial, spatiotemporal, and random effects; Hofner et al. 2014a) and correlated covariates (i.e., parameter estimates are regularized), identifies the most relevant subset among potentially many covariates (i.e., variable selection), and evaluates competing representations of continuous covariates (e.g., linear vs. nonlinear effects of covariates; Kneib et al. 2009, Maloney et al. 2012).

Wintering sea ducks (Tribe: Mergini) represent an appropriate group of marine organisms, and their distribution and abundance an appropriate application, to demonstrate this model-based approach. Most sea ducks that inhabit North America migrate from high-latitude arctic and sub-arctic breeding areas to mid-latitude temperate wintering areas (Silverman et al. 2013, Bowman et al. 2015). During the ca. 10-month non-breeding (non-territorial) period of the annual cycle, some species of sea ducks can be highly mobile (Loring et al. 2013) and congregate in very large flocks (thousands of birds) at sites where prey are abundant and accessible (Flint 2013). Large-scale ocean characteristics (e.g., Flint 2013), climatic conditions (e.g., Zipkin et al. 2010), and static or persistent habitat features (e.g., bathymetry, substrate, current and frontal systems; Hyrenbach et al. 2000) typically associate with large-scale patterns in distribution of marine birds (e.g., Nur et al. 2011). At regional and local scales, however, the abundance and distribution of prey, and thus predators, can be extremely ephemeral and dynamic (Hyrenbach et al. 2000, Ortega Cisneros et al. 2011), even relating to prevailing meteorological conditions (e.g., Garthe et al. 2009). Thus, sea duck distribution and abundance is spatially and temporally dynamic yet expected to be driven by biophysical covariates (Oppel et al. 2009, Zipkin et al. 2010). Population sizes of most of North American sea duck species are believed below historic levels and with considerable uncertainty in population trajectories (Bowman et al. 2015). However, despite the anthropogenic vitiation of coastal marine ecosystems (Lotze et al. 2006, Halpern et al. 2008) and the role of sea ducks as top consumers in these ecosystems, the environmental factors associated with their distribution and abundance remain notably understudied. Furthermore, evidence from Europe's numerous offshore wind facilities portends possible negative consequences on sea duck populations of imminent offshore wind development in the northeastern United States (Larsen and Guillemette 2007, Langston 2013). Thus there exists considerable relevance in quantifying the associations between the spatiotemporal dynamics of sea ducks and the biophysical characteristics of the marine environment.

Our specific objectives were to demonstrate a flexible, model-based approach that (1) provides spatially- and temporally-explicit estimates of sea duck abundance and distribution while accommodating many of the challenges inherent in animal survey data; and (2) describes the associations between biophysical features and sea duck distribution and abundance in one of the most important migrating and wintering areas in the western Atlantic.

Methods

Study area and species.—We conducted fieldwork throughout Nantucket Sound in Massachusetts, USA (Figure 1). Our study area encompassed ca. 1,500 km^2, was relatively shallow (generally < 20 m deep), and included some of the most important sea duck wintering habitat in the western Atlantic (White et al. 2009, Silverman et al. 2013). The primary species of sea ducks found in Nantucket Sound were Common Eider (*Somateria mollissima*; hereafter eider), Black Scoter (*M. americana*), Surf Scoter (*M. perspicillata*), and White-winged Scoter (*Melanitta deglandi*), and Long-tailed Duck (*Clangula hyemalis*). Additionally, approximately 62 km^2 of Horseshoe Shoal in northwestern Nantucket Sound is fully permitted for wind energy (Figure 1), which could affect sea duck use of this important wintering area (Drewitt and Langston 2006, Langston 2013, Winiarski et al. 2014).

Aerial strip transect surveys.—During the winters (late October to mid-April) of 2003-2005, we conducted 30 (2003/2004: 13, 2004/2005: 10, 2005/2006: 7) standardized aerial strip-transect surveys (e.g., Pollock et al. 2006, Certain and Bretagnolle 2008) throughout an 1,100 km^2 study area in Nantucket Sound (Figure 1). Surveys occurred primarily from November - March (n = 27), with occasional October (n = 1) or April (n = 2) surveys. During each survey, we flew along 15 parallel (ca. 2.5 km apart), roughly north-south transects (Figure 1) using a high-wing, twin-engine aircraft (Cessna Skymaster 337) at an average altitude of 152 m and speed of 167 kmh-1 (90 kts). This altitude allowed us to identify most birds at the sea surface and reduced sea disturbance (i.e., flushing birds to another part of the study area and, thus, potential double counting). The airspeed was the slowest at which the aircraft could safely fly. Surveys occurred only on days with wind speeds ≤ 15 kts and good visibility (>15 km). Surveys occurred between 0900 to 1600 h (average duration of ~ 2.5 h) to ensure that birds had completed their post-dawn movements (e.g., Davis 1997) but had yet to initiate pre-sunset movements from feeding to roosting areas; this time window also reduced glares due to low sun angles.

On each survey flight, two observers used their unaided eyes to continuously detect individuals or flocks, identifying sea ducks to species with the aid of binoculars as needed. Observers monitored the sea surface on their side of the plane in a ca. 91 m wide transect between ca. 56 - 147 m from the plane; observers could not see beneath the plane out to 56 m on each side. The narrow strip width ensured birds were detectable and identifiable with the naked eye and allowed us to avoid situations in which ducks were too abundant or spread over too wide an area to count accurately. This last point was especially relevant because sea ducks occasionally congregated in extremely large aggregations. We thus assumed inconsequential perception bias (i.e., few missed individuals present to be counted; Marsh and Sinclair 1989) given that our flight altitude and narrow transect width (e.g., Certain and Bretagnolle 2008) within the strip transect (Buckland et al. 2012). Our transect dimensions resulted in the sampling of approximately 6% (median; 68.4 km^2) of the study area during a survey.

Observers verbally communicated duck sightings to a recorder, who entered the number of ducks, species, behavior (i.e., on the water or flying), and geographic location into a laptop computer linked to the plane's onboard GPS; we report results only from ducks observed sitting on the water. We subsequently consolidated counts for each species (eider and Long-

tailed Duck) or species group (scoter) into 2.25km^2 segments (Figure 1); this resolution (1.5 km x 1.5 km) corresponded approximately to the coarsest level of resolution of biophysical covariates (see below).
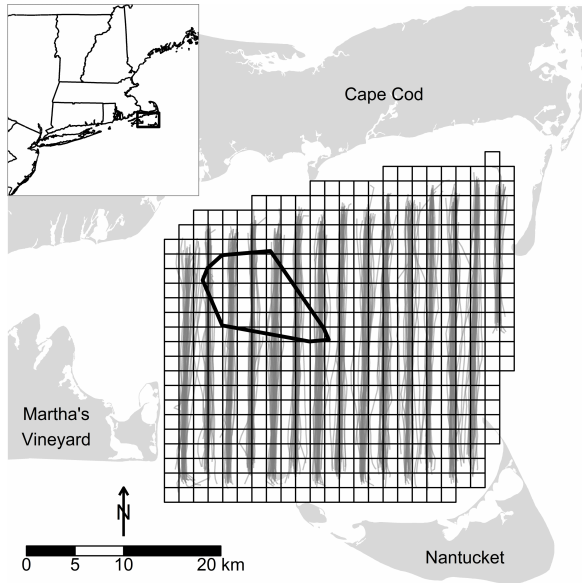

Figure 1. Actual aerial strip transect tracks (gray lines) during winter (October - April, 2003 - 2005) sea duck surveys (n = 30) in Nantucket Sound, Massachusetts, US. The grid indicates the extent of the 1100 km^2 study area and its division into 504 2.25km^2 segments. The polygon in northwest Nantucket Sound indicates the 62 km^2 area of permitted wind energy development on Horseshoe Shoal.

Modeling approach.—We related spatiotemporal variation in sea duck occupancy (i.e., probability of presence) and abundance to potentially relevant biophysical and spatiotemporal covariates. We applied a negative binomial hurdle model that separately modeled (1) the probability of occurrence of at least one individual (hereafter, the occupancy model) in a given segment using a logistic regression model and (2) the abundance of sea ducks in that segment conditional on their presence (hereafter, the count model) using a truncated negative binomial model. The prevalence of zero counts (e.g., 75% of eider segment observations) prompted the use of a hurdle model. We generated separate hurdle models for each sea duck group (i.e., eider, scoter, and Long-tailed Duck; Zipkin et al. 2010).

To model sea duck occupancy, we used a generalized additive model (GAM; Hastie and Tibshirani 1990, Wood 2006) that flexibly accommodated potential nonlinearities between covariates and sea duck occupancy. We used generalized additive models for location, scale, and shape (GAMLSS; Rigby and Stasinopoulos 2005) to fit the count models; these models accommodated potential nonlinearities between covariates and the conditional mean and overdispersion of sea duck abundance.

We implemented GAMs and GAMLSS in a gradient descent boosting framework (Mayr et al. 2012, Hofner et al. 2015b). More specifically, we used component-wise functional gradient descent boosting to fit the models. Briefly, one begins by computing the negative gradient of a user-specified loss function (typically the negative log-likelihood of the underlying model);

4

the negative gradient can be viewed as a proxy for residuals. We used the negative binomial log-likelihood as the loss function for occupancy models and the (negative) truncated negative binomial log-likelihood as the loss function for count models. In the next step, all modeling alternatives (i.e., functional forms of the covariates relative to the response, called base-learners; Hofner et al. 2014b) are fitted separately to the negative gradient and a fraction of the best fitting base-learner is added to the current model fit. The negative gradient is then reevaluated at the current model fit and the procedure iterated until for a fixed number of iterations, m_{stop}, specified by the user (see Bühlmann and Hothorn 2007 for additional details). Boosted GAMLSS models are fitted similarly, although in each iteration the negative gradient is computed separately with respect to each GAMLSS parameter (e.g., mean and overdispersion) while holding other parameters as fixed constants. Base-learners are then fitted to the resulting parameter-specific negative gradient vector and model updates are computed separately for each GAMLSS parameter as for boosted GAMs (see Mayr et al. 2012 for additional details).

As only the single best-fitting base-learner is selected in each iteration, the algorithm thus integrates intrinsic selection of the most relevant covariates and their functional form (i.e., some base-learners may never be selected). Variable selection is further fostered by stopping the algorithm prior to convergence to maximum likelihood estimates (early stopping; Mayr et al. 2012, Maloney et al. 2012). Furthermore, early stopping aims at maximizing predictive accuracy while avoiding model overfitting. We used 25-fold subsampling to determine the optimal stopping iteration for each model. Specifically, we randomly drew (without replacement) 25 samples of size $n/2$ from the original data set. We used the selected sample to estimate the model and the balance of the data in each sample to determine the out-of-bag prediction accuracy (empirical risk) measured by the negative log-likelihood of each model; the optimal stopping iteration ($\hat{m}_{\text{stop}}$) is the iteration with the lowest average empirical risk. In boosted GAMLSS models we used multi-dimensional subsampling to determine the stopping iteration for each of the GAMLSS parameters while allowing for potentially different model complexities in the parameters; a detailed explanation of this cross-validation (subsampling) scheme is given in Hofner et al. (2014a).

Despite these agreeable features, boosting methods typically produce "rich" models relying to some extent on many base-learners (Hofner et al. 2015a). Thus, we additionally applied stability selection (Meinshausen and Bühlmann 2010, Shah and Samworth 2013) to identify those base-learners, and thus covariates, most commonly selected in each model while preserving an upper bound of $\alpha \sim 0.06$ for the per-comparison error rate; Appendix 1 and Hofner et al. (2015b) provide more details.

Covariates.—We evaluated biophysical covariates expected to influence the distribution, abundance, and movements of sea ducks or, more likely, the distribution and availability of their benthic prey (e.g., mollusks and crustaceans; see Appendix 2); we did not have information related directly to the distribution of preferred prey. Biophysical covariates could be characterized as strictly spatial or temporal effects (i.e., varying only among segments or within or among winters, respectively) or as spatiotemporal effects (i.e., varying among segments and within or among winters; see Appendix 2). Additionally, we included interactions that allowed the effects of two covariates to vary over time within a given winter (see Appendix

2). We standardized (i.e., mean centered and scaled) all continuous covariates. Appendix 3 describes a function to visualize the spatial and temporal distribution of covariates in Nantucket Sound.

For each continuous covariate, we specified two base-learners: a linear base-learner and a penalized nonlinear base-learner (i.e., P-splines; Eilers and Marx 1996, Schmid and Hothorn 2008). We dummy-coded categorical covariates, resulting in K - 1 linear base-learners for a K-level covariate. To address potential spatial autocorrelation, we included a smooth surface function of the spatial coordinates of segment centers (Kneib et al. 2008); this surface comprised four base-learners – linear base-learners for the easting and northing, their linear interaction, and a penalized nonlinear tensor product P-spline (Kneib et al. 2008, 2009, Maloney et al. 2012). We also allowed this surface to vary over (interact with) time within a winter. Hofner et al. (2014b) describe base-learners in detail.

The decomposition of continuous covariates into centered linear and penalized orthogonal nonlinear base-learners allowed us to weight base-learners equally and thus allowed unbiased model choice (i.e., prevented the preferential selection of smooth base-learners; Kneib et al. 2009, Hofner et al. 2011). Specifically, we restricted each base-learner to a single degree of freedom and omitted the intercept term from each base-learner (Kneib et al. 2009, Hofner et al. 2011). Consequently, we added an explicit linear base-learner for the model intercept. The occupancy and count models had the following structure (see Appendix 2 for descriptions of abbreviated covariates):

$$
\begin{aligned}
g(\cdot) =& int + time + f(time) + SSTw + f(SSTw) + SSTm + f(SSTm) + SSTrel + \\
& f(SSTrel) + SSTrel \cdot time + f(SSTrel, time) + SBT + f(SBT) + NAOw + \\
& depth + f(depth) + depth \cdot time + f(depth, time) + d2land + f(d2land) + \\
& chla + f(chla) + cdom + f(cdom) + f(cdom, chla) + meanphi + f(meanphi) + \\
& SAR + f(SAR) + tidebmean + f(tidebmean) + tidesd + f(tidesd) + strat + \\
& f(strat) + ferry + y2004 + y2005 + xkm + ykm + xkm \cdot ykm + \\
& f(xkm, ykm) + xkm \cdot time + ykm \cdot time + xkm \cdot ykm \cdot time + \\
& f(xkm, ykm) \cdot time + obs\_window + f(obs\_window).
\end{aligned}
\tag{1}
$$

In our occupancy models, $g(\cdot)$ is $g(\pi_{\text{sea ducks}})$, the occupancy probability of a given duck species in a segment and g is the logit link. In our count models, $g(\cdot)$ took two forms within the GAMLSS framework – the (conditional) mean count of sea ducks, $g(\mu_{\text{sea ducks}})$, and the (conditional) overdispersion in sea duck counts, $g(\sigma_{\text{sea ducks}})$; g is the log link in both cases. Base-learners denoted as $f(\cdot)$ indicate the penalized nonlinear deviations from the corresponding linear base-learner (e.g., $f(time)$). The explicit intercept (int) was a necessary byproduct of our decomposition of base-learners (see above; Kneib et al. 2009, Hofner et al. 2011). Small values of *obs_window*, our measure of survey effort (see Appendix 2), in some segments impaired the estimability of models including *obs_window* as an offset; we thus included it as a covariate.

To evaluate the approximate explanatory power of our final models, we calculated a pseudo R^2 measure of the explained variation (Nagelkerke 1991, Maloney et al. 2012).

All analyses were conducted in R (Version 3.1.3; R Core Team 2014) with the add-on packages gamboostLSS (Hofner et al. 2015b), mboost (Hothorn et al. 2015), and stabs (Hofner and Hothorn 2015). The data and code for reproducing this manuscript and analyses are given as an online electronic supplement at *TBD*.

Results

As the occupancy and the hurdle model are independent, we fitted two models per sea duck species. Bootstrapped empirical risk suggested that occupancy models for all species converged on the maximum likelihood estimates (i.e., occupancy models failed to stop early; see Appendix 4). Conversely, bootstrapped empirical risks prescribed early stopping for both parameters (i.e., conditional mean and overdispersion) in all count models (see Appendix 4). Consequently, final occupancy models generally contained more covariates than did count models, particularly among stably selected covariates and their interactions. This suggests that the effects of the environmental variables on sea duck occupancy were more complex than their effect on sea duck abundance. Nonetheless, final occupancy and count (conditional mean) models included only a subset (12% to 38%) of the 48 base-learners initially specified for selection (see equation 1). This translated into the selection of only 8-10 (of 23) covariates and their interaction in occupancy models and 3-6 covariates and their interactions in conditional count (mean) models (Figure 2).

Sea duck occupancy.—The suite of stably selected covariates associated with occupancy was relatively consistent among sea duck species (Figure 2). When interpreting Figure 2, the influence of univariate effects (i.e., lines) on the response is reflected in the range of the effect over the Y-axis and, due to standardization, can be compared among species and covariates within a model. For example, monthly sea surface temperature (SST_m) associated more strongly with eider occupancy than did distance to land (*d2land*) because it spanned a larger range of the Y-axis (Figure 2). However, monthly sea surface temperature (SST_m) associated much more strongly with occupancy of Long-tailed Duck than eider for the same reason (Figure 2). Covariate interactions, illustrated with bivariate plots, are similarly comparable within a model. Only the general association (i.e., positive or negative) with the additive predictor is given for factor variables. Comparing univariate, bivariate, and categorical effects is accomplished using the detailed covariate plots for eider, scoter, and Long-tailed Duck (Appendices 5 - 7, respectively); these plots are ordered in descending order of effect size within a model, and provide values for the partial contribution to the additive predictor for covariates and their interactions.

Spatiotemporal effects (i.e., occupancy associated with the *xkm-ykm* location of segments and the change over time within winter) were the dominant explanatory feature in sea duck occupancy models, although the spatiotemporal patterns of occupancy varied considerably among species (Figure 2). Sea duck occupancy increased, but at a decreasing rate, with survey effort (*obs_window*) in a given segment (Figure 2). Among biophysical covariates, sea duck

occupancy increased at intermediate monthly sea surface temperature (SST_m), particularly (and at somewhat warmer temperatures) in Long-tailed Ducks. Sea duck occupancy also increased modestly at greater distances from land (*d2land*), although eider occupancy was lowest at intermediate distances. Scoter and especially eider occupancy was greater in areas with coarser sediments (i.e., smaller *meanphi*; Figure 2). Furthermore, eider occupancy associated negatively with chromomorphic dissolved organic material (*cdom*) and positively with sea floor surface:planimetric area (*SAR*), our measure of the topographic variability of the sea floor (Figure 2); scoter occupancy likewise related to *SAR* and *cdom*, but in the opposite direction in both cases (Figure 2). Scoter occupancy was modestly greater in deeper waters (*depth*); Long-tailed Duck occupancy was greatest in shallow waters early in the winter but in deeper waters later in the winter (Figure 2). Other effects were relatively minor and inconsistent among species.

Spatially-explicit maps of median occupancy over the course of the study (Figure 3, top row) convey the dominant influence of strictly spatial (i.e., segment *xkm-ykm* location) effects. The result is quite distinct patterns of occupancy among species despite the relative similarity of occupancy associations with biophysical covariates (Figure 2). Occupancy was typically highest for eider in northwest and southwest Nantucket Sound, in interior Nantucket Sound for scoter, and in northeast and south Nantucket Sound for Long-tailed Duck (Figure 3, top row). All species relatively avoided the western edge of the Sound northeast of Martha's Vineyard (see Figure 1). Generally, the areas of highest occupancy exhibited the lowest relative variability (Figure 3, bottom row), defined as the median absolute deviation (MAD) of occupancy relative to median occupancy within a segment (a measure analogous to the coefficient of variation).

Sea duck conditional abundance and overdispersion.—In contrast to occupancy, fewer covariates associated consistently with the abundance of sea ducks, conditional on their presence, and the suite of stably selected effects varied more markedly among sea duck species (Figure 2; see also Appendices 5 - 7). However, most selected variables fell within the suite of covariates selected in occupancy models. The primary exception to this pattern was the increase in eider and scoter abundance, conditional on presence, associated with relatively warm or cool sea surface temperatures (SST_{rel}; Figure 2). Spatial effects (*xkm-ykm*) were again the dominant explanatory feature in scoter and Long-tailed Duck, but they were not selected in the eider model (Figure 2). For a given species, biophysical covariates selected in occupancy and count models exhibited broad agreement with the exception of sediment grain size (*meanphi*) in scoter and more complex relationships between eider conditional abundance and dissolved organic material (*cdom*) and sea floor topography (*SAR*; Figure 2).

Spatially-explicit patterns of median conditional abundance (Figure 4, top row) did not necessarily reflect patterns of occupancy (Figure 3, top row). Some areas of Nantucket Sound exhibited mutually high conditional abundance and occupancy for a given species (e.g., eider in the southwest, scoter in the interior, and Long-tailed Ducks in parts of the northeast). In some cases, however, conditional abundance was low despite relatively high occupancy (e.g., eider in the northeast and Horseshoe Shoal, scoter in the northeast and southeast, and Long-tailed Duck along the northern margin). Conversely, other areas of Nantucket Sound exhibited lower occupancy but sea ducks, when present, were more abundant (e.g.,

eider along the eastern margin, and scoter and Long-tailed Duck in the southwest). As in occupancy models, sea ducks were relatively absent from the middle western margin of Nantucket Sound (i.e., northeast of Martha's Vineyard; see Figure 1). In contrast to sea duck occupancy, howver, areas of highest conditional sea duck abundance typically exhibited the highest relative variability over time (Figure 4, bottom row).

Overdispersion in conditional sea duck abundance also varied with biophysical covariates, although there was less consistency in the associated covariates among species (Figure 2; see also Appendices 5 - 7). Variability (i.e., overdispersion) in sea duck counts was heterogeneous in space (Appendix 8; Figure 8.1, top row) and time (Appendix 8; Figure 8.2, bottom row) in Nantucket Sound, particularly for eider and scoter (as indicated by the magnitude of the overdispersion parameter values).

Expected sea duck abundance.—We consolidated occupancy and conditional count models (see, e.g., Equation 6 in Zeileis et al. 2008) to estimate expected (average) sea duck abundance in the study area over the survey period (Figure 6, top row). Final models of expected sea duck abundance explained moderate amounts of variation in observed counts of eider, scoter, and Long-tailed Duck (pseudo R^2 = 0.31, 0.48, and 0.32, respectively). Conditional abundance (Figure 4) strongly influenced the spatially-explicit patterns of expected abundance. Sea duck species exhibited relatively distinct patterns of abundance in Nantucket Sound. Eider were most abundant, and relatively consistently so over time, in southwestern Nantucket Sound (Figure 5); eider were also relatively abundant in northeastern part of the sound but less consistently so as evidenced by the relatively high MAD/median abundance over time (Figure 5, bottom row). Scoter were also most abundant, occasionally extremely so, in southwestern Nantucket Sound, although this was also the area of highest relative variation in scoter abundance (Figure 5); relatively high abundances of scoter also occurred in interior Nantucket Sound (Figure 5). Long-tailed Ducks were consistently most abundant in northeastern Nantucket Sound, as well as along its southern margin (Figure 5). No species' highest abundances occurred in the permitted Nantucket Shoal area, although expected eider and scoter abundances were consistently elevated in some parts of the Shoal (west and southeast, respectively; Figure 5).

Summing the spatially explicit estimates of expected sea duck abundance (i.e., Figure 5) provides an estimate of total animal abundance in a 1.5 km x 180 m transect through all segments in the study area. We elected to compare the total number of each sea duck species observed in aerial strip transects with the corresponding total (summed) expected abundance in surveyed segments for each of the 30 aerial surveys (Figure 6). Our models tended to overestimate sea duck abundance when the actual numbers of sea ducks were relatively low, although overestimation was typically less than an order of magnitude. Additionally, scoter abundance was occasionally extreme relative to typical counts and somewhat prone to underestimation during these extreme counts. Nonetheless, the general adherance of observed and predicted abundance to a line of unit slope indicated that it may be reasonable to estimat sea duck abundance for the entire study area based on observed sea duck densities in transects.

Temporal dynamics in wintering sea ducks.—The MAD/median maps (bottom rows of Figures 3-5) emphasize something important but perhaps not obvious – that our spatially-explicit

estimates of occupancy, abundance, and overdispersion invariably change over time, either explicitly via the selection of a within- or among-winter temporal effect (*time* and *y2004/y2005*, respectively) or implicitly via the selection of biophysical covariates that change within or among winters. The temporal dynamic of the wintering sea duck system in Nantucket Sound was one of its most striking attributes. Appendix 9 illustrates the dynamic nature of occupancy and average abundance in Nantucket Sound using scoter data during the 2005 winter.
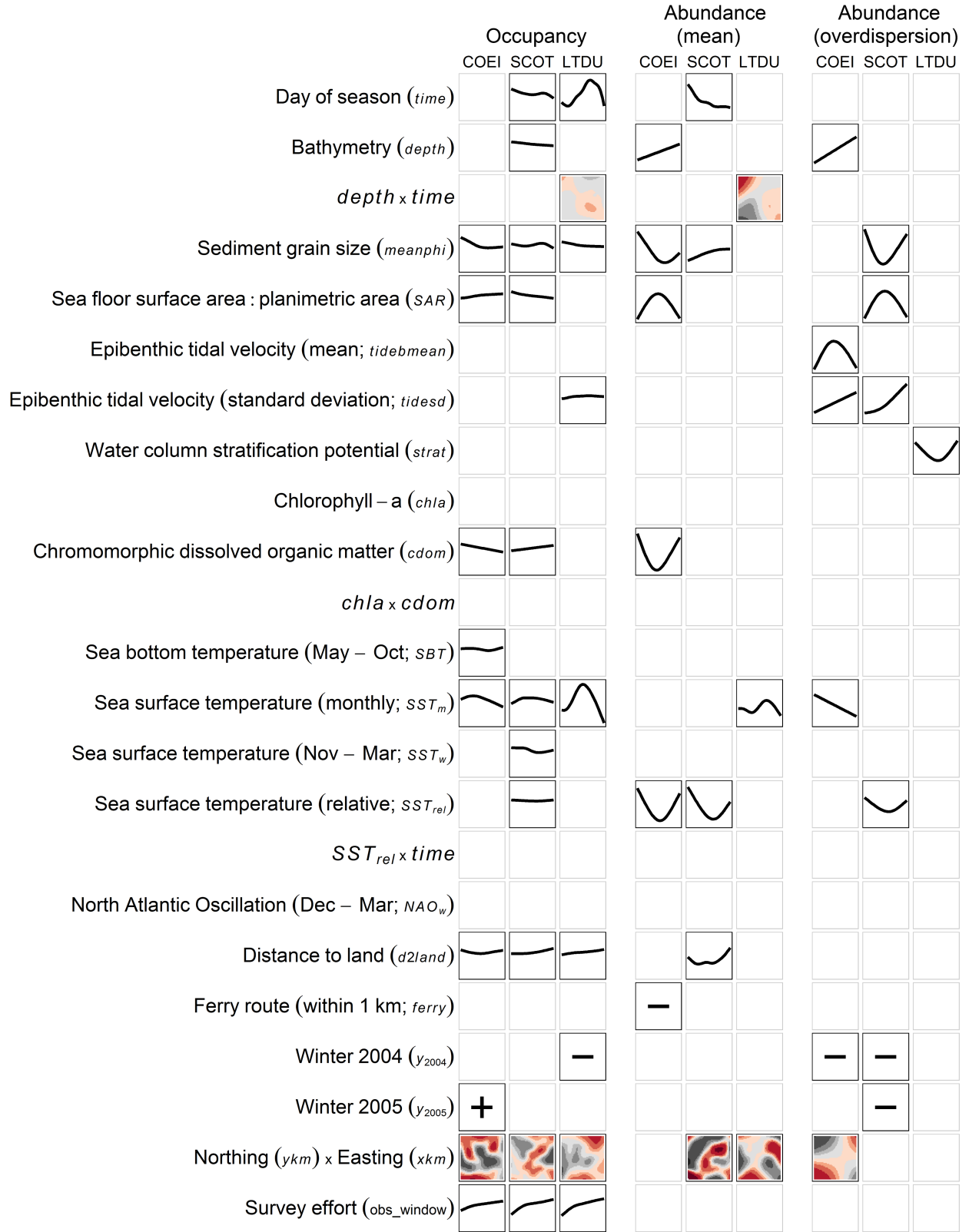
Occupancy | Abundance (mean) | Abundance (overdispersion)

COEI SCOT LTDU   COEI SCOT LTDU   COEI SCOT LTDU

Day of season ($time$)

Bathymetry ($depth$)

$depth \times time$

Sediment grain size ($meanphi$)

Sea floor surface area : planimetric area (SAR)

Epibenthic tidal velocity (mean; $tidebmean$)

Epibenthic tidal velocity (standard deviation; $tidesd$)

Water column stratification potential ($strat$)

Chlorophyll $-$ a ($chla$)

Chromomorphic dissolved organic matter ($cdom$)

$chla \times cdom$

Sea bottom temperature (May $-$ Oct; SBT)

Sea surface temperature (monthly; SST_m)

Sea surface temperature (Nov $-$ Mar; SST_w)

Sea surface temperature (relative; SST_{rel})

$SST_{rel} \times time$

North Atlantic Oscillation (Dec $-$ Mar; NAO_w)

Distance to land ($d2land$)

Ferry route (within 1 km; $ferry$)

Winter 2004 (y_{2004})

Winter 2005 (y_{2005})

Northing (ykm) $\times$ Easting (xkm)

Survey effort (obs_window)

Figure 2. Marginal functional plots for stably selected covariates in the occupancy (probability of presence) and conditional abundance (mean and overdispersion of count model) of Common Eider (COEI), scoter (SCOT), and Long-tailed Duck (LTDU) in Nantucket Sound

during winters 2003 - 2005. Plots illustrate the partial contribution to the additive predictor of a covariate holding all other covariates at their mean. Within a model, univariate plots (i.e., lines) share a Y-axis scale, enabling direct comparisons of effect sizes among covariates and species. For bivariate plots, the Y-axis and X-axis reflect the first and second variables listed in the interaction, respectively; colors indicate the direction and magnitude of the partial contribution (blacks = negative, reds = positive; darker colors = larger effect) and are likewise comparable within a model. Northing by easting effects are given only at 31 December. For factor variables, only the general association (i.e., positive or negative) with the additive predictor is given. Covariate abbreviations correspond to Equation 1.
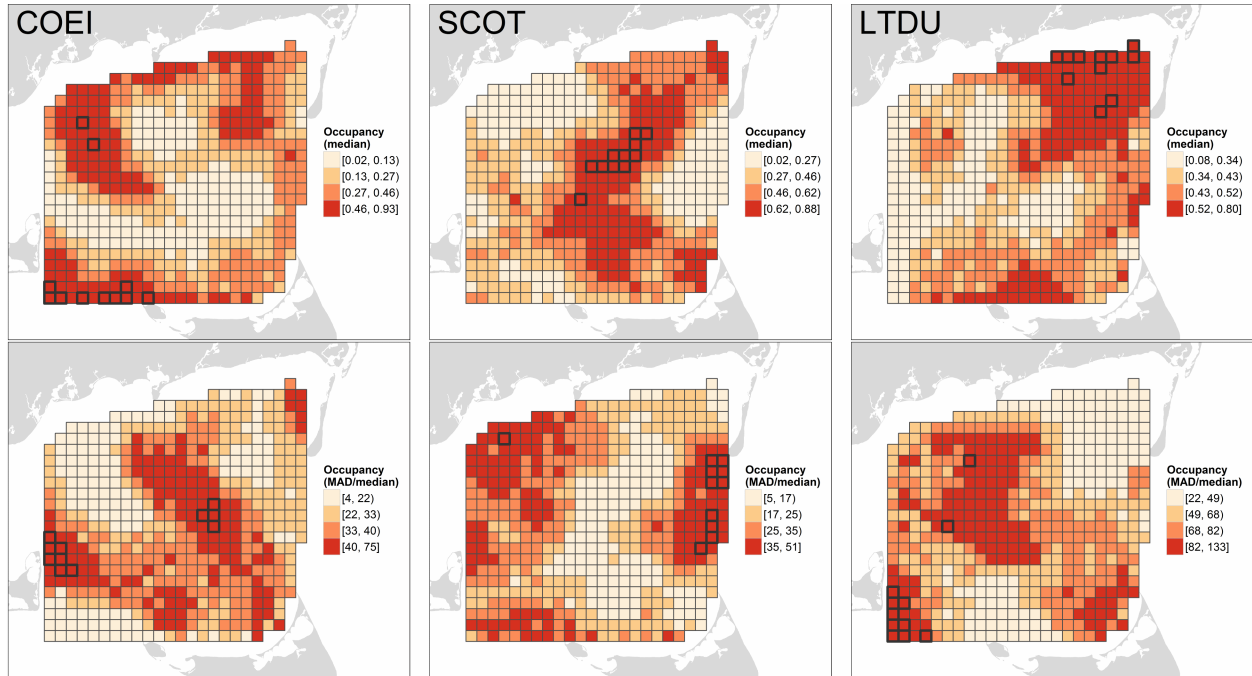
Figure 3. Median (first row) and median absolute deviation (MAD) relative to median (%; second row) of Common Eider (COEI), scoter (SCOT), and Long-tailed Duck (LTDU) occupancy in Nantucket Sound based on predictions at 10 dates evenly-spaced from 15 November through 1 April in three winters (2003 - 2005). Predicted values are categorized based on their quartiles; segments with predictions $\geq$ 98th percentile are outlined in black. Occupancy values represent the predicted probability of sea duck presence in a 1.5 km x ca. 180 m transect through the segment.
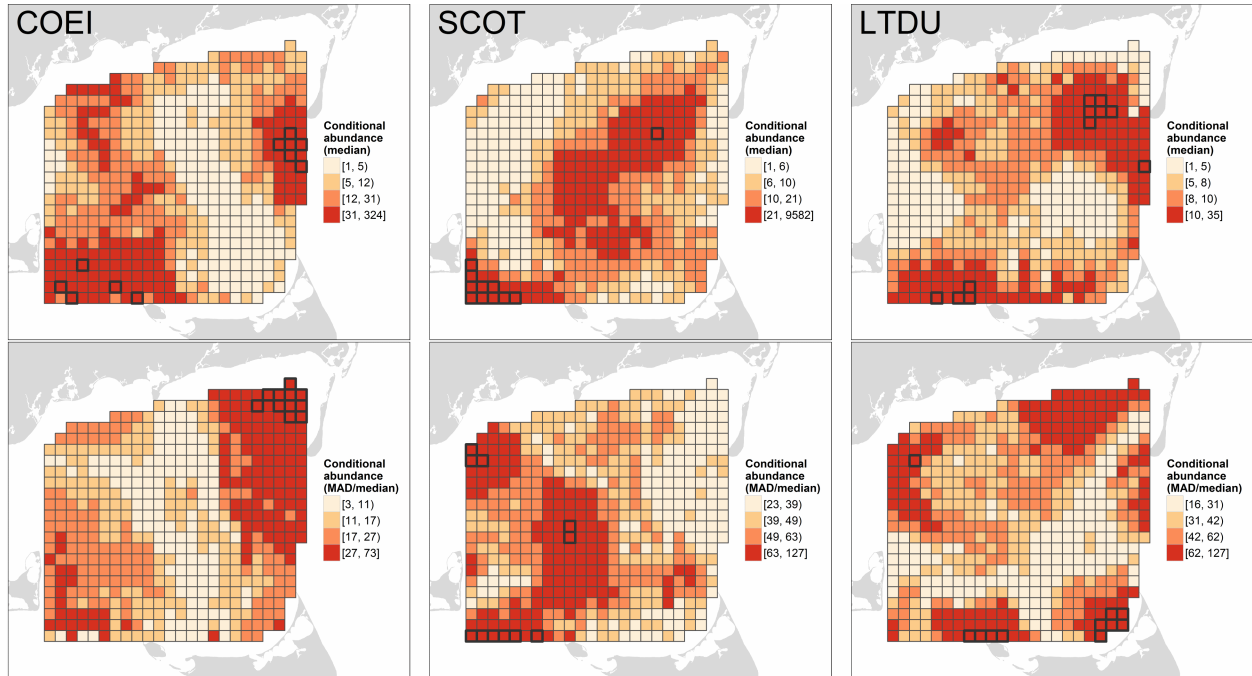
Figure 4. Median (first row) and median absolute deviation (MAD) relative to median (%; second row) of Common Eider (COEI), scoter (SCOT), and Long-tailed Duck (LTDU) conditional abundance in Nantucket Sound based on predictions at 10 dates evenly-spaced from 15 November through 1 April in three winters (2003 - 2005). Predicted values are categorized based on their quartiles; segments with predictions $\geq$ 98th percentile are outlined in black. Conditional abundance values represent the expected number of sea ducks, conditional on their presence, in a 1.5 km x ca. 180 m transect through the segment.
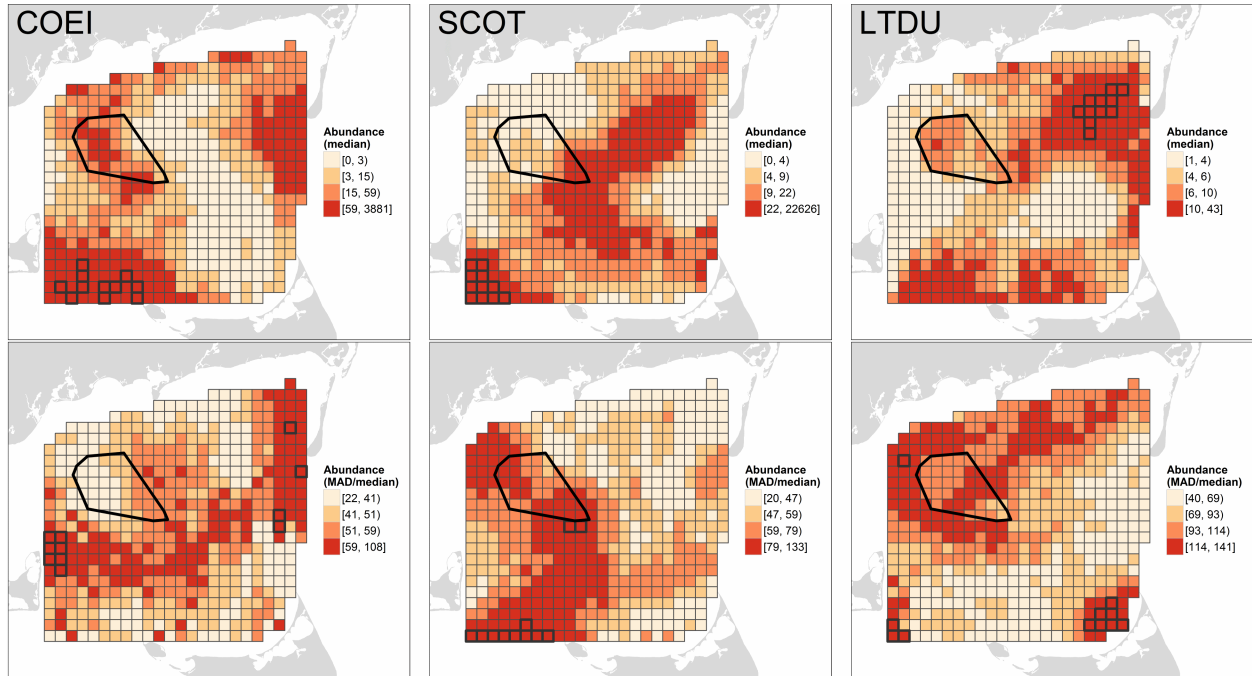
Figure 5. Median (first row) and median absolute deviation (MAD) relative to median (%; second row) of Common Eider (COEI), scoter (SCOT), and Long-tailed Duck (LTDU) average abundance in Nantucket Sound based on predictions at 10 dates evenly-spaced from 15 November through 1 April in three winters (2003 - 2005). Predicted values are categorized based on their quartiles; segments with predictions $\geq$ 98th percentile are outlined. Conditional abundance values represent the expected number of sea ducks in a 1.5 km x ca. 180 m transect through the segment. The polygon in the northwest corner of Nantucket Sound indicates the location of permitted wind energy development on Horseshoe Shoal.
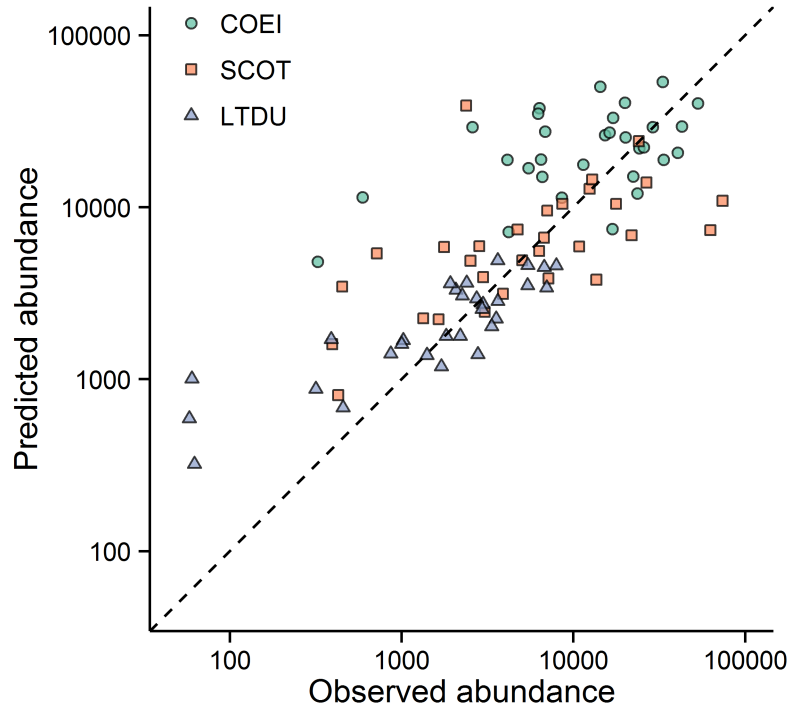
Figure 6. Observed vs. predicted total abundance of Common Eider (COEI), scoter (SCOT), and Long-tailed Duck (LTDU) during 30 aerial surveys of Nantucket Sound. The dashed line indicates a 1:1 relationship between predicted and observed abundances in surveyed segments; points below and above this line indicate predicted abundances below (underestimateion) and above the observed totals (overestimation), respectively.

Discussion (expanded outline)

Please suggest places where you feel citations are needed

We demonstrated a flexible model-based approach to evaluating the environmental associations of animal distribution and abundance based on multiyear replicated surveys. The boosted GAMLSS framework offered several useful features, including (1) the ability to model all parameters of the conditional distribution (e.g., conditional mean and overdispersion) as a function of covariates, (2) integrated variable reduction and selection among many covariates, and (3) integrated model selection via the simultaneous consideration of competing functional covariate forms (e.g., linear vs. non-linear). Additionally, this framework allowed us to incorporate smooth effects to efficiently account for spatiotemporal trends in the data poorly explained by other covariates and use stability selection to identify those covariates and their functional forms most consistently associated with animal distribution and abundance. **(Benjamin, anything you'd like to add to this?)**

Boosted GAMLSS models revealed disparate spatiotemporal distributions and abundances of sea duck species in Nantucket Sound. **More info?** Biophysical associations with sea duck occupancy were relatively consistent among species, although spatial effects dominated and resulted in disparate patterns of occupancy. Biophysical associations with sea duck conditional abundance were more species-specific but, with the exception of eider, dominated by spatial effects. **Probably need more detail on relationships here? Anything in particular to focus on?** Of note is the spatiotemporal heterogeneity in the overdispersion parameter in all species. This contrasts sharply with the typical assumption in the modeling of counts – that overdispersion is fixed – and suggests that in practice it is useful to consider that other parameters of the assumed response distribution vary with covariates (as GAMLSS models allow) and not simply the conditional mean. In other words, certain biophysical or ecological features might influence the abundance of organisms and the variability in their abundance relative to the expectations of the underlying model (negative binomial in this case).

Care must be taken when inferring the generality of associations between sea duck occupancy and abundance. It is implied that our discussion of these biophysical factors applies only at the scale used in this study, which is rather coarser than the scale of sea duck foraging. It has been shown that certain covariates may associate with marine bird abundance or behavior at certain scales and not at others (Logerwell and Hargreaves 1996). Specifically to sea ducks, for example, Zipkin et al. (2010) found the North Atlantic Oscillation important in all these species at a continental scale whereas we found no support at the scale of this study. Our estimates of the spatiotemporal abundance of sea ducks in Nantucket Sound were controlled largely by information on the conditional abundance of sea ducks. That is, while occupancy plays an important role in describing species distribution, it does not necessarily adequately reflect animal abundance, suggesting that occupancy models alone (e.g., Winiarski et al. 2014, Flanders et al. 2015) may be inadequate for assessing risk from anthropogenic disturbances or describing the distribution of marine species **(or more general)**.

The general agreement between predicted and observed sea duck counts suggested a fit

sufficient to extrapolate total sea duck abundance to segments (actually, transects of segments) unobserved in the study area, one of the primary advantages of model-based methods. Nonetheless, we elected not to extrapolate total abundance estimates of sea ducks to all Nantucket Sound for several reasons. First, sea duck abundance changed dynamically throughout a winter season (e.g., Appendix 9). This extensive temporal variability in sea duck abundance and distribution is often ignored **(I say this because I rarely see people modeling within-season effects)** as a potential source of variation and may complicate the assessment of anthropogenic effects (e.g., before/after studies) or long-term monitoring programs; **(revise to make intelligible)** for example, potentially something as trivial as inconsistency in the timing of surveys could influence pre/post-impact comparisons or obscure annual trends. Second, that fitted models contained considerable unexplained variation suggested that abundance estimates would contain considerable uncertainty, especially considering that the pseudo-R^2 measure (1) was estimated from the data used to fit the model (c.f., Maloney et al. 2012) and (2) may be misleading because much of the explanatory power in most models was attributable to marginal spatial effects (i.e., *xkm-ykm* position) and not known biophysical features. This unexplained variation and the predominance of marginal spatial effects suggested we were likely missing important variable(s) relevant to the distribution of sea ducks or we need better biophysical proxies for dynamic sea duck prey distributions or concurrent prey distribution information (e.g., Vaitkus and Bubinas 2001, Kaiser et al. 2006, Žydelis et al. 2009, Cervencl and Alvarez Fernandez 2012, Cervencl et al. 2014), although this is typically considerably more difficult to characterize at appropriate scales and does not guarantee improved predictive accuracy (Grémillet et al. 2008, Torres et al. 2008, Benoit-Bird et al. 2013). Moreover, the importance of strictly spatial covariates complicated the generalization of the model fit to all but the immediate study area. Finally, unknown and unquantified availability bias (e.g., ducks diving during survey; Marsh and Sinclair 1989) possibly resulted in underestimated sea duck occupancy (although we expect aggregating observations into 2.25 km^2 segments reduced the impact on occupancy) and almost certainly underestimated abundance. Model predictive ability and generality may be improved by incorporating mechanistic considerations (e.g., Megrey et al. 2007, Kearney and Porter 2009, Smolik et al. 2010) relevant to sea ducks and their prey.

Maybe a paragraph of the possible drawbacks that moves into extensibility? BENJAMIN: please vet this Current drawbacks of the approach include the substantial requisite computing power and time, particularly when precision estimates for the fit are to be derived. Ditto for generating confidence intervals (and thus approximate tests of significance) for functional relationships, although this is a characteristic of all penalization approaches and stability selection offers a considerably more speedy heuristic to formal tests of significance. Some difficulty in specifying a good candidate model, i.e., which effects should be included and how? **(BENJAMIN: please elaborate.** But this framework is also very extensible — the approach used herein could readily be modified to generate precision estimates of model fits and derived parameters **(BENJAMIN: we need to discuss this more)**, incorporate observation models (i.e., detection functions; distance sampling), use newer smoothing approaches (Scott-Hayward et al. 2013), accommodate double hurdle models that accommodate zero inflation and extreme counts (e.g., so called "double-hurdle"" model; E. Balderama, pers. comm.; see also Katz et al. (2005)), and incorporate simple random

effects.

Literature cited

Andrewartha, H. G., and L. C. Birch. 1986. The ecological web: More on the distribution and abundance of animals. University of Chicago Press, Chicago, Illinois, USA.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. Ecological Modelling 200:1–19.

Benoit-Bird, K. J., B. C. Battaile, S. A. Heppell, B. Hoover, D. Irons, N. Jones, K. J. Kuletz, C. A. Nordstrom, R. Paredes, R. M. Suryan, C. M. Waluk, and A. W. Trites. 2013. Prey patch patterns predict habitat use by top marine predators with diverse foraging strategies. PLoS ONE 8:e53348.

Bertrand, S., R. Joo, C. Arbulu Smet, Y. Tremblay, C. Barbraud, and H. Weimerskirch. 2012. Local depletion by a fishery can affect seabird foraging. Journal of Applied Ecology 49:1168–1177.

Borchers, D. L., S. T. Buckland, and W. Zucchini. 2002. Estimating animal abundance: Closed populations. Springer, London.

Bowman, T. D., E. D. Silverman, S. G. Gilliland, and J. B. Leirness. 2015. Status and trends of North American sea ducks: Reinforcing the need for better monitoring. Pages 1–27 *in* J.-P. L. Savard, D. V. Derksen, D. Esler, and J. M. Eadie, editors. Ecology and Conservation of North American Sea Ducks. Studies in Avian Biology. CRC Press, New York, NY, USA.

Boyd, C., R. Castillo, G. L. Hunt, A. E. Punt, G. R. VanBlaricom, H. Weimerskirch, and S. Bertrand. 2015. Predictive modelling of habitat selection by marine predators with respect to the abundance and depth distribution of pelagic prey. Journal of Animal Ecology.

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

Bühlmann, P., and T. Hothorn. 2007. Boosting algorithms: Regularization, prediction and model fitting. Statistical Science 22:477–505.

Certain, G., and V. Bretagnolle. 2008. Monitoring seabirds population in marine ecosystem: The use of strip-transect aerial surveys. Remote Sensing of Environment 112:3314–3322.

Cervencl, A., and S. Alvarez Fernandez. 2012. Winter distribution of Greater Scaup *aythya marila* in relation to available food resources. Journal of Sea Research 73:41–48.

Cervencl, A., K. Troost, E. Dijkman, M. de Jong, C. J. Smit, M. F. Leopold, and B. J. Ens. 2014. Distribution of wintering Common Eider *somateria mollissima* in the Dutch Wadden Sea in relation to available food stocks. Marine Biology 162:153–168.

Cunningham, R. B., and D. B. Lindenmayer. 2005. Modeling count data of rare species: Some statistical issues. Ecology 86:1135–1142.

Davis, J., W.E. 1997. He Nantucket oldsquaw flight: New England's greatest bird show? Bird Observer 25:16–22.

Drewitt, A. L., and R. H. W. Langston. 2006. Assessing the impacts of wind farms on birds. Ibis 148:29–42.

Eilers, P. H. C., and B. D. Marx. 1996. Flexible smoothing with B-splines and penalties (with discussion). Statistical Science 11:89–121.

Flanders, N. P., B. Gardner, K. J. Winiarski, P. W. C. Paton, T. Allison, and A. O'Connell. 2015. Using a community occupancy model to identify key seabird areas in southern New England Nicholas P. Flanders1,2*, Beth Gardner1, Kristopher J. Winiarski3, Peter W. C. Paton4, Taber Allison5, Allan F. O'Connell6. Marine Ecology Progress Series 533:in press.

Flint, P. L. 2013. Changes in size and trends of North American sea duck populations associated with North Pacific oceanic regime shifts. Marine Biology 160:59–65.

Garthe, S., N. Markones, O. Hppop, and S. Adler. 2009. Effects of hydrographic and meteorological factors on seasonal seabird abundance in the southern North Sea. Marine Ecology Progress Series 391:243–255.

Grémillet, D., S. Lewis, L. Drapeau, C. D. Van Der Lingen, J. A. Huggett, J. C. Coetzee, H. M. Verheye, F. Daunt, S. Wanless, and P. G. Ryan. 2008. Spatial match–mismatch in the Benguela upwelling zone: Should we expect chlorophyll and sea-surface temperature to predict marine predator distributions? Journal of Applied Ecology 45:610–621.

Halpern, B. S., S. Walbridge, K. A. Selkoe, C. V. Kappel, F. Micheli, C. D'Agrosa, J. F. Bruno, K. S. Casey, C. Ebert, H. E. Fox, R. Fujita, D. Heinemann, H. S. Lenihan, E. M. P. Madin, M. T. Perry, E. R. Selig, M. Spalding, R. Steneck, and R. Watson. 2008. A global map of human impact on marine ecosystems. Science 319:948–952.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. CRC Press, Boca Raton, FL, USA.

Hofner, B., and T. Hothorn. 2015. Stabs: Stability selection with error control, R package version 0.5-1.

Hofner, B., L. Boccuto, and M. Göker. 2015a. Controlling false discoveries in high-dimensional situations: Boosting with stability selection. BMC Bioinformatics in press.

Hofner, B., T. Hothorn, T. Kneib, and M. Schmid. 2011. A framework for unbiased model selection based on boosting. Journal of Computational and Graphical Statistics 20:956–971.

Hofner, B., A. Mayr, and M. Schmid. 2014a. gamboostLSS: An R package for model building and variable selection in the GAMLSS framework.

Hofner, B., A. Mayr, N. Fenske, and M. Schmid. 2015b. gamboostLSS: Boosting methods for GAMLSS models, R package version 1.1-3.

Hofner, B., A. Mayr, N. Robinzonov, and M. Schmid. 2014b. Model-based boosting in R: A hands-on tutorial using the R package mboost. Computational Statistics 29:3–35.

Hothorn, T., P. Buehlmann, T. Kneib, M. Schmid, and B. Hofner. 2015. Mboost: Model-based boosting, R package version 2.4-2.

Hothorn, T., P. Bühlmann, T. Kneib, M. Schmid, and B. Hofner. 2010. Model-based boosting 2.0. Journal of Machine Learning Research 11:2109–2113.

Hyrenbach, K. D., K. A. Forney, and P. K. Dayton. 2000. Marine protected areas and ocean basin management. Aquatic Conservation: Marine and Freshwater Ecosystems 10:437–458.

Kaiser, M. J., M. Galanidi, D. A. Showler, A. J. Elliott, R. W. G. Caldow, E. I. S. Rees, R. A. Stillman, and W. J. Sutherland. 2006. Distribution and behaviour of Common Scoter *melanitta nigra* relative to prey resources and environmental parameters. Ibis 148:110–128.

Katz, R. W., G. S. Brush, and M. B. Parlange. 2005. Statistics of extremes: Modeling ecological disturbances. Ecology 86:1124–1134.

Kearney, M., and W. Porter. 2009. Mechanistic niche modelling: Combining physiological and spatial data to predict species' ranges. Ecology Letters 12:334–350.

Kinlan, B., C. Menza, and F. Huettmann. 2012. Chapter 6: Predictive modeling of seabird distribution patterns in the New York Bight. Pages 87–224 *in* C. Menza, B. Kinlan, D. Dorfman, M. Poti, and C. Caldow, editors. A biogeographic assessment of seabirds, deep sea corals and ocean habitats of the New York Bight: Science to support offshore spatial planning. NOAA Technical Memorandum NOS NCCOS 141. NOAA, Silver Spring, MD.

Kirk, M., D. Esler, S. A. Iverson, and W. S. Boyd. 2008. Movements of wintering Surf Scoters: Predator responses to different prey landscapes. Oecologia 155:859–867.

Kneib, T., T. Hothorn, and G. Tutz. 2009. Variable selection and model choice in geoadditive regression models. Biometrics 65:626–634.

Kneib, T., J. Müller, and T. Hothorn. 2008. Spatial smoothing techniques for the assessment of habitat suitability. Environmental and Ecological Statistics 15:343–364.

Langston, R. H. W. 2013. Birds and wind projects across the pond: A UK perspective. Wildlife Society Bulletin 37:5–18.

Larsen, J. K., and M. Guillemette. 2007. Effects of wind turbines on flight behaviour of wintering common eiders: Implications for habitat use and collision risk. Journal of Applied Ecology 44:516–522.

Logerwell, E. A., and N. B. Hargreaves. 1996. The distribution of sea birds relative to their fish prey off Vancouver Island: Opposing results at large and small spatial scales. Fisheries Oceanography 5:163–175.

Loring, P. H., P. W. C. Paton, S. R. McWilliams, R. A. McKinney, and C. A. Oviatt. 2013. Densities of wintering scoters in relation to benthic prey assemblages in a North Atlantic estuary. Waterbirds 36:144–155.

Lotze, H. K., H. S. Lenihan, B. J. Bourque, R. H. Bradbury, R. G. Cooke, M. C. Kay, S. M. Kidwell, M. X. Kirby, C. H. Peterson, and J. B. C. Jackson. 2006. Depletion, degradation, and recovery potential of estuaries and coastal seas. Science 312:1806–1809.

MacArthur, R. H. 1972. Geographical ecology: Patterns in the distribution of species. Joanna Cotler Books, New York, NY, USA.

Maloney, K. O., M. Schmid, and D. E. Weller. 2012. Applying additive modelling and gradient boosting to assess the effects of watershed and reach characteristics on riverine assemblages. Methods in Ecology and Evolution 3:116–128.

Marsh, H., and D. F. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. The Journal of Wildlife Management 53:1017–1024.

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. Ecology Letters 8:1235–1246.

Maurer, B. A. 1999. Untangling ecological complexity: The macroscopic perspective. University Of Chicago Press, Chicago, Illinois, USA.

Mayr, A., N. Fenske, B. Hofner, T. Kneib, and M. Schmid. 2012. Generalized additive models for location, scale and shape for high-dimensional data - a flexible approach based on boosting. Journal of the Royal Statistical Society: Series C (Applied Statistics) 61:403–427.

Megrey, B. A., K. A. Rose, R. A. Klumb, D. E. Hay, F. E. Werner, D. L. Eslinger, and S. L. Smith. 2007. A bioenergetics-based population dynamics model of Pacific herring (*clupea harengus pallasi*) coupled to a lower trophic level nutrient–phytoplankton–zooplankton model: Description, calibration, and sensitivity analysis. Ecological Modelling 202:144–164.

Meinshausen, N., and P. Bühlmann. 2010. Stability selection (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72:417–473.

Nagelkerke, N. J. D. 1991. A note on a general definition of the coefficient of determination. Biometrika 78:691–692.

Nur, N., J. Jahncke, M. P. Herzog, J. Howar, K. D. Hyrenbach, J. E. Zamon, D. G. Ainley, J. A. Wiens, K. Morgan, L. T. Ballance, and D. Stralberg. 2011. Where the wild things are: Predicting hotspots of seabird aggregations in the California Current System. Ecological Applications 21:2241–2257.

Oppel, S., A. N. Powell, and D. L. Dickson. 2009. Using an algorithmic model to reveal individually variable movement decisions in a wintering sea duck. Journal of Animal Ecology 78:524–531.

Ortega Cisneros, K., A. J. Smit, J. Laudien, and D. S. Schoeman. 2011. Complex, dynamic combination of physical, chemical and nutritional variables controls spatio-temporal variation

of sandy beach community structure. PLoS ONE 6:e23724.

Pollock, K. H., H. D. Marsh, I. R. Lawler, and M. W. Alldredge. 2006. Estimating animal abundance in heterogeneous environments: An application to aerial surveys for dugongs. Journal of Wildlife Management 70:255–262.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Richards, S. A. 2008. Dealing with overdispersed count data in applied ecology. Journal of Applied Ecology 45:218–227.

Rigby, R. A., and D. M. Stasinopoulos. 2005. Generalized additive models for location, scale and shape (with discussion). Journal of the Royal Statistical Society: Series C (Applied Statistics) 54:507–554.

Robinson, L. M., J. Elith, A. J. Hobday, R. G. Pearson, B. E. Kendall, H. P. Possingham, and A. J. Richardson. 2011. Pushing the limits in marine species distribution modelling: Lessons from the land present challenges and opportunities. Global Ecology and Biogeography 20:789–802.

Rose, G. A., and W. C. Leggett. 1990. The importance of scale to predator-prey spatial correlations: An example of Atlantic fishes. Ecology 71:33–43.

Schmid, M., and T. Hothorn. 2008. Boosting additive models using component-wise P-splines. Computational Statistics & Data Analysis 53:298–311.

Scott-Hayward, L. a. S., M. L. Mackenzie, C. R. Donovan, C. G. Walker, and E. Ashe. 2013. Complex region spatial smoother (CReSS). Journal of Computational and Graphical Statistics 23:340–360.

Shah, R. D., and R. J. Samworth. 2013. Variable selection with error control: Another look at stability selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75:55–80.

Silverman, E. D., D. T. Saalfeld, J. B. Leirness, and M. D. Koneff. 2013. Wintering sea duck distribution along the Atlantic Coast of the United States. Journal of Fish and Wildlife Management 4:178–198.

Smolik, M., S. Dullinger, F. Essl, I. Kleinbauer, M. Leitner, J. Peterseil, L.-M. Stadler, and G. Vogl. 2010. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. Journal of Biogeography 37:411–422.

Torres, L. G., A. J. Read, and P. Halpin. 2008. Fine-scale habitat modeling of a top marine predator: Do prey data improve predictive capacity. Ecological Applications 18:1702–1717.

Vaitkus, G., and A. Bubinas. 2001. Modelling of sea duck spatial distribution in relation to food resources in Lithuanian offshore waters under the gradient of winter climatic conditions. Acta Zoologica Lituanica 11:288–302.

White, T. P., R. R. Veit, and M. C. Perry. 2009. Feeding ecology of Long-Tailed Ducks *clangula hyemalis* wintering on the Nantucket Shoals. Waterbirds 32:293–299.

Winiarski, K. J., D. L. Miller, P. W. Paton, and S. R. McWilliams. 2014. A spatial conservation prioritization approach for protecting marine birds given proposed offshore wind energy development. Biological Conservation 169:79–88.

Wood, S. N. 2006. Generalized additive models: An introduction with R. Chapman; Hall/CRC, Boca Raton, FL, USA.

Zeileis, A., C. Kleiber, and S. Jackman. 2008. Regression models for count data in R. Journal of Statistical Software 27:1–25.

Zipkin, E. F., B. Gardner, A. T. Gilbert, A. F. O. Jr, J. A. Royle, and E. D. Silverman. 2010. Distribution patterns of wintering sea ducks in relation to the North Atlantic Oscillation and local environmental characteristics. Oecologia 163:893–902.

Žydelis, R., D. Esler, M. Kirk, and W. Sean Boyd. 2009. Effects of off-bottom shellfish aquaculture on winter habitat use by molluscivorous sea ducks. Aquatic Conservation: Marine and Freshwater Ecosystems 19:34–42.