Statistical guidelines for assessing marine avian hotspots and coldspots: a case study on wind energy development in the U.S. Atlantic Ocean

Elise F. Zipkin[a], Brian P. Kinlan[b,c], Allison Sussman[a,d], Diana Rypkema[e], Mark Wimer[d] and Allan F. O'Connell[d]

[a]Michigan State University
Department of Integrative Biology
East Lansing, MI 48824

[b]NOAA National Ocean Service
National Centers for Coastal Ocean Science (NCCOS)
SSMC-4, N/SCI-1
1305 East-West Hwy
Silver Spring, MD 20910-3281

[c]CSS-Dynamac, Inc.
10301 Democracy Lane, Suite 300
Fairfax, VA 22030

[d]USGS Patuxent Wildlife Research Center
12100 Beech Forest Rd.
Laurel MD 20708

[e]Stanford University
Department of Biology
Stanford, CA 94305

Running title: Sampling guidelines for seabirds

Corresponding Author: Elise F. Zipkin, Email: ezipkin@msu.edu, Phone: 517-884-8039

Word count: 6788

Number of Tables: 1

Number of Figures: 3

Number of Appendices: 5

46 **Abstract**

47 Estimating patterns of habitat use is challenging for marine avian species because seabirds tend
48 to aggregate in large groups and it can be difficult to locate both individuals and groups in vast
49 marine environments. We developed an approach to estimate the statistical power of discrete
50 survey events to identify species-specific hotspots and coldspots of long-term seabird abundance
51 in marine environments. We illustrate our approach using historical seabird data from survey
52 transects in the U.S. Atlantic Ocean Outer Continental Shelf (OCS), an area that has been
53 divided into "lease blocks" for proposed offshore wind energy development. For our power
54 analysis, we examined whether discrete lease blocks within the region could be defined as
55 hotspots (3x mean abundance in the OCS) or coldspots (1/3x) for individual species within a
56 given season. For each of 25 species, we determined which of eight candidate statistical
57 distributions (ranging in their degree of skewedness) best fit seasonal count data. We then used
58 the selected distribution and estimates of regional prevalence to calculate and map statistical
59 power to detect hotspots and coldspots, and estimate the p-value from Monte Carlo significance
60 tests that specific lease blocks are in fact hotspots or coldspots relative to regional average
61 abundance. The power to detect species-specific hotspots was higher than that of coldspots for
62 most species because species-specific prevalence was relatively low (mean: 0.111; SD: 0.110).
63 The number of surveys required for adequate power (>0.6) was large for most species (tens to
64 hundreds) using this hotspot definition. Regulators may need to accept higher proportional effect
65 sizes, combine species into groups, and/or broaden the spatial scale by combining lease blocks in
66 order to determine optimal placement of wind farms. Our power analysis approach provides a
67 general framework for both retrospective analyses and future avian survey design and is
68 applicable to a broad range of research and conservation problems.

Keywords: Marine spatial planning, Model selection, Power analysis, Seabirds, Sampling design, Wind energy development

1. Introduction

Understanding the distribution and abundance patterns of marine species is important not only to address fundamental ecological questions on species habitat use and movement but also to evaluate potential impacts of human activities, such as energy development, on marine populations and communities (Louzao et al. 2006, Nur et al. 2011). Offshore renewable energy development is increasingly common in both Europe and the United States with potential long-term consequences for marine species (Garthe and Hüppop 2004). Wind farms can cause declines in seabird populations through direct impacts from collision (Hüppop et al. 2006) or indirect impacts such as displacement due to disturbance and habitat loss or disruption of migratory pathways (Drewitt and Langston 2006). Evaluating the potential consequences of alternative energy development necessitates a clear understanding of species spatial distributions, abundances, and habitat use to identify sensitive areas in need of protection (Huettman and Diamond 2001, Ford et al. 2004). One important way to reduce risks associated with offshore energy facilities is through scientifically informed marine spatial planning processes that identify and avoid areas that are seabird "hotspots" (high use areas). It is equally useful to determine "coldspot" locations (areas of low use) where wind farms can be safely implemented with minimal risks to seabirds.

There are several difficulties in identifying species hotspots and coldspots in marine environments. The first is that sampling in the ocean, particularly in offshore areas, is expensive

92 and logistically difficult due to remote survey locations and variable climatic conditions.
93 Although seabird sampling methodology is relatively standardized, data can be collected using
94 either aircraft or ships and continuous or discrete transects (Tasker et al. 1984). Additionally, the
95 number and duration of studies is much smaller as compared to terrestrial locations, such that it
96 is difficult to use any one survey effort to determine hotspot/coldspot locations and combining
97 data requires standardizing across sampling discrepancies (Smith et al. 2014). The second issue
98 is that seabird populations tend to have patchy aggregations with extremely skewed distributions
99 (Beauchamp 2011). Thus, typical statistical distributions that are used to model counts (e.g.,
100 Poisson, negative binomial) may not be appropriate for seabird data (Zipkin et al. 2014). The
101 disparate data on seabirds and the uncertainty on how to model available data creates a challenge
102 for identification of consistent long-term patterns in occurrence and abundance of marine birds.
103 We present a framework for assessing species hotspots and coldspots – including the
104 necessary amount of data – which accounts for the extreme skewedness observed in seabird
105 count data. We apply our approach to data from the Outer Continental Shelf (OCS) of the
106 Atlantic Ocean in the eastern United States, a proposed area for offshore wind energy
107 development (Bowes and Allegro 2012). Ongoing research efforts have focused on compiling
108 all available seabird data in the OCS into the Atlantic Seabird Compendium (O'Connell et al.
109 2009), allowing an unprecedented opportunity to examine species' uses of the marine
110 environment. Detailed spatio-temporal models of the occurrence and abundance of birds and
111 other highly mobile species in the offshore marine environment are challenging (Smith et al.
112 2014). Our purpose here is not to create such a complicated model, but instead to develop a
113 general framework that can be applied with a minimum of input data to provide a first-order
114 estimate of retrospective and prospective statistical power to guide interpretation of past data

115 collection efforts and planning of future surveys. Although we focus our approach on seabirds in
116 the Atlantic Ocean for the specific topic of wind energy development, our framework should be
117 useful in identifying hotspots/coldspots for other animal species that aggregate (e.g., insects,
118 fish).
119
120 **2. Material and methods**
121 Our objective is to determine the number of surveys required for sufficient statistical
122 power to detect whether the long-term mean of a species count of individuals (0,1,2,….,100,…)
123 in standardized surveys at a given location is larger (i.e., a hotspot) or smaller (i.e., a coldspot)
124 than some *a priori* reference mean by a meaningful amount. The terms hotspot and coldspot
125 have held a variety of interpretations in the scientific community and popular literature. In this
126 case, we define a location as a species-specific hotspot if the mean count of individuals (i.e.,
127 abundance) of that species, conditional on presence, is more than three times the mean of the
128 non-zero counts in some predefined reference region. We similarly define a species-specific
129 coldspot as a location where the mean count of individuals of that species, conditional on
130 presence, is less than one-third the mean of the non-zero counts in some predefined reference
131 region. Other proportional effect sizes could easily be substituted, as appropriate. Our analyses
132 are focused on defining hotspots and coldspots for individual species based on their prevalence
133 in a region and their abundance at specific locations within seasons. Other metrics, such as
134 species richness or community composition, could be used for defining hotspots/coldspots but
135 are not considered here.
136 We assume that the abundance of a given species at a particular location in time is the
137 outcome of a two-component random process known as a hurdle model (Mullahy 1986). In a

138 hurdle model, abundance is 0 with probability 1-∅, and non-zero with probability ∅ (also
139 referred to as the occurrence probability) according to a Bernoulli distribution. If abundance is
140 non-zero, then the count of individuals (i.e., the group sizes 1,2,3,…) is distributed according to a
141 discrete probability mass function with positive integer support.
142 Using this modeling framework, we can calculate the probability of detecting a
143 hotspot/coldspot given that a location is a hotspot/coldspot for a specific number of sampling
144 events. Conversely, we can determine the number of sampling events that are necessary to
145 detect a hotspot/coldspot with a certain level of power. With spatially referenced count data, we
146 can also use the mean of a location's counts and the number of surveys that have been conducted
147 to calculate a p-value for evaluation of the null hypothesis that the location is not a
148 hotspot/coldspot. To do this, we must determine for each species: 1) its prevalence (occurrence
149 probability) in the reference region (for the Bernoulli portion of the hurdle model) and 2) the
150 discrete probability distribution that best describes the distribution of non-zero counts (i.e., the
151 species' group sizes) within the reference region (for the abundance component of the hurdle
152 model). We then implement a one-sample Monte Carlo significance test (Hope 1968; section
153 2.3) to test for either hotspots/coldspots at given sampling locations using the estimate of
154 prevalence (as a surrogate for the ∅ parameter), the mean of the fitted distribution (as a surrogate
155 of the mean for the reference region), and the parameter estimates from the fitted discrete
156 statistical distribution that describes the non-zero counts.
157
158 ***2.1. Atlantic Seabird Compendium***
159 The data for each seabird species come from the Atlantic Seabird Compendium, which
160 contains the largest collection of scientific seabird surveys conducted within the Atlantic Ocean

161 (O'Connell et al. 2009). We defined our reference region as the Outer Continental Shelf (OCS),
162 the area currently being considered for renewable energy leasing by the Bureau of Ocean Energy
163 Management (BOEM). This area has been divided into 48 446 lease blocks that are roughly 5
164 km^2 in area (Appendix A, Figure A1).

165 The raw data consist of ship-based and aerial visual observations along fixed-width
166 survey-transects recording the species and number of birds seen in each discrete time strip, or at
167 each location along continuous time strips. Observers were generally trained to avoid double
168 counting individuals but survey-specific observation errors are unknown. We used a total of 32
169 datasets that were collected between 1978-2010, 28 of which were ship-based while the
170 remaining 4 were conducted from fixed-wing aircraft (Appendix A, Table A1, Figures A2-A6).
171 Most of the surveys (28 total; 24 ship-based and 4 aerial) were conducted using the continuous
172 time strip method. The four discrete time strip surveys were all ship-based and generally
173 conducted for fixed 15-minute intervals on ships traveling at approximately 10 knots. We
174 segmented all continuous time strip survey data (both ship-based and aerial) into transects of
175 4.63km, equivalent to the distance covered by a ship moving at 10 knots for 15 minutes, to
176 standardize the data across the two survey platforms and to match the discrete time surveys. We
177 eliminated all transect segments shorter than 60% (2.78km) of this distance, and any discrete
178 time strip surveys shorter than 10 minutes (n=209 removed transects). This allowed the
179 remaining discrete and continuous time strip transect segments to be compared on an
180 approximately common basis, "15-minute-ship-survey-equivalents." The resulting data
181 consisted of 44 176 transects that covered our reference region (the OCS) with approximately
182 84% having lengths of 4.63km (and the remainder having lengths no less than 2.78km). Each
183 standardized transect segment was then assigned to a BOEM lease block based on its centroid,

184 such that all count data from a specific transect was assumed to have been observed in the
185 corresponding lease block. All count data for a single species were then summed for each
186 transect, date, dataset combination. We tabulated the number of samples for each lease block
187 within each season and assumed that if a transect was flown/cruised and a given species was not
188 recorded then it was not present (because none of the surveys recorded species absences).
189 Although it is likely that this standardization did not fully resolve all differences among survey
190 platforms and protocols, we believe that it accounted for the major differences among surveys.
191 Because species habitat uses and aggregations can vary throughout the year, we analyzed
192 the count data for each species separately by season (spring = March 1 to May 31; summer =
193 June 1 to August 31; fall = September 1 to November 30; winter = December 1 to February
194 28/29) and only considered counts where individuals were identified to species (approximately
195 88% of records in the data). We detected no temporal or spatial correlation in observations of
196 the same species on repeated occasions within seasons (observations were usually separated by
197 at least several days) using a semivariogram analysis of the log-transformed data (Kinlan et al.
198 2012), and thus did not include temporal or spatial components in our analyses. For each
199 species/season combination, we assumed that a species' prevalence was the proportion of
200 occurrences within a season relative to the total number of transects surveyed within that season
201 (i.e., number of occurrences divided by number of transects surveyed).
202
203 ***2.2. Model fitting and selection***
204 To identify candidate distributions for the non-zero component of the hurdle model, we
205 conducted an extensive literature review of recent and historical papers that attempted to
206 statistically describe or model animal group sizes or counts of individuals. Identifying

207 appropriate statistical distributions for analyzing count data of animal populations is an ongoing
208 area of interest in ecology and can be particularly challenging for seabirds, where both single
209 individuals and large aggregations are frequently observed (Bonabeau et al. 1999, Griesser et al.
210 2011, Zipkin et al. 2010). There are several studies that discuss the ecological and statistical
211 principles as to why aggregations of animals occur in nature (e.g. Beauchamp 2011, Clauset et
212 al. 2009, Jovani et al. 2008, Ma et al. 2011, Niwa 2003, Okubo 1986), much of which is
213 summarized in Zipkin et al. (2014).
214 Based on this literature review, we identified a set of eight discrete probability
215 distributions that could describe the non-zero counts of seabird data (i.e., the non-zero
216 component of the hurdle model). These candidate distributions span the spectrum of possible
217 mean to variance ratios that could be observed in animal data (Table 1). Five of these
218 distributions naturally have positive integer support: geometric, logarithmic, discrete power law
219 (which we refer to as the zeta), discrete power law with exponential cutoff (referred to as the zeta
220 decay), and Yule-Simon (referred to as the Yule). The other three distributions, Poisson,
221 negative binomial, and discretized lognormal, include 0 in their natural support set and were
222 truncated for use in the non-zero component of the hurdle model. The degree of skewedness for
223 these distributions is ranked as follows (from most heavy tailed to least): zeta ≈ Yule > zeta
224 decay > discretized lognormal > logarithmic ≈ negative binomial > geometric > Poisson.
225 We fit each of the eight candidate distributions (Table 1) to available reference data from
226 the Atlantic Seabird Compendium using maximum likelihood estimation (MLE) in the program
227 R (version 2.13.2; R Development Core Team 2011). We used the non-zero count data only
228 from species that had more than 200 observations for a season because it is difficult to
229 distinguish between competing models when sample sizes are small (Beauchamp 2011, Clauset

et al. 2009, Myers and Pepin 1990). We used the VGAM package (Yee 2010) to estimate parameters for the positive Poisson, positive negative binomial, geometric, logarithmic, zeta, and the Yule. We used the methods and code in Clauset et al. (2009) to estimate the parameters for the truncated discretized lognormal and zeta decay distributions. For model selection purposes, we calculated the log-likelihood of each model fit and ranked the models according to Akaike's Information Criterion corrected for finite (i.e., small) sample sizes (AICc; Burnham and Anderson 2002). We then used the Vuong closeness test (Vuong 1989) to compare the top model (i.e., the model with the lowest AICc value) to the fits of the statistical distributions that were ranked second and third. The model with the lowest AICc that was also estimated to be significantly better than the next top models ($p < 0.1$) was selected for use as a reference distribution. We did not conduct power analyses in cases where the Vuong test indicated that the top models performed equally well because this indicates that there was insufficient data to adequately determine the appropriate count distribution; but we note that a model-averaging approach could be implemented in such cases.

The maximum likelihood parameter estimates for the top model were used to define the null hypothesis distribution for calculation of the mean count (i.e., the expected count in a lease block during one sampling event conditional on occurrence) in the reference region as well as subsequent significance tests and power analyses. Most distributions used only one parameter, which we altered to reach the specified effect sizes for the alternative hypothesis tests in the power analyses. In cases where the top distribution had two parameters (i.e., the negative binomial, discretized lognormal, and zeta decay distributions), one parameter (the second parameter listed in Table 1) was held constant at its estimated value, while the other was adjusted to give the desired effect size, measured as the ratio of the mean under the alternative hypothesis

253 to the mean under the null hypothesis. Thus, we assumed that the mean of the distribution
254 changes only as a function of the first parameter, whereas the second parameter is a shape
255 parameter that remains unchanged for a given species in a season. We ensured the validity of
256 this assumption by checking for correlations between the first and second parameter of each
257 distribution type. No significant correlations were detected ($p>0.05$) in our datasets.
258
259 ***2.3. Power analyses***
260 We used the estimate of prevalence (for the Bernoulli component of the model) and the
261 best fitting discrete distribution (for the count component of the model) for each species/season
262 combination to: 1) calculate the power to detect a hotspot of effect size three times the
263 conditional (on presence) reference mean (as estimated using the discrete count distribution) and
264 a coldspot of effect size one-third of the conditional reference mean in each lease block on the
265 Atlantic OCS, given the actual number of surveys that had occurred in that lease block; and 2)
266 estimate the p-value for independent significance tests of the sample mean of each surveyed
267 lease block against one-tailed hotspot/coldspot alternative hypotheses (i.e., the probability that a
268 lease block with a specific number of transects and mean count is a hotspot/coldspot for a given
269 species within a season). We chose to focus on the mean as our test statistic for abundance data,
270 because the long-term mean count of individuals in a discrete spatial unit is often a desired
271 quantity for an environmental impact assessment. However, other test statistics focusing on
272 different aspects of the distribution (e.g., median, quantile, or extreme value statistics) could be
273 relevant for specific questions and would likely have different power characteristics.
274 Because the test statistic is the mean of a potentially small sample, the distribution of the
275 null hypothesis is not readily available in closed form for many of the candidate distributions.

Therefore, we derived the critical value (i.e., 3x the mean for hotspots and 1/3x the mean for coldspots) for the chosen significance level (α=0.05; i.e., Type I error rate) using a Monte Carlo method. Given the sample size, *M*, for a lease block we estimated the critical values by drawing 10 000 samples of *M* random variates from the combined hurdle model using the prevalence in the region for the Bernoulli component (i.e., probability of a non-zero count) and the probability distribution determined from the model selection procedure using appropriate random number generators. In all cases, we held the prevalence constant at its estimated value. We then calculated the conditional sample mean for each of the 10 000 samples and found the 1-α quantile of the simulated distribution of sample means to estimate the upper critical value and the α quantile to calculate the lower critical value. The null hypothesis is rejected at the α significance level if the observed sample mean exceeds the upper critical value (hotspot case) or is less than the lower critical value (coldspot case). We generated power curves for each species/season combination showing power for the actual sample sizes that were encountered in the cumulative historical data. Through this approach, effect sizes are introduced via the non-zero component of the hurdle model. Thus we assume that differences in the mean arise through a multiplicative effect on the non-zero component of the hurdle model and not a change in prevalence. However, this approach can be easily modified to consider other cases in which differences arise as a consequence of changes in occurrence probability, or when both occurrence and non-zero abundance processes change simultaneously.

A similar procedure can be used to derive Monte Carlo p-values for the same one-tailed hypothesis tests. For the hotspot case, the p-value is equal to the proportion of simulated sample means (conditional on presence) that are greater than or equal to the observed sample mean (also conditional on presence). For the coldspot case, the p-value is equal to the proportion of

299 simulated sample means that are less than or equal to the observed sample mean. Model
300 assumptions and their implications are expanded upon in Appendix B.
301
302 **3. Results**
303 ***3.1. Model fitting and selection***
304 There were a total of 74 species/season combinations (19 in spring, 18 in summer, 22 in
305 fall, 15 in winter) that had at least 200 recorded observations in the OCS (Appendix C, Table
306 C1). The prevalence of these species ranged from 0.016 to 0.419 (mean: 0.111; SD: 0.110)
307 across all seasons, with mean species prevalence being twice as high in the winter as compared
308 to the other three seasons: 0.182 (SD: 0.124) in winter, 0.094 (SD: 0.88) in spring, 0.087 (SD:
309 0.099) in summer, and 0.098 (SD: 0.108) in fall (Appendix C, Table C2). We fit the eight
310 discrete probability distributions to the non-zero count data for each of these species/season
311 combinations. It was possible to find the maximum likelihood parameter estimates for all
312 distributions with the species/season combinations, except the Poisson in nine cases and the
313 negative binomial in seven cases because models did not converge. There was surprising
314 consistency in the results with the discretized lognormal having the lowest AICc for 54 of the
315 species/season combinations (Appendix C, Table C2). The Yule was the second most common
316 distribution (10 species/season), followed by the zeta decay (4 species/season), negative
317 binomial (3 species/season), the logarithmic (2 species/season) and the geometric (1
318 species/season).
319 There were only 41 species/season combinations (11 in the winter, 11 in the spring, 7 in
320 the summer, 12 in the fall) where the top model (as estimated using AICc) had a significantly
321 better fit then the next best model according to the Vuong pairwise closeness test ($p < 0.1$). Of

322 the cases where there was a clear best fitting model, the discretized lognormal distribution was
323 the top model for 38 species/season combinations; the Yule, negative binomial, and zeta decay
324 distributions each had the best fit in one case (Appendix C, Table C2). The Yule distribution had
325 a significantly better fit than all other distributions for the Great Black-backed Gull data in the
326 winter. However, the mean of the Yule distribution is undefined when the parameter is
327 estimated to be less than one, as was the case here. As such, we excluded the winter Great
328 Black-backed Gull data from further analysis.
329
330 ***3.2. Power analyses***
331 In all cases, the power to detect hotspots was greater than the power to detect coldspots
332 for lease blocks with low sample sizes (i.e., approximately less than 25-30 samples; Figure 1;
333 Appendix C, Figure C2). For species that had low prevalence (less than 15-19%), the power to
334 detect hotspots was typically greater throughout the observed range (1-79 in winter, 1-59 in
335 spring, 1-81 in summer, 1-61 in fall) of sampling events in lease blocks. This is a logical result
336 given how coldspots are defined as lease blocks with less than a third of the average count
337 conditional on presence. Under this definition, it is more difficult to detect whether a location is
338 truly a coldspot for species that have low occurrence in the reference region. For species with
339 higher prevalence, the ability to detect coldspots increased quickly with additional sampling
340 events such that it was easier to detect coldspots than hotspots for locations with a large number
341 of samples (i.e., approximately greater than 30-40 samples; Appendix C, Figure C2). Again, this
342 result has an intuitive explanation: observing a small number of individuals (or none at all) from
343 a species that is reasonably common over a large number of sampling events suggests that a
344 location may be a coldspot, whereas multiple detections of that same, common species does not

345 necessarily indicate that a location is a hotspot, especially considering that counts of seabirds
346 come from highly skewed distributions (see model fitting results).
347 In general, the power to detect hotspots and coldspots within lease blocks in the OCS was
348 low for most species, even for locations with large numbers of samples. More than half of the
349 species/season combinations had less than 50% maximum power of detecting both hotspots and
350 coldspots (Appendix C). The average ability to detect coldspots among species was lower than
351 that of hotspots across all seasons (Appendix D) and was less than 40% for even the most
352 heavily sampled regions when all seasons were combined (Figure 2). However, some individual
353 species had reasonable power across the region, which allows for examination of species patterns
354 of hotspots and coldspots in the OCS. For example, Northern Gannets had a fairly high
355 prevalence (>20%) and sufficient data for analyses in winter, spring, and fall. Using our power
356 analysis approach, we were able to determine that most of Nantucket Sound is a coldspot for
357 Northern Gannets in the winter but tends to be a mix of hot and coldspots for the species in the
358 spring and fall. Similarly, there are large areas approximately 150-250 km off the east coast of
359 the United States from Delaware to Rhode Island that are hotspots for Northern Gannets during
360 the spring (Figure 3). Additional maps of other species' hot and coldspots in the Atlantic OCS
361 are presented in Appendix E.
362
363 **4. Discussion**
364 We developed a general method for defining species-specific hotspots and coldspots of
365 abundance for marine birds and for assessing the significance and statistical power to detect
366 these locations. Our approach: 1) can serve as the basis for the design of statistically robust
367 surveys to detect departures from regional average patterns of abundance/occurrence and 2)

368 represents a methodology for using existing marine bird survey data to assess the state of
369 knowledge about relative hotspots and coldspots in offshore areas. The power analysis
370 framework accurately accounts for the extreme variation observed in seabird count data, where
371 both individuals and large aggregations are detected, and can additionally accommodate data
372 from disparate surveys. Researchers and conservation managers can use our results to determine
373 the number of surveys necessary to detect hot and coldspots for a particular species using the
374 species-specific power curves (Appendix C). Furthermore, as we demonstrated with our case
375 study in the Atlantic OCS, all available data in a particular region can be combined to map out
376 hotspot and coldspot locations for seabirds using this approach (e.g., Figure 3 and Appendix E).
377 We illustrated our approach for the specific task of defining hotspots and coldspots of
378 species abundances at the spatial scale that management decisions are made for wind energy
379 development (using one set of proportional effect sizes, 3x and 1/3x, as an example). We found
380 that adequate statistical power (>0.6) for even the most prevalent species is not achieved at these
381 proportional effect sizes until the number of sampling events exceeds 40 (within a particular
382 season). For less frequently observed species, more than 100 independent sampling events are
383 necessary to characterize lease blocks as hotspots or coldspots (Appendix C). Studies examining
384 offshore wind energy development in Europe have similarly found that survey efforts must be
385 intensive to detect even large changes in seabird abundance between pre- and post- construction
386 (Maclean et al. 2012, Vanermen et al. 2015). Management decisions on turbine construction are
387 likely to be made on the spatial scale of lease blocks and as such, it is a necessary first step to
388 understand the statistical power of surveys on that fine spatial scale. However, in many cases it
389 is not be practical to conduct such large numbers of sampling events.

A number of possible remedies could be implemented to increase power for hotspot or coldspot detection. First, data on species could be combined so as to determine group- or guild-specific hotspot and coldspot locations. Although information is lost at the species level in this approach, statistical power is higher and analyses could include inferences on rare species, which were excluded in our study. A pooled approach also allows data on unidentified individuals to be incorporated into analyses, increasing sample sizes. Examining broad groups such as gulls, loons, seaducks, or even more general, bottom-feeders, provides information on community level hotspots and coldspots, which is equally important for species conservation and management. Second, the spatial grain of power analyses could be expanded. Although we did not detect spatial correlation at the lease block level, predictive modeling of long-term average seabird distributions in this region have found spatial autocorrelation at scales of up to 10-15 km (Kinlan et al. 2012). Furthermore, typical offshore wind projects are likely to be larger than the 5x5 km (25 km^2) lease blocks. Power analyses conducted on aggregations of 2x2 or 3x3 BOEM lease blocks (i.e., areas of 100-225 km^2) may thus be acceptable for management decisions and would provide improved statistical power. Finally, other metrics such as species prevalence or maximum abundance can be used in future power analyses. The low power to detect hotspot and coldspots based on mean abundances is partially due to the extreme skewedness observed in group sizes. Simplifying the data to presence/absence or modifying the analysis approach to explicitly account for variation in prevalence could increase power. Definitions of hot and coldspots could also be modified to be more stringent (e.g., 10x, 1/10x) thereby increasing power to detect such larger proportional effect sizes.

Our power analysis approach represents a straightforward and effective way to assess the amount of data necessary to determine areas of high and low use for a broad range of marine

avian species. Although our method is necessarily simplified so as to be widely applicable, it suffices for a first order analysis of historical seabird data and should additionally be useful in the design of future surveys and conservation planning. When applying our power analysis method, researchers will need to decide the appropriate spatial and temporal scales relevant to their specific questions and determine both the appropriate effect sizes and reference region(s) for biologically meaningful hotspots and coldspots. For example, reference regions might be stratified by biogeographic breaks, distance-from-shore, bathymetric bands, or on boundaries of major management areas. Results of the power analysis and subsequent guidelines for the appropriate number of surveys will certainly depend on the scientific questions and possibly also on management or regulatory decisions. Wind energy development in the OCS of the U.S. Atlantic Ocean will undoubtedly impact marine species distributions and could alter population-level abundances. Our power analysis revealed that it may be difficult to assess the effects of turbines on seabirds at the spatial scale of lease blocks, the proposed planning unit. Using our methodology, future analyses can leverage existing data to identify locations that require greater survey effort and maximize the probability of detecting locations with irregular abundance/occurrence patterns.

Acknowledgements

We are grateful to Robert Rankin (NOAA) for analytical support, and Jocelyn Brown-Saracino (U.S. Department of Energy), David Bigger (BOEM), and James Baldwin (U.S. Department of Agriculture, Forest Service) for reviews and comments on earlier versions of this manuscript. This project was funded by the BOEM, Office of Renewable Energy Programs through Interagency Agreement M12PG00068 with the U.S. Department of Commerce, NOAA,

National Centers for Coastal Ocean Science. BK was supported by NOAA Contract No. DG133C07NC0616 with CSS-Dynamac, Inc. DR was funded by the NOAA Ernest F. Hollings Scholarship Program. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

References

Beauchamp, G. 2011. Fit of aggregation models to the distribution of group sizes in Northwest Atlantic seabirds. Marine Ecological Progress Series, 425, 261-268.

Bonabeau, E., Dagorn, L. and Freon, P. 1999. Scaling in animal group-size distributions. Proceedings of the National Academy of Sciences, 96, 4472–4477.

Bowes, C. and Allegro, J. 2012. The turning point for Atlantic offshore wind energy: time for action to create jobs, reduce pollution, protect wildlife, and secure America's energy future. Report produced by the National Wildlife Federation, pp. 56

Burnham, K.P. and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information-theoretic approach, 2nd edn. Springer-Verlag, New York.

Clauset, A., Shalizi, C.R. and Newman, M.E.J. 2009. Power-law distributions in empirical data. SIAM Review, 51, 661-703.

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

Ford, R.G., Ainley, D.G., Casey, D.G., Keiper, C.A., Spear, L.B. and Balance, L.T. 2004. The biogeographic patterns of seabirds in the central portion of the California current. Marine Ornithology, 32, 77-96.

458 Garthe, S. and Hüppop, O. 2004. Scaling possible adverse effects of marine wind farms on
459 seabirds: developing and applying a vulnerability index. Journal of Applied Ecology, 41,
460 724-734.
461 Griesser, M., Ma, Q., Webber, S., Bowgen, K. and Sumpter, D.J.T. 2011. Understanding animal
462 group-size distributions. PLoS ONE, 6, doi:10.1371/journal.pone.0023438.
463 Hope, A.C.A. 1968. A simplified Monte Carlo significance test procedure. Journal of the Royal
464 Statistical Society Series B (Methodological), 30, 582-598.
465 Huettman, F. and Diamond, A.W. 2001. Seabird colony locations and environmental
466 determination of seabird distribution: a spatially explicit breeding seabird model for the
467 Northwest Atlantic. Ecological Modelling, 141, 261-298.
468 Hüppop, O., Dierschke, J., Exo, K.-M., Fredrich, E. and Hill, R. (2006), Bird migration studies
469 and potential collision risk with offshore wind turbines. Ibis, 148: 90–109.
470 Jovani, R., Serrano, D., Ursua, E. and Tella, J.L. 2008. Truncated power laws reveal a link
471 between low-level behavioural processes and grouping patterns in a colonial bird. PLoS
472 ONE, 3 e1992, doi:10.1371/journal.pone.0001992.
473 Kinlan, B.P., Zipkin, E.F., O'Connell, A.F. and Caldow, C. 2012. Statistical analyses to support
474 guidelines for marine avian sampling: final report. U.S. Department of the Interior, Bureau
475 of Ocean Energy Management, Office of Renewable Energy Programs, Herndon, VA. OCS
476 Study BOEM 2012-101. NOAA Technical Memorandum NOS NCCOS 158. xiv+77 pp.
477 Louzao, M., Hyrenbach, K.D., Arcos, J.M., Abelló, P., Sola, L.G.D. and Oro, D. 2006.
478 Oceanographic habitat of an endangered Mediterranean procellariiform: implications for
479 marine protected areas. Ecological Applications, 16, 1683-1695.

480 Ma, Q., Johansson, A. and Sumpter, D.J.T. 2011. A first principles derivation of animal group
481 size distributions. Journal of Theoretical Biology, 283, 35-43.
482 Maclean, I.M.D., Rehfisch, M.M., Skov, H., Thaxter, C.B. 2012. Evaluating the statistical power
483 of detecting changes in the abundance of seabirds at sea. Ibis, 155, 113–126.
484 Mullahy. J. 1986. Specification and testing of some modified count data models. Journal of
485 Econometrics, 33, 341–365.
486 Murphy, K.R., Myors, B. and Wolach, A. 2008. Statistical power analysis: a simple and general
487 model for traditional and modern hypothesis tests, 3rd Edn. Routledge Academic, New
488 York.
489 Myers, R., and Pepin, P. (1990). The robustness of lognormal-based estimators of abundance.
490 Biometrics. 46, 1185-1192.
491 Niwa, H.S. 2003. Power-law versus exponential distributions of animal group sizes. Journal of
492 Theoretical Biology, 224, 451–457.
493 Nur, N., Jahncke, J., Herzog, M. P., Howar, J., Hyrenbach, K. D., Zamon, J. E., Ainley, D.G.,
494 Wiens, J.A., Morgan, K., Balance, L.T. and Stralberg, D. 2011. Where the wild things
495 are: predicting hotspots of seabird aggregations in the California Current
496 System. Ecological Applications, 21, 2241-2257.
497 O'Connell, Jr., A.F., Gardner, B., Gilbert, A.T. and Laurent, K. 2009. Compendium of avian
498 occurrence information for the continental shelf waters along the Atlantic coast of the
499 United States (database section: seabirds). A final report for the U.S. Department of the
500 Interior, Minerals Management Service, Atlantic OCS Region, Herndon, VA. 50 pp.
501 Contract No. M08PG20033.
502 Okubo, A. 1986. Dynamical aspects of animal grouping. Advances in Biophysics, 22, 1-94.

503 R development Core 2011. R: A language and environment for statistical computing, R
504 Foundation for Statistical Computing, Vienna, Austria.
505 Smith, M.A., Walker, N.J., Free, C.M., Kirchhoff, M.J., Drew, G.S., Warnock, N. and
506 Stenhouse, I.J. 2014. Identifying marine Important Bird Areas using at-sea survey data.
507 Biological Conservation, 172, 180-189.
508 Tasker, M., Jones, P.H., Dixon, T., and Blake, B.F. 1984. Counting seabirds at sea from ships: a
509 review of methods employed and a suggestion for a standardized approach. The Auk,
510 101, 567-577.
511 Vanermen, N., Onkelinx, T., Verschelde, P., Courtens, W., Van de walle, M., Verstraete, H.,
512 Stienen, E.W.M. 2015. Assessing Seabird Displacement at Offshore Wind Farms: Power
513 Ranges of a Monitoring and Data Handling Protocol. Hydrobiologia, 1-13.
514 Vuong, Q.H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses.
515 Econometrica, 57, 307-333.
516 Yee, T.W. 2010. The VGAM package for categorical data analysis. Journal of Statistical
517 Software, 32, 1-34.
518 Zipkin, E.F., Gardner, B. Gilbert, A., O'Connell, A.F., Royle, J.A. and Silverman, E.D. 2010.
519 Distribution patterns of wintering sea ducks in relation to the North Atlantic Oscillation
520 and local environmental characteristics. Oecologia, 163, 893-902.
521 Zipkin, E.F., Leirness, J.B., Kinlan, B.P., O'Connell, A.F., and Silverman, E.D. 2014. Fitting
522 statistical distributions to sea duck count data: implications for survey design and
523 abundance estimation. Statistical Methodology, 17, 67-81.

Table 1. Parameters and probability mass functions for the eight candidate distributions used for model fitting. In all cases, the support is defined for positive integers, $x \in \{1,2,3, \ldots\}$. Specifications of all distributions are taken from Yee (2010) except for the discretized lognormal (which is truncated so as not to include 0 in the support) and zeta decay (zeta with exponential cutoff), which are specified as in Clauset et al. (2009). Symbols are explained in the notes column.

530
531

Distribution	Parameters	Probability mass function	Notes
Positive Poisson	$\lambda > 0$	$\frac{\frac{\lambda^x}{x!}e^{-\lambda}}{1-e^{-\lambda}}$	λ is both the mean and the variance
Positive negative binomial	$\mu > 0$ $k > 0$	$\frac{\left(\frac{\Gamma(x+k)}{x!\,\Gamma(k)}\right)\left(\frac{\mu}{\mu+k}\right)^x\left(\frac{k}{\mu+k}\right)^k}{1-\left(\frac{k}{\mu+k}\right)^k}$	μ is the mean and 1/k is the dispersion of the corresponding untruncated negative binomial distribution. $\Gamma()$ denotes the gamma function.
Geometric	$0 < p \leq 1$	$p(1-p)^{x-1}$	1/p is the mean
Logarithmic	$0 < p < 1$	$\frac{-1}{ln(1-p)}\frac{p^x}{x}$	$\frac{-1}{ln(1-p)}\frac{p}{1-p}$ is the mean
Positive discretized lognormal	μ $\sigma > 0$	$\frac{\frac{exp\left(-\frac{(ln\,(x-0.5)-\mu)^2}{2\sigma^2}\right)}{(x-0.5)\sqrt{2\pi\sigma^2}} - \frac{exp\left(-\frac{(ln\,(x+0.5)-\mu)^2}{2\sigma^2}\right)}{(x+0.5)\sqrt{2\pi\sigma^2}}}{\sqrt{\frac{2}{\pi\sigma^2}}\;exp\left(-\frac{(ln\,(0.5)-\mu)^2}{2\sigma^2}\right)}$	μ is the mean and σ is the standard deviation of the corresponding continuous, untruncated lognormal distribution. Note that μ and σ are expressed in natural log-transformed units from the original scale. *exp()* denotes the exponential function, *ln()* denotes the natural logarithm function.
Zeta (Discrete Power Law)	$a > 0$	$\frac{1}{x^{a+1}}\Big/\sum_{n=1}^{\infty}\frac{1}{n^{a+1}}$	a is the exponent of the distribution. n is a variable used in the summation. The infinite series summation in the denominator is Riemann's zeta function.
Zeta with exponential cutoff	$a > 0$ $\lambda \geq 0$	$\left(\frac{1}{x^{a+1}\exp(\lambda x)}\right)\Big/\sum_{n=1}^{\infty}\frac{1}{n^{a+1}\exp(\lambda n)}$	a is the exponent of the distribution, and λ is the exponential rate of decay of the power law tail. n is a variable used in the summation. The infinite series summation in the denominator must be approximated numerically.
Yule	$a > 0$	$\frac{a\,\Gamma(x)\Gamma(a+1)}{\Gamma(x+a+1)}$	a is the shape parameter of the distribution, and behaves similarly to the a parameter of zeta and zeta with exponential cutoff distributions. $\Gamma()$ denotes the gamma function.

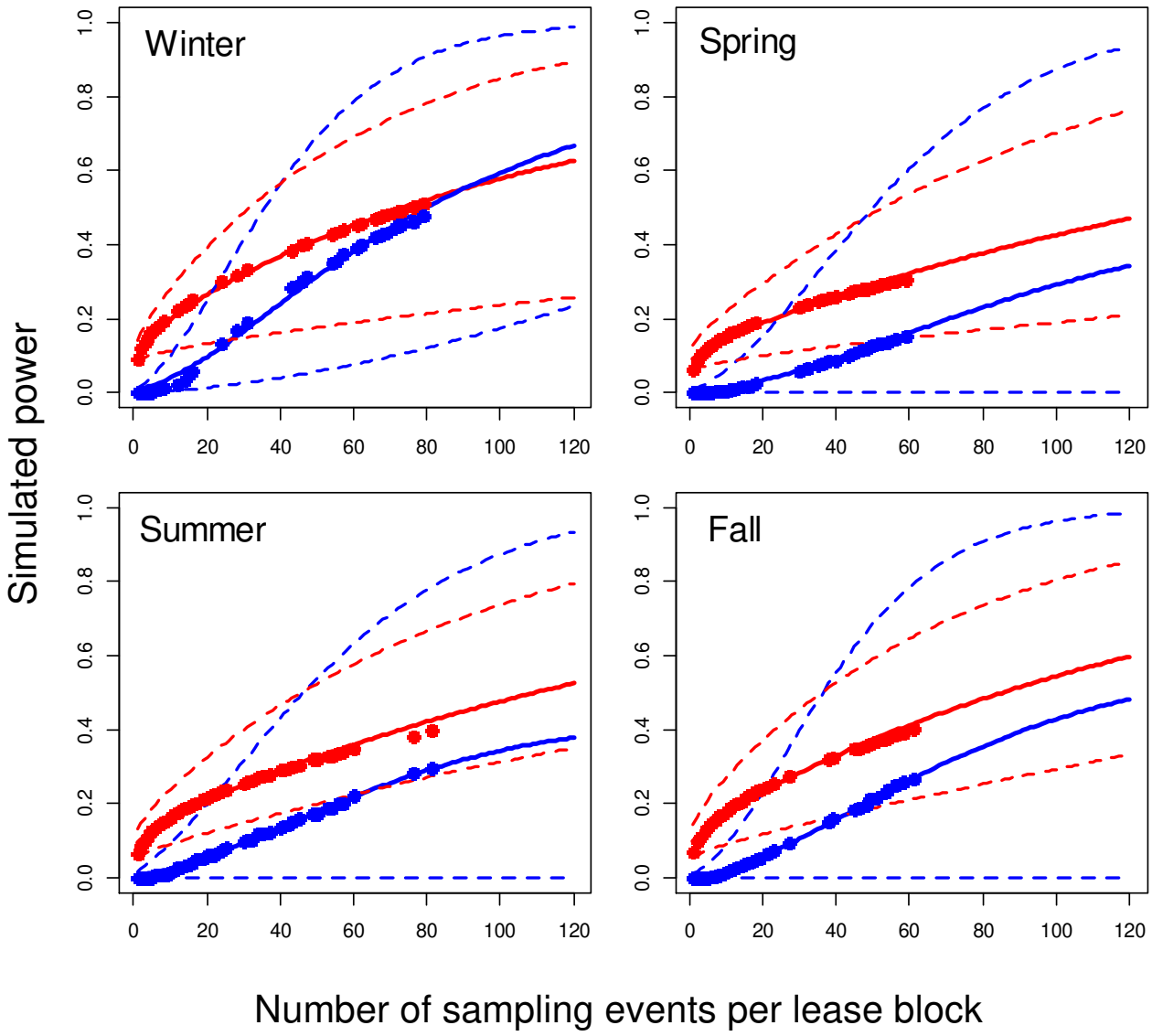
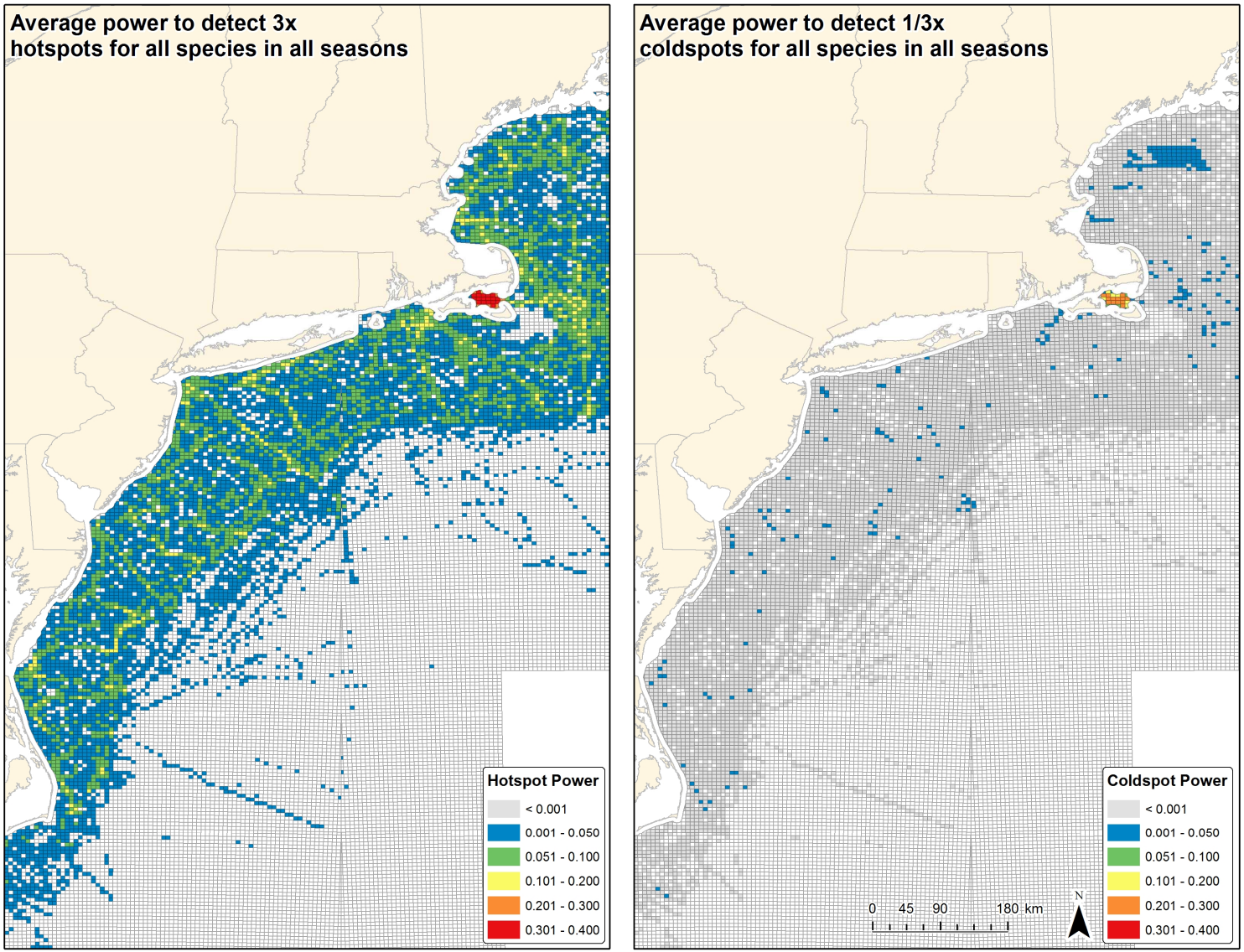


Figure 1. Average simulated power for hotspots (red lines) and coldspots (blue lines) and 95% confidence intervals (dashed lines) for all species in each of the four seasons. We generated the individual species power curves using the approach specified in the methods and then approximated the mean and 95% confidence intervals across species by fitting a regression with

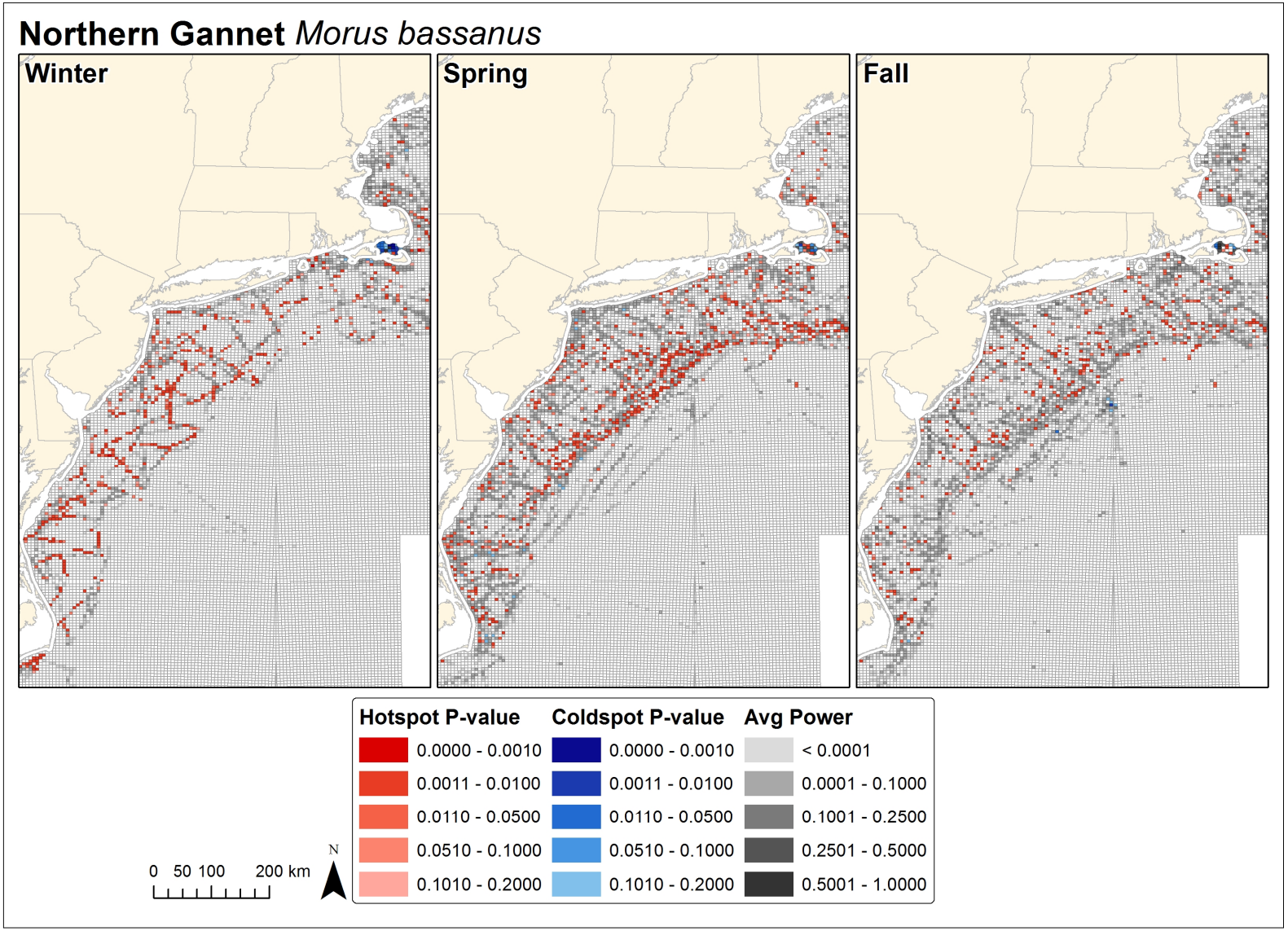
541 a probit transformation (Murphy et al. 2008) to each of the individual species' results, which
542 smoothed the curves. This allowed us to summarize the general patterns across species and
543 generate predictions beyond the range of available sampling data. The red and blue circles
544 indicate the (non-smoothed) mean across species of the simulated power for hotspots and
545 coldspots, respectively, as estimated using the species' power curves presented in Appendix C.

546



547

548 **Figure 2.** Average power to detect hotspots (left panel) and coldspots (right panel) for all species/season combinations.


Northern Gannet Morus bassanus
Winter
Spring
Fall
Hotspot P-value
0.0000 - 0.0010
0.0011 - 0.0100
0.0110 - 0.0500
0.0510 - 0.1000
0.1010 - 0.2000
Coldspot P-value
0.0000 - 0.0010
0.0011 - 0.0100
0.0110 - 0.0500
0.0510 - 0.1000
0.1010 - 0.2000
Avg Power
< 0.0001
0.0001 - 0.1000
0.1001 - 0.2500
0.2501 - 0.5000
0.5001 - 1.0000
0 50 100 200 km
N

549

Figure 3. The p-values (≤0.2) for lease blocks within the OCS that indicate possible hotspot (shades of red) and coldspot (shades of blue) locations for Northern Gannets in the winter, spring, and fall seasons. Blocks in shades of grey show the average power to detect whether a lease block is a hotspot or coldspot for Northern Gannets in instances where the hot/coldspot p-value was greater than 0.2. Thus the darkest grey shading indicates lease blocks not identified as significant hotspots or coldspots, and for which we can be confident in that result because there was relatively high power to detect a hotspot or coldspot, had it existed. Light grey shading indicates lease blocks not identified as significant hotspots or coldspots, but for which there was little or no statistical power. The darkest blue lease blocks can be regarded as the most significant coldspots, the darkest red lease blocks as the most significant hotspots, and the darkest grey blocks as places most likely to be neither hotspots nor coldspots. Blank (white) polygons indicate lease blocks in which no presences of this species were observed. Additional maps of other species' hot and coldspots are presented in Appendix E.