EVALUATING SURVEY DURATION WITH COMMUNITY ATTRIBUTES OF MARINE BIRDS IN THE SOUTH ATLANTIC BIGHT

J. CHRISTOPHER HANEY1 & DAVID S. LEE2†

1Defenders of Wildlife, 1130 17th Street, NW, Washington, DC 20036 (chaney@defenders.org)
2Tortoise Reserve, P.O. Box 7082, White Lake, NC 2837
†Deceased

SUMMARY


The expense of conducting dedicated marine bird surveys creates incentives to optimize sampling effort. We appraised spatiotemporal effort in two shipboard surveys off the southeastern coast of the United States using analyses of seabird community structure. By applying randomization routines from both data-analytic (curve-fitting) and sampling-theoretic methods (bootstrap, jackknife, coverage), we assessed sampling adequacy for estimating numerical and biomass dominance, species accumulation rates and species richness. Maximum-likelihood estimators indicated that each survey had been carried out long enough for cumulative species richness ($S_{obs}$) to reach an apparent asymptote. Data stratification by season did not reduce the asymptotic-based estimates of minimum survey effort overall. Three estimators of total species richness ($\delta_{max}$) were equal to or less than values actually observed ($S_{obs} = 53$ species in each survey). Five other estimators of $\delta_{max}$ exceeded $S_{obs}$ by 5%–18%, suggesting few species remained undetected by the time our surveys ended. Indeed, after >600 additional observation days, only five more offshore species were detected (all extralimitally rare), a tally that fell within error terms for estimators noted above. The two avifaunal surveys examined here were two to six times longer than required (36–65 days) to detect community dominants (species comprising ≥95% of numerical abundance and biomass). We conclude that sampling effort as appraised by community attributes was minimally adequate in both surveys. Measures of saturation in community structure can be applied while inventories are in progress to complement analyses of sample size based on statistical power, and may help minimize costs associated with over-sampling in dedicated surveys for marine birds.

Key words: survey adequacy, sampling effort, South Atlantic Bight, western Atlantic Ocean, seabird community structure, species accumulation rate, species richness

INTRODUCTION


Regardless of purpose, the challenging logistics require critical evaluation of the methodology used in marine bird surveys (Briggs et al. 1985). Substantial time, personnel and costs are devoted to these survey programs, some of which span decades (Ainley et al. 1996, Veit et al. 1996, Ribic et al. 1997). Fees levied for dedicated ship use can exceed $5,000–$15,000 per day (Dufour 2012). Given the high financial costs, surveys are often done from ships of opportunity, which risks compromising the sample representation. Considerations of survey design thus beg scrutiny: How much survey effort is adequate? What is the minimum temporal and/or spatial coverage necessary to achieve optimal efficiency in offshore surveys? What consequences arise from using different appraisal methods for evaluating this adequacy?

Answers to these questions depend on the intent of the surveys and the nature of the particular research issue addressed (Ainley et al. 2012). Seabird responses to temporal or spatial changes in their environment might be evaluated at the focal individual (Votier et al. 2010), population (Oedekoven et al. 2001) or multi-species level (Ballance et al. 1997). Therefore, different response variables may be required to address spatial pattern (Fritz et al. 2003, Johnston et al. 2014), numerical abundance (Maclean et al. 2013) or community structure (Wiens et al. 1996).

In preliminary stages, survey programs for marine birds are often exploratory, intended to describe community composition for estimating feeding requirements (Schneider & Hunt 1982) or to assess species’ vulnerability to oil spills (Seip et al. 1991, Begg et al. 1997). Species richness and other community-level metrics may also be used as the primary response variable in some environmental impact assessments (Wiens et al. 1996). Confidence that most species have been detected in a survey is a fundamental prerequisite to any community-level analysis (Cam et al. 2002).

Here we use attributes of seabird community structure to retroactively examine minimal survey adequacy for two shipboard
STUDY AREA AND METHODS

Marine bird surveys were conducted in the South Atlantic Bight off the southeastern United States (Fig. 1). The survey in the northern portion of the bight off North Carolina was carried out mostly in an area of steep, converging bathymetric contours adjacent to the Outer Banks (OB). The survey in the south-central portion of the bight was conducted over the broad and shallow continental shelf in the large indentation of the coastline termed the Georgia Embayment (GE).

Regional oceanography throughout this region is dominated by the Gulf Stream, which flows over the Blake Plateau just seaward of and adjacent to the continental shelf break (Olson et al. 1983). Along-shore current fluctuations promote episodically high but localized productivity in the offshore (Yoder et al. 1981, Atkinson & Targett 1983). Other distinguishing features of the South Atlantic Bight include high, turbid river discharges in the Georgia Embayment (Blanton & Atkinson 1983), greater temporal changes in offshore than in coastal winds (Weisberg & Pietrafesa 1983), and strong seasonal influence of atmospheric cooling on continental shelf water masses (Atkinson et al. 1983). These and other factors lead to substantial patchiness, i.e. high local variation in seabird distributions across the region (e.g. Lee 1995, Haney 1986, Hass 1997). Spatial dispersion of both surveys examined here was sufficient to represent most (if not all) of this environmental variation, from middle shelf marine regions proceeding seaward (see Results).

Survey techniques

Standardized counts for seabirds were available from 223 days of dedicated shipboard surveys using continuous counts off the Outer Banks (Moser & Lee 2012), and 143 days of opportunistic shipboard surveys at intervals using 300 m strip transects in the Georgia Embayment (Haney 1986). Data analyzed here were collected entirely in the open ocean (≤200 km from shore), outside sounds, bays, and inlets. OB surveys were carried out from 1975 to 1989 and GE surveys from 1982 to 1985. Surveys were conducted during all months of the year, but effort was not allocated proportionately across seasons (Table 1).

OB surveys recorded seabird locations throughout the day, and these were summed as daily totals within plotted transects of varying lengths and periods. GE surveys originally recorded bird count data within 15 min periods (Haney 1986). Seabird numbers from the GE survey were therefore re-binned into daily totals to achieve grain-size compatibility (Hortal et al. 2006) between the two surveys. Non-quantitative information on species composition was also gathered during each survey. Although this extra information was not compatible with most of the numerical analyses examined here, we report the additional species detected in the joint total of all species observed in both surveys (these totals are distinguished within and between surveys in Appendix 1, available on the website).

### Table 1

<table>
<thead>
<tr>
<th>Season</th>
<th>Outer Banks</th>
<th>Georgia Embayment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of days</td>
<td>Percent of total</td>
</tr>
<tr>
<td>Winter (Nov–Mar)</td>
<td>43</td>
<td>19.3</td>
</tr>
<tr>
<td>Transitional (Apr–May, Sep–Oct)</td>
<td>72</td>
<td>32.3</td>
</tr>
<tr>
<td>Summer (Jun–Aug)</td>
<td>108</td>
<td>48.4</td>
</tr>
<tr>
<td>Total</td>
<td>223</td>
<td>100.0</td>
</tr>
</tbody>
</table>

* Seasons are differentiated on the basis of regional marine climate (Atkinson et al. 1983).
Data treatment

Survey data were organized into two large matrices (one for each region) consisting of individual dates, species and number of individuals per species. Each matrix was initially configured with the survey dates in chronological order. Matrix manipulations were the bases for three types of analyses, most of which used various randomization routines to evaluate simultaneously the number of species and either species’ abundances or presence/absence. Because statistical estimation of the number of species is notoriously resistant to a single solution (Foggo et al. 2003, Hortal et al. 2006, Reichert et al. 2010), we screened both data-analytic and sampling-theoretic approaches (Bunge & Fitzpatrick 1993). Performance among estimators was evaluated by conducting a regression analysis of the observed with the predicted accumulation curve, then comparing the amount of variance explained among the different model types. Standard notation is used throughout the text for species richness estimators (Chazdon et al. 1998).

Data-analytic methods. Data-analytic methods included both descriptive and diagnostic techniques. A general property of counting species is that more new species are detected during the initial stages of surveys, so relationships between species number and effort were examined first as curvilinear functions (Tjørve 2003). To remove influence of sample order in shaping the actual curves, we used an iterative routine to calculate average species richness (S) over the entire curve. Curve smoothing was accomplished by random permutations of the one-day survey samples with a procedure set to n = 1000 runs (PISCES 1.2; Henderson & Seaby 1997). Curves depicting number of species in actual survey order were depicted with n = 1 run.

Asymptotic functions can be suitable for some species accumulation curves (Miller & Wiepert 1989, soberon & llorente 1993, Keating et al. 1998). If an asymptote for total species (S≤) can be estimated, the corresponding level of survey effort can be interpolated from the curve’s horizontal axis. A parametric maximum-likelihood estimator (Raaijmakers 1987) was used to calculate whether an asymptote was evident. Sampling was assumed complete if the asymptotic estimate was equal to or less than the observed value of S (Colwell & Coddington 1994). This diagnostic routine was applied incrementally to larger combinations of randomly shuffled one-day survey samples (n = 1 000 runs each) until this “stopping rule” (sense Scharff et al. 2003) indicated that sampling was adequate.

Accurate estimation of Smax may be compromised unless accumulation curves are derived from a homogeneous species pool (Hortal et al. 2006, Reichert et al. 2010). As a prelude to other estimations of Smax (see below), we compared mean randomized accumulation curves (n = 1000 runs) to a hypothetical curve expected if all individual birds recorded over all survey samples had been assigned randomly to individual samples (Colwell & Coddington 1994). If the expected curve (using the function described by Coleman et al. 1982) lies above the observed species accumulation curve, the samples are of diverse origin (e.g. Flather 1996). If heterogeneity was detected, subsamples of species richness were re-binned and then analyzed separately by season to examine whether survey efficiency could be improved by this stratification. Improvement was indicated if asymptotes were identified for the separate accumulation curves, and if fewer cumulative survey-days (relative to the cumulative total) would have been required by this alternative allocation of survey days.

Sampling-theoretic methods. Sampling-theoretic methods included estimation of both maximum species richness as well as functions that predicted the curve (or rate) of species accumulation. Survey adequacy was also evaluated by comparing differences (if any) between the total species richness observed and the maximum species richness that was estimated (Smax). Because no single model has been found to be universally suitable for species richness data (Bunge & Fitzpatrick 1993), a variety of non-parametric models were tested. Models included bootstrap, jackknife, and coverage (the latter were both incidence-based and abundance-based [ACE]; Chazdon et al. 1998).

Two different Chao estimators of Smax were used. An abundance-based estimator and variance for species richness (Chao 1984), respectively, are given by the following equations:

\[ S_{\text{Chao}} = S_{\text{obs}} + (F_1^2 / 2F_2), \]
\[ \text{var}(S_{\text{Chao}}) = F_2(4G^2 / 4 + G^3 + G^2 / 2), \]

Where \( S_{\text{obs}} \) is the total number of species observed in all samples pooled, \( F_1 \) is the number of species that have exactly 1 individuals when all samples are pooled, and \( G = F_1 / F_2 \). An incidence-based estimator and variance of species richness (Chao 1987), respectively, are given by the following equations:

\[ S_{\text{Chao2}} = S_{\text{obs}} + (Q^2 / 2Q_2), \]
\[ \text{var}(S_{\text{Chao2}}) = F_2(4G^2 / 4 + G^3 + G^2 / 2), \]

In this estimator, \( Q_1 \) is the number of species that occur in exactly one sample (uniques), and \( Q_2 \) is the number of species that occur in exactly two samples (duplicates).

Two additional coverage estimators were used for species richness based on abundance (Chao & Lee 1992). The general form of these ACE estimators is given by the following equation:

\[ S_{\text{ACE}} = S_{\text{obs}} + S_{\text{rare}} + F_1G_{\text{ACE}}^2 + F_rG_{\text{ACE}}^2, \]

Where \( S_{\text{obs}} \) is the number of abundant species (each with more than 10 individuals) when all samples are pooled, \( S_{\text{rare}} \) is the number of rare species (each with 10 or fewer individuals) when all samples are pooled.

Therefore:

\[ C_{\text{ACE}} = 1 - (F_1 / N_{\text{rar}},) \]
\[ N_{\text{rar}} = \sum_{i=1}^{10} F_i, \]
\[ G_{\text{ACE}} = \max(S_{\text{max}}, \sum_{i=1}^{10} (i-1)F_i / (N_{\text{rar}})), \]

as computed in Henderson & Seaby (1997).

The bootstrap estimator of Smax used here was that described by Smith & van Belle (1984). Jackknife estimators of species richness included both the first-order (Burnham & Overton 1978, Heltsh e & Forrester 1983) and the second-order, robust-limiting forms (Burnham & Overton 1979, Palmer 1991). Non-parametric estimates of Smax were computed with n = 143–1000 runs of randomly shuffled one-day samples from each survey. Computational routines for these and related estimators are available from Colwell (2014), Henderson & Seaby (1997), and Hines et al. (1999).

Community dominance. Products of mean body mass (Appendix 1, available on the website) and the number of individuals for each species recorded were used to calculate the total abundance and biomass of the seabird community pooled over all samples collected across the entire duration in each of the two surveys. This procedure
formed the basis of two retrospective analyses aimed at measuring saturation in community dominance.

In the first analysis, we computed the minimum number of survey days required to detect only those species that composed 95% of the total abundance and biomass. This approach involved ranking species by their importance values (numbers or contributions to total biomass) until the cumulative total came to 95%, then identifying the shortest duration that would have detected these same species, given the actual start date and sampling sequence used in each survey.

In the second analysis, Monte Carlo trials were employed with different start dates assigned randomly to hypothetical surveys. Using subsamples of species richness from the real surveys, hypothetical surveys were “conducted” for the same duration identified in the analysis described above. This resampling approach therefore simulated the extent to which alternative start dates (as surveys can only progress forward) might have influenced identification of the community dominants.

**Post-survey evaluation of effort and cost**

To appraise whether our surveys were carried out long enough using compositional criteria, we compared the species richness totals we observed (and their statistical estimators) to the cumulative species totals detected later as the consequence of two decades of additional offshore marine bird observations conducted in the same region. Between 1988 and the present, one of us (D.S.L.), as well as commercial operators, ran various one-day excursions in the South Atlantic Bight. The longest-running excursions have been those off North Carolina by Brian Patteson (http://seabirding.com/), most of which were conducted off the Outer Banks coastline (Fig. 1).

We assessed the species composition recorded in these additional observations (n = 619 survey days from 1988 to 2013) for any offshore species that had not been recorded during our earlier surveys. We interpolated the number of additional survey days that were required to detect each additional species. Then we derived a range of equivalent costs for dedicated ship time for this added survey effort based on the average 2012 daily charge rates for coastal- ($5000–$10000) and regional-capable vessels ($10000–$15000) that are suitable for outer continental shelf surveying, and operated expressly for such purposes under the Atlantic fleet of the University National Oceanographic Laboratory System (Dufour 2012).

### Table 2

<table>
<thead>
<tr>
<th>Technique</th>
<th>Outer Banks</th>
<th>Georgia Embayment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Species richness, ( S_{\text{max}} \pm SE )</strong></td>
<td><strong>Model fit ((r^2)^a)</strong></td>
<td><strong>Species richness, ( S_{\text{max}} \pm SE )</strong></td>
</tr>
<tr>
<td>Actual survey</td>
<td>53.0 ± 0.0</td>
<td>53.0 ± 0.0</td>
</tr>
<tr>
<td>Maximum-likelihood Coverage 1</td>
<td>51.1 ± 0.0</td>
<td>0.787</td>
</tr>
<tr>
<td>Coverage 1 (Chao &amp; Lee 1992)</td>
<td>53.0 ± 0.0</td>
<td>0.998</td>
</tr>
<tr>
<td>Coverage 2 (Chao &amp; Lee 1992)</td>
<td>53.0 ± 0.0</td>
<td>0.998</td>
</tr>
<tr>
<td>Chao1 (1984: abundance-based)</td>
<td>59.3 ± 0.4</td>
<td>0.996</td>
</tr>
<tr>
<td>Chao2 (1987: incidence-based)</td>
<td>56.0 ± 0.1</td>
<td>0.906</td>
</tr>
<tr>
<td>First-order jackknife</td>
<td>59.0 ± 0.2</td>
<td>0.990</td>
</tr>
<tr>
<td>Second-order jackknife</td>
<td>58.0 ± 3.1</td>
<td>–</td>
</tr>
<tr>
<td>Bootstrap</td>
<td>56.2 ± 0.0</td>
<td>0.984</td>
</tr>
</tbody>
</table>

*Values for model fit for each estimation technique are based on least-squares regressions of the observed on the estimated species accumulation curves, as generated from selected sampling-theoretic techniques.*
RESULTS

Species richness

Both surveys recorded a total of 53 seabird species, but this number was achieved in 80 fewer sampling days in the GE (Fig. 2). Maximum-likelihood estimators identified asymptotes for each of the two survey’s seasonally combined daily samples. The asymptote for the OB survey corresponded to 46 species and 88 days of surveys. The asymptote for the GE survey corresponded to 52 species and 135 days of surveys. Distinct curve shapes of each asymptote, however, indicated that the underlying statistical functions for the maximum-likelihood estimators were different (Fig. 3). Of estimators examined, the maximum-likelihood estimator provided the poorest fit to the entire observed accumulation curves (Table 2), tending to greatly overshoot $S$ during the earliest stages of these surveys (Fig. 3).

Although species accumulation curves were asymptotic for each survey, additional diagnostic tests indicated that the daily samples collected during each survey were heterogeneous (Fig. 4). Differences between the observed and homogeneous community curves ranged from about 5 to 17 species, and 5 to 18 species, during the first 20 days of the OB and GE surveys, respectively. Despite this heterogeneity, stratifying samples by season did not materially reduce the asymptote-based estimates of minimum survey effort. Except for the OB during summer, no asymptotes were identified for any of the other single-season species accumulation curves (Table 3). Even so, values for $\hat{S}_{\text{max}}$ were within a single species of the observed richness for the two remaining seasons in the OB area. Observed and estimated values of $\hat{S}_{\text{max}}$ in the GE were closest in summer ($<2$ species), despite relatively few days of sampling during that season (Table 1). Asymptotic estimators for $\hat{S}_{\text{max}}$ during winter and transitional seasons in the GE were at least 3 species greater, however, than the values of species richness observed.

Three of eight estimators for species richness were equal to or less than the species richness actually observed in the surveys (Table 2). Coverage estimators gave highly accurate point estimates for $\hat{S}_{\text{max}}$ in each survey, and explained more than 99% of the variance in species richness along the entire accumulation curves. The bootstrapped estimator for $\hat{S}_{\text{max}}$ gave the next most accurate estimate of total species richness in each survey. The bootstrap also explained high variance (>98%) in species richness along the entire accumulation curve. In general, non-parametric estimates of species richness in the GE were slightly higher than comparable values from the OB, but low precision in these estimators casts doubt on attaching particular biological significance to the differences (Table 2).

Species dominance

As few as 18 species contributed ≥95% of all individual seabirds ultimately recorded during the GE survey (Table 4). Based on the

Fig. 3. Minimum survey effort interpolated from an asymptotic estimate of total species richness, $\hat{S}_{\text{max}}$, for each of two surveys of marine birds off the coast of the southeastern United States. Asymptotes were calculated with a maximum-likelihood estimator (Raaijmakers 1987).

Fig. 4. Test for community heterogeneity in the species pool for two surveys of marine birds conducted off the coast of the southeastern United States. Curves for a homogeneous community are based on the method described by Coleman et al. (1982). The observed species accumulation curve was derived from $n = 1000$ randomized iterations of one-day survey samples.

actual chronological sequence of the OB survey and the shortest possible detection time for these 18 dominants, a minimum of 66 consecutive survey days, or 46% of this survey program’s total duration, would have been required to record dominant species. With a random start date, a hypothetical OB survey with duration 66 days had a 95% likelihood of recording all 18 dominants (n = 20 trials). As few as 16 species contributed ≥95% of all seabird biomass detected in the GE survey. Based on actual chronological sequence and the shortest possible detection time to observe these dominants, a minimum of 65 consecutive survey days, or 45% of this survey program’s total duration, would have been required to detect these species. With random start dates, a hypothetical GE survey with duration 65 days had a 95% likelihood of recording these same 16 dominants (n = 20 trials).

As few as 14 species contributed ≥95% of all individual seabirds ultimately recorded during the OB survey (Table 4). Based on actual chronological sequence of the OB survey and the shortest possible detection time for these 14 dominants, a minimum of 43 consecutive survey days, or 19% of this survey program’s total duration, would have been required to record these species. With a random start date, a hypothetical OB survey of duration 43 days had only a 45% likelihood of recording all 14 dominants (n = 20 trials). A hypothetical OB survey with a duration equivalent to that used in the GE analysis above (66 days), however, would have a 60% likelihood of recording these 14 dominants (n = 20 trials). As few as 11 species contributed ≥95% of all seabird biomass in the OB survey. Based on the actual chronological sequence of the OB survey and the shortest possible detection time for these 11 dominants, a minimum of 43 consecutive survey days, or 19% of this survey program’s total duration, would have been required to detect these species. With a random start date, a hypothetical OB survey with duration 43 days had a 55% likelihood of recording these 14 dominants (n = 20 trials).

Consequences of “missed” species

The joint total for both South Atlantic Bight surveys was 60 species, not including those species detected outside the quantitative portions of the surveys analyzed here (Appendix 1, available on the website). As of 2014, 60 species represent about 75% of a total pool (sensu Cam et al. 2000) consisting of all aquatic bird species known to occur from the region’s marine waters (e.g. see http://www.carolinabirdclub.org/brc/checklist_of_North_Carolina_birds.html). Most of the species we “missed” during our surveys were lacking either as a result of revised taxonomy (e.g. Patteson & Armistead 2004) or because particular taxa (e.g. certain sea ducks, grebes, larids) inhabit primarily neritic waters very near shore, habitats that we did not sample effectively from ships. However, one primarily offshore alcid known from the region was certainly missed during the quantitative phases of our surveys (Atlantic puffin Fratercula arctica), and five offshore procellariiforms ultimately recorded in the region went undocumented in our surveys (Table 5).

If added to our either of our survey totals, these six taxa (11% of each survey’s $S_{obs}$) bring the total to 59 species, a value nearly identical to several point estimates for $S_{max}$ or within the confidence limits of other estimators (Table 2). With the exception of Fratercula arctica, which is casual to rare in the region (Lowther et al. 2002), all other species detected after the mid-1980s are extralimital to the South Atlantic Bight. The second species added, Swinhoe’s Storm-Petrel Oceanodroma monorhis, took 117 additional days to detect, a duration that hypothetically would cost $585 000–$1 755 000 using conventional research ships (Table 5). It took 579 additional days to detect the 59th species, Black-browed Albatross Thalassarche melanophrys, a survey duration equivalent to at least $289 500 in dedicated ship costs.

DISCUSSION

Survey duration in the South Atlantic Bight

As far as we are aware, ours is the first attempt to test various quantitative attributes of community structure to assess adequacy of survey duration for marine birds. Whether we gauged adequacy by species richness or community dominance, we conclude that total duration for each South Atlantic Bight marine bird survey was minimally adequate. Indeed, community-level measures indicate...
that our surveys had been conducted unnecessarily long in the context of our research goals. Setting aside extralimital species, detecting only community dominants could have been accomplished in as few as 36 to 65 total survey days (Table 4). This level of effort corresponds to an estimated nominal cost of $180 000 to about $1 000 000 in dedicated ship time. Judged by the same measures of community dominance, however, our surveys were conducted two to six times longer than necessary. Such supplemental effort in one survey alone could equate to overspending on dedicated ship use by as much as $2 800 000.

Substantially less effort was required to detect community dominants than to achieve saturation in species richness (Tables 2, 4). Simulations indicated a high probability of detecting regularly occurring community dominants with less than half the survey effort actually expended in GE. Off the OB, however, simulations indicated only modest probability (≤60%) of detecting dominants using similar survey durations. This disparity might be explained by differences between the two surveys with respect to spatial or temporal allocation of sampling across the continental shelf. Also, the OB survey devoted less effort in winter during its final stages, whereas simulations revealed a high likelihood of detecting only community dominants had the survey begun day 100. If community metrics are contemplated as stopping rules (Scharff et al. 2003), we recommend that investigators: (1) survey evenly throughout each season to increase the likelihood of temporal representation, and (2) adopt quantitative criteria that yield a high probability (e.g. ≥95%) of detecting all dominants. Monte Carlo trials or other resampling methods can be used to test such dominance-based stopping rules (cf. Magnussen et al. 2010).

Although we detected substantial heterogeneity in our daily samples of species richness (Fig. 4), we nevertheless ultimately found asymptotes to \( S_{\text{max}} \) (Fig. 3). We attribute this somewhat unexpected finding to the peculiar variability that is common to marine bird survey data. Numerical variability in species richness can stem from several sources, including any differences in total time counting seabirds among survey days. Local seabirds are also patchily distributed due to behavioral (Haney et al. 1992), trophic (Moser & Lee 2012) and oceanographic factors (Hass 1997), any of which may affect a species total by day. Weather can also alter survey date, duration and visibility of detection. Nevertheless, based on the observed saturation in species richness (e.g. Fig. 3; see also Table 5), we inferred that spatial coverage of our surveys was adequate (cf. Longino & Colwell 1997) to represent primary habitats in our study area (e.g. inner vs. outer shelf, Gulf Stream eddies, etc.).

Disparities between scales of data collection and environmental patchiness create noise from binning data into sampling frames of essentially arbitrary grain size (Haney & Solow 1992). Although this noise might increase differences in species richness observed among samples, it need not greatly influence the species accumulation curve. Indeed, several species richness estimators (e.g. ACE, Chao1, jackknife, bootstrap) have been shown to be robust and precise despite variations in the grain size elected for the sampling frame (e.g. Hortal et al. 2006).

As in other species inventories (Longino & Colwell 1997), stratification did not reduce our asymptotic-based estimates for total survey effort (Table 3). This may be because seabird occurrence and species composition are partially independent of the seasons that we used for stratification; seabirds in the South Atlantic Bight originate from tropical, subtropical and temperate regions in both southern and northern hemispheres (Lee & Booth 1979, Lee 1995). Sub-adults of several species may reside at sea in this portion of the Atlantic for years before breeding. Some seabird communities may thus conform to those special cases in which a well-known or -studied taxon (Soberon & Llorente 1993) leads to well-behaved statistical properties in the underlying species richness data (e.g. Keating et al. 1998).

Choosing a community-based estimator for appraising survey duration

In this study, we used asymptotic estimates for species richness in two regional marine bird communities strictly as an informal, practical and descriptive guide to evaluate whether survey duration

---

**TABLE 5**

Duration and hypothetical costs associated with the extra effort required to find additional seabird species not detected during two prior, quantitative surveys in the South Atlantic Bight, United States

<table>
<thead>
<tr>
<th>Species</th>
<th>Additional survey days(^a)</th>
<th>Cost per species</th>
<th>Cumulative cost to survey</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Minimum(^b)</td>
<td>Maximum(^b)</td>
<td>Minimum(^b)</td>
</tr>
<tr>
<td>Atlantic Puffin <em>Fratercula arctica</em></td>
<td>2</td>
<td>$10 000</td>
<td>$30 000</td>
</tr>
<tr>
<td>Swinhoe’s Storm-Petrel <em>Oceanodroma monorhis</em></td>
<td>117</td>
<td>$575 000</td>
<td>$1 725 000</td>
</tr>
<tr>
<td>Atlantic Yellow-nosed Albatross <em>Thalassarche chlororhynchos</em></td>
<td>155</td>
<td>$190 000</td>
<td>$570 000</td>
</tr>
<tr>
<td>Black-bellied Storm-Petrel <em>Fregetta tropica</em></td>
<td>268</td>
<td>$565 000</td>
<td>$1 695 000</td>
</tr>
<tr>
<td>European Storm-Petrel <em>Hydrobates pelagicus</em></td>
<td>409</td>
<td>$705 000</td>
<td>$2 115 000</td>
</tr>
<tr>
<td>Black-browed Albatross <em>Thalassarcha melanophrys</em></td>
<td>542</td>
<td>$665 000</td>
<td>$1 995 000</td>
</tr>
</tbody>
</table>

\(^a\) Number of days required to detect species additional to those recorded during both the OB and GE surveys summarized here.

\(^b\) Based on daily usage fees for a small-capacity vessel in the University National Oceanographic Laboratory System (UNOLS) fleet (Dufour 2012), capable of short-duration surveys (≤1 week) in coastal and shelf waters.

\(^c\) Based on daily usage fees for a medium-capacity vessel in the UNOLS fleet (Dufour 2012), capable of medium-duration surveys (1–3 weeks) in shelf waters.
was minimally adequate. Before our two surveys, the seabird fauna in this portion of the western Atlantic Ocean were very poorly known. Community inventories carry an inherent risk that some species, typically the rarest, go undetected (Gotelli & Colwell 2001, Dorazio et al. 2006, Chao et al. 2009). Observations after our surveys ended confirmed that only species extralimital to the South Atlantic Bight had gone undetected (Table 5). For purposes of an environmental assessment, none of these species could be deemed at population risk within our study regions (and, in some cases globally, e.g., *Fregata tropica*). Moreover, our survey durations were sufficient to detect the presence of arguably the most endangered seabird in the northwest Atlantic Ocean, the Bermuda Petrel *Pterodroma cahow* (see Appendix 1, available on the website).

In addition, we have shown that surveys carried out longer than necessary, at least from the standpoint of merely adding extralimital species, may carry exceedingly high costs. Using dedicated ships in the South Atlantic Bight to detect ever more species would have cost hundreds of thousands of dollars per species (Table 5). The cumulative costs to merely add more species to the community composition would have been even more expensive, exceeding millions of dollars. Thus, from a fiscal standpoint, there is great incentive to sample seabird communities with surveys that are both pragmatic and efficient.

More formally, the number of species detected in a survey \( S_{\text{obs}} \) is typically expected to be less than the number of species actually present \( S_{\text{max}} \). Unless differences between \( S_{\text{obs}} \) and \( S_{\text{max}} \) are small, or biases that deflate \( S_{\text{max}} \) remain constant, comparisons of relative species richness across space or time are problematic (Conroy & Noon 1996, Nichols et al. 1998). Ideally, detection probabilities for species in a community should be very close to 1 and should not vary greatly across sites or times (Boulinnie et al. 1998). If detection biases are large (e.g., observability varies among species; Barbraud & Thiebot 2009), comparing species richness between surveys may be unwise. The probability of detecting all species in the South Atlantic Bight was likely close to 1, given that there are virtually no diving species in the offshore; that is, there is “no place to hide” for birds in open marine habitat. Because failures to detect bird species can be attributed to fundamental rarity instead of a truly low rate of detectability (Dorazio et al. 2006), we conclude that the differences between \( S_{\text{obs}} \) and \( S_{\text{max}} \) illustrated in our study arose largely from estimator performance.

Diverging performance (Table 2) in the various estimators for \( S_{\text{max}} \) was hardly unexpected. Because of unique properties of the individual data sets that underlie species richness, identifying a consistently accurate estimator has proven elusive (e.g. Esty 1986, O’Hara 2005). After considering bias, accuracy and precision (Hellman & Fowler 1999, Walther & Moore 2005), the performance of \( S_{\text{max}} \) depends on inherent attributes of community structure (Foggo et al. 2003), prevalence of rare species (Reichert et al. 2010) and extent to which the true number of species in the community is known (Ugland et al. 2003).

With the data generated from our study, more than one type of estimator gave fairly accurate predictions of the point estimate, \( S_{\text{max}} \), within and between the two surveys, some with confidence limits (Table 2) that encompassed addition even of exceedingly rare species observed well after our surveys ended (Table 5). Because our focus here was exploratory and practical, however, we are reluctant to endorse particular quantitative estimators for all situations. Maximum likelihood was useful only for \( S_{\text{max}} \), not the entire accumulation curve (Table 2, Fig. 3). Coverage estimators are “better understood and…behaved” (Bunge & Fitzpatrick 1993), and they delivered accurate estimates for \( S_{\text{obs}} \) in our surveys (Table 2). The Chao, (1984) ACE and the first-order jackknife were almost exactly accurate for \( S_{\text{max}} \) (Table 2), once the additional species recorded after our surveys were included (Table 5). The non-parametric estimators are less biased and more precise than species accumulation curves (Brose et al. 2003). They have also performed well in other species inventories (Walther & Morand 1998, Foggo et al. 2003, Walther & Moore 2005), including surveys of bird communities (Walther & Martin 2001).

Survey duration in other contexts

We regard our approach for evaluating survey duration as representing the “floor” rather than “ceiling” for effort needed to characterize key species in a seabird community within a defined study area. Whereas minimum survey effort that fails to meet statistical criteria for community composition could be difficult to justify, additional considerations govern whether effort is sufficient for other purposes, including the frequency of sampling. For example, surveys spaced approximately three to five days apart have been found to be statistically independent, and surveys conducted over one to three years captured much of the inter-annual variance for marine habitats off the eastern United States (see Kinlan et al. 2012).

Choice of appropriate sample size is always a critical part of research planning (Morrison 1988, Dale et al. 1991). Our proposed cut-off criteria for survey duration (e.g. Fig. 3) are suitable only when community composition is the response variable (Wiens et al. 1996, Votier et al. 2004, Ainley et al. 2012). If population trends are the focus, statistical power analysis might guide decisions on whether enough samples were collected (e.g. Hatch 2003, Kinlan et al. 2012). Detecting offshore population changes may prove difficult if power gains cannot be achieved through extending duration, frequency and spatial extent of surveys. Maclean et al. (2013) report limited ability to circumvent this problem in one survey program. In such cases, other response variables, including changes to species composition (Wiens et al. 1996, Tittensor et al. 2010) or spatial pattern (Lapeña et al. 2010), may help inform impact assessments for marine birds in offshore marine environments. Algorithms based on bagged decision trees can also identify consistent aggregations or “hotspots” of seabird abundance (e.g. Nur et al. 2011, Santora et al. 2011).

Evaluating survey duration using community attributes could inform the selection of sampling lengths in preliminary or pilot settings elsewhere. One potential application is for identifying the minimum survey duration for impact assessments at offshore wind energy projects (Garthe & Hüppop 2004, Drewitt & Langston 2006, Watts 2010). Before siting of wind power installations, detecting bird species most susceptible to avoidance behavior (Desholm & Kahlert 2005), collision risk (Johnston et al. 2014) and habitat loss from turbines or other physical structures (Fox & Petersen 2006, Hüppop et al. 2006) is vital. Individual sites considered for offshore wind development are typically far smaller than our survey areas (Fig. 1), so we would not expect that minimum-length surveys would require as many as 65 days in order to detect community dominants over such local scales. To avoid any temporal bias in species detections, seasonal representation should be addressed explicitly.
Appraising survey duration using community-level criteria is also relevant to planning new large-scale, region-wide surveys. The Gulf of Mexico remains the least-studied offshore region for seabirds in the continental United States. Despite an immense infrastructure for offshore energy in the region (>4000 offshore oil and gas production platforms; Dismukes 2010), long-term, baseline seabird surveys are still inexplicably lacking for the entire Gulf. Given parallels between the South Atlantic Bight and Gulf of Mexico’s marine environments and seabird faunas, and given adequate spatial dispersion, we recommend that future Gulf surveys be conducted for at least 36–65 total days. Duration could be iteratively tested while such surveys were in progress using tools that rely upon any of several saturation criteria available (e.g. Table 2).

ACKNOWLEDGEMENTS

J.C.H. acknowledges support from the University of Georgia, Skidaway Institute of Oceanography, South Carolina Wildlife and Marine Resources Department, the NOAA/National Marine Fisheries Service, Burleigh-Stoddard Fund, Sheldon Fund, and NSF grants OCE81-10707 to L.R. Pomeroy and OCE81-17761 to G.-A. Paffenholz. D.S.L. received support from the US Fish & Wildlife Service (Contract 14-16-0009-84-985), the North Carolina Sea Grant Program (North Carolina State University), National Undersea Research Center (University of North Carolina-Wilmington), the US Army Corps of Engineers (Wilmington, North Carolina District), the Department of the Navy, the Outer Continental Shelf Office, NOAA (North Carolina), and private funds. M. Socci assisted with data entry and management of D.S.L.’s offshore surveys. C. Zotter prepared the figures. We thank R.K. Colwell, R.H. Day, P.A. McGillivary, D.J. Thornhill, D.G. Ainley, and an anonymous reviewer for their helpful comments on earlier versions of this manuscript. Mention of commercial products or services cited in this manuscript does not imply any endorsement or responsibility for the opinions, ideas, data, or products presented at those locations, or guarantee the validity of the information provided.

REFERENCES


