OCS Study
BOEM 2012-101

NOAA Technical Memorandum NOS NCCOS 158

Statistical Analyses to Support Guidelines for Marine Avian Sampling

Final Report

U.S. Department of the Interior
Bureau of Ocean Energy Management
Office of Renewable Energy Programs

U.S. Department of Commerce
National Oceanic & Atmospheric Administration
National Centers for Coastal Ocean Science

OCS Study
BOEM 2012-101

NOAA Technical Memorandum NOS NCCOS 158

Statistical Analyses to Support Guidelines for Marine Avian Sampling

Final Report

Authors

Brian P. Kinlan[1]
Elise F. Zipkin[2]
Allan F. O'Connell[2]
Christopher Caldow[1]

December 2012
Herndon, VA

Prepared under Interagency Agreement
M12PG00068
by
[1]National Oceanic & Atmospheric Administration, National Ocean Service
National Centers for Coastal Ocean Science
Center for Coastal Monitoring and Assessment, Biogeography Branch
1305 East-West Hwy, SSMC-4, N/SCI-1
Silver Spring, MD 20910

In cooperation with
[2]United States Geological Survey
Patuxent Wildlife Research Center
12100 Beech Forest Rd
Laurel, MD 20708

Published by
U.S. Department of the Interior
Bureau of Ocean Energy Management
Office of Renewable Energy Programs

BOEM
Bureau of Ocean Energy Management

U.S. Department of Commerce
National Oceanic & Atmospheric Administration
National Centers for Coastal Ocean Science

DISCLAIMER

This report was prepared under an interagency agreement between the Bureau of Ocean Energy Management (BOEM) and the National Oceanic and Atmospheric Administration (NOAA), National Ocean Service (NOS), National Centers for Coastal Ocean Science (NCCOS). This report has been technically reviewed by BOEM, NOAA, and two external expert reviewers and it has been approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of BOEM or NOAA, nor does mention of trade names or commercial products constitute endorsement or recommendation for use. It is, however, exempt from review and compliance with BOEM editorial standards.

REPORT AVAILABILITY

The report is available for download on the Bureau of Ocean Energy Management, Office of Renewable Energy Programs website. You may request the report from either the Bureau of Ocean Energy Management or the National Technical Information Service. The contact information is:

U.S. Department of the Interior
Bureau of Ocean Energy Management
Office of Renewable Energy Programs
381 Elden Street, HM 1328
Herndon, VA 20170
Phone: (703) 787-1329
Fax: (703) 787-1708

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161
Phone: (703) 605-6040
Fax: (703) 605-6900
Email: bookstore@ntis.gov

CITATION

Kinlan, B.P., E.F. Zipkin, A.F. O'Connell, and C. Caldow. 2012. Statistical analyses to support guidelines for marine avian sampling: final report. U.S. Department of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs, Herndon, VA. OCS Study BOEM 2012-101. NOAA Technical Memorandum NOS NCCOS 158. xiv+77 pp.

ACKNOWLEDGEMENTS

We are grateful to Mark Wimer (U.S. Geological Survey [USGS]) and Allison Sussman (USGS) for technical assistance, Diana Rypkema (Cornell University) and 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 document. This project was funded by the Bureau of Ocean Energy Management, Office of Renewable Energy Programs through Interagency Agreement M12PG00068 with the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, National Centers for Coastal Ocean Science. Brian Kinlan was supported by NOAA Contract No. DG133C07NC0616 with Consolidated Safety Services, Inc.

ABOUT THE COVER

Cover photo (storm-petrel flock) courtesy of David Pereksta (BOEM). Used with permission.

ABSTRACT

Interest in development of offshore renewable energy facilities has led to a need for high-quality, statistically robust information on marine wildlife distributions. A practical approach is described to estimate the amount of sampling effort required to have sufficient statistical power to identify species-specific “hotspots” and “coldspots” of marine bird abundance and occurrence in an offshore environment divided into discrete spatial units (e.g., lease blocks), where “hotspots” and “coldspots” are defined relative to a reference (e.g., regional) mean abundance and/or occurrence probability for each species of interest. For example, a location with average abundance or occurrence that is three times larger the mean (3x effect size) could be defined as a “hotspot,” and a location that is three times smaller than the mean (1/3x effect size) as a “coldspot.” The choice of the effect size used to define hot and coldspots will generally depend on a combination of ecological and regulatory considerations. A method is also developed for testing the statistical significance of possible hotspots and coldspots. Both methods are illustrated with historical seabird survey data from the USGS Avian Compendium Database.

Our approach consists of five main components:

1. A review of the primary scientific literature on statistical modeling of animal group size and avian count data to develop a candidate set of statistical distributions that have been used or may be useful to model seabird counts.

2. Statistical power curves for one-sample, one-tailed Monte Carlo significance tests of differences of observed small-sample means from a specified reference distribution. These curves show the power to detect "hotspots" or "coldspots" of occurrence and abundance at a range of effect sizes, given assumptions which we discuss.

3. A model selection procedure, based on maximum likelihood fits of models in the candidate set, to determine an appropriate statistical distribution to describe counts of a given species in a particular region and season.

4. Using a large database of historical at-sea seabird survey data, we applied this technique to identify appropriate statistical distributions for modeling a variety of species, allowing the distribution to vary by season. For each species and season, we used the selected distribution to calculate and map retrospective statistical power to detect hotspots and coldspots, and map p-values from Monte Carlo significance tests of hotspots and coldspots, in discrete lease blocks designated by the U.S. Department of Interior, Bureau of Ocean Energy Management (BOEM).

5. Because our definition of hotspots and coldspots does not explicitly include variability over time, we examine the relationship between the temporal scale of sampling and the proportion of variance captured in time series of key environmental correlates of marine bird abundance, as well as available marine bird abundance time series, and use these analyses to develop recommendations for the temporal distribution of sampling to adequately represent both short-term and long-term variability.

We conclude by presenting a schematic “decision tree” showing how this power analysis approach would fit in a general framework for avian survey design, and discuss implications of model assumptions and results. We discuss avenues for future development of this work, and recommendations for practical implementation in the context of siting and wildlife assessment for offshore renewable energy development projects.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

1.1 Motivation

We begin with an illustration of the characteristic statistical noisiness of seabird count data and the challenges this presents for identification of "hotspots" and "coldspots". First, we define a "hotspot" of seabird abundance as a discrete spatial unit where the long-term mean abundance of a given bird species is substantially larger than some reference value (e.g., the regional average abundance), where "substantially" is determined by some pre-specified effect size that has biological or regulatory meaning. Figure 1 shows maps of the mean counts of three species of seabirds in standardized visual transect surveys (data are from the USGS Avian Compendium Database and are described in greater detail in section 2.6). Means are reported on a spatial grid defined by the U.S. Department of Interior (DOI) Bureau of Ocean Energy Management (BOEM) lease blocks. Note that certain grid cells stand out as apparent "hotspots" (or "coldspots"), that is, the mean values in those cells are much larger (or smaller) than the regional mean. However, because we expect random variation in the number of birds observed at a given location on any given sampling occasion, it is impossible to tell whether any of these cells are truly hotspots or coldspots without also considering the number of independent repeat surveys that occurred at each location. The purpose of this study is to develop a simple approach for identifying which cells on maps like the ones shown in Figure 1 have been sampled adequately to draw conclusions about their status as a "hotspot" or "coldspot" relative to some reference mean. In addition to enabling retrospective analysis of maps of previous surveys like those shown in Figure 1, such a method will also enable prospective planning of sampling to ensure adequate ability to discriminate hotspots and coldspots for a given species, based on knowledge about the statistical properties of that species' statistical variation in abundance.

In a manner directly analogous to our definition of abundance hotspots and coldspots, we can also define hotspots and coldspots of species occurrence probability. Indeed, if one considers observations in which the species of interest is absent, then it is critical to consider a species' occurrence probability along with its abundance when present in order to design surveys to detect hotspots and coldspots. We therefore consider methods of survey design that consider both occurrence probability and abundance, separately and jointly.

Detailed spatio-temporal models of the occurrence and abundance of birds and other highly mobile species in the offshore marine environment can be extremely complex. Our purpose here is not to create such a complex model, but instead to develop a simple, general framework that can be applied with a minimum of input data to provide a first-order estimate of statistical power. Therefore, we will make a number of simplifying assumptions to arrive at a pragmatic approach to power estimation for hotspot and coldspot surveys in marine birds. Section 4 of this document discusses the implications and limitations of some of these assumptions and addresses how additional complexity can be incorporated into this power analysis framework.

Finally, it is important to point out that many different definitions of hotspots and coldspots have been proposed and different definitions may be useful in different contexts. In this study, we are focused on single-species abundance and occurrence probability. We do not consider diversity, species richness, total abundance across all species, multi-species occurrence probabilities, or any other multi-species aggregation metrics, although these can be important in some contexts, for example biodiversity conservation and ecosystem function. We also do not explicitly

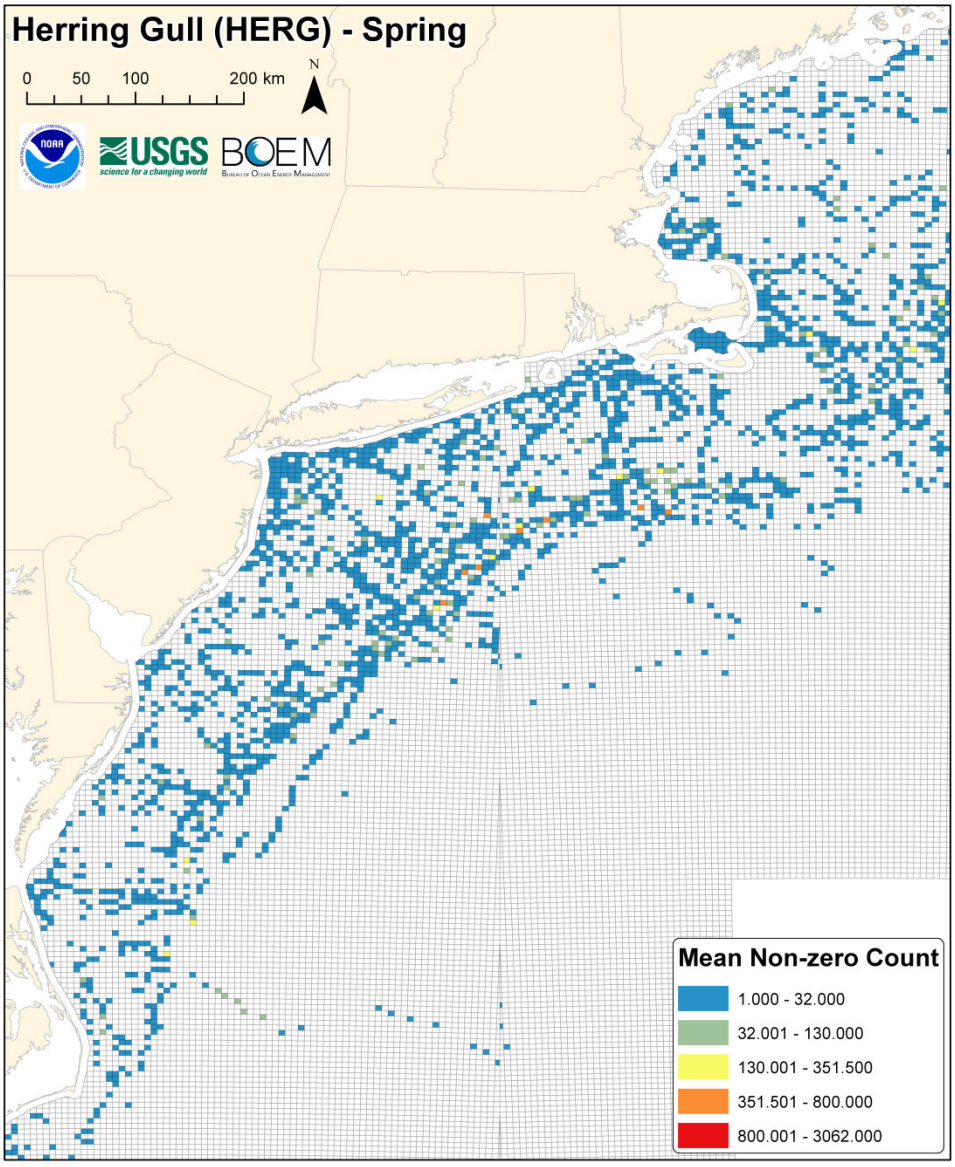


Figure 1. Maps depicting sample means, by BOEM lease block (grid cells), of all non-zero counts of marine bird species listed in Table 3 from standardized 15-minute-ship-survey-equivalent visual transects. Based on data from multiple surveys compiled in the USGS Avian Compendium database (1978-2010; see Table 2). Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was not observed.

a. Herring Gull (Spring). 3,828 surveys with non-zero counts (41,959 total individuals). 2,791 BOEM Atlantic OCS lease blocks with at least one non-zero observation.

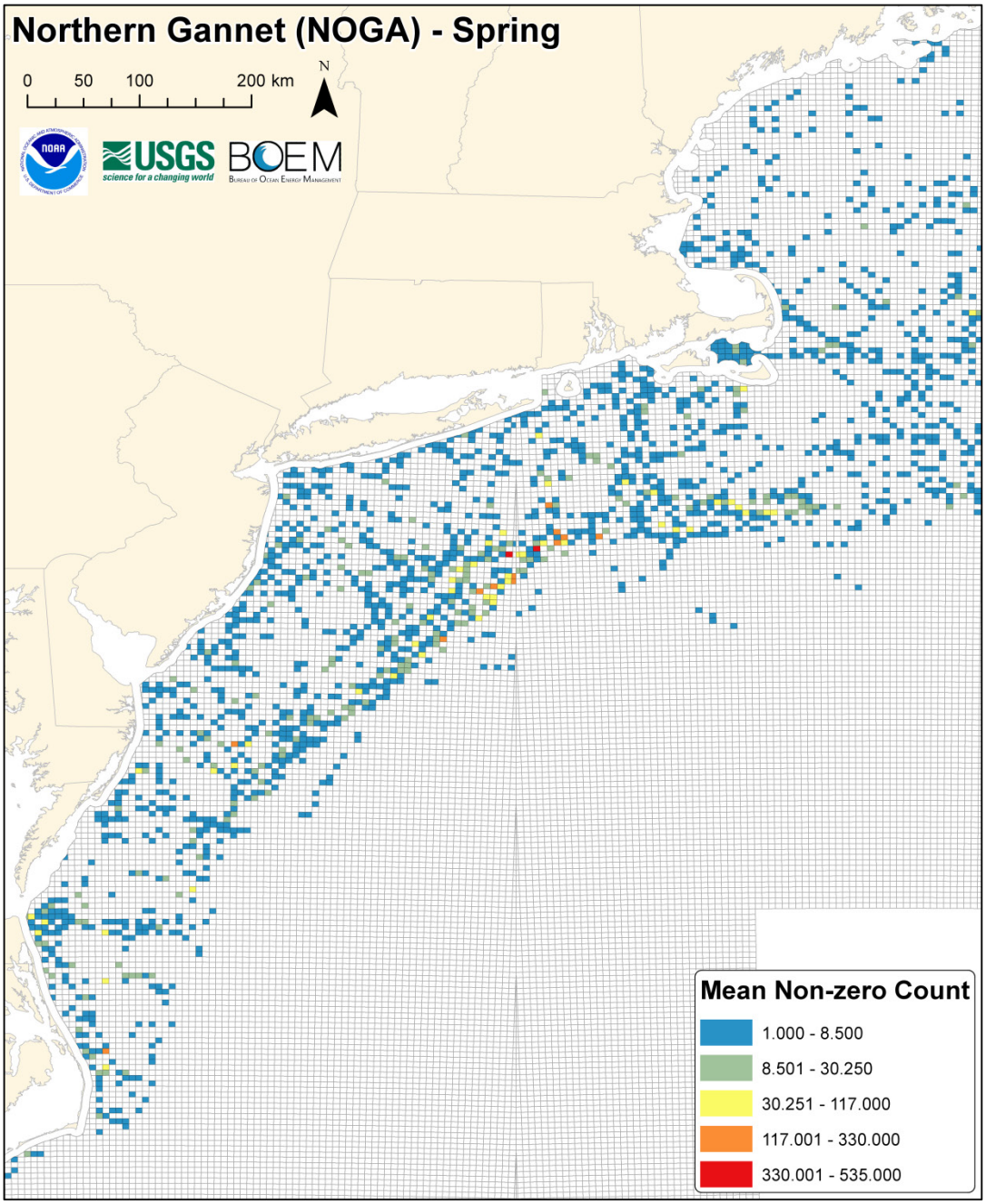


Figure 1.

b. Northern Gannet (Spring). 2,749 surveys with non-zero counts (21,114 total individuals). 2,014 BOEM Atlantic OCS lease blocks with at least one non-zero observation.

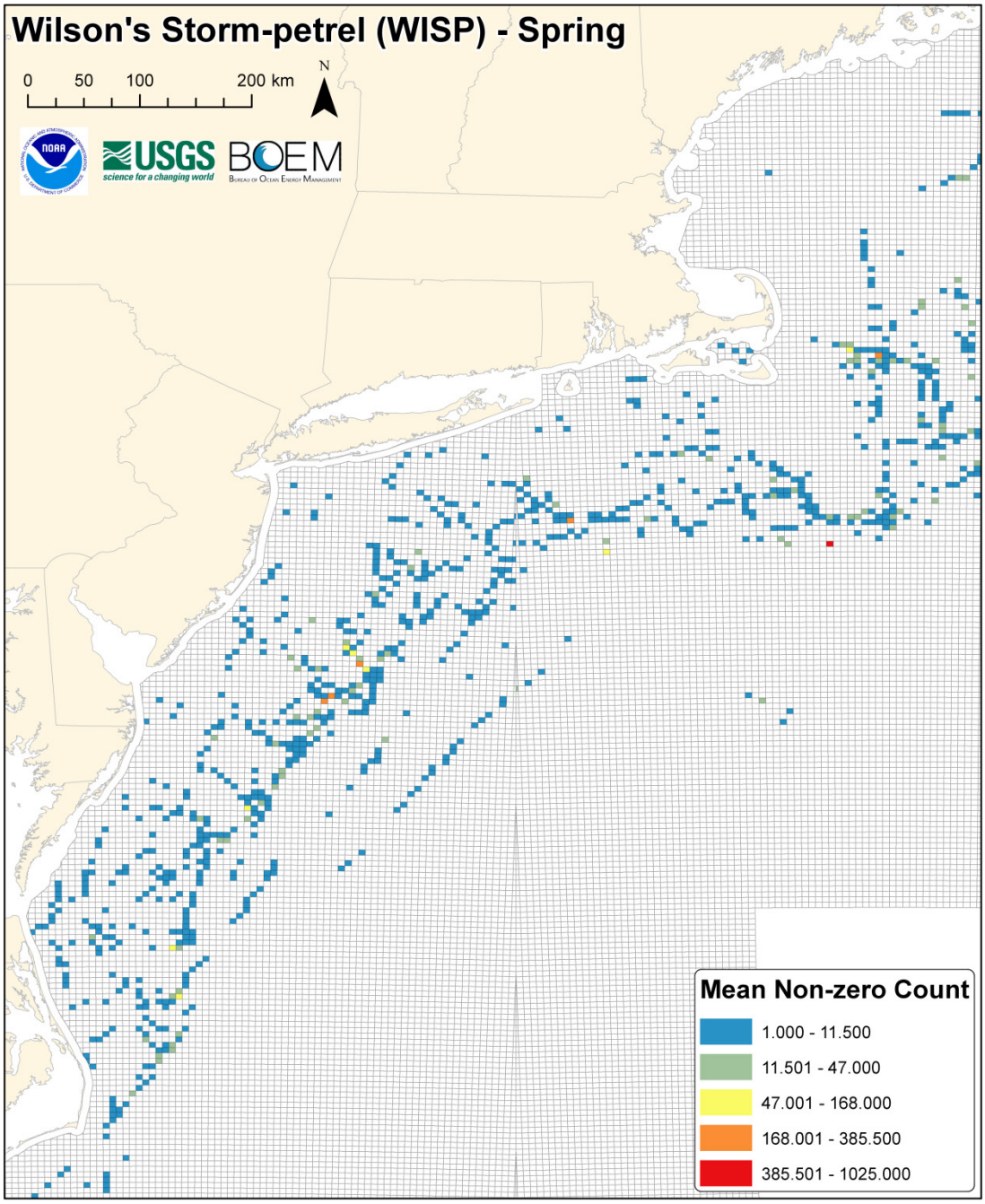


Figure 1.

c. Wilson's Storm-Petrel (Spring). 1,476 surveys with non-zero counts (12,092 total individuals). 1,244 BOEM Atlantic OCS lease blocks with at least one non-zero observation.

consider the length of time for which a hotspot or coldspot must be detectable in order to consider it "persistent". Our method does require multiple surveys to identify a statistically robust hotspot or coldspot, but those surveys could in theory occur within a single year if many independent surveys were conducted in a discrete spatial unit over a short period of time. If long-term persistence of hotspots or coldspots is of interest, then one could apply the method developed here in temporal strata to evaluate whether statistically robust hotspots or coldspots recur in the same location over long periods of time. In sections 2.10, 2.11, 3.7 and 3.8, we describe analyses of temporal variability in relevant characteristics of the ocean environment and in marine bird abundance, and examine implications of these analyses for allocating sampling effort over time.

1.2 Basic Model and Assumptions

Consider an offshore coastal region divided into a grid of discrete spatial units ("grid cells") for purposes of regulation, leasing, and biological monitoring. We first assume that there is a standardized survey protocol available with which to sample bird abundance, in which birds are counted along fixed-width, fixed-time transects, and that the standardized transect length is appropriate to characterize discrete spatial units of the chosen grid size. For simplicity, the location of a transect is taken as its centroid, and the transect is assumed to sample the grid cell in which this centroid falls.

We then assume that number of birds of a given species that would be counted by a standardized survey conducted in a given grid cell at any instant is the outcome of a type of two-component random process known as a hurdle model (Mullahy 1986), in which the abundance is 0 with probability 1-$\emptyset$, or non-zero with probability $\emptyset$, according to a Bernoulli($\emptyset$) distribution, and non-zero abundances are distributed according to a discrete probability mass function with positive integer support, such as a zero-truncated Poisson. We further assume that this model directly reflects the underlying variation in bird abundance; in other words, we ignore any effects of the observation process such as detection errors (i.e., detectability), incorrect counts, lumping, misidentification of species, and approximation of group sizes. The non-zero count component of the hurdle model is assumed to follow a discrete probability distribution with positive integer support that can be determined by a model selection process (see sections 2.1 and 2.2). The random processes are assumed independent from grid cell to grid cell (no spatial correlation), and successive samples from a given grid cell are assumed independent (no temporal correlation). Finally, we assume that the distribution of a given grid cell is stationary, meaning that the parameters do not change over time.

Given these assumptions, for a given bird species in a given grid cell, we want to determine how many surveys are required to:

Case (1) Have sufficient statistical power to detect whether the long-term mean of the **non-zero counts** in the grid cell is larger or smaller than some *a priori* reference mean by a meaningful amount.

Case (2) Have sufficient statistical power to detect whether the **probability of occurrence** in the grid cell is larger or smaller than some *a priori* reference probability of occurrence by a meaningful amount.

Case (3) Have sufficient statistical power to detect whether the long-term mean of the **unconditional counts** (i.e., the mean including zero and non-zero counts) in the grid cell is larger or smaller than some *a priori* reference mean by a meaningful amount.

For simplicity, we treat each of these three cases separately in this document. We will first develop methods to address cases (1) and (2). We will then show how results from (1) and (2) can be combined, using the hurdle model concept described above, to address case (3). These methods are described in detail in the next section of this document (section 2). Section 3 and the Digital Supplements (described in Appendix A) present the results of an application of the methodology developed in section 2. In section 4, we discuss assumptions, limitations, caveats, and recommendations for application of the proposed methodology.

Statistical power is defined as the probability of rejecting the null hypothesis of a statistical test when it is false; that is, when some alternative hypothesis holds. Calculation of statistical power requires specification of:

i. the test statistic and test procedure on which hypothesis testing will be based
ii. the Type I error rate (the significance threshold, or desired probability of rejecting the null hypothesis when it is true)
iii. the distribution of the test statistic under the null hypothesis
iv. the distribution of the test statistic under the alternative hypothesis (which involves specifying the effect size: the difference in the test statistic between the null and alternative hypotheses)
v. sample size

We will first describe how power is calculated for case (1) described above—power to detect differences in mean non-zero counts. For the test statistic (i), we use the sample mean, and for the test procedure, a one-sample Monte Carlo significance test (Hope 1968; see section 2.3) based on simulating a sample from a known reference distribution (the null hypothesis). We use one-tailed tests whose direction depends on whether we are testing for a hotspot or a coldspot. Unless otherwise specified, we use a Type I error rate (ii) of 0.05, from which we can calculate the upper (for hotspot tests) or lower (for coldspot tests) critical value of the null distribution. For the null hypothesis distribution (iii), we assume the hurdle model as described above, with mean of the non-zero distribution equal to some reference mean specified *a priori*, and any additional parameters of the reference distribution also specified *a priori* (we will refer to these additional parameters as nuisance parameters). For the alternative hypothesis distribution (iv), we assume the same distributional form as the null hypothesis, and the same nuisance parameters, but with the location parameter adjusted to give the desired effect size. We specify effect sizes multiplicatively in terms of the reference distribution mean; if the reference mean is denoted by μ, then a "3x" effect size indicates an alternative hypothesis with a mean of 3μ, and would correspond to a power calculation for the scenario in which the grid cell in question is a hotspot whose mean is at least three times larger than the *a priori* reference mean. Similarly, a "$\frac{1}{3}$x" effect size would indicate a power calculation for a coldspot whose mean is at least three times smaller than the *a priori* reference mean. Finally, the sample size (v) is provided as an input to the power calculation, allowing for example, the calculation of prospective power vs. sample size curves (see section 2.5), or calculation of retrospective power given actual sample sizes (see section 2.7).

Given all of this information and access to appropriate random number generators, it is straightforward to estimate power by direct simulation of the test procedure (see sections 2.3 and 2.4). Hotspot (or coldspot) detection power can be estimated as the frequency of simulated values under the alternative hypothesis that are greater than or equal to the upper critical value (or less than or equal to the lower critical value) of the null hypothesis distribution given the Type I error rate. We choose a direct simulation method for calculating power to allow for maximum flexibility in choice of a distribution for the second component of the hurdle model and accurate calculations for small sample sizes. The distribution of the sample mean is not available in closed form for all distributions and a normal approximation based on the central limit theorem cannot be relied upon for small sample sizes.

An important question that arises is the specification of the reference distribution, which determines both the distributional form and the parameters of the null and alternative hypotheses. In practice this reference distribution may be derived from the literature, from prior knowledge of the species' distribution, from previously collected data, and/or by averaging data over some reference region that is much larger than the grid cells being evaluated. We will illustrate the latter approach: selecting and fitting the reference distribution based on available data from a region of interest, and then taking it as known for purposes of the subsequent power analysis. Although this violates the underlying one-sample assumption that the null hypothesis distribution is known *a priori* without sampling error, the error in resulting power calculations is negligible as long as the reference dataset is sufficiently large relative to the sample (data falling in a single grid cell).

In case (2), we are interested in power to detect hotspots and coldspots defined on the basis of occurrence probability. This is a much more straightforward problem, because the appropriate statistical test to use is the one-sample Fisher's Exact Test, a well-studied and classical statistical test. This test can be applied as long as we are willing to assume that the occurrence probability at a given location does not vary over time and is not temporally autocorrelated. Under these conditions, the null hypothesis is a binomial(p,k) distribution with probability p equal to the reference probability of occurrence, and the number of samples k equal to the sample size. The alternative hypothesis is a binomial distribution where p is multiplied by the desired effect size. If we then specify the type I error rate and sample size, the power of this test can be well-approximated by an analytical formula (Bennett and Hsu 1960).

Case (3), the power to detect differences in the unconditional mean (mean including zero and non-zero counts), can be addressed by simply combining the two approaches described above for case (1) and case (2) using the hurdle model. Case (1) represents the non-zero component of the hurdle model and case (2) represents the Bernoulli process that generates zeros. The only question that arises is whether we assume that differences in the unconditional mean arise through differences in the non-zero component, or in the occurrence probability component of the hurdle model. In this paper, we treat only the former case, where differences in the mean arise through a multiplicative effect on the non-zero component of the hurdle model. It would be trivial, however, to generalize this work to consider other cases in which differences arise as a consequence of changes in occurrence probability, or both occurrence and non-zero abundance processes change simultaneously.

2.0 METHODS

2.1 Identification of Candidate Distributions

To identify candidate distributions for the non-zero component of the hurdle model described in section 1.2 (case (1)), we searched the peer-reviewed scientific literature using Google Scholar and ISI Web of Science for a selection of recent papers that attempted to statistically describe or model bird group sizes or counts of birds observed in timed surveys in discrete spatial units. We also included papers that studied distributions of group sizes in other highly mobile species that form aggregations, such as fish.

The most challenging problem we face is characterizing a count distribution with an extreme variance to mean ratio, as is often observed in seabird data (Zipkin et al. 2010, Zipkin et al. 2012). Seabirds are often unevenly and unpredictably distributed (Caraco 1980, Certain et al. 2007, Silverman et al. 2001); for example, counts often include many zeros (Hall 2000, Martin et al. 2005) and distributions of count data can be extremely right skewed (Bonabeau et al. 1999, Griesser et al. 2011). Identifying appropriate statistical distributions for analyzing count data of animal populations is an ongoing area of interest in ecology.

For reasons based on first principles and for convenience, the Poisson distribution has frequently been used to model skewed count data (Caraco 1980) and is popular in modeling avian species (e.g., Fujisaki et al. 2008, Link and Sauer 2007). Yet the inherent assumption that the variance equals the mean often does not hold for many seabird species, which are known to form large flocks. The negative binomial distribution, which allows the variance to exceed the mean, is used as an alternative to the Poisson to characterize the count distributions for species where spatial aggregation is known to occur (e.g., Beauchamp 2011, Cohen 1972, Wood 1985). The negative binomial distribution is the result of a Poisson-Gamma mixture and converges to the Poisson distribution as the shape parameter, k, approaches infinity (Table 1). Okubo (1986) recommended the geometric distribution – a discrete analog to the exponential distribution and also a special case of the negative binomial where the shape parameter equals one – to handle extremely large group sizes and demonstrated its applicability for a number of taxa including birds. Empirical evidence suggests, however, that the negative binomial and geometric models do not adequately capture observed distributions of counts for some populations, especially those that are found in very large group sizes, such as some fish and bird species. Ma et al. (2011) derived a logarithmic distribution from first principles based on rules for when individuals should join and leave groups; this model has outperformed the Poisson and negative binomial distributions in studies of house sparrows (Griesser et al. 2011) and seabirds (Jovani et al. 2008). Ma et al. (2011) additionally pointed out that the logarithmic distribution can be derived as a limiting case of the negative binomial distribution as the shape parameter (k, Table 1) approaches zero (see also Quenouille 1949), placing it in the context of other distributions used to model ecological count data.

More recently, the power law distribution (also known as the zeta distribution in its discrete form) has been proposed for modeling group sizes when the variance to mean ratio is much larger than can be accommodated by the aforementioned models (Bonabeau and Dagorn 1995, Bonabeau et al. 1999). Several studies have demonstrated that the power law distribution fits well to a number of empirical examples including populations of fish, seabirds, and mammals

Table 1

Parameters and probability mass functions for the eight candidate distributions. In all cases, the support is $x \in \{1,2,3,\ldots\}$ (that is, the distributions are defined for positive integers). Specifications of all distributions are as in the VGAM package in R (Yee 2010) except for the discretized lognormal and zeta with exponential cutoff which are specified as in Clauset et al. (2009). Symbols used in probability mass functions are explained in the Notes column.

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
Discretized lognormal (truncated such that $x \in \{1,2,3,\ldots\}$)	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.

(Clauset et al. 2009, Beauchamp 2011, Jovani et al. 2008, Keitt and Stanley 1998, Sjoberg et al. 2000). However, the power law distribution (using ecologically relevant parameter ranges) is capable of producing extremely large counts (e.g., in the millions; Clauset et al. 2009), which are not realistic for most seabird species. The power law can be combined with an exponentially decaying function (Niwa 2003) to address this problem; such distributions are referred to as power law with exponential cutoff, power law with exponential decay, or simply power law exponential distributions. Ma et al. (2011) pointed out that the logarithmic distribution itself is a discrete form of a power law distribution with an exponential cutoff, where the power law exponent is -1 and the upper tail decays exponentially above a cutoff that is directly related to the average group size experienced by an individual. Bonabeau et al. (1999) also presents mechanistic models of group size that lead to power law distributions with exponential decay.

Other heavy-tailed distributions exist and should be considered in a model selection context before concluding that "power law-like" behavior observed in empirical data necessarily indicates a power law distribution (Clauset et al. 2009). These include the Yule and the discretized lognormal distributions, which themselves can be viewed, respectively, as limiting distributions of stochastic preferential attachment or multiplicative growth processes (Clauset et al. 2009, Mitzenmacher 2004). The lognormal distribution has a long history in ecology (e.g., Preston 1948) and a diversity of other fields (Limpert et al. 2001) where it often arises as a plausible alternative to other heavy-tailed distributions like power laws (e.g., in birds; Allen et al. 2001). One classical generative process for a lognormal distribution is the multiplicative stochastic growth process first proposed by Gibrat (1931), in which the size of an entity changes by successive multiplicative random effects; if the multiplicative random effects are independent and lognormally distributed, then the size distribution will be lognormal. The lognormal distribution arises even more generally as a direct consequence of the Central Limit Theorem for products of random variables; any process that involves the product of a sufficiently large number of independent and identically distributed random variables having any distribution with finite mean and variance has a limiting lognormal distribution. Thus, a discretized lognormal distribution of counts could arise from a variety of plausible ecological mechanisms.

Based on this literature survey, a set of eight candidate distributions were identified to describe the distribution of non-zero counts of seabird data, i.e. the non-zero component of the hurdle model described in section 1.2, case (1). These candidate distributions, their parameters, and probability mass functions are listed in Table 1. Some of these distributions naturally have positive integer support (geometric, logarithmic, zeta, zeta with exponential cutoff, and Yule), whereas others include 0 in their natural support set and must be truncated to positive integer support for use in the non-zero component of any hurdle model (Poisson, negative binomial, discretized lognormal). We refer to the discrete power law distribution as the zeta distribution and the discrete power law with exponential cutoff as the zeta exponential.

2.2 Model Fitting and Selection

To derive the reference distribution for non-zero counts to be used in power-analysis for a particular species, we fit each of the eight candidate distributions (Table 1) to available reference data using maximum likelihood estimation (MLE) in the program R (version 2.13.2; R development Core 2011). We used the VGAM package (Yee 2010) to estimate parameters for the positive Poisson, positive negative binomial, geometric, and logarithmic distributions. We

used the methods and code provided in Clauset et al. (2009) to estimate the parameters for the truncated discretized lognormal, the zeta, zeta exponential, and the Yule distributions. Together, these packages define probability mass functions, cumulative probability functions, and maximum likelihood fitting methods for each candidate distribution, and account for zero-truncation when required.

For model selection purposes, we calculated the log-likelihood of each model using the Yee (2010) and Clauset et al. (2009) methods and R code. We used the likelihoods to calculate Akaike's Information Criterion corrected for finite (i.e., small) sample sizes (AICc), which we then used to rank the models (Burnham and Anderson 2002). The model with the lowest AICc was selected for use as a reference distribution and we compared the fit of the top distribution to the fits of the distributions that were ranked 2nd and 3rd using the Vuong closeness test (Vuong 1989). For the models with the lowest AICc values we additionally conducted one-sample Kolmogorov-Smirnov tests to evaluate the null hypothesis that the data could have been drawn from the specified distribution (Sokal and Rohlf 2012). This allowed us to evaluate whether the top-ranked distributions by AICc adequately described the observed data.

The maximum likelihood parameter estimates for the top model were used to define the null hypothesis distribution for subsequent significance tests and power analysis. Most distributions used only one parameter, which we altered to give the specified effect sizes for the alternative hypothesis tests in the power analyses. In cases where the best-fitting distribution had two parameters (e.g., the negative binomial, discretized lognormal, and zeta exponential 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 to the mean under the null hypothesis. This approach requires an assumption that the mean of the distribution changes only as a function of the first parameter, while the second parameter is a shape parameter that remains unchanged for a given species, perhaps for a given region or season. Implications of this assumption are discussed further in section 4.

2.3 Monte Carlo Significance Testing Procedure

As discussed in section 1.2, power analysis requires specification of the test statistic, the significance testing procedure, and the significance level (Type I error rate) for which power is to be evaluated. We have chosen to focus on the mean as our test statistic for abundance data, because the long-term mean count of birds of a particular species in a discrete spatial unit is often a desired quantity for environmental impact assessment. However, it should be noted that other test statistics focusing on other aspects of the distribution could be relevant for specific questions (e.g. median, quantile, or extreme value statistics), and would likely have different power characteristics.

For case (1) described in section 1.2, we use the sample mean, m, as the test statistic to evaluate the one-sample null hypothesis Ho: $\mu=m$, where μ denotes the mean of the reference distribution. We consider two possible one-tailed alternative hypotheses, corresponding to the hotspot case (Ha: $\mu<m$) and the coldspot case (Ha: $\mu>m$). Unless otherwise specified, we use a Type I error rate of $\alpha=0.05$. Because the test statistic is the mean of a possibly small sample, the distribution of the null hypothesis is not readily available in closed form for many of the candidate distributions. Therefore, we derive the critical value for the chosen significance level by a Monte

Carlo method (Hope 1968). Given the sample size, *M*, the upper critical value is estimated by drawing *N* samples of *M* random variates from the reference distribution using an appropriate random number generator, calculating the sample mean for each of the *N* samples, and finding the 1-α quantile of the simulated distribution of sample means. The lower critical value is estimated by finding the α quantile of the same distribution. 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).

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 that are greater than or equal to the observed sample mean. For the coldspot case, the p-value is equal to the proportion of simulated sample means that are less than or equal to the observed sample mean.

In the case of occurrence probability (section 1.2, case (2)), we apply one-sample Fisher's Exact Tests to test the significance of deviations of occurrence probability in a given grid cell from the reference probability (Sokal and Rohlf 2012).

In the case of the full hurdle model (section 1.2, case (3)), we use the same procedure described above for case (1), but simulations use a binomial random number generator to implement the Bernoulli component of the hurdle model process.

All critical value and p-value simulations used N=10000 or more Monte Carlo realizations. Random number generators were implemented in R as described in Yee (2010) and Clauset et al. (2009).

2.4 Monte Carlo Power Estimation

We follow a Monte Carlo, or direct simulation approach to estimate statistical power of the hotspot/coldspot significance tests described in section 2.3. In addition to the choice of test statistic, test procedure, and Type I error rate discussed above, power estimates require specification of the sample size (*M*), reference distribution, and effect size to be used to construct the alternative hypothesis. For case (1), we specify the alternative hypothesis such that it has the same form and nuisance parameters as the reference distribution, but has a mean that differs from the reference distribution mean by a multiplicative effect size. We then simulate *N* realizations of the sample mean under the alternative hypothesis: *N* samples of *M* random variates are drawn from the alternative hypothesis distribution and the mean of each sample is calculated. For the hotspot case, the Monte Carlo power estimate is equal to the proportion of simulated sample means that are greater than the upper critical value of the null hypothesis distribution (found as described in section 2.3). For the coldspot case, the Monte Carlo power estimate is equal to the proportion of simulated sample means that are less than the lower critical value of the null hypothesis distribution.

For case (2), we use the power formula for the one-sample Fisher's Exact Test, as implemented in the Matlab R2012b Statistics Toolbox (The Mathworks, Natick, MA) to calculate the power of this test for different effect sizes (Bennett and Hsu 1960).

For case (3), the full hurdle model, power is calculated using the same procedure described for case (1), but adding a simulation of the Bernoulli process of the hurdle model using a binomial

random number generator. Effect sizes are introduced via the non-zero component of the hurdle model in the same manner described for case (1).

2.5 Power Curves

For six of the eight candidate distributions in Table 1 (positive Poisson, positive negative binomial, geometric, logarithmic, truncated discretized lognormal, and zeta exponential), we used the Monte Carlo procedures described in sections 2.3 and 2.4 to estimate power at a range of sample sizes, from 1 to 100 surveys, and a range of multiplicative effect sizes (coldspot effect sizes:1/3x, 1/2x, 2/3x; hotspot effect sizes: 1.5x, 2x, 3x) for a given reference mean. We repeated these simulations for a representative range of reference means. This resulted in a set of power vs. sample size curves for each of these six candidate distributions (Digital Supplement A). The purpose of these curves is to facilitate prospective power analysis; given an estimate of the reference mean and the effect size of interest, one can use these curves to determine the sample size needed to achieve a desired level of power to detect a hotspot or a coldspot.

We also generated power curves for the one-sample Fisher's Exact Test for differences in occurrence probability (Digital Supplement B), and for the full hurdle model (section 1.2, case (3); Digital Supplement C). In the latter case, we generated power curves for different combinations of both reference non-zero mean and reference prevalence.

Preliminary experimentation with fits of the eight candidate distributions to real seabird data from the U.S. Atlantic indicated that the Yule and zeta distributions did not have finite means for the parameter values typical of observed seabird distributions ($\alpha<1$). Thus, we did not calculate power curves for these distributions, but still include them in the model fitting process. When one of these two models is selected as the best-fitting model and the mean is not finite (because the parameter $\alpha<1$), the implication is that sample-mean-based test statistics are not a reliable way to test for hotspots and coldspots, because the upper tail of the distribution is too "heavy". Instead, non-parametric statistics such as median tests may be more appropriate. Alternatively, removal of trends by accounting for covariates will sometimes reduce the skew of such fat-tailed distributions enough to make the mean well-defined. Or, for simplicity, one could simply use the next best-fitting model and accept that the power estimate will be less accurate.

2.6 Data

To illustrate the power analysis and significance testing methods described in this document, we used at-sea seabird count data extracted from the USGS Avian Compendium Database (O'Connell et al. 2009; Table 2 and Digital Supplement D). The raw data consisted of ship-based and aerial visual observations along fixed-width survey transects, recording the species and number of birds seen in each discrete time transect segment, or at each location along continuous transects. We used a total of 32 datasets that were collected from 1978-2010, 28 of which were ship-based while the remaining four were conducted from fixed-wing aircraft. Most of the surveys (28 total) were conducted using the continuous transects method. The four discrete-time surveys were ship-based and were generally conducted for fixed 15-minute intervals on ships traveling at approximately 10 knots. We segmented all continuous transect survey data into transects of 4.63km, equivalent to the distance covered by a ship moving at 10 knots for 15 minutes in an effort to standardize the data. We eliminated all transect segments shorter than 60% of this distance, and any discrete time surveys shorter than 10 minutes, such that 209

Table 2

Datasets used for analyses. Data from these surveys were extracted from the USGS Avian Compendium Database (O'Connell et al. 2009) and standardized to 15-minute ship survey equivalent transect segments as described in section 2.6.

			Year		Number of transect segments surveyed[c] (15-minute-ship-survey-equivalents[d])				
Source Dataset ID[a]	Platform	Method[b]	Start	End	Total	Spring	Summer	Fall	Winter
BarHarborWW05	Boat	cts	2005	2005	575	0	482	93	0
BarHarborWW06	Boat	cts	2006	2006	650	0	471	179	0
CapeHatteras0405	Boat	cts	2004	2005	275	0	154	0	121
CapeWindAerial	Aerial	cts	2002	2004	2175	528	520	538	589
CapeWindBoat	Boat	cts	2002	2003	119	67	45	7	0
CDASMidAtlantic	Aerial	cts	2003	2003	66	66	0	0	0
CSAP	Boat	dts	1980	1988	23753	7043	6587	6556	3567
EcoMonAug08	Boat	cts	2008	2008	370	0	370	0	0
EcoMonAug09	Boat	cts	2009	2009	350	0	350	0	0
EcoMonAug10	Boat	cts	2010	2010	307	0	302	5	0
EcoMonFeb10	Boat	cts	2010	2010	238	0	0	0	238
EcoMonJan09	Boat	cts	2009	2009	312	0	0	0	312
EcoMonMay07	Boat	cts	2007	2007	383	346	37	0	0
EcoMonMay09	Boat	cts	2009	2009	470	170	300	0	0
EcoMonMay10	Boat	cts	2010	2010	485	233	252	0	0
EcoMonNov09	Boat	cts	2009	2009	354	0	0	354	0
EcoMonNov10	Boat	cts	2010	2010	309	0	0	309	0
GeorgiaPelagic	Boat	dts	1982	1985	2127	675	677	551	224
HatterasEddyCruise2004	Boat	cts	2004	2004	93	0	93	0	0
HerringAcoustic06	Boat	cts	2006	2006	195	0	0	195	0
HerringAcoustic07	Boat	cts	2007	2007	220	0	0	220	0
HerringAcoustic08	Boat	cts	2008	2008	623	0	0	623	0
HerringAcoustic09Leg1	Boat	cts	2009	2009	100	0	0	100	0
HerringAcoustic09Leg2	Boat	cts	2009	2009	196	0	0	196	0
HerringAcoustic09Leg3	Boat	cts	2009	2009	223	0	0	223	0
MassAudNanAerial	Aerial	cts	2002	2006	2029	375	274	467	913
NOAAMBO7880	Boat	dts	1978	1979	6341	1396	2353	1868	724
PlattsBankAerial	Aerial	cts	2005	2005	744	0	744	0	0
SEFSC1992	Boat	cts	1992	1992	30	0	0	0	30
SEFSC1998	Boat	cts	1998	1998	37	0	37	0	0
SEFSC1999	Boat	cts	1999	1999	27	0	0	27	0
TOTALS	**ALL**	**ALL**	**1978**	**2010**	**44176**	**10899**	**14048**	**12511**	**6718**

[a]The Source Dataset ID can be used to look up datasets in Digital Supplement D, Table D1, which gives detailed additional background information about each survey. Table D1 lists several additional datasets; these additional datasets are available but did not contain any segments that fell within the BOEM lease block area.
[b]Survey method: cts, continuous-time strip transects; dts, discrete-time strip transects
[c]Counts exclude segments whose midpoint falls outside BOEM lease blocks (i.e., segments inshore of 3nmi state waters boundary or outside U.S. Exclusive Economic Zone were excluded), and any partial segments that were less than 60% of standard transect segment length (i.e., only segments >2.778km in length were included).
[d]A15-minute-ship-survey-equivalent is defined as the distance a ship would travel in 15 minutes at 10 knots.

transects were removed from our data. This allowed the remaining discrete time and continuous time transect segments to be compared on an approximately common basis, "15-minute-ship-survey-equivalents." This left us with a total of 44,176 transects that covered our reference region (the BOEM Atlantic Outer Continental Shelf [OCS] lease blocks) with approximately 84% having approximate lengths of 4.63km (and the remainder having lengths no less than 2.78km). Although it is likely that this standardization did not fully resolve all differences among survey platforms and protocols, we consider it acceptable for a first-order analysis of retrospective power based on historical survey effort.

We extracted data for all species in the database and chose species with at least 200 observations in a given season to model (Table 3). We selected three species to use as examples in the main body of this report: (a) Herring Gull (HERG), (b) Northern Gannet (NOGA), (c) Wilson's Storm-Petrel (WISP) (Figure 1). Full results for all species and seasons with sufficient data are given in the Digital Supplements (described in Appendix A). HERG, NOGA, and WISP were chosen because they are three of the most abundant species present in the study area in Spring, and so illustrate the best-case performance of the methods.

Data were clipped based on standardized transect segment midpoints to the Atlantic OCS, as defined by BOEM lease block coverage. Each standardized transect segment was assigned to a BOEM lease block based on its centroid. We tabulated the number of samples and the mean of non-zero counts in each BOEM block in which a species was sighted. To reduce temporal dependence of samples, we analyzed data separately by season. Spring is defined as March 1 to May 31, Summer is defined as June 1 to August 31, Fall is defined as September 1 to November 30, and Winter is defined as December 1 to February 28/29 of a calendar year. Within a season, we observed no obvious temporal autocorrelation in observations of the same species on repeated occasions (observations were usually separated by at least several days), and so neglected temporal autocorrelation in analyses. The effects of temporal autocorrelation, if present, are discussed further in section 4. We did not explicitly account for spatial autocorrelation, and instead assume that the spatial scale of the analysis (the size of the discrete spatial unit) has been chosen appropriately to match the scale of spatial autocorrelation. Other approaches are possible, but would add complexity to the method. We discuss the implications of assumptions on spatial scale and correlation in section 4, below.

Table 4 gives the total number of transect segments surveyed and number of transects in which each species was observed in each season. The Herring Gull was observed on 3828 transects (at least one individual), while the Northern Gannet and Wilson's Storm-Petrel were observed on 2749 and 1476 transects, respectively. Herring Gulls were observed in 2791 unique BOEM lease blocks and when present, were observed between 1-25 times in a given lease block. Northern Gannets were observed in 2014 unique BOEM lease blocks and when present, were observed between 1-23 times in a given lease block. Wilson's Storm Petrel were observed in 1244 unique BOEM lease blocks and when present, were observed between 1-5 times in a given lease block. Statistics for other species and seasons are given in Table 4.

Table 3

List of species analyzed. Four-letter species codes in the first column are generally used in place of the full common or scientific name. The “Modeled?” column indicates whether there were sufficient data available to model the species in the indicated season. Only species with >200 sightings in a season were modeled.

				Modeled?			
Species code	Common name	Scientific name	Family	Spring	Summer	Fall	Winter
aush	Audubon's Shearwater	*Puffinus lherminieri*	Procellariidae	No	Yes	Yes	No
blki	Black-Legged Kittiwake	*Rissa tridactyla*	Laridae	Yes	No	Yes	Yes
blsc	Black Scoter	*Melanitta americana*	Anatidae	Yes	No	Yes	Yes
bogu	Bonaparte's Gull	*Chroicocephalus philadelphia*	Laridae	No	No	No	Yes
coei	Common Eider	*Somateria mollissima*	Anatidae	Yes	Yes	Yes	Yes
colo	Common Loon	*Gavia immer*	Gaviidae	Yes	Yes	Yes	Yes
cosh	Cory's Shearwater	*Calonectris diomedea*	Procellariidae	No	Yes	Yes	No
cote	Common Tern	*Sterna hirundo*	Sternidae	Yes	Yes	Yes	No
dove	Dovekie	*Alle alle*	Alcidae	No	No	Yes	Yes
gbbg	Great Black-Backed Gull	*Larus marinus*	Laridae	Yes	Yes	Yes	Yes
grsh	Great Shearwater	*Puffinus gravis*	Procellariidae	Yes	Yes	Yes	No
herg	Herring Gull	*Larus argentatus*	Laridae	Yes	Yes	Yes	Yes
lagu	Laughing Gull	*Leucophaeus atricilla*	Laridae	Yes	Yes	Yes	No
lesp	Leach's Storm-Petrel	*Oceanodroma leucorhoa*	Hydrobatidae	No	Yes	Yes	No
ltdu	Long-tailed Duck	*Clangula hyemalis*	Anatidae	Yes	Yes	Yes	Yes
nofu	Northern Fulmar	*Fulmarus glacialis*	Procellariidae	Yes	Yes	Yes	Yes
noga	Northern Gannet	*Morus bassanus*	Sulidae	Yes	Yes	Yes	Yes
poja	Pomarine Jaeger	*Stercorarius pomarinus*	Stercorariidae	No	No	Yes	No
razo	Razorbill	*Alca torda*	Alcidae	Yes	Yes	Yes	Yes
reph	Red Phalarope	*Phalaropus fulicarius*	Scolopacidae	Yes	No	No	No
rtlo	Red-Throated Loon	*Gavia stellata*	Gaviidae	Yes	No	Yes	Yes
sosh	Sooty Shearwater	*Puffinus griseus*	Procellariidae	Yes	Yes	No	No
susc	Surf Scoter	*Melanitta perspicillata*	Anatidae	Yes	Yes	Yes	Yes
wisp	Wilson's Storm-Petrel	*Oceanites oceanicus*	Hydrobatidae	Yes	Yes	Yes	No
wwsc	White-Winged Scoter	*Melanitta fusca*	Anatidae	Yes	Yes	Yes	Yes

Table 4a

Summary of species data and best-fitting model of non-zero counts for the Spring season. Species codes are as in Table 3. Footnotes are given below Table 4d. 10,899 standardized transect segments were used to calculate prevalence.

Species code	Season	Maps created?[a]	Number of Observations[b]	Total Number Observed[c]	Prevalence[d]	Reference mean[d]	Best Fitting Model (by AICc[e])	K-S statistic[f]	K-S statistic p-value[g]
herg	Spring	Yes	3828	41959	0.351	9.514	Discretized lognormal	0.014	0.465
noga	Spring	Yes	2749	21114	0.252	6.877	Discretized lognormal	0.017	0.431
gbbg	Spring	No	2422	22527	0.222	9.301	Yule	0.044	0.000*
nofu	Spring	Yes	1700	22149	0.156	10.839	Discretized lognormal	0.008	1.000
wisp	Spring	Yes	1476	12092	0.135	6.222	Discretized lognormal	0.019	0.669
colo	Spring	Yes	976	2449	0.090	2.508	Discretized lognormal	0.011	1.000
ltdu	Spring	Yes	770	66676	0.071	40.517	Discretized lognormal	0.027	0.647
sosh	Spring	No	691	5375	0.063	8.114	Discretized lognormal	0.013	1.000
susc	Spring	No	652	15303	0.060	26.731	Discretized lognormal	0.042	0.206
blki	Spring	No	568	2223	0.052	4.041	Discretized lognormal	0.024	0.905
coei	Spring	Yes	541	82582	0.050	166.5	Discretized lognormal	0.039	0.378
grsh	Spring	Yes	510	3934	0.047	7.335	Discretized lognormal	0.027	0.857
wwsc	Spring	Yes	469	4358	0.043	9.212	Discretized lognormal	0.026	0.903
razo	Spring	No	457	2962	0.042	6.474	Negative binomial	0.031	0.787
lagu	Spring	No	394	852	0.036	2.151	Discretized lognormal	0.011	1.000
cote	Spring	Yes	362	1729	0.033	4.652	Discretized lognormal	0.015	1.000
reph	Spring	Yes	361	84170	0.033	213.77	Discretized lognormal	0.067	0.081
rtlo	Spring	No	312	742	0.029	2.212	Zeta exponential	0.056	0.283
blsc	Spring	No	243	2950	0.022	13.303	Discretized lognormal	0.035	0.934

Table 4b

Summary of species data and best-fitting model of non-zero counts for the Summer season. Species codes are as in Table 3. Footnotes are given below Table 4d. 14,048 standardized transect segments were used to calculate prevalence.

Species code	Season	Maps created?[a]	Number of Obser-vations[b]	Total Number Observed[c]	Preval-ence[d]	Refer-ence mean[d]	Best Fitting Model (by AICc[e])	K-S statistic[f]	K-S statistic p-value[g]
wisp	Summer	Yes	5529	72611	0.394	11.120	Discretized lognormal	0.021	0.012
grsh	Summer	Yes	3926	104432	0.279	12.332	Discretized lognormal	0.016	0.260
gbbg	Summer	Yes	2255	8721	0.161	3.284	Discretized lognormal	0.015	0.684
herg	Summer	No	1943	7934	0.138	4.083	Yule	0.043	0.001
cosh	Summer	No	1561	8382	0.111	4.736	Discretized lognormal	0.020	0.555
sosh	Summer	No	1151	26077	0.082	22.656	Yule	0.015	0.962
lesp	Summer	No	860	3853	0.061	3.829	Discretized lognormal	0.013	0.999
cote	Summer	Yes	715	3872	0.051	4.995	Discretized lognormal	0.034	0.384
noga	Summer	No	651	1339	0.046	2.057	Yule	0.009	1.000
lagu	Summer	Yes	558	1871	0.040	3.274	Discretized lognormal	0.014	1.000
nofu	Summer	No	492	3031	0.035	4.973	Discretized lognormal	0.034	0.637
ltdu	Summer	Yes	486	9400	0.035	20.524	Discretized lognormal	0.020	0.990
susc	Summer	No	437	7024	0.031	16.097	Negative binomial	0.030	0.829
coei	Summer	No	348	30984	0.025	152.61	Discretized lognormal	0.037	0.719
colo	Summer	No	343	552	0.024	1.610	Geometric	0.016	1.000
aush	Summer	No	316	915	0.022	2.896	Yule	0.030	0.935
wwsc	Summer	No	279	2250	0.020	8.069	Logarithmic	0.061	0.245
razo	Summer	Yes	253	1842	0.018	7.422	Discretized lognormal	0.037	0.886

Table 4c

Summary of species data and best-fitting model of non-zero counts for the Fall season. Species codes are as in Table 3. Footnotes are given below Table 4d. 12,511 standardized transect segments were used to calculate prevalence.

Species code	Season	Maps created?[a]	Number of Observations[b]	Total Number Observed[c]	Prevalence[d]	Reference mean[d]	Best Fitting Model (by AICc[e])	K-S statistic[f]	K-S statistic p-value[g]
herg	Fall	Yes	5088	42441	0.407	7.243	Discretized lognormal	0.019	0.049
grsh	Fall	Yes	4101	61596	0.328	13.689	Discretized lognormal	0.009	0.914
gbbg	Fall	Yes	3640	27240	0.291	6.473	Discretized lognormal	0.023	0.039
noga	Fall	Yes	2635	11044	0.211	3.960	Discretized lognormal	0.009	0.984
blki	Fall	Yes	1675	9972	0.134	5.297	Discretized lognormal	0.017	0.731
cosh	Fall	Yes	1210	6421	0.097	5.001	Discretized lognormal	0.017	0.868
nofu	Fall	Yes	1151	6634	0.092	4.700	Discretized lognormal	0.013	0.986
wisp	Fall	No	820	4663	0.066	5.687	Yule	0.047	0.051
colo	Fall	No	759	1587	0.061	2.094	Discretized lognormal	0.008	1.000
ltdu	Fall	No	755	17456	0.060	26.381	Discretized lognormal	0.557	0.000*
lagu	Fall	Yes	690	2706	0.055	3.907	Discretized lognormal	0.747	0.228
susc	Fall	Yes	688	23414	0.055	34.036	Negative binomial	0.027	0.677
coei	Fall	No	554	48067	0.044	83.989	Zeta exponential	0.042	0.292
wwsc	Fall	No	534	7547	0.043	15.236	Discretized lognormal	0.031	0.682
poja	Fall	No	533	742	0.043	1.392	Logarithmic	0.006	1.000
cote	Fall	No	431	3637	0.034	9.080	Discretized lognormal	0.032	0.754
razo	Fall	Yes	329	2709	0.026	8.254	Discretized lognormal	0.031	0.919
blsc	Fall	Yes	315	5032	0.025	16.699	Discretized lognormal	0.040	0.707
rtlo	Fall	Yes	282	899	0.023	2.946	Zeta exponential	0.047	0.554
dove	Fall	No	229	2502	0.018	11.438	Discretized lognormal	0.023	1.000
lesp	Fall	No	221	521	0.018	2.361	Discretized lognormal	0.023	1.000
aush	Fall	No	206	773	0.016	3.752	Yule	0.041	0.888

Table 4d

Summary of species data and best-fitting model of non-zero counts for the Winter season. Species codes are as in Table 3. Footnotes are given below. 6,718 standardized transect segments were used to calculate prevalence.

Species code	Season	Maps created?[a]	Number of Observations[b]	Total Number Observed[c]	Prevalence[d]	Reference mean[d]	Best Fitting Model (by AICc[e])	K-S statistic[f]	K-S statistic p-value[g]
herg	Winter	Yes	2817	22978	0.419	6.581	Discretized lognormal	0.017	0.362
blki	Winter	Yes	2745	26918	0.409	8.094	Discretized lognormal	0.029	0.022
gbbg	Winter	No	2450	27719	0.365	11.314	Yule	0.044	0.000*
noga	Winter	Yes	1904	13503	0.283	5.226	Discretized lognormal	0.021	0.371
nofu	Winter	Yes	1281	17241	0.191	11.219	Discretized lognormal	0.027	0.304
ltdu	Winter	Yes	1277	60860	0.190	39.860	Discretized lognormal	0.015	0.949
susc	Winter	Yes	1008	40571	0.150	43.806	Discretized lognormal	0.023	0.634
coei	Winter	Yes	862	152448	0.128	253.856	Discretized lognormal	0.034	0.264
razo	Winter	No	848	6983	0.126	8.607	Discretized lognormal	0.028	0.511
wwsc	Winter	Yes	822	13398	0.122	15.957	Discretized lognormal	0.020	0.887
colo	Winter	Yes	803	1706	0.120	2.094	Discretized lognormal	0.009	1.000
dove	Winter	Yes	423	2759	0.063	6.450	Discretized lognormal	0.022	0.989
blsc	Winter	No	403	5753	0.060	13.971	Zeta exponential	0.037	0.640
bogu	Winter	No	351	2972	0.052	8.467	Yule	0.017	1.000
rtlo	Winter	No	341	667	0.051	1.956	Yule	0.013	1.000

[a]Power analysis and related maps were only produced when Vuong closeness tests indicated that the leading model was clearly better than the runner-up model (as ranked by AICc). See Digital Supplements F and G for maps.
[b]Number of 15-minute-ship-survey-equivalent transect segments in which at least one individual of the species was observed.
[c]Total number of individuals observed summed over all transect segments surveyed this season.
[d]Prevalence is the proportion of all standardized transect segments this season in which this species was observed. Reference mean is the mean of the best-fitting distribution to all count data for this species. Since fitted count distributions are all >0, the reference mean refers to the mean conditional on presence (i.e., the average of non-zero counts). The unconditional mean (accounting for zeros when a species is absent) can be found by multiplying the prevalence by the reference mean.
[e]AICc, Akaike's Information Criterion corrected for small sample size.
[f]Nonparametric one-sample Kolmogorov-Smirnov test statistic for discrete distributions, measuring the maximum difference between the empirical cumulative distribution of the data and the theoretical cumulative distribution of the selected model.
[g]P-value for the K-S statistic. The null hypothesis is that the data could have been a sample from the selected distribution. Therefore, non-significant p-values suggest that the selected model is a good fit. Significant p-values are starred (*) (using α=0.05 and Bonferroni-corrected significance threshold for 74 tests=0.05/74=0.000676). In these cases the data suggest significant departures from the selected model.

2.7 Species-specific Power Maps and Curves

For each modeled species/season combination (Table 3), we fit and selected the best candidate distribution as described in section 2.2 using the data from Atlantic OCS as the reference region (example fits are shown in Figure 2 and full fit information is given in Digital Supplement E). Taking this best-fitting model as the reference distribution, we then calculated the power to detect a hotspot of effect size 3x and a coldspot of effect size 1/3x in each BOEM lease block on the Atlantic OCS, given the number of surveys that had occurred in that lease block. The reference mean was calculated by averaging 1,000,000 random draws from the reference distribution, except in the case of the Yule and zeta distributions, which generally did not have a finite mean; in these cases, the sample mean of all non-zero data was used as the reference mean. Power was not calculated for lease blocks that had not been surveyed and we did not include data where no individuals from that species were observed. Thus a hotspot (coldspot) is defined as a lease block where the mean, given that the species is present, is >3x (<1/3x) the reference mean, which is also conditional on presence. The resulting power estimates were mapped for an example region in the Mid-Atlantic, and are presented in Figures 3-5 and in Digital Supplement F. We then generated power curves for each example species showing power for each of the actual sample sizes that were encountered in the historical dataset (Figure 6 and Digital Supplement F). Digital Supplement F also contains summary maps showing the number of occurrences and the mean of non-zero counts for each mapped species, and for all species combined in each season and over all seasons in which the species occurred (see Appendix A for details).

We repeated these procedures for the full hurdle model (section 1.2, case (3)), which jointly considers both zero and non-zero counts, and report the results of these power analyses in Digital Supplement G. Digital Supplement G also contains summary maps showing the number of times each lease block was surveyed (by season and overall) and the mean of all counts (including zeros) for each mapped species, and for all species combined in each season and over all seasons (see Appendix A for details).

2.8 Species-specific Significance Maps

Using the same best-fitting reference distribution for each example species to specify the null hypothesis, we followed the procedure in section 2.3 to estimate the p-value for independent Monte Carlo significance tests of the sample mean of each surveyed BOEM lease block against one-tailed hotspot and coldspot alternative hypotheses. We then produced combined maps of p-values for potential hotspots and coldspots (Figures 7-9 and Digital Supplements F and G). In these combined maps, blocks that were not identified as a potential hot or coldspot are shaded to indicate how confident we can be in that result, based on the average of power to detect a 3x hotspot or a 1/3x coldspot. The symbology of these maps is described in detail in the associated figure captions. *P*-value maps for the non-zero conditional model for the three example species are presented in Figures 7-9. Digital Supplement F presents non-zero conditional model (section 1.2, case (1)) results for the rest of the species/seasons, and Digital Supplement G presents p-value maps for the full hurdle model (section 1.2, case (3)).

The purpose of these maps is to allow spatial planning blocks (e.g., BOEM lease blocks) to be separated into several qualitative categories based on power analysis and significance testing using available datasets. The darkest blue lease blocks can be regarded as the most significant

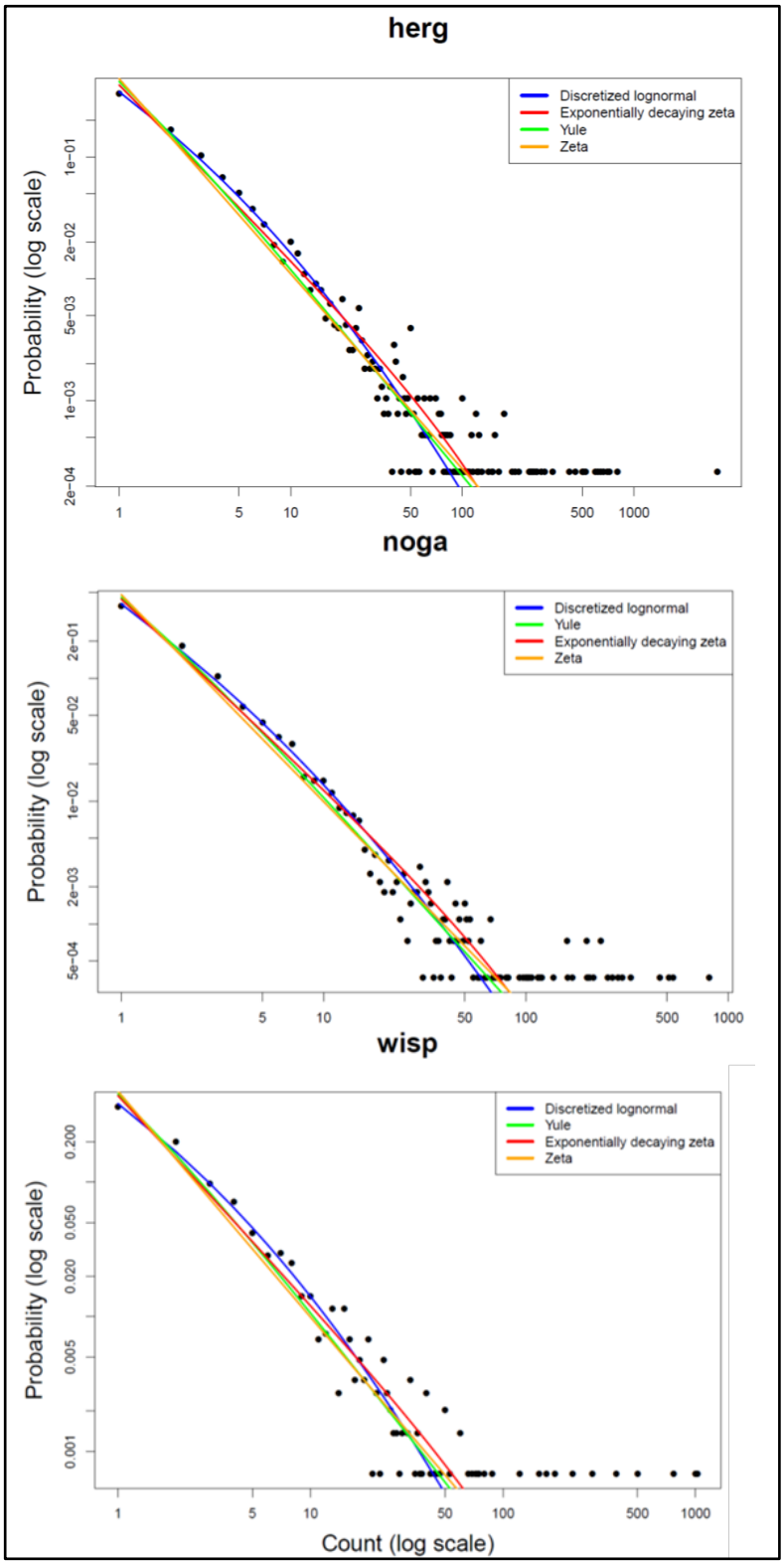


Figure 2. Maximum likelihood model fits (lines) and observed probabilities (black dots) for non-zero count data for the three example species. Fits are shown for the top four models, ranked from lowest to highest AICc.

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. Light grey shading indicates lease blocks not identified as significant hotspots or coldspots, but for which there was little or no power to detect a hotspot or coldspot, had it existed. Finally, blank (white) polygons indicate lease blocks that were not surveyed.

Several caveats apply to these maps. First, they reflect only the available datasets that were used in this study (Table 2, Digital Supplement D). Second, hotspot (coldspot) significance does not consider whether high (low) abundances persisted across years or occurred in the same year. If interannual persistence is of concern, the temporal distribution of the data should be examined. Finally, a very high p-value threshold (p<0.2) has been chosen to flag lease blocks as possible hot or coldspots. Thus the shaded blocks in Figures 7-9 (especially those with the lightest red or blue shading) represent only potential hot or coldspots, many of which are likely to be false positives. This issue is compounded by the fact that p-values have not been corrected for the large number of simultaneous tests performed (two tests for each lease block that was surveyed in this season). This is particularly true of p-value maps for the full hurdle model (Digital Supplement G), where many more simultaneous statistical tests were performed per map because of the larger number of lease blocks considered. The number of false-positives will be correspondingly higher in the full model p-value maps. The most significant values (darkest red and blue) are more reliable, but will still contain some false positives. Similarly, the lightest grey cells have the highest chance of being false negatives, whereas the darkest grey cells have the lowest chance of being false negatives.

Using the underlying data in a geographic information system, the p-value threshold for flagging lease blocks as potential hotspots or coldspots could be adjusted to balance the risk of false positives and false negatives for a particular application.

2.9 Summaries of Species-specific Power Curves and Maps

To examine and display the general patterns evident in species-specific power analyses, we generated statistical summaries of species-specific power curves and species-specific power maps.

Each species' power curve was first approximated by fitting a regression model. In agreement with theory (Murphy et al. 2008), power curves were found to be approximately linear when power was transformed with a Probit transformation (inverse normal cumulative distribution function, $\Phi^{-1}(p)$) and sample size with a square root transformation. We thus used ordinary linear regression to estimate the following model for each simulated power curve for which sufficient non-zero points were available:

$$\text{Probit}(p) = \Phi^{-1}(p) = a + b \times \sqrt{n}$$

Power curves fitted to simulation results were then evaluated at sample sizes ranging from 1 to 200 and back-transformed to the original units of power vs. number of samples. This was done for 3x hotspot and 1/3x coldspot power for both the conditional (non-zero) and full (zero and non-zero) models. For each of these cases, we then plotted the median power as a function of sample size by season (Figure 10), where the median was calculated using all available species power curves for that season. The median power curves (solid lines in Figure 10) were plotted

with the 2.5th and 97.5th percentiles (dashed lines) to show the variability in power curves among species (Figure 10).

We also computed summary statistic maps (average, minimum, and maximum) of 3x hotspot and 1/3x coldspot power for both conditional and full models, for all species in each season, and for all species combined over all seasons.

For the conditional model case, which depends on a species being present, an average/minimum/maximum power value was calculated for each BOEM lease block that was sampled in the season and for which at least one of the analyzed species occurred. Species which were analyzed in a season but did not occur in a particular block contributed a power value of 0 to the average/minimum/maximum calculation. In these conditional model summary maps (Figures 11a, 11b and Digital Supplement F), lease blocks that did not have at least one occurrence of one of the analyzed species in a given season are shown as "no data" (blank grid cells). The all-species/all-season summary maps were created by averaging (or taking minimum, maximum) of the four seasonal power summary maps. In this case, blocks for which at least one occurrence of an analyzed species happened in more than zero but less than four seasons are counted as 0 power for seasons in which there were no occurrences. "No data" or blank grid cells represent lease blocks in which none of the analyzed species occurred in any season.

For the full model case, an average/minimum/maximum power value was calculated for every block sampled in a season (Figures 12a, 12b and Digital Supplement G). The all-species/all-season summary maps were created by averaging (or taking minimum, maximum) of the four seasonal power summary maps. In this case, blocks surveyed in more than than zero but less than four seasons are counted as 0 power for seasons in which they were not surveyed. "No data" or blank grid cells represent lease which were never surveyed in any season.

In addition to these statistical summaries, the logical process flow of the power analysis methodology was summarized in a flowchart-style "decision-tree," indicated how the various components of the process fit together, decisions that must be made at each step, and external information that is needed (Figure 13).

2.10 Analyses of Environmental Time Series

Changes in the ocean environment can exert a strong influence on marine bird occurrence and abundance (Tremblay et al. 2009), and environmental variables such as sea surface temperature (SST) and surface chlorophyll-a concentration (*chl*) have been found to be important predictors of seabird occurrence and abundance in the U.S. Atlantic (Kinlan et al. 2012). Since our definition of hotspots and coldspots does not explicitly account for the possibility of environmental variability, we analyzed time series of known correlates (SST, *chl*) of marine bird abundance to determine the temporal extent and resolution of sampling necessary to capture observed variance.

To characterize the proportion of total variance in a time series as a function of temporal scale, we use a statistical plot known as a semivariogram (or variogram, for short), which plots the average variance between pairs of observations separated by a given amount of time (the time lag) (Deutsch & Journel 1998). At longer time lags, the variance observed tends to approach that

of the overall sample variance. At shorter time lags, the variance is lower due to temporal autocorrelation of observations. The empirical variogram, $\gamma(h)$, is calculated as:

$$\gamma(h) = \frac{1}{2N(h)} \cdot \sum_{i=1}^{N(h)} \left(z(t_i) - z(t_i + h)\right)^2$$

where *h* is the time lag, *N(h)* is the number of pairs of observations available at that lag, and *z(t)* are sample data at time t. In practice, we use lag intervals and *h* is the midpoint of the time interval. To aid interpretation, we plot the relative semivariance *r*(*h*), which is the semivariance expressed as a fraction of the total sample variance; a reference line plotted at *r*(*h*)=1 indicates 100% of the sample variance. Because we are only able to estimate the variance at each time lag, the value of *r*(*h*) will not be precisely equal to 1 at long time lags, but it will tend to approach and fluctuate around 1 (Deutsch & Journel 1998).

Daily time series of 3-day composite night-and-day SST and 3-day composite sea surface chlorophyll-a concentration were obtained from the NOAA CoastWatch U.S. East Coast dataset, derived from the MODIS instrument on board the Aqua satellite, available on the internet at the following locations (Foley 2012):

SST: http://coastwatch.pfeg.noaa.gov/erddap/info/erdMEssta3day/index.html

Chl: http://coastwatch.pfeg.noaa.gov/erddap/info/erdMEchla3day/index.html

The time series covered the period from July 5, 2002 to September 11, 2011 (approximately 9 years) at a horizontal pixel resolution of approximately 1.1km. Time series were extracted in four regions of interest chosen based on the approximate location of BOEM's Wind Energy Areas as of October, 201 (Figures 14a, 15a). Within each region, relative semivariograms were calculated in the time direction using Matlab R2011b (The Mathworks, Natick, MA). *Chl* was $\log_{10}(x+1)$ transformed for approximate normality and homoscedasticity prior to analysis. The seasonal cycle was removed prior to variogram analysis by subtracting the monthly climatology for each region. Observations were binned in one day intervals with midpoints ranging from 0.5 to 1677.5 ± 0.5days (approximately 4.5 years). The resulting variograms reflect the average pattern of temporal variance in each region of interest (Figures 14b-d, 15b-d).

Because of the length of the satellite data series, the maximum time scale that could be resolved in the SST and *chl* analyses was 4.5 years. To examine longer time scales of variation, we considered 65-year time series (1948-2012) of two well-characterized regional ocean/atmosphere climate indices known to correlate with changes in spatio-temporal patterns of marine bird occurrence and abundance, the North Atlantic Oscillation (NAO) and the Atlantic Multidecadal Oscillation (AMO) (Veit and Montevecchi 2006, Tremblay et al 2009). Monthly time series of both indices were obtained from NOAA's Earth System Research Laboratory, and are available on the internet at the following locations:

NAO: http://www.esrl.noaa.gov/psd/data/correlation/nao.data

AMO: http://www.esrl.noaa.gov/psd/data/correlation/amon.us.data

The North Atlantic Oscillation was low-pass filtered with a simple rectangular 5 month running mean filter to remove high-frequency variability. Semivariograms were calculated for both time series, allowing assessment of timescales of long-term regional climate variability to a maximum time lag of 20 years. Observations were binned in one month intervals with midpoints ranging from 0.5 to 240.5 ± 0.5 months (approximately 20 years).

2.11 Analyses of Marine Bird Abundance Time Series

It is also of interest to directly analyze patterns of temporal variance in marine bird abundance, rather than indirect environmental correlates. Analyses of long-term temporal autocorrelation in species-specific marine bird abundances are needed to inform the allocation of sampling effort over time, and analyses of short-term temporal autocorrelation (day-to-day within a season) are needed to inform the selection of sampling intervals to maximize statistical independence of surveys. However, repeated surveys of the same discrete area at our scale of interest (BOEM lease blocks) are relatively rare, and the skewed and zero-inflated nature of marine bird count distributions further complicates temporal correlation analyses. It is for this reason that the analyses of environmental correlates of bird occurrence and abundance previously described (section 2.10) are important.

Nevertheless, by integrating many datasets, the USGS Avian Compendium database contains some repeat survey information for a subset of BOEM lease blocks, which enables temporal autocorrelation analyses for some species, albeit with less precision in the resulting semivariograms than for the much more densely sampled environmental time series.

Count data for repeat surveys of the same species in the same BOEM block in standardized transect segments (section 2.6) were extracted from the USGS Avian Compendium Database, counts were $\log_{10}(x+1)$ transformed to improve normality and homoscedasticity, and temporal semivariogram analysis was conducted as described in section 2.10 for each species with sufficient data in each season.

We conducted two separate variogram analyses of the repeat survey avian count data, one to address long-term variability among years (time scales 1-10 years), and the other to address short-term variability within seasons (time scales of 1-60 days). The long-term variability analysis binned observations in 30-day intervals with midpoints ranging from 1 to 3631 ± 15days (approximately 10 years). The short-term variability analysis binned observations in 1-day intervals with midpoints ranging from 1 to 60 ± 2 days.

3.0 RESULTS

3.1 Power Curves

Digital Supplement A shows power vs. sample size curves for six of the eight candidate distributions (the six distributions with finite means at reasonable parameter values) for the non-zero conditional model (section 1.2, case (1)). For each distribution, curves were generated for a range of reference means, and for each reference mean, curves are shown for a range of multiplicative effect sizes relative to that reference mean. Curves show that the required sample size to achieve a given level of power to detect a given effect size depends on a species' average abundance and the type of distribution. These types of curves are intended to serve as useful guides for prospective power analysis and survey design.

Several general features of the power curves are notable. Power to detect a hotspot of a given multiplicative magnitude (e.g., 3x) is not necessarily the same as power to detect a coldspot of the same multiplicative magnitude (e.g., 1/3x). Moreover, the relationship between power and the reference mean is dependent on the type of distribution: for the Poisson and negative binomial, power increases for larger reference means, all else being equal, whereas for the

geometric and logarithmic power to detect differences in the non-zero mean decreases for more abundant species. However, it is important to note that these features of the power curves are dependent in part on our assumption that the second (shape or dispersion) parameter of two-parameter distributions remains constant and changes in the mean occur only through the first parameter. If the parameters change jointly with the mean, then the relative shapes of the power curves could change. Future work should carefully explore observed parameter correlations in multi-parameter distributions fitted to a variety of real datasets.

Digital Supplement B shows power vs. sample size curves for tests of occurrence probability hotspots and coldspots (section 1.2, case (2)), for a range of values of the reference prevalence (Figure B1). Figure B2 illustrates how the expected number of non-zero observations relates to occurrence probability, and can serve as a guide to the number of surveys required to achieve a certain number of detections of a species. This may be useful in the case of planning studies of rare species, in particular.

Digital Supplement C shows power vs. sample size curves for six of the eight candidate distributions (the six distributions with finite means at reasonable parameter values) for the full hurdle model (section 1.2, case (3)). Curves are shown for various combinations of both the non-zero reference mean and the reference prevalence. Note that the allowance for zero observations in the full hurdle model reduces the power achieved for a given sample size. This effect is particularly pronounced for rare (i.e., low prevalence) species.

3.2 Model fitting and selection

Table 4 shows the top-ranked distribution and goodness-of-fit Kolmogorov-Smirnov (K-S) statistic for each species. Table 5 shows maximum likelihood parameter estimates for each of the eight candidate distributions, for each of the three example species data sets. Digital Supplement E (Table E1, Figures E1-74) shows model fits and fit statistics for the remaining species/season combinations. For each species, the distributions are ranked from lowest to highest AICc. For all three example species (Table 5) and the majority of all species/seasons combinations (Table 4, Digital Supplement E), the best-fitting model selected by the AICc method was the discretized lognormal distribution. K-S tests generally indicated the data were consistent with the best-fitting distributions (null hypotheses that data were a sample from the fitted distribution were not rejected), with a few exceptions noted in Table 4. A significant K-S test statistic indicates that the data deviate from what would be expected if they were drawn from the fitted distribution, and suggest that an alternative model (not in the candidate model set) might be more appropriate. In some cases, this may indicate non-stationarity (trends in space and/or time). Addition of covariates to remove trends may improve the fit of the candidate distributions to residuals. For the three example species (Table 5) and the vast majority of all species/season combinations (Digital Supplement E), the negative binomial, geometric, and Poisson did not fit as well as the discretized lognormal (based on AICc), which is interesting given that Poisson and negative binomial are two of the most commonly used distributions for modeling avian count data.

Figure 2 shows the top four model fits for each of the three example species overlaid on data frequencies, on log-frequency vs. log-count axes. Similar plots for other species/seasons are shown in Digital Supplement E. For the three examples, after the discretized lognormal, the closest alternative models were the zeta exponential, the Yule, and the zeta in all cases, although

Table 5

Model fitting and selection example: maximum likelihood estimates of best-fitting parameters of each candidate distribution to non-zero counts for three example species, with AICc and log-likelihood values. For each species, the models are ranked from lowest to highest AICc.

		Parameter estimates		AICc Rank	AICc	Log-Likelihood
Herring Gull (Spring)						
	Discretized lognormal	μ=0.138	σ=1.857	1	20473.03	-10234.51
	Zeta exponential	a=0.422	λ=0.006	2	20644.87	-10320.43
	Yule	a=0.711		3	20699.00	-10348.50
	Zeta	a=0.599		4	20884.84	-10441.42
	Logarithmic	p=0.976		5	21214.38	-10606.19
	Negative binomial	μ=0.206	k=0.005	6	21231.33	-10613.67
	Geometric	p=0.091		7	25628.78	-12813.39
	Poisson	λ=10.961		8	157322.40	-78660.20
Northern Gannet (Spring)						
	Discretized lognormal	μ=0.367	σ=1.870	1	13042.51	-6519.253
	Yule	a=0.835		2	13114.06	-6556.027
	Zeta exponential	a=0.526	λ=0.008	3	13116.56	-6556.278
	Zeta	a=0.684		4	13230.19	-6614.093
	Logarithmic	p=0.962		5	13605.86	-6801.929
	Negative binomial	μ=0.281	k=0.012	6	13632.13	-6814.064
	Geometric	p=0.130		7	16334.22	-8166.111
	Poisson	λ=7.677		8	70701.25	-35349.62
Wilson's Storm-Petrel (Spring)						
	Discretized lognormal	μ=0.009	σ=1.683	1	7004.017	-3500.005
	Yule	a=0.836		2	7067.586	-3532.792
	Zeta exponential	a=0.539	λ=0.006	3	7090.827	-3543.409
	Zeta	a=0.680		4	7144.087	-3571.042
	Logarithmic	p=0.965		5	7386.321	-3692.159
	Negative binomial	μ=0.259	k=0.010	6	7400.027	-3698.010
	Geometric	p=0.122		7	8974.693	-4486.345
	Poisson	λ=8.190		8	48571.69	-24284.84

the ranking differed from species to species. Pairwise Vuong tests indicated that the discretized lognormal was significantly closer to the true model than any of these alternative distributions for all species ($p<0.05$). It is notable that these alternative distributions are all of the power law type with exponents $\alpha < 1$. In our parameterization of the power law distributions (Table 1), these distributions have infinite variance when $\alpha < 2$ and infinite mean when $\alpha < 1$. Thus, sample mean-based power analysis would not be appropriate for fitted parameter ranges. The discretized lognormal has comparatively less probability in the upper tail, and more probability for moderate counts. We use the discretized lognormal with parameter estimates in Table 5 as the reference distribution for all subsequent power analyses and significance tests for the three example species. For other species/season combinations, we used the best-fitting distribution identified in Table 4 and detailed in Digital Supplement E. The discretized lognormal distribution consistently arose as the best-fitting distribution, with some exceptions (Table 4).

To focus on the most robust results, power analyses and significance tests were only carried out when pairwise Vuong tests indicated that the top-ranked model was significantly closer to the true model than its closest competitor. The third column of Table 4 indicates the species/season combinations that passed this test ("Maps created? = "Yes" if the Vuong test was passed). Note that species maps in Digital Supplements F and G follow the same ordering as in Table 4, but with species that did not pass the Vuong test omitted. Note also that the statistical power of the Vuong test decreases as the number of non-zero observations for a species decreases, so that species with low prevalence are less likely to have one clear "winner" among the candidate models. Of species with reference prevalences <10%, only 40% passed the Vuong test and were mapped, whereas 80% of species with reference prevalences >10% were mapped. There are other options to handle cases where no clear winner is identified by the Vuong test. For example, one could run the power analysis using all plausible models, and calculate a weighted multi-model average the results, using Akaike model weights derived from AICc values (Burnham and Anderson 2002). Multi-model p-values for significance tests of hotspots and coldspots could be calculated for these species in a similar way.

3.3 Species-specific Power Maps

Using the selected best maximum likelihood fit for each species as the reference distribution (Table 4, Digital Supplement E), we calculated species-specific power maps on the BOEM lease block grid (Figures 3-5, Digital Supplements F and G). Under the conditional model (section 1.2, case (1)), figures 3, 4, and 5 show the estimated power to detect a hotspot at least 3x the reference mean (Figures 3a, 4a, 5a) or coldspot 1/3x the reference mean (Figures 3b, 4b, 5b) in BOEM lease blocks in the Mid-Atlantic region. These maps are based on available historical survey effort from the USGS Avian Compendium database for the three example species. Similar conditional model power maps for the remaining modeled species/seasons are given in Digital Supplement F, with species ordered as in Table 4. Power maps for the full hurdle model (section 1.2, case (3)) are given in Digital Supplement G.

For Herring Gull in Spring (Figure 3), conditional power to detect a 3x hotspot ranges from 17% to 55%, and conditional power to detect a 1/3x coldspot ranges from 0% to 80%. For Northern Gannet in spring (Figure 4), conditional power to detect a 3x hotspot ranges from 16% to 45%, and conditional power to detect a 1/3x coldspot ranges from 0% to 80%. For Wilson's Storm-Petrel in spring (Figure 5), conditional power to detect a 3x hotspot ranges from 18% to 28%,

and conditional power to detect a 1/3x coldspot ranges from 0% to 26%. In general, the region of highest power is concentrated in Nantucket Sound where intensive survey efforts have been conducted.

The same general features are evident in the all-species/all-seasons summary maps of average hot and coldspot power under the conditional (Figure 11, Digital Supplement F) and full hurdle (Figure 12, Digital Supplement G) models. Average power to detect 3x hotspots and 1/3x hotspots is generally <10% for most lease blocks, and coldspot power is lower than hotspot power for these blocks. Coldspot power was close to zero in many more lease blocks than for hotspot power (light gray shading, Figures 11b and 12b). However, there are several regions of moderate to high power (20 to 65%), including Nantucket Sound. In these better-sampled areas average coldspot power was similar to or greater than average hotspot power.

Differences between the conditional model (Figure 11) and full model (Figure 12) are controlled by two competing factors: for a given sample size, full model power is always less than conditional model power, but the species is unlikely to be observed in all samples. However, because the full model considers all surveys, rather than only those in which the species of interest is present, the actual power estimate may be higher for the full model.

3.4 Species-specific Power Curves

Species-specific power curves from the conditional model (Figures 3-5, Digital Supplement F) show that power to detect a 1/3x coldspot starts lower than power to detect a 3x hotspot for very small sample sizes, but increases more rapidly with the number of samples. 50% (80%) power to detect 1/3x coldspots is attained with approximately 10 (25) repeat surveys for Herring Gulls under the conditional model, whereas these power levels would require many more samples for the 3x hotspot case: 23 surveys to reach 50% power, and >50 surveys to reach 80% power. Patterns for Northern Gannet and Wilson's Storm-Petrel are similar. Given its lower frequency of occurrence in this season (relatively few non-zero abundances), very low conditional power was achieved for Wilson's Storm-Petrel (<30%; but note that x-axis only ranges up to 5 surveys because no lease block contained more than 5 WISP presences in the historical dataset).

The summary power curves (Figure 10) confirm the general finding that power to detect 1/3x coldspots is lower than power to detect 3x hotspots for small sample sizes, but then rapidly increases to become higher than hotspot power for larger sample sizes. There is substantial variability among species power curves within a season, with the widest range in power curves generally occurring in spring and winter. There were some differences in seasonal median power curves; for example, power tended to be lower in spring and higher in fall. However, these differences in median power curves were much smaller than the species-to-species variation. Seasonal differences in median power curves are most likely driven by differences in the composition of modeled species, although they may also be related to behavioral differences associated with different seasons (migration, breeding, foraging) and to seasonal differences in the spatio-temporal distribution of ocean habitat.

3.5 Species-specific Significance Maps

Using the selected best maximum likelihood fit for each species as the reference distribution (Table 4, Digital Supplement E), we calculated species-specific p-values for one-sample Monte

Carlo significance tests for both conditional model (Figures 7-9, Digital Supplement F) and full model (Digital Supplement G) cases. These are presented in combined hotspot/coldspot p-value maps, in which lease blocks with p-values<0.2 are displayed in blue (coldspot) or red (hotspot) shading, with darker shading corresponding to higher statistical significance. All such lease blocks should be regarded as potential rather than certain hotspot/coldspots, as discussed in section 2.8. It is highly likely that some of the potential hot/coldspots are false positives.

Consistent with the generally low power across the region for all species, the number of p-values that would be deemed significant at the 0.05 level is relatively small for all species. The number of significant grid cells would be even lower if adjustments to significance thresholds were made for multiple testing. However, for all species, there were at least some lease blocks for which historical survey data could be used to identify a hotspot or coldspot with reasonable confidence (darkest blue and red shading in Figures 7-9). Other grid cells could be positively identified as "neither hot nor cold spots," because no hot or coldspot was detected and power was adequate (darkest grey shading in Figures 7-9). Although the vast majority of lease blocks could not confidently be labeled as a hotspot, coldspot, or neither at the 3x (1/3x) effect size level, certain well-sampled areas (e.g., Natucket Sound) illustrate that it is possible to achieve reasonable power to detect and identify hotspots and coldspots with realistic sampling programs.

The example significance maps for Herring Gull (Figure 7) and Northern Gannet (Figure 8) also illustrate a potential pitfall of this method when strong onshore-offshore or regional gradients in abundance are present. For these two species, abundances are consistently higher offshore than they are onshore. Because the reference distribution has been defined using the entire mid-Atlantic outer continental shelf as the reference region, the offshore areas are much more likely to be identified as hotspots. The solution to this problem lies in defining reference regions appropriately to the question one wishes to ask. For some purposes, the identification of offshore areas as "hotspots" may be appropriate. However, for other purposes, one may wish to define separate "nearshore" and "offshore" reference regions and identify hot/coldspots relative to those reference regions. This issue is discussed further in section 4.5.

3.6 Decision Tree

It is useful to consider how the procedures described in section 2 of this document would fit into a more general framework for power analysis. To do this, we developed a schematic "decision tree" (Figure 13), organized around the question "How many independent surveys are needed to have adequate power to detect a hot/coldspot of a given species in a given grid cell?" We have so far described the implementation of Components A, B, C, D, E, F, G, and H of the decision tree (lettering of components follows Figure 13). Component A, the question to be answered, is described in sections 1.1 and 1.2. In our example for the U.S. Mid-Atlantic, the answer to the question posed in Component B is "yes," because the region has been sampled before. We proceed to develop a model selection method to address Component C by identifying a candidate distribution that adequately describes the data. If the answer to C had been "no," the same model selection technique could be applied to data from a nearby or similar region, as stated in Component D. Component E states the required inputs for power analysis; we have described these in detail in section 1.2 and explained how we derived them for our example species in section 2 and associated tables and figures. The 4th input listed under Component E, effect sizes,

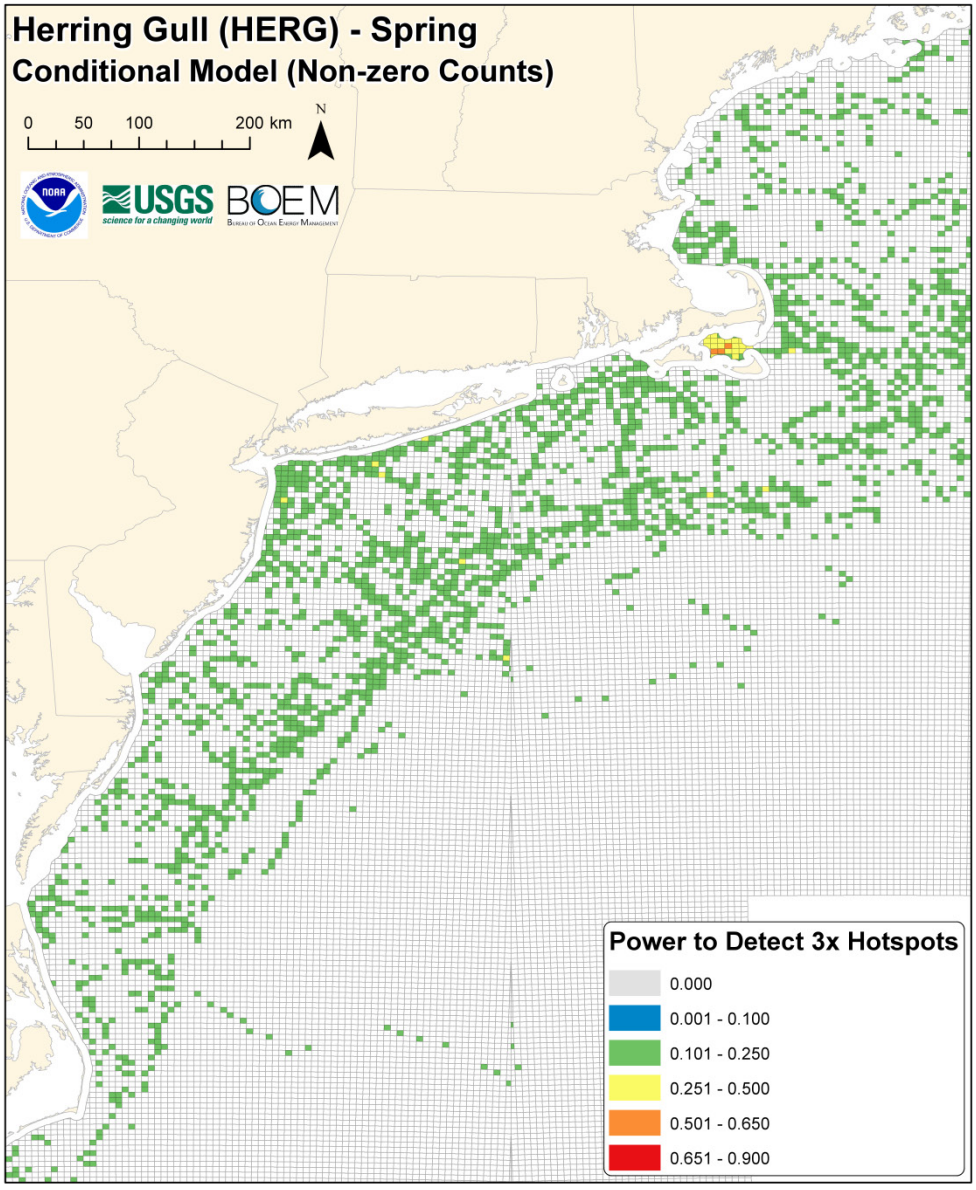


Figure 3a. Herring Gull (Spring): Map of estimated power to detect a 3x hotspot of non-zero abundance as defined in section 1.2, case (1), based on the number of surveys conducted in each BOEM lease block extracted from the USGS Avian Compendium Database as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed Power analysis used the top-ranked reference distribution (Tables 4, 5), with a reference mean of 9.58 individuals per 15-minute-ship-survey-equivalent transect.

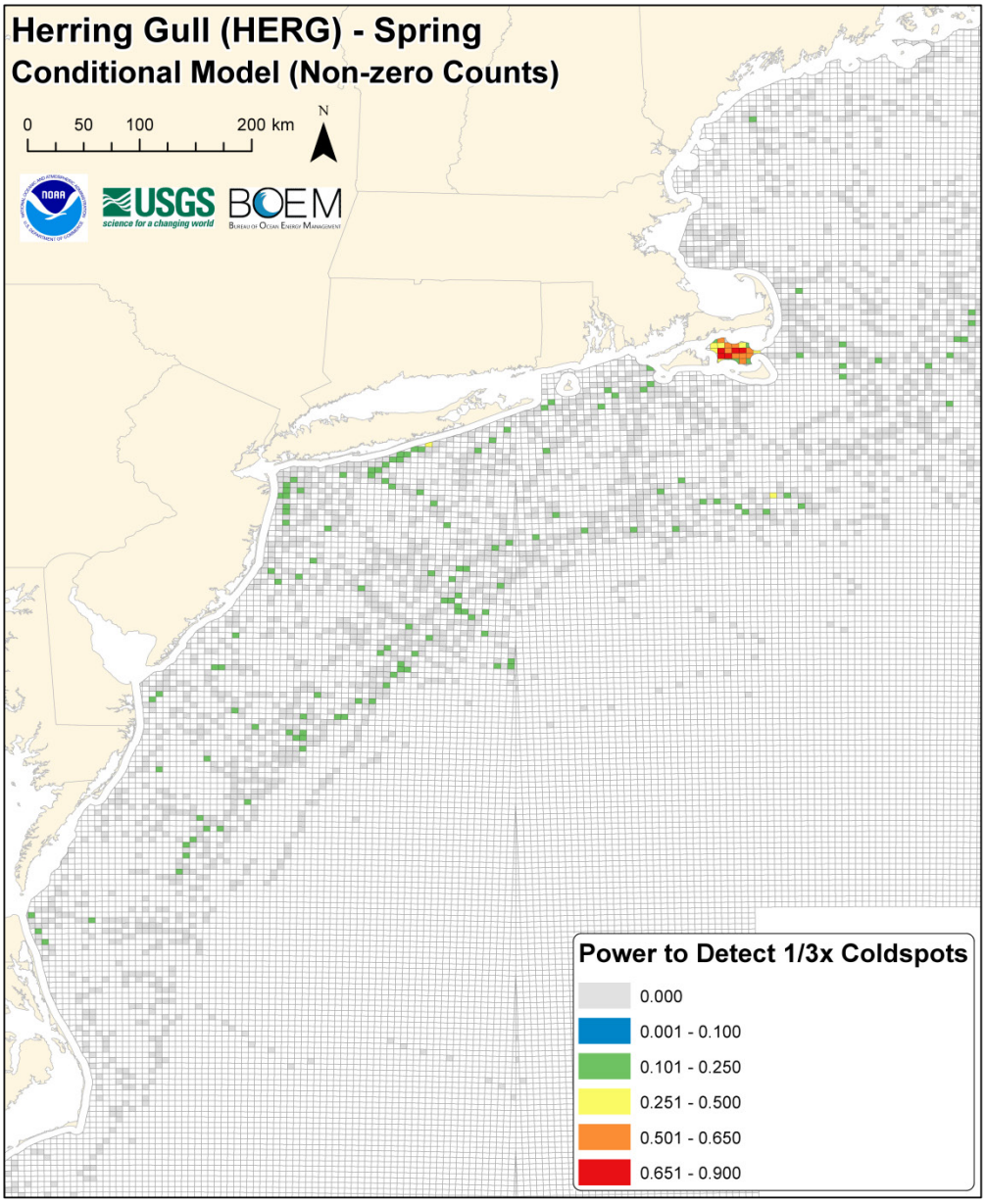


Figure 3b. Herring Gull (Spring): Map of estimated power to detect a 1/3x coldspot of non-zero abundance as defined in section 1.2, case (1), based on the number of surveys conducted in each BOEM lease block extracted from the USGS Avian Compendium Database as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed Power analysis used the top-ranked reference distribution (Tables 4, 5), with a reference mean of 9.58 individuals per 15-minute-ship-survey-equivalent transect.

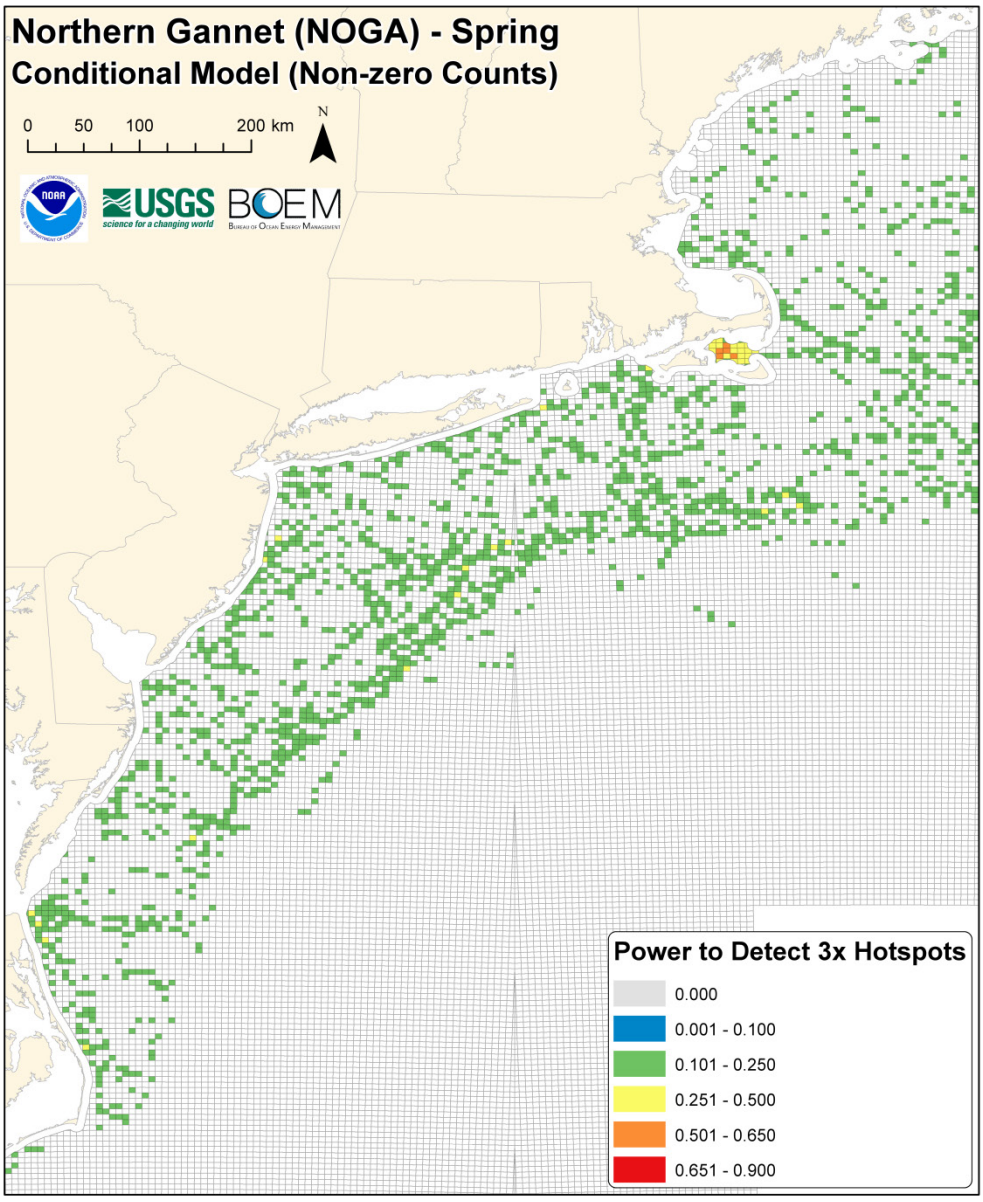


Figure 4a. Northern Gannet (Spring): Map of estimated power to detect a 3x hotspot of non-zero abundance as defined in section 1.2, case (1), based on the number of surveys conducted in each BOEM lease block extracted from the USGS Avian Compendium Database as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed. Power analysis used the top-ranked reference distribution (Tables 4, 5), with a reference mean of 11.6 individuals per 15-minute-ship-survey-equivalent transect.

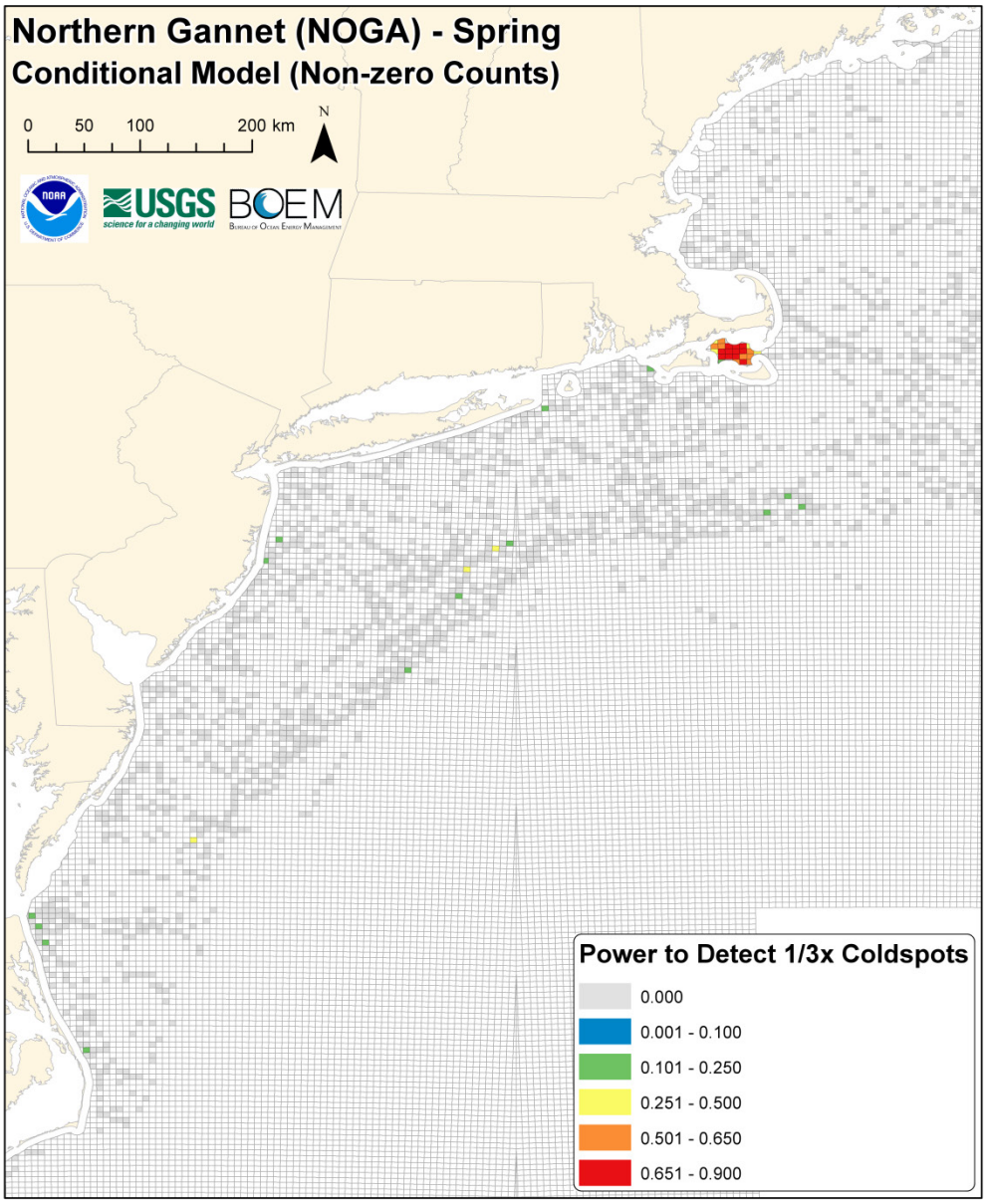


Figure 4b. Northern Gannet (Spring): Map of estimated power to detect a 1/3x coldspot of non-zero abundance as defined in section 1.2, case (1), based on the number of surveys conducted in each BOEM lease block extracted from the USGS Avian Compendium Database as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed. Power analysis used the top-ranked reference distribution (Tables 4, 5), with a reference mean of 11.6 individuals per 15-minute-ship-survey-equivalent transect.

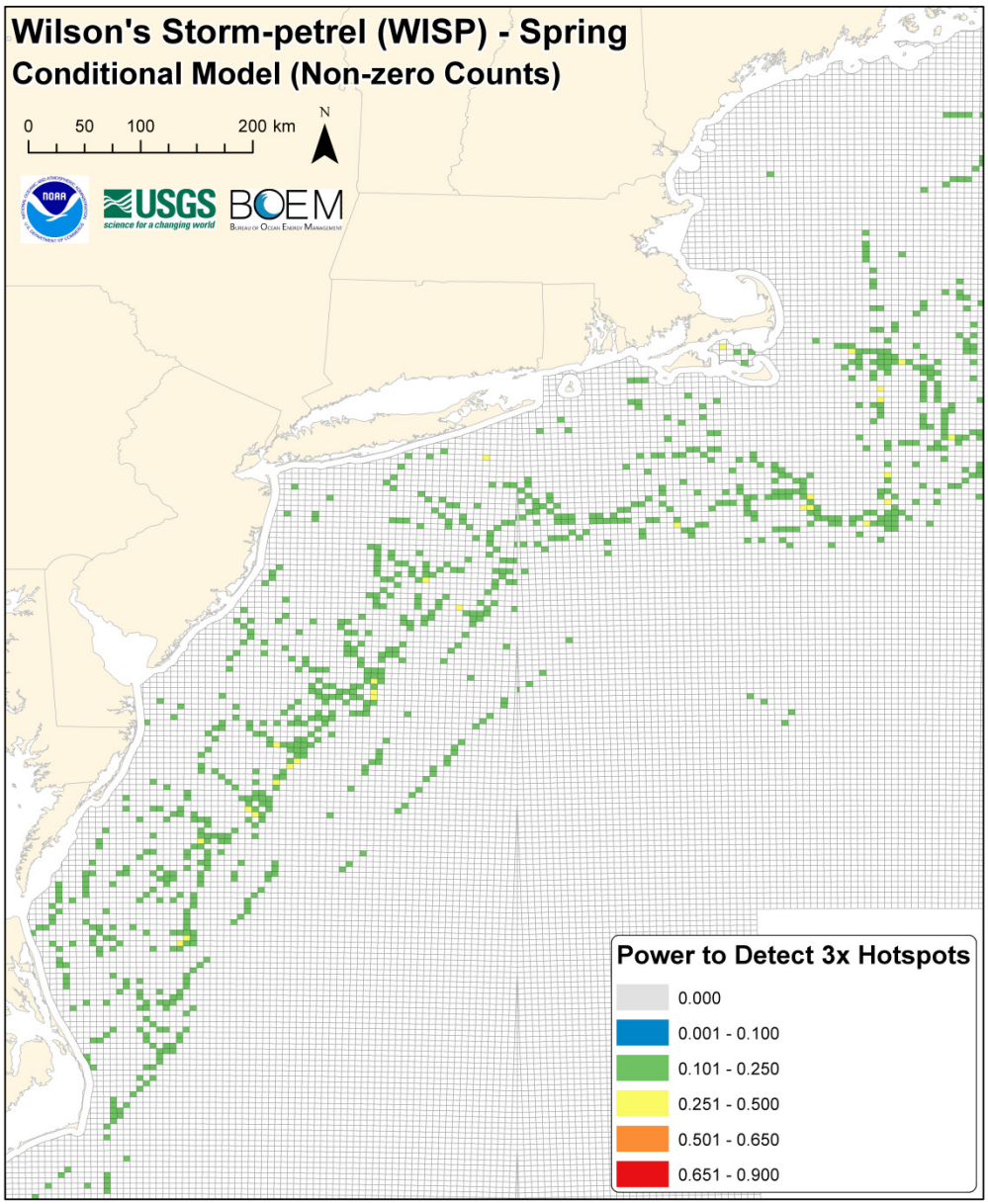


Figure 5a. Wilson's Storm-Petrel (Spring): Map of estimated power to detect a 3x hotspot of non-zero abundance as defined in section 1.2, case (1), based on the number of surveys conducted in each BOEM lease block extracted from the USGS Avian Compendium Database as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed. Power analysis used the top-ranked reference distribution (Tables 4, 5), with a reference mean of 6.24 individuals per 15-minute-ship-survey-equivalent transect.

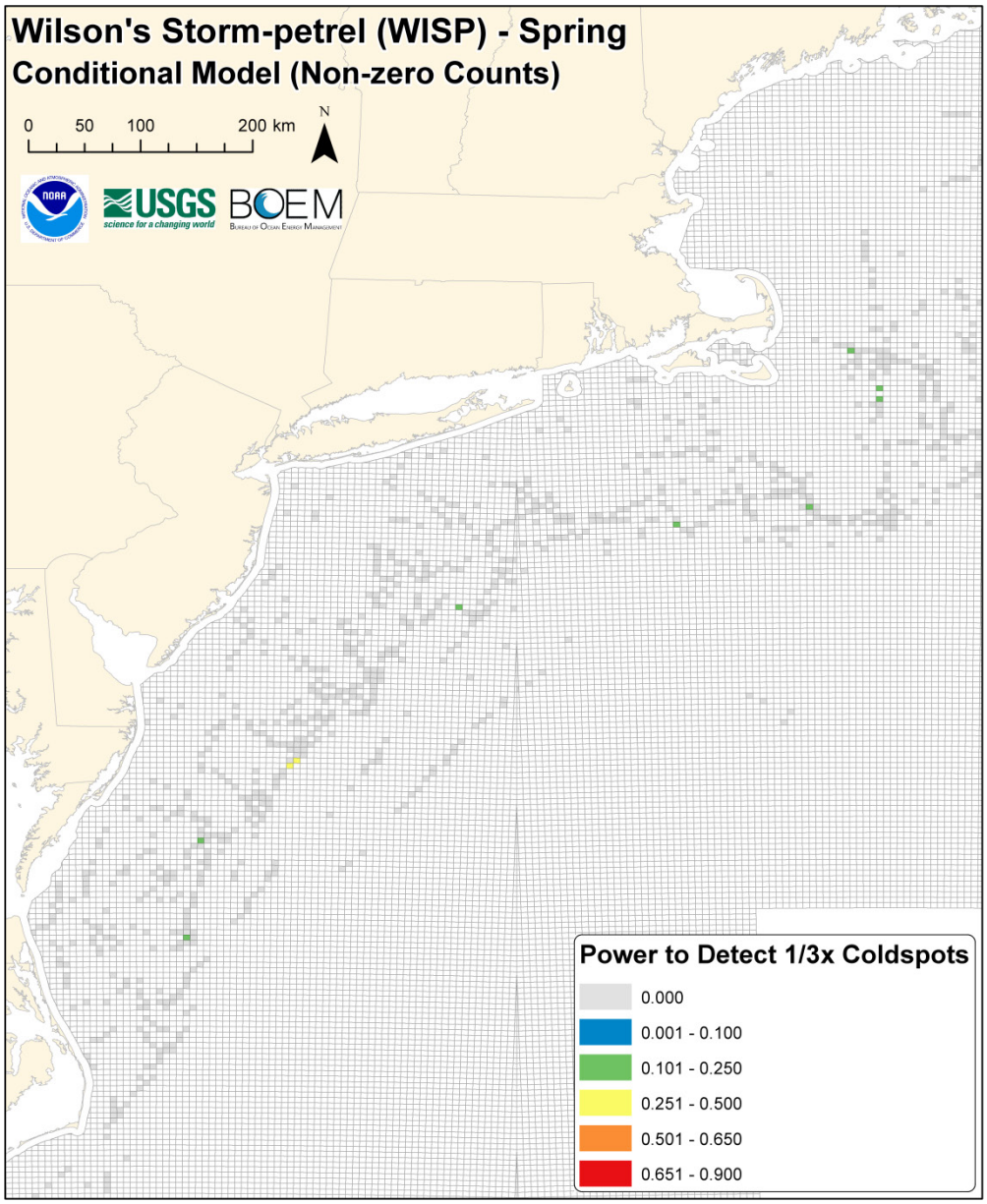
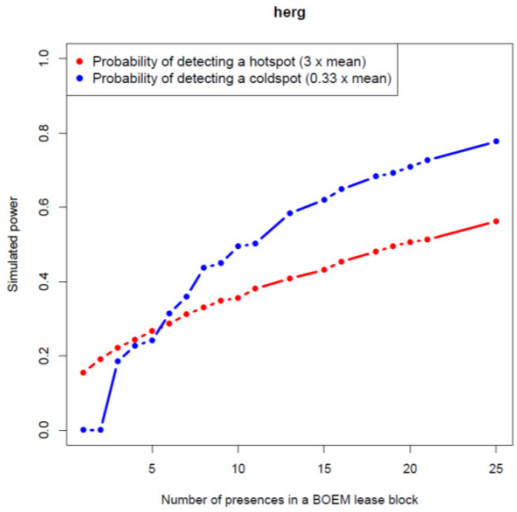
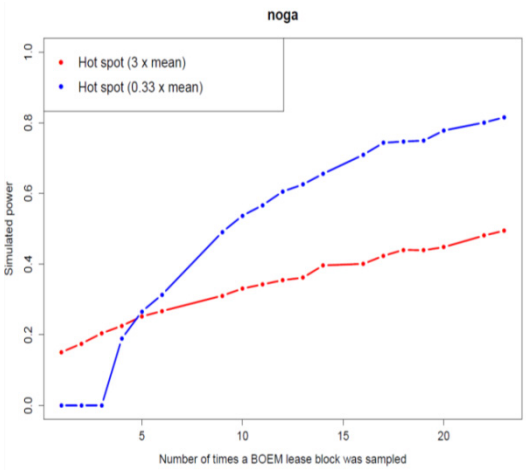


Figure 5b. Wilson's Storm-Petrel (Spring): Map of estimated power to detect a 1/3x coldspot of non-zero abundance as defined in section 1.2, case (1), based on the number of surveys conducted in each BOEM lease block extracted from the USGS Avian Compendium Database as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed. Power analysis used the top-ranked reference distribution (Tables 4, 5), with a reference mean of 6.24 individuals per 15-minute-ship-survey-equivalent transect.

a)



b)



c)

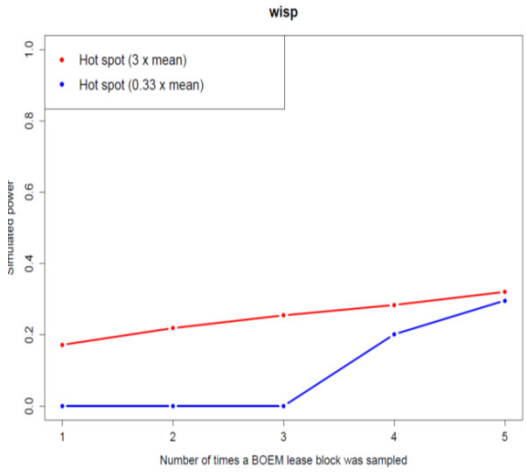


Figure 6. Power vs. sample size curves for a) Herring Gull, b) Northern Gannet, and c) Wilson's Storm-Petrel based on the number of surveys conducted in BOEM lease blocks in the USGS Avian Compendium database in Spring. Power curves assumed the top-ranked reference distribution (Tables 4, 5), and show power to detect a 3x hotspot (red lines) or a 1/3x coldspot (blue lines) of non-zero abundance as defined in section 1.2, case (1).

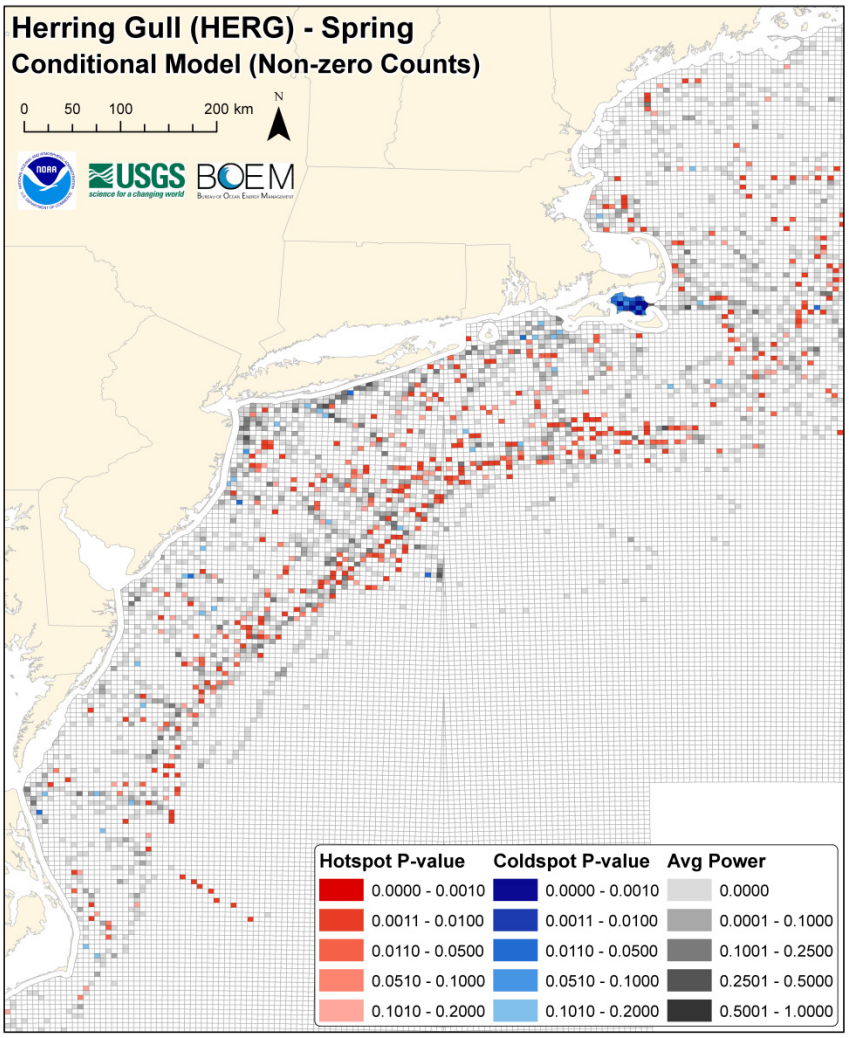


Figure 7. Herring Gull (Spring): Combined map of hotspot (red) and coldspot (blue) significance test p-values, based on one-sample, one-tailed (hotspot) Monte Carlo significance tests of the mean non-zero count in each lease block compared to the reference mean. Darker shading indicates greater statistical significance. Lease blocks that did not approach statistical significance ($p>0.2$) are shown in grey, with the intensity of the shading proportional to the average of 3x hotspot and 1/3x coldspot power values for that cell. That is, 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. In contrast, light grey shading indicates lease blocks not identified as significant hotspots or coldspots, but for which there was little or no power to detect a hotspot or coldspot, had it existed. The darkest blue lease blocks can therefore 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.

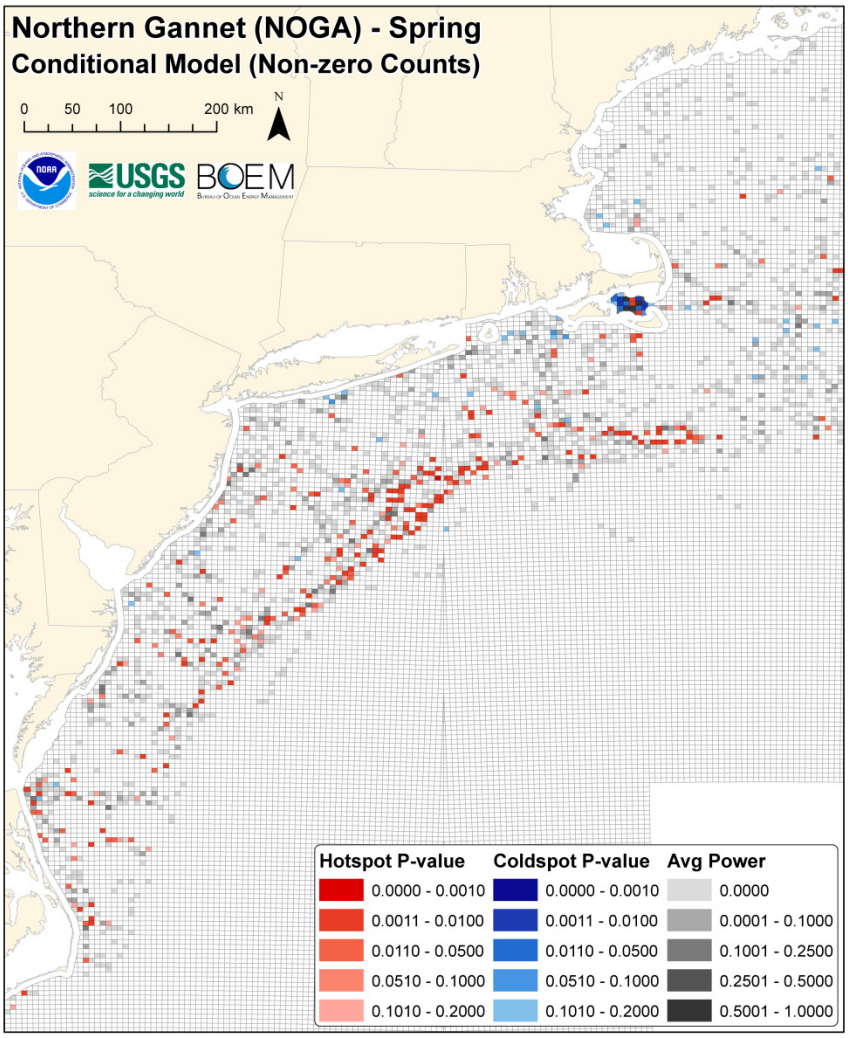


Figure 8. Northern Gannet (Spring): Combined map of hotspot (red) and coldspot (blue) significance test p-values, based on one-sample, one-tailed (hotspot) Monte Carlo significance tests of the mean non-zero count in each lease block compared to the reference mean. Darker shading indicates greater statistical significance. Lease blocks that did not approach statistical significance ($p>0.2$) are shown in grey, with the intensity of the shading proportional to the average of 3x hotspot and 1/3x coldspot power values for that cell. That is, 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. In contrast, light grey shading indicates lease blocks not identified as significant hotspots or coldspots, but for which there was little or no power to detect a hotspot or coldspot, had it existed. The darkest blue lease blocks can therefore 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.

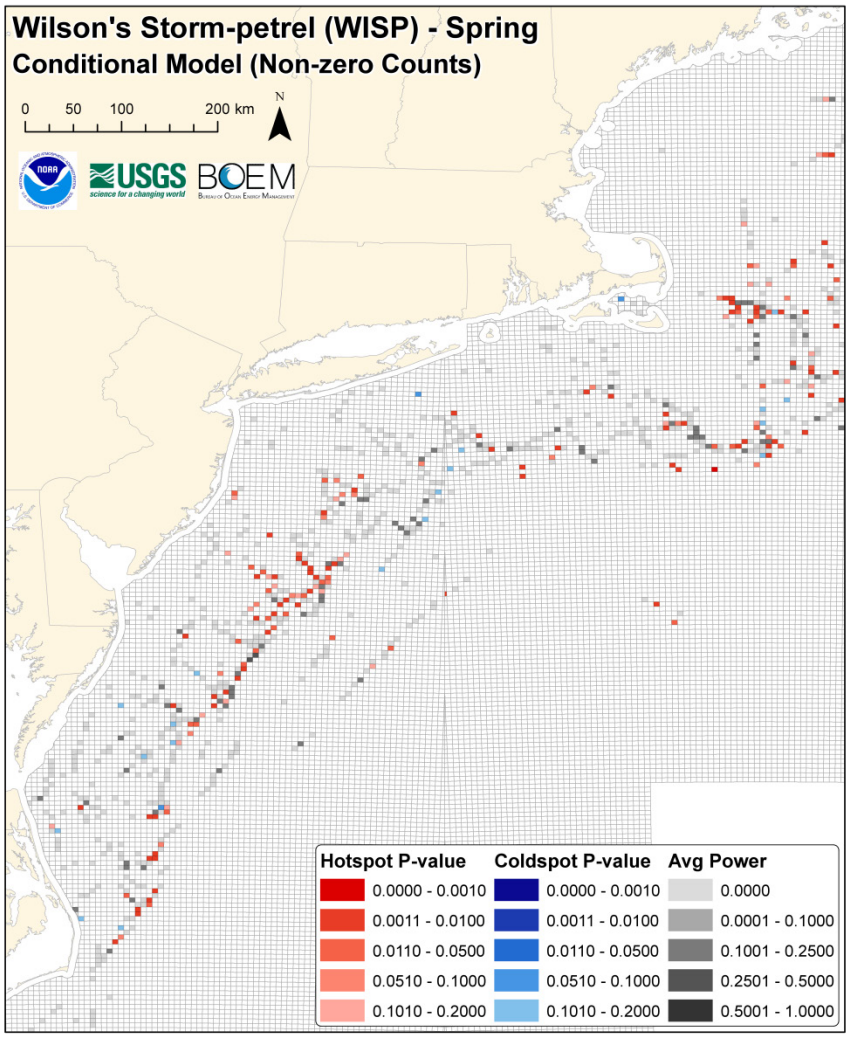


Figure 9. Wilson's Storm-Petrel (Spring): Combined map of hotspot (red) and coldspot (blue) significance test p-values, based on one-sample, one-tailed (hotspot) Monte Carlo significance tests of the mean non-zero count in each lease block compared to the reference mean. Darker shading indicates greater statistical significance. Lease blocks that did not approach statistical significance ($p>0.2$) are shown in grey, with the intensity of the shading proportional to the average of 3x hotspot and 1/3x coldspot power values for that cell. That is, 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. In contrast, light grey shading indicates lease blocks not identified as significant hotspots or coldspots, but for which there was little or no power to detect a hotspot or coldspot, had it existed. The darkest blue lease blocks can therefore 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.

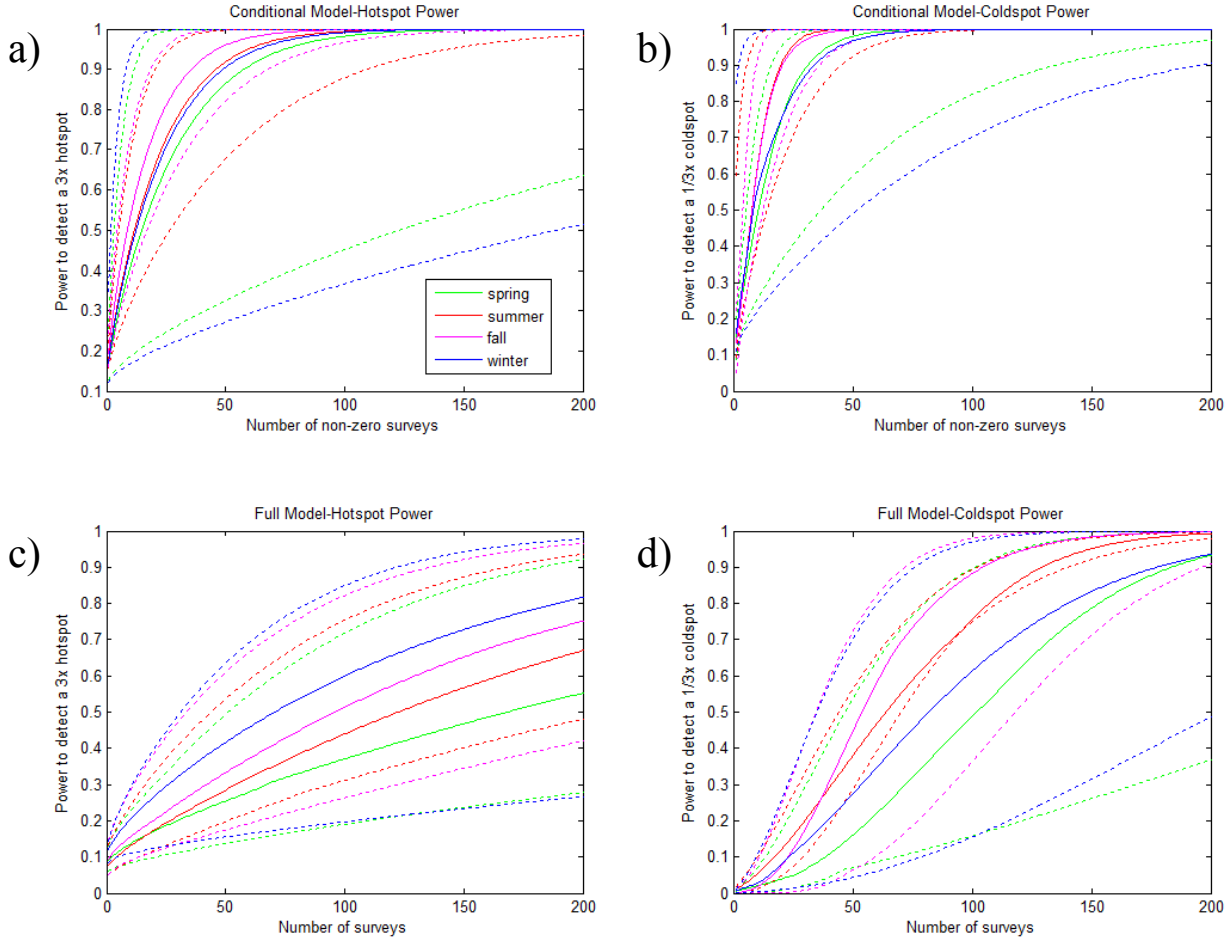


Figure 10. Summary of species-specific power curves. Simulated power vs. sample size curves (Figure 6, Digital Supplements F and G) were approximated by regression for each species in each season (colors, see legend in panel a) and the resulting curves are summarized here by plotting the median value of power (solid lines) and the 95% range (97.5th and 2.5th percentiles, dashed lines) versus sample size.

a) Conditional model, 3x hotspot power. Some species power curves had too few non-zero points to be included (spring: grsh, nofu, reph, wisp; summer: *none*; fall: cosh, nofu; winter: dove, nofu).

b) Conditional model, 1/3x coldspot power. Some species power curves had too few non-zero points to be included (spring: cote, grsh, nofu, reph, wisp; summer: *none*; fall: cosh, nofu; winter: dove, nofu).

c) Full model, 3x hotspot power.

d) Full model, 1/3x coldspot power. Some species power curves had too few non-zero points to be included (spring: coei, cote, grsh, reph, wwsc; summer: cote, lagu, ltdu, razo; fall: blsc, lagu, razo, rtlo, susc; winter: *none*).

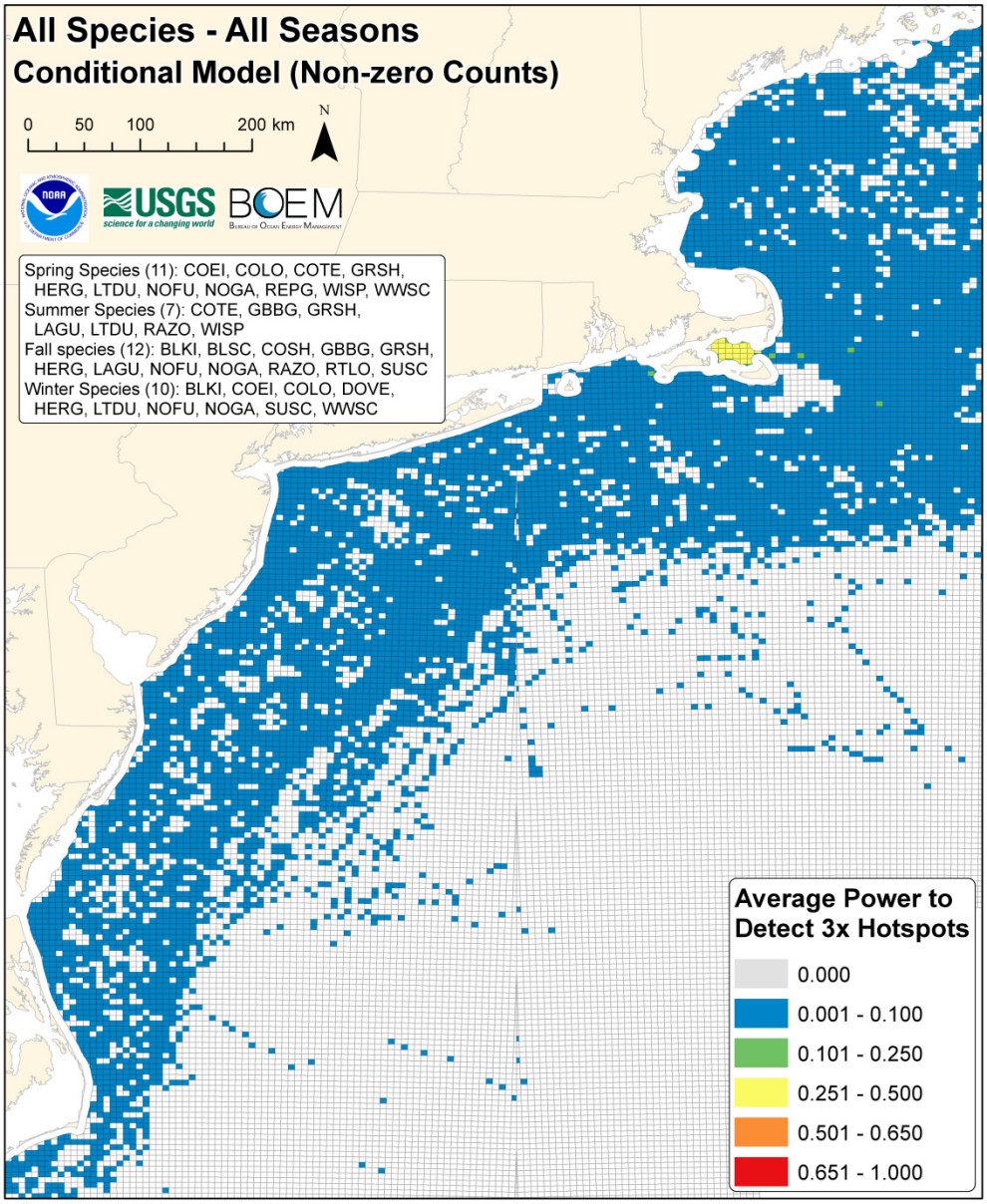


Figure 11a. Conditional (non-zero count) model: average power to detect a 3x hotspot, averaged over all modeled species in all modeled seasons as described in section 2.9. Based on data extracted from the USGS Avian Compendium Database, as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed.

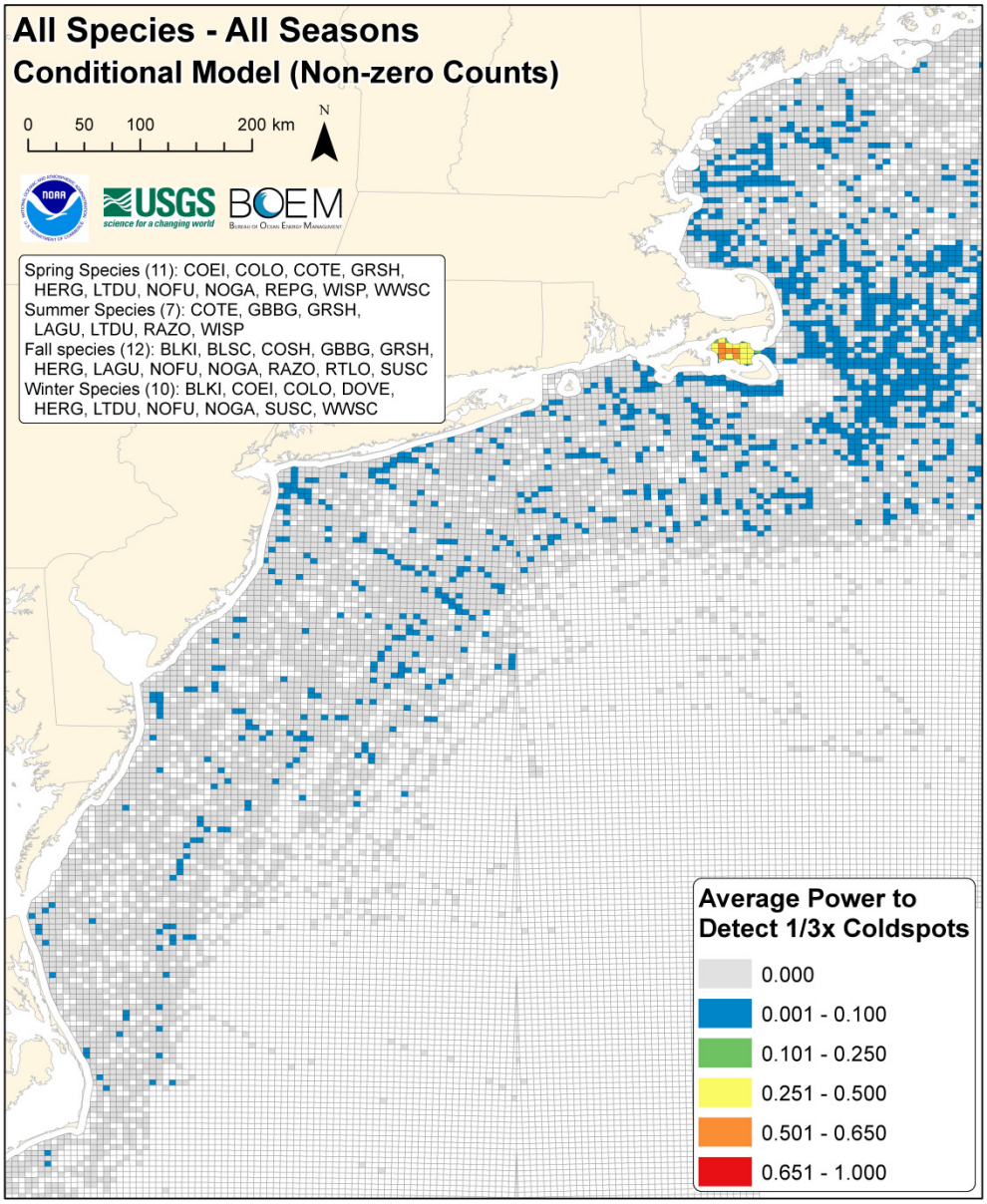


Figure 11b. Conditional (non-zero count) model: average power to detect a 1/3x coldspot, averaged over all modeled species in all modeled seasons as described in section 2.9. Based on data extracted from the USGS Avian Compendium Database, as described in section 2.6. Blank cells indicate BOEM lease blocks that were either not surveyed or where the species was never observed.

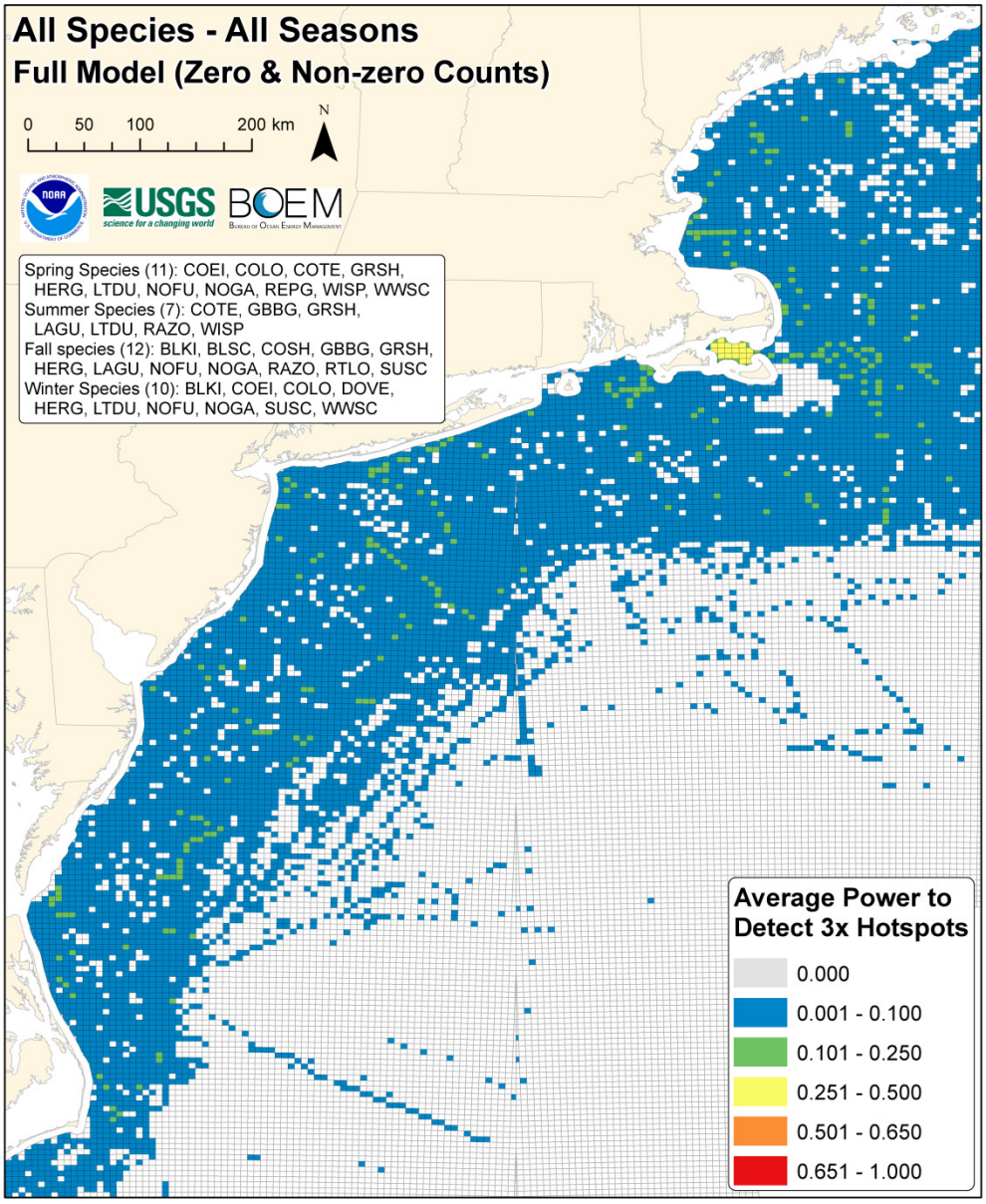


Figure 12a. Full hurdle (zero and non-zero count) model: average power to detect a 3x hotspot, averaged over all modeled species in all modeled seasons, as described in section 2.9. Based on data extracted from the USGS Avian Compendium Database, as described in section 2.6. Blank cells indicate BOEM lease blocks that were not surveyed in any season.

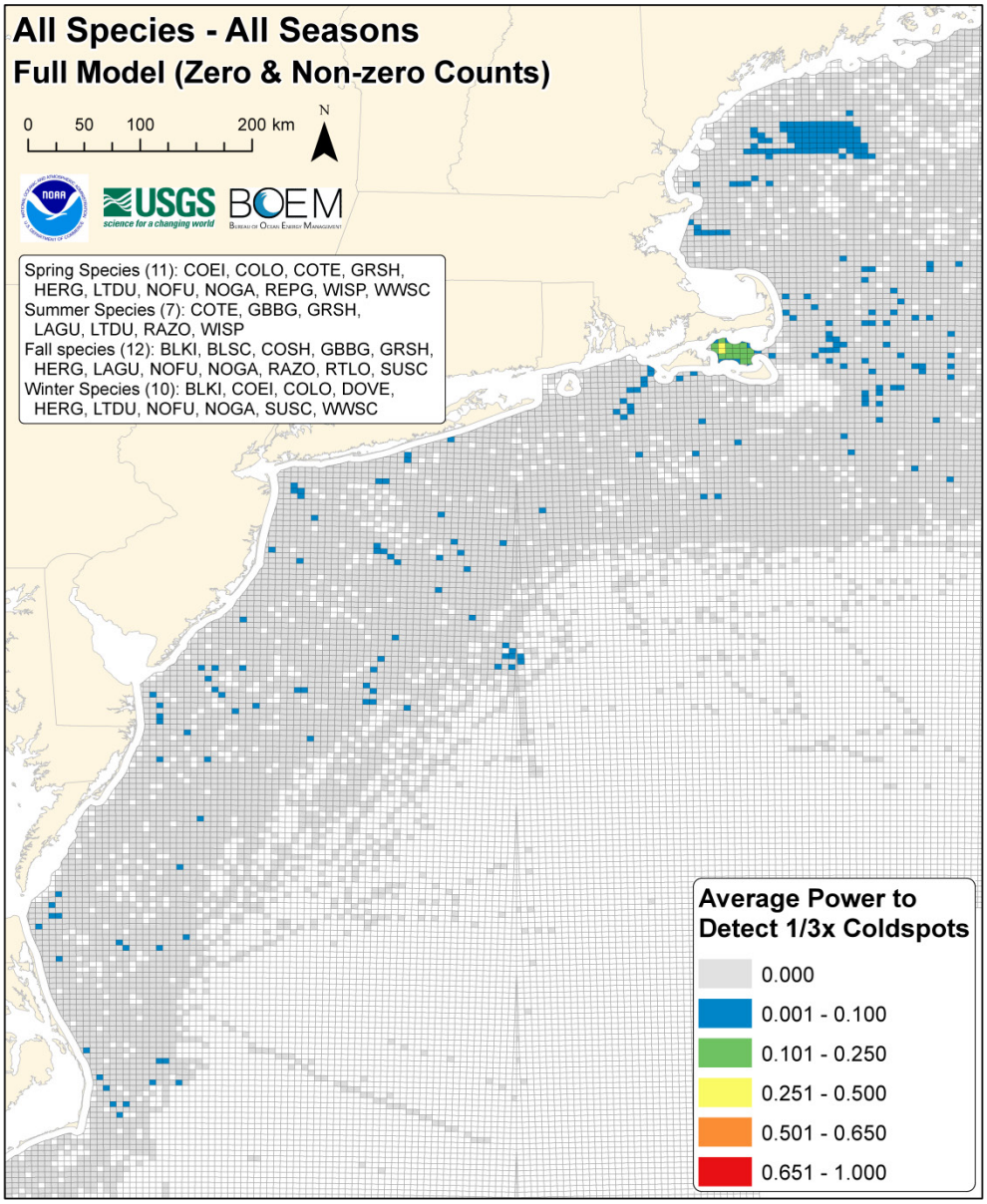


Figure 12b. Full hurdle (zero and non-zero count) model: Average power to detect a 1/3x coldspot, averaged over all modeled species in all modeled seasons as described in section 2.9. Based on data extracted from the USGS Avian Compendium Database, as described in section 2.6. Blank cells indicate BOEM lease blocks that were not surveyed in any season.

depends in part on user specification of the problem to be addressed (how big of an effect size matters for the ecological or regulatory question of interest?).

Component F describes alternative choices when an adequate candidate distribution or suitable data for fitting a candidate distribution cannot be identified. The first two choices listed under Component F (try another distribution not listed here or implement a more complex model) are beyond the scope of this study, but would be fairly straightforward to implement; they would simply require additional time and investment for the species of interest. These represent challenges for future work. What should we do when no data are available to select and fit a reference distribution, none of the candidate distributions produce an adequate fit, or issues like spatial and temporal correlation and environmental variability demand a more complex model to assess power? The 3rd and 4th choices listed under Component F (select default distribution/parameters, or use best-fitting of the candidates with some precautionary adjustment to power estimates) provide reasonable options by which one could proceed with an approximate power analysis using the best available information. Indeed, the frequency with which the discretized lognormal distribution was identified as the best-fitting model provides a strong basis for choosing this model in the absence of other information, or when model selection criteria (AICc, Vuong tests) yield ambiguous results (Zipkin et al. 2012).

Components G and H correspond to the power analysis simulation method described in sections 1.2 and 2.4 (Component G relates to section 1.2, case(1); Component H relates to section 1.2, case(2); together Components G and H relate to section 1.2, case (3)). This modular representation of the process illustrates how the candidate set of distributions and associated power simulation modules (G1 through G8) could easily be expanded to accommodate additional distributions.

A useful final step might be to produce maps of the number of *additional* surveys, beyond the historical survey effort, that are needed to achieve a certain level of power in each grid cell. This is stated in Component I of the decision tree, and corresponds to the "output" of the decision tree process. (Of course, there are many other useful outputs produced along the way, including the power maps, power curves, and significance maps described in detail in sections 3.3, 3.4 and 3.5). Such maps can easily be produced with the information provided in this report: one would simply look up the required number of surveys to achieve the desired power to detect the desired effect size on power curves like those shown in Figure 6 (and given for each modeled species in Digital Supplements F and G), and then display the difference between this number and the actual numbers of non-zero presences (case (1), Digital Supplement F) or the actual number of surveys conducted for each species (case (3), Digital Supplement G).

Lastly, it is important to note that the red boxes in the decision tree (Components (i), (ii), and (iii)) indicate external information that depends on the end-user's goals and specifications. With regard to Component (i), the appropriate reference region in relation to which hotspots and coldspots are defined will depend on the nature and scope of the ecological and regulatory questions to be answered (see section 4.5). With regard to Component (ii), it is unlikely than any model will be a perfect representation of reality, and therefore the question of how good of a fit is adequate arises (see section 4.3); although statistics can help answer this question, ultimately the tradeoff between level of investment in model development and improved decision-making must be evaluated by the user. With regard to Component (iii), the definition of adequate statistical power depends on the relative costs of requiring additional surveys vs. making type II

errors (failing to detect hotspots or coldspots when they are in fact present). This will often depend on issues such as the ecology, regulatory status, or vulnerability of the species of interest, sensitivity of the area being studied, the type of impact being evaluated, and the type of regulatory question being addressed. It is important to consider these externalities early in the sampling design process so that the power analysis can be properly parameterized to address relevant ecological and regulatory questions.

3.7 Sampling to Capture Temporal Variance: Environmental Time Series

Variogram analysis of daily time series of satellite sea surface temperature (SST, Figure 14) and sea surface chlorophyll-a concentration (*chl*, Figure 15), both previously shown to be important correlates of marine bird occurrence and abundance in this region (Kinlan et al. 2012), reveals that nearly all of the variance that would be observed in any ~5 year period between 2002 and 2011 accumulates within 1 year of observation. 70-80% of variance in de-seasoned SST accumulates within 7-10 days for all four regions (Figure 14). Similarly, 40-85% of variance in de-seasoned, $\log_{10}(x+1)$-transformed *chl* accumulates within 7-10 days (Figure 15). To the extent that these environmental variables correlate with a particular species' occurrence and abundance, time scales of short-term autocorrelation are expected to be short (less than 7-10 days) and the majority of interannual variance observed in any given ~5 year period is expected to be well-characterized by 1 year of sampling.

Because these satellite records are relatively short, we also analyzed longer monthly time series (1948-2012) of regional oceanic/atmospheric climate indices known to correlate with long-term variation in marine bird abundance and occurrence, the North Atlantic Oscillation (NAO; Figure 16a) and Atlantic Multidecadal Oscillation (AMO; Figure 16b). Both climate time series showed similar patterns of accumulation of temporal variance (Figure 16). These patterns were consistent with what was observed in the shorter SST and *chl* time series, in that a large proportion of the variance (about 50-70%) accumulates in the first 1-2 years, followed by little increase in variance out to 8-9 years. However, the longer time series allowed resolution of variance patterns at larger time lags than the SST and *chl* analyses, and these reveal substantial additional variability (20-40% of the total) accumulating at time scales from 9 to 15 years and longer. This is indicative of the decadal-scale ocean climate variability that is well-documented for this region of the Northwest Atlantic (e.g., Enfield et al. 2001, Veit and Montevecchi 2006). It is interesting to note, however, the distinct gap in time scales of variance accumulation: 1-2 years for the first plateau in variance vs. 13-15 years for the second plateau. If this time period is indicative of future conditions, the implication is that little additional variance in ocean climate will be observed if a 1-2 year sampling program is extended for an additional 6-7 years, but decadal variability could be captured by repeating annual sampling every 10-15 years or so.

3.8 Sampling to Capture Temporal Variance: Abundance Time Series

Variogram analysis of time series of repeat marine bird counts on standardized transects with midpoints in the same BOEM lease block reveal patterns that are generally consistent with those observed in the environmental time series (Figures 17, 18). The variograms of marine bird counts are, of course, much noisier due to lower sample sizes, the irregular and scattered nature of the time series, and the skewed and zero-inflated data. However, consistent with the SST, *chl*, NAO, and AMO analyses, long-term (interannual) patterns of variability for time scales of <10

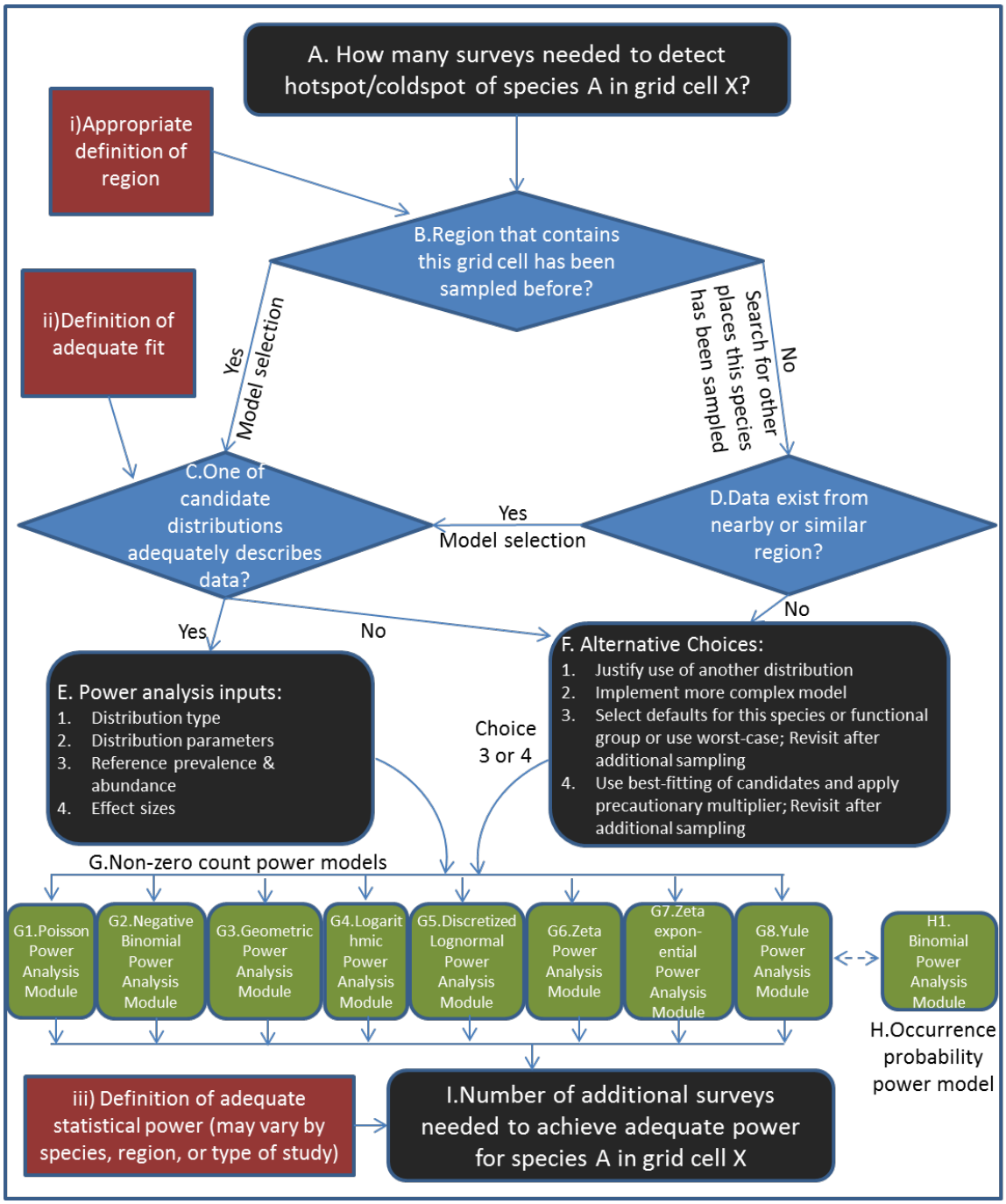


Figure 13. Schematic of decision tree for determining number of surveys required in a discrete spatial unit according to the methods described in this study. Red boxes indicate external information, defined by the end-user. Black boxes indicate inputs and outputs. Blue diamonds represent decisions based on the data and model results. Green boxes represent modules of the power analysis process. Labels (A-I and i-iii) are used to indicate each component of the decision tree for easy reference. See section 3.6 in text.

years (the maximum possible with the length of available time series) indicate that >50% of variance is captured within the first season of sampling for nearly all species analyzed, and the 100% reference line (all variance captured) is generally crossed within 1-3 years (Figure 17). Fall abundance exhibited the least interannual variability (1 year of sampling generally sufficient), winter abundance exhibited the most interannual variability (2-3 years of sampling required for some species), and summer and spring were in between (1-3 years).

Short-term patterns of temporal variability in marine bird count data were also consistent with analyses of environmental time series (Figure 18). For most species, variance in abundance accumulated rapidly within less than 3-5 days, reaching 25-100% of the long-term total (typically, 50-60%), similar to the pattern observed for SST and *chl* variability. Little additional accumulation of variance was observed beyond 3-15 days, a pattern that is also strikingly similar to that observed in the environmental data. These results suggest marine bird abundance decorrelates rapidly with time in this region, supporting our assumption that repeat surveys are approximately independent as long as they are separated by a few days or longer. They also suggest that a season could be well-characterized by repeat surveys conducted ~3 days or more apart and spread over a 10-30 day period.

4.0 DISCUSSION

The approach outlined and demonstrated in this document is designed to facilitate estimates of statistical power as a function of survey intensity without requiring complex species-, region- or survey-specific models. Given this goal, we have necessarily made assumptions that ignore some known sources of real-world complexity. In this section we briefly discuss some of the issues involved in judging the appropriateness of our approach for a particular situation and how violations of assumptions may affect our results. We also provide recommendations for and notes on possible applications of results, and draw conclusions from our example power analyses and analyses of temporal variability.

It is important to note that application of this method to generate products or guidelines that will be used for management or decision-making purposes will involve assumptions that need to be clearly stated, and each application should evaluate the extent to which assumptions may be violated and the likely impact on conclusions.

4.1. Choosing Between the Non-Zero Conditional Model and the Full Hurdle Model

We have presented two types of models that can be used to derive power curves and maps for avian count data: the conditional model (section 1.2, case (1); Figures 3, 4, 5, 6, 7, 8, 9, 10a, 10b, 11 and Digital Supplements A and F), applicable to non-zero count data, and the full hurdle model (section 1.2, case (3); Figures 10c, 10d, 12 and Digital Supplements C and G), applicable to count data in which zeros are recorded when species are not seen in a standardized survey time/area. Which model one chooses to use depends in part on the type of data available, and in part on the question being asked. In general, the full hurdle model should be applied to any dataset that includes zeros. In situations where zeros are not reliably recorded and cannot be inferred based on survey protocols, the conditional model may be the only option for power analysis. Certain types of sampling, survey design, and data management schemes might give rise to such a situation. Perhaps more importantly, the conditional model may be useful in the

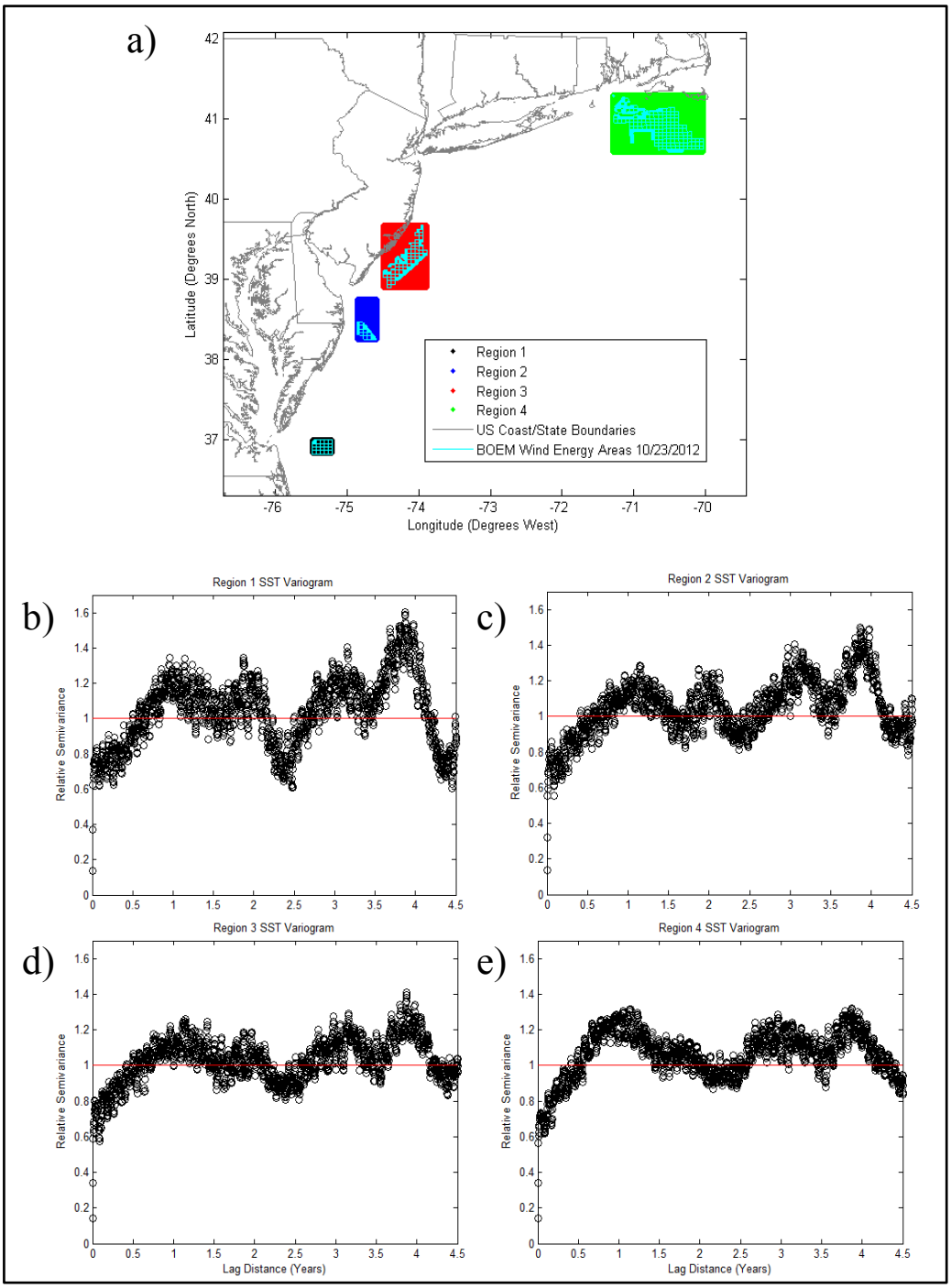


Figure 14. Temporal extent of sampling needed to capture intraannual to interannual environmental variability, as inferred from sea surface temperature (SST) time series in indicated regions (panel a), from NOAA Coastwatch MODIS Aqua 1km daily 3-day composite night and day SST (Foley 2012). The seasonal cycle was removed by subtracting the monthly climatology. See section 2.10 for details;
b) SST variogram for region 1. The relative semivariance (i.e., fraction of total variance) is plotted for increasing time lag distances (temporal scales, measured in years). The red horizontal reference line indicates the sample variance. Little additional variance is encountered at increasing temporal scales once the relative semivariance crosses this line; c) SST variogram for region 2; d) SST variogram for region 3; e) SST variogram for region 4.

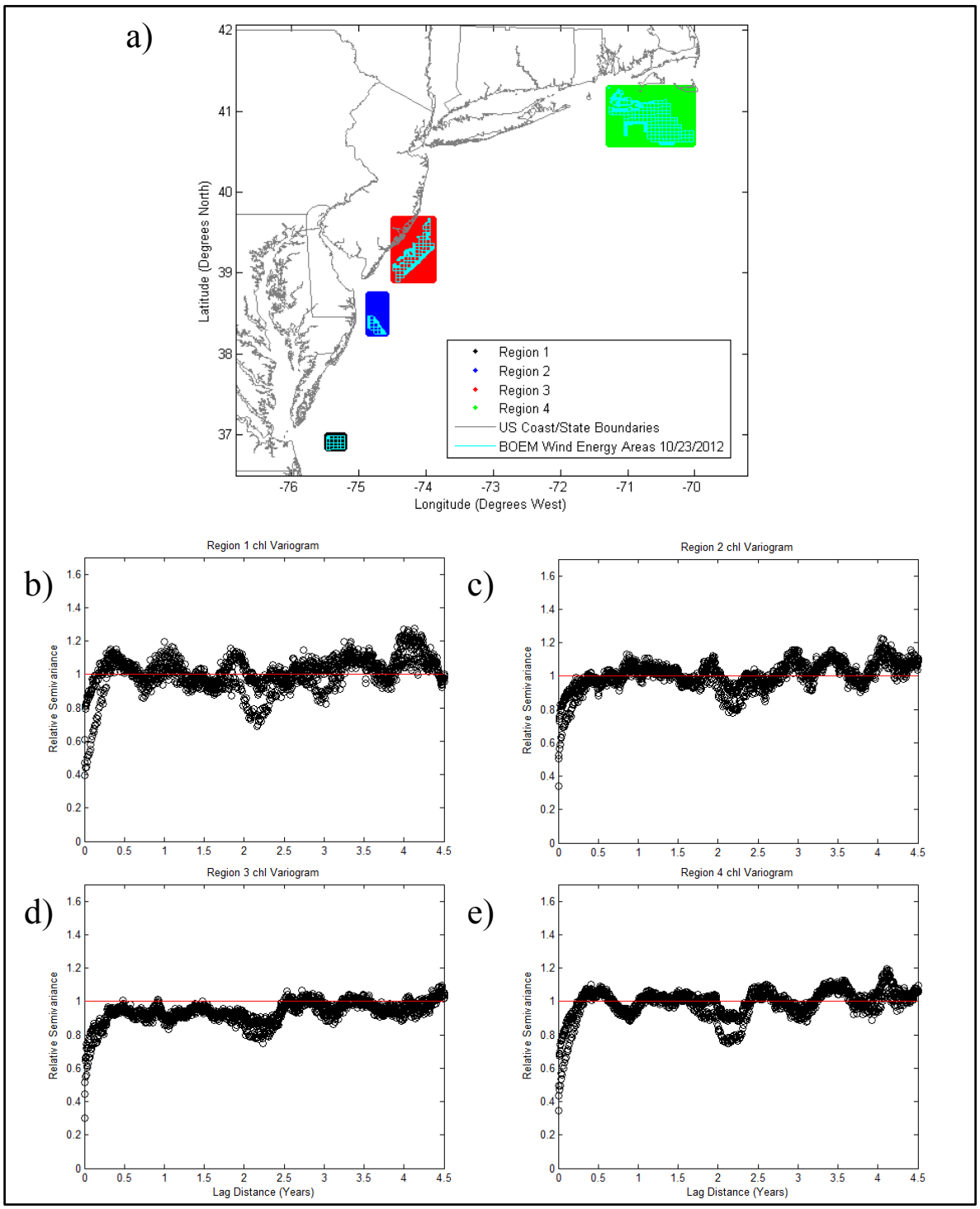


Figure 15. Temporal extent of sampling needed to capture intraannual to interannual environmental variability, as inferred from sea surface chlorophyll-a concentration (*chl*) time series in indicated regions (panel a), from NOAA Coastwatch MODIS Aqua 1km daily 3-day composite chl-a (Foley 2012). The seasonal cycle was removed by subtracting the monthly climatology. *Chl* data were $\log_{10}(x+1)$ transformed prior to analysis. See section 2.10 for details; b) *Chl* variogram for region 1. Variogram interpretation is as described in Figure 14 caption; c) *Chl* variogram for region 2; d) *Chl* variogram for region 3; e) *Chl* variogram for region 4.

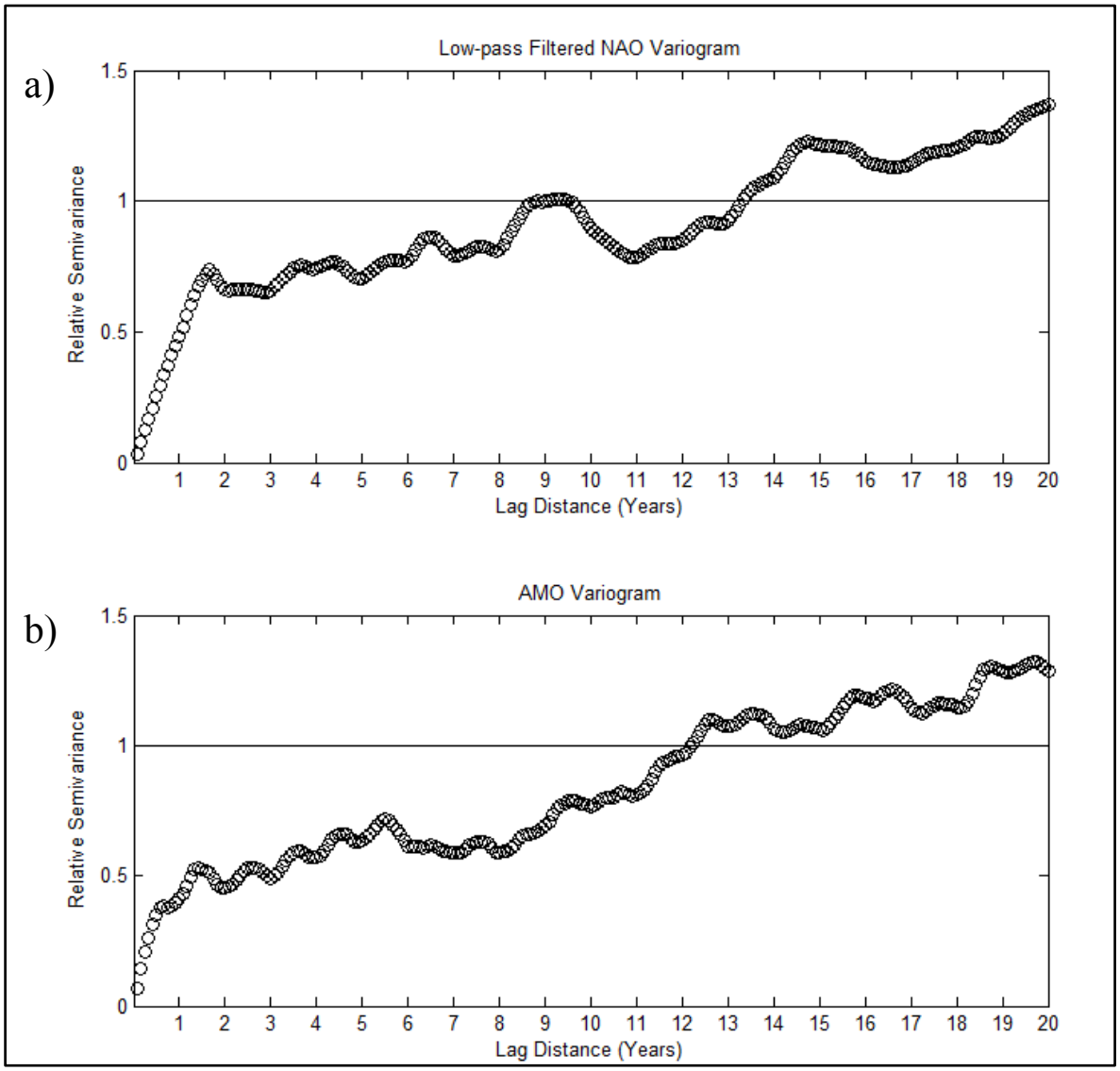


Figure 16. Temporal extent of sampling needed to capture interannual to decadal environmental variability, as inferred from monthly time series of indices of regional climate variability (1948-2012). See section 2.10 for details. The relative semivariance (i.e., fraction of total variance) is plotted for increasing time lag distances (temporal scales, measured in years). The black horizontal reference lines indicate the sample variance. Little additional variance is encountered at increasing temporal scales once the relative semivariance crosses these lines. Note that for both panels the semivariance reaches a sill (plateaus) after time lags of 1-3 years and does not undergo another large increase until 9-10 years, plateauing again beyond 13-15 years.

a) Variogram of the North Atlantic Oscillation (NAO) index. The raw NAO index was low-pass filtered with a simple rectangular 5 month running mean to remove short-term variability); NAO index data are available at: http://www.esrl.noaa.gov/psd/data/correlation/nao.data

b) Variogram of the Atlantic Multidecadal Oscillation (AMO) index; AMO index data are available at: http://www.esrl.noaa.gov/psd/data/correlation/amon.us.data

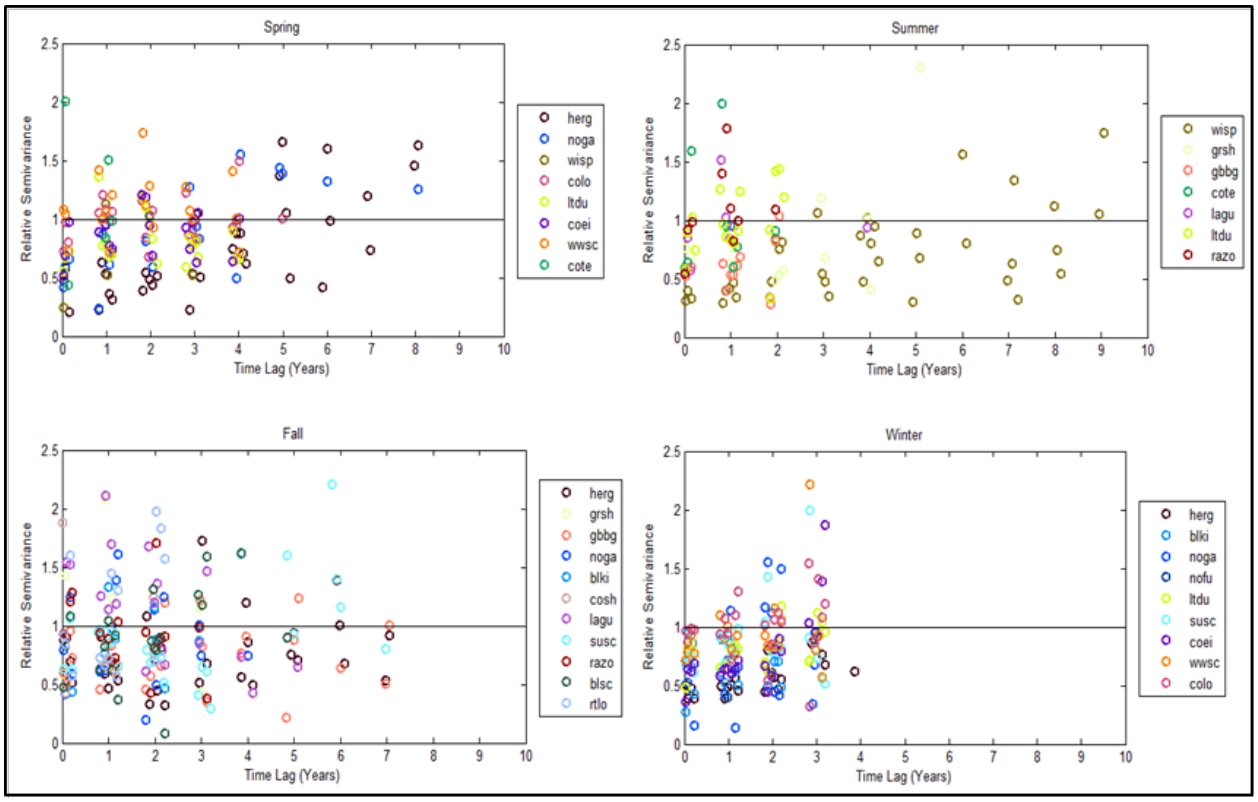


Figure 17. Long-term (interannual) variance in observed relative abundance of marine birds in BOEM lease blocks on the Atlantic OCS. Relative semivariance of $\log_{10}(x)$-transformed species-specific marine bird counts on standardized transects conducted within the same BOEM lease block versus time lag (in years) between surveys. Points are only shown if at least 20 pairs of observations were available to estimate semivariance at the given time lag. See section 2.11 for details. Colors indicate species, as shown in the legends to the right of each panel (the same color is used for a given species in all panels in which it occurs). Four-letter species codes are as in Table 3. The relative semivariance (i.e., fraction of total variance) is plotted for increasing time lag distances (temporal scales, measured in years). The black horizontal reference lines indicate the sample variance. Little additional variance in bird counts is encountered at increasing temporal scales once the relative semivariance crosses these reference lines. Note that for the majority of species, the reference line is crossed within 1-3 years, with >50% of variance captured within 1 year for nearly all analyzed species. This is consistent with the time series analysis of environmental correlates in Figures 15-16. However, on the basis of Figure 16, additional variance might be expected for some species at longer times scales (9-15 years) not resolved by this analysis.

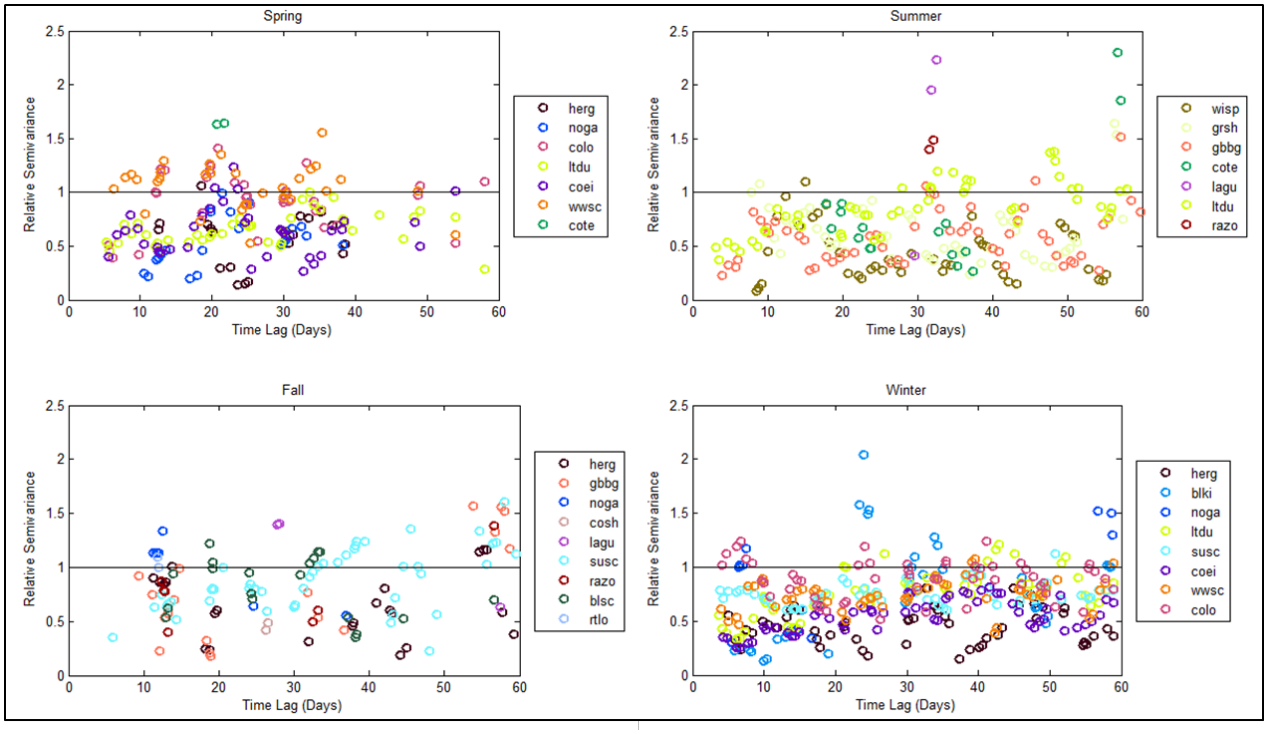


Figure 18. Short-term (intra-seasonal) variance in observed relative abundance of marine birds in BOEM lease blocks on the Atlantic OCS. Relative semivariance of $\log_{10}(x)$-transformed species-specific marine bird counts on standardized transects conducted within the same BOEM lease block versus time lag (in days) between surveys. Points are only shown if at least 20 pairs of observations were available to estimate semivariance at the given time lag. See section 2.11 for details. Colors indicate species, as shown in the legends to the right of each panel (the same color is used for a given species in all panels in which it occurs). Four-letter species codes are as in Table 3. The relative semivariance (i.e., fraction of total variance) is plotted for increasing time lag distances (temporal scales, measured in days). The black horizontal reference lines indicate the sample variance. Because this analysis focuses on short time scales, the semivariance may not rise above the reference line within the short (60 day) maximum time scale studied. However, note that for the majority of species, variance is already more than 25% of the reference variance (which is based on 5-15 years of data) within less than 3-5 days (the lower limit of resolution of the analysis). Variance approaches a stable range of values (i.e., values remain similar for the rest of the 60-day period) within 5 days for most species, and within 10-15 days for nearly all species. This suggests that short-term repeat samples spaced by at least 3-5 days will often be effective as independent or nearly-independent surveys within a season.

case of less common species for which non-zero observations occur so infrequently that the full model power curves rise very slowly with sample size. In these instances, the conditional model represents an upper bound on the power that could be achieved if surveys were properly stratified or targeted such that the species of interest was observed nearly 100% of the time. For example, the precise environmental conditions associated with habitat occupancy by a particular species could be determined and surveys conducted only in times/places where those conditions existed, or a survey protocol could be initiated in which timed transects only started when a species was sighted. The conditional model results could serve as a useful tool for planning such surveys. Finally, the conditional model may also be the best choice when analyzing statistical power and hotspot/coldspot significance for management questions in which the number of birds that occur when they are seen, not how often they are seen, is the relevant question.

In some cases, the relative abundance (count) of birds is not as relevant as the probability of occurrence. In these cases, neither the conditional nor the full hurdle model is necessary, and the occurrence-probability-based power analysis and tests of hotspot/coldspot significance can be applied (section 1.2, case (2); Digital Supplement B). We have not included occurrence-probability-based hotspot/coldspot power maps in this report, but note that these could be particularly useful for less common species for which count data are insufficient to identify hotspots or coldspots, and/or when the management question of interest focuses on the presence/absence of a species rather than the number of individuals affected.

4.2. Appropriateness of Mean-based Test Procedures

It is clear from our results and a large body of previous work (e.g., Bonabeau et al. 1999, Griesser et al. 2011) that avian count data is highly right skewed and in many cases "heavy-tailed" compared to commonly-used distributions such as the Poisson and Negative Binomial (in other words, there is a higher probability of very large counts than would be expected under these distributions). This raises the question of whether the sample mean is the most relevant statistic on which to base analyses of ecological impacts. Although the mean is attractive in terms of its ease of interpretation, it clearly does not completely characterize the entire distribution and in particular is very sensitive to the upper tails of the distribution, which are hard to measure with small sample sizes. In the extreme case of power-law-type distributions with $\alpha<1$, the mean is not even well-defined. Thus for some purposes, tests based on the median, quantiles, or extreme value statistics of the distribution might be more relevant and also have more attractive statistical properties. We recommend exploring other types of test statistics in future work.

In the case of two-parameter distributions, we made the simplifying assumption that the second (shape or dispersion) parameter remained constant for a given species in a given season as the mean of the distribution changed. In other words, the alternative hypotheses for power analysis were formed by adjusting only the first parameter to achieve the desired effect size. Where possible, the validity of this assumption should be evaluated. One way to do this is by checking for correlations in the first and second parameter of each distribution type in our example datasets (Table 4). No significant correlations are detected ($p>0.05$, regressions not shown), supporting our assumption. However, a stronger test of this assumption would be possible for some species by analyzing many different spatial and/or temporal subsets of the domain: does the maximum likelihood estimate of the second parameter remain relatively constant as the mean

abundance varies, or do the first and second parameters change jointly? In the latter case, the relationship between the first and second parameters and the mean could be explicitly parameterized based on these additional analyses. This approach is worth exploring in future work, but is beyond the scope of the current study.

4.3. Model Selection

Power estimates presented in this study are based on the assumption that the chosen distributional model holds, where the chosen model is the model with the lowest AICc from a set of candidates. In reality, it is unlikely that the selected distribution model fits perfectly, and the resulting uncertainty in power estimates should be borne in mind. Where several models appear to fit equally well (as may often occur when limited data are available for fitting a reference distribution) it may be appropriate to conduct power analyses under each alternative distribution and examine the range of power estimates and sample sizes that would be indicated under the different distributional assumptions. We used pairwise Vuong closeness tests to identify cases where the top-ranked model (in terms of AICc) is not significantly better than competitors. In the case of a statistical "tie" between distributions, power analyses could be repeated with multiple alternative distributions and the average or most conservative estimate of power could be used.

Since our model selection method is based on relative ranking by AICc, rather than an absolute measure of model performance, it is possible that even the best-fitting model will not be a particularly good representation of the data. We used one-sample Kolmogorov-Smirnov tests to help identify situations where this occurs, but for low sample sizes these tests have little power to distinguish whether or not a model is appropriate. This highlights the value of meta-analyses of large datasets, which might reveal patterns in the types of distributions most appropriate for particular types of species and environments, and allow *a priori* model selection in data-poor cases. Such studies might also reveal additional candidate distributions that would expand the range of distribution shapes that could be accommodated.

4.4. Observation Process

As formulated here, our model assumes that survey counts directly reflect true underlying abundances, ignoring any features of the observation process (e.g., detectability or process error) that could influence count distributions. In reality, observed count distributions are a reflection of both the underlying biological state and the observation process. For example, in aerial bird surveys at sea, singles and pairs have a higher probability of being undetected (Pollock and Kendall 1987), whereas flocks with more birds are typically undercounted (Pearse et al. 2008). Beauchamp (2011) noted that rough conditions at sea could bias counts and possibly alter which statistical distribution fits best to observed flock sizes. Further exploration of the counting process and the relationship of the observed counts to actual group sizes could lend insight into how the observed distributions might differ from the true underlying distributions of the biological process. Explicitly accounting for the effects of the observation process, such as by including covariates, detection functions, and upper limits imposed by the size of the observation unit, may lead to more accurate estimates of survey power. However, these gains would come at the cost of increased model complexity and decreased generality. Simulation studies might help lend insight into how severely detectability and observer bias problems could be expected to influence our results.

How sensitive are our results to the critical issue of detectability? Formal sensitivity analysis is beyond the scope of this study, but we can consider the problem of detectability in a straightforward way by observing the effect of changes in reference prevalence on the power curves shown in Digital Supplement C. Power to detect a given effect size obviously decreases for a given sample size when prevalence decreases. However, the effect of changing prevalence on the shape of the curves is fairly subtle over a wide range of prevalence values. Imperfect detectability would have an effect on these power curves similar to that of reducing prevalence. For example, the difference between curves with reference prevalence of 0.01 can be compared to those with reference prevalence of 0.02 to get a sense of the effect of 50% detectability. While imperfect detectability reduces power, it does not fundamentally change the nature of our results.

4.5. Spatial Scale and Structure

Seabird distributions are commonly characterized by spatial structure, ranging from small-scale spatial autocorrelation (patchiness), to large-scale gradients and trends. The methods described here do not explicitly incorporate either of these features. The first issue, spatial autocorrelation, is relevant to choosing the appropriate grid scale on which to conduct a power analysis for hotspot or coldspot detection. If the grid cell size is much smaller than the scale of autocorrelation (hotspots or coldspots are bigger than the grid cells), then power estimates from our method may be too low, as the number of surveys will be considered grid-cell-by-grid-cell rather than summed over the entire hot/coldspot. If the grid cell size is much larger than the scale of autocorrelation, then power estimates may be biased low, because data from the hot/coldspot will be contaminated with data from adjacent areas that are not hot/coldspots, reducing the effect magnitude. As an alternative to changing grid size, models that explicitly include spatial autocorrelation could be used—but these are much more complex to formulate, fit, analyze, and generalize.

The second issue, large-scale gradients and trends, is primarily an issue for choice of the reference distribution. If one chooses a reference distribution based on data collected over a strong spatial gradient, then the choice and parameters of the distribution could be biased. Moreover, results of hot/coldspot significance tests will be trivial as grid cells located at the low end of the spatial gradient will predictably be identified as coldspots and those at the high end of the gradient as hotspots. One solution to this problem is to carefully stratify the regions for which reference distributions are specified and power analyses are conducted to reduce the impact of trends and gradients. For example, regions could be stratified based on biogeographic breaks and onshore-offshore zones. This would involve more analytical effort to fit models and conduct power analyses for each regional stratum, but would be substantially easier than the alternative, which is to fit a trend model for each species prior to analysis and incorporate this trend model into the power analysis.

In the context of siting an offshore wind facility, there are several implications of the scale issues discussed above. First, the lease block scale used in the example applications presented here may not be the most appropriate scale. Choice of scale should involve consideration of the size of the project, and the scale of spatial autocorrelation ("patch size") of the long-term average seabird abundance surface. Typical offshore wind project size might be 5x5 or 6x6 nmi, and it will often be desirable for analyses of impacts to include a "buffer zone" around the project. Moreover, predictive modeling of long-term average seabird distributions in this region have found typical

spatial autocorrelation scales of 10-15km (5.4-8.1 nmi) (Kinlan et al. 2012). Both of these pieces of information suggest that power analyses conducted on aggregations of 2x2 or 3x3 BOEM lease blocks (i.e., 6x6 or 9x9 nmi rectangles) might be more appropriate. Analysis at this scale would likely improve statistical power to detect a given effect size, provided the spatial autocorrelation estimates of 10-15km are appropriate. The issue of large-scale trends and gradients must also be considered. Practically speaking, this means that it will be very important to specify the geographic region and habitat that is to be used as a context (reference region) for each species of interest for determining whether the observed abundance at a given location represents a relative hotspot or coldspot. The most important gradients to account for will be onshore-offshore gradients and regional biogeographic differences (e.g., north vs. south of Cape Hatteras and Cape Cod).

To answer power questions about particular aggregations of blocks (e.g. a wind energy area [WEA]) simulations could be specifically tailored to answer the questions of interest. For example, if the question is "what is the power to detect a single hotspot at some unknown location in the WEA?", one could design a simulation to estimate a power curve for this problem. Alternatively one might ask "what is power to detect a hotspot given that there are, on average, a certain number of hotspots in the WEA?". Answering this question would require a different simulation. Although insight into these questions can be gained from the power curves and power maps at the lease block scale, simulations specific to a given WEA or lease block aggregation pattern are necessary for precise estimates of power for any given scenario.

4.6. Temporal Scale and Structure

The method described in this study assumes that samples at a given location are independent in time. This assumption could be violated in three primary ways: (1) if surveys are conducted very close to one another in time and are autocorrelated, (2) if data from different seasons are pooled together for species' whose abundances or group size distributions change seasonally, or (3) if data are pooled over strong interannual to interdecadal trends.

Short-term non-independence, (1), can be dealt with relatively easily by ensuring that surveys are conducted a sufficient amount of time apart. Our analyses of marine bird count time series in BOEM lease blocks that were repeatedly surveyed over short time scales (Figure 18), although limited, suggest that 5-15 days is nearly always sufficient to ensure that successive surveys of the same lease block will not be highly autocorrelated, and 3-5 days, or even less, is probably acceptable for most species. Of course, where a particular species is of interest and detailed repeat survey studies are available, the pattern of temporal autocorrelation should be studied in more detail. If repeat surveys are found to be autocorrelated, the results of this study can still be applied but the effective number of independent surveys will need to be adjusted to account for autocorrelation (Cressie 1993). Finally, biases can also be introduced by the time of day at which observations conducted for species with strong diel variation in behavior, a factor that should be considered in survey design for a particular species. We have not considered within-day variation in this study.

Predictable seasonal fluctuations, (2), are usually the dominant source of temporal variability in seabird time series for a particular region. It would be inappropriate to pool data across seasonal fluctuations for identification of reference distributions or for power analysis. We have accounted for seasonal effects in this study by conducting power analyses separately on a

seasonal basis, using definitions of seasons appropriate to the species and region of interest. The temporal analysis of variation in marine bird abundance (Figures 17, 18) suggests that our definition of seasons was effective in isolating seasonal variability, but this should be evaluated carefully for each new region to which this method is applied.

It is also inappropriate to pool data in the presence of strong long-term trends (3). Such trends may arise from long-term changes in population status or climate variability. Recommendations for dealing with temporal trends are similar to those for spatial trends. In some cases, one may be able to remove the effects of trends by stratifying by climate regime (e.g. phases of climate cycles like the North Atlantic Oscillation). Otherwise, trends could be fit and incorporated directly into the power analysis. Our analyses of long-term variation in marine bird abundance (for periods of 1-10 years) suggested trends were not a major issue (Figure 17), but the analysis of longer time series of climate indices (allowing resolution of time scales from 10 to 20 years) suggests more care should be taken to account for possible trends in data separated by more than 10 years.

These considerations lead to some straightforward recommendations for temporal design of surveys to assist in offshore wind siting and environmental assessment. First, in general, surveys should be conducted in all seasons in which the species of interest is present, and be spread across one to three years. Repeat surveys of the same location within the same season should be conducted at least a day or two and ideally three or more days apart if they are to count as independent surveys; more if there is evidence of longer-term correlation. Surveys should adequately cover both breeding and non-breeding seasons and locations, and analyses should be stratified or conducted separately on qualitatively different seasons, locations, or populations. The same holds true for adequately sampling other key life history phases such as migration. Finally, surveys may need to be repeated periodically if there is evidence for a major shift in ocean climate and/or a change in the large-scale distribution of the species of interest at longer time scales. However, assuming the patterns of the previous ~65 years continue to hold (an assumption that admittedly might need to be re-evaluated in light of global climate change), repeating 1-2 years of survey work at 10-15 year intervals would be adequate to characterize variability due to ocean/atmosphere climate fluctuations.

4.7. Selecting Species for Which this Method Will Apply

Obviously, this approach may not be appropriate for all species. For example, a candidate species would need to satisfy some minimum biological requirements (e.g., it aggregates, timing of those aggregations, persistence). Even if a species satisfies the biological requirement, there are minimum data requirements, too (number of transects, number of non-zero transects, number of unique transects, etc.). It would be useful to list the candidate species and to identify those that satisfy the data requirements. Third, even if a species satisfies the biological and information requirements, it still may not be tractable to conduct enough surveys for that species.

A particular challenge arises in the case of species for which we have little prior knowledge of the overall pattern of spatial distribution, and so have little ability to establish reliable reference regions or account for regional and onshore-offshore trends. The types of statistical techniques described here are no substitute for a detailed knowledge of the natural history of each species, including its overall regional and global pattern of distribution. Telemetry studies will be a particularly valuable complement to hotspot analyses, as they provide detailed insights into

individual behavior and habitat usage, and can reveal variation among individuals, habitat areas outside of the focal area, fine-scale timing of movements, and other important information not captured by ship-based and aerial at-sea surveys.

In general, the technique presented here performs best for more common species for which some observational data is already available. For very rare species and/or data-poor species, there will always be challenges. A combination of approaches will likely be necessary. In developing guidelines, regulators may want to consider categorizing birds by commonness/rarity and data availability and making different recommendations on those bases.

5.0 SUMMARY

We have developed and illustrated a simple, general method for defining species-specific hotspots and coldspots of occurrence and abundance for marine birds, and for assessing the significance and statistical power to detect these hot and coldspots. Given information about a species' regional occurrence and abundance patterns, this method can serve as the basis for general guidelines for the design of robust surveys to detect departures from regional average patterns of abundance and occurrence.

It should be emphasized that the power maps, power curves, and significance tests shown in this report and the Digital Supplements (listed in Appendix A) are intended only as examples of the application of this method, rather than as a definitive power analysis intended for operational use. The user should bear in mind that the spatial distribution of information in maps is dependent on the input data used. There are a variety of reasons that some datasets may not be reflected in these maps: some datasets existed but were not available to us, others were excluded because they were not of a consistent high scientific quality, and others may not yet been collected or made available at the time of this analysis.

The end-user will also need to decide the appropriate effect sizes to define biologically meaningful hot and coldspots (3x and 1/3x effect sizes chosen here are only illustrative), spatial and temporal scale and extent appropriate to management and regulatory decisions, and appropriate definitions of the data from which reference means and distributions are defined. Results of the power analysis and subsequent guidelines for the appropriate number of surveys to conduct in any given instance will depend on both scientific and regulatory decisions that influence these parameters.

Analyses of temporal variability, using environmental and seabird data from the U.S. Mid-Atlantic continental shelf, suggest that most of the interannual variance in relevant environmental correlates of seabird occurrence and abundance (sea surface temperature [SST] and surface chlorophyll concentration [*chl*] from satellite remote sensing) and in relative abundance of birds BOEM lease blocks, will be captured by surveys spread over 1 to 3 years. Some species in certain seasons may require longer periods to establish a baseline, but for most of the species analyzed this period appears sufficient to capture a large percentage of the variance at sub-decadal scales. Relatively little additional variance would be expected if 1-3 year surveys were extended to 4, 5, or even 10 years for most species. Analysis of regional ocean and atmosphere climate indices (North Atlantic Oscillation, Atlantic Multidecadal Oscillation) for the period from 1948 to 2012 support this conclusion, but also reveal that 20-40% additional variance in ocean climate variability occurs at decadal or greater time scales. To the extent that

it translates into variance in seabird occurrence and abundance, capturing this variability would require long-term sampling programs (20-40 years).

At short time scales, within a season, temporal autocorrelation drops off quickly for both environmental variables (SST, *chl*) and relative abundance of birds observed in repeated surveys on the same BOEM lease block. For most species/season combinations studied, surveys spaced 3-5 days or more apart will exhibit sufficient variance to be considered approximately independent.

Taken together, the results of this study represent a methodology for: a) using existing marine bird survey data to assess the state of knowledge about relative hotspots and coldspots of marine bird abundance and occurrence in offshore areas; b) planning future marine bird surveys in offshore areas to leverage existing data, and maximize probability of detecting any hotspots/coldspots of abundance/occurrence probability that may exist in discrete spatial planning blocks; and c) distributing sampling effort in time to ensure adequate representation of environmental and ecological variance.

Literature Cited

Allen, A.P., B.L. Li, and E.L. Charnov. 2001. Population fluctuations, power laws and mixtures of lognormal distributions. Ecol. Lett. 4:1-3.

Beauchamp, G. 2011. Fit of aggregation models to the distribution of group sizes in Northwest Atlantic seabirds. Mar. Ecol. Prog. Ser. 425:261-268.

Bennett, B.M. and P. Hsu. 1960. On the power function of the exact test for the 2 × 2 contingency table. Biometrika 47:393-398.

Bonabeau, E. and L. Dagorn. 1995. Possible universality in the size distribution of fish schools. Phys. Rev. E 51:R5220-R5223.

Bonabeau, E., L. Dagorn, and P. Freon. 1999. Scaling in animal group-size distributions. Proc. Natl. Acad. Sci. 96:4472–4477.

Burnham, K.P. and D.R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach, second ed. Springer-Verlag. 488 pp.

Caraco, T. 1980. Stochastic dynamics of avian foraging flocks. Am. Nat. 115:262-275.

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

Certain, G., E. Bellier, B. Planque, and V. Bretagnolle. 2007. Characterising the temporal variability of the spatial distribution of animals: an application to seabirds at sea. Ecogr. 30:695-708.

Cohen, J.E. 1972. Markov population processes as models of primate social and population dynamics. Theor. Popul. Biol. 3:119-134.

Cressie, N.A. 1993. Statistics for spatial data. John Wiley & Sons. Ontario, Canada.

Deutsch, C.V. and A.G. Journel. 1998. GSLIB: Geostatistical Software Library: and User's Guide, second ed. Oxford University Press. New York, NY. 369 pp.

Enfield, D.B., A. M. Mestas-Nunez and P.J. Trimble. 2001. The Atlantic multidecadal oscillation and it's relation to rainfall and river flows in the continental U.S. Geophys. Res. Lett. 28:2077-2080.

Foley, D. 2012. CoastWatch ERDDAP Datasets erdMEssta3day and erdMEchla3day: MODIS Aqua 3 day composite SST and Chl-a. NOAA NESDIS, NOAA SWFSC ERD. http://coastwatch.pfeg.noaa.gov/erddap/info/erdMEchla3day/index.html, http://coastwatch.pfeg.noaa.gov/erddap/info/erdMEchla3day/index.html, Accessed July 10, 2012. Email contact: dave.foley@noaa.gov

Fujisaki, I., E.V. Pearlstine, and M. Miller. 2008. Detecting population decline of birds using long-term monitoring data. Popul. Ecol. 50:275-284.

Gibrat, R. 1931. Les Inégalités Economiques. Librairie du Recueil Sirey. Paris, France.

Griesser, M., Q. Ma, S. Webber, K. Bowgen, and D.J.T. Sumpter. 2011. Understanding animal group-size distributions. PLoS ONE 6 doi:10.1371/journal.pone.0023438.

Hall, D. 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics 56:1030-1039.

Hope, A.C.A. 1968. A simplified Monte Carlo significance test procedure. Journal of the Royal Statistical Society Series B (Methodological) 30:582-598.

Jovani, R., D. Serrano, E. Ursua, and J.L. Tella. 2008. Truncated power laws reveal a link between low-level behavioural processes and grouping patterns in a colonial bird. PLoS One 3 e1992, doi:10.1371/journal.pone.0001992

Keitt, T.H. and H.E. Stanley. 1998. Dynamics of North American breeding bird populations. Nature 393:257-260.

Kinlan, B.P., C. Menza, and F. Huettmann. 2012. Predictive Modeling of Seabird Distribution Patterns in the New York Bight. Chapter 6 in C. Menza, B.P. Kinlan, D.S. Dorfman, M. Poti and C. Caldow (eds.). 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. Silver Spring, MD. 224 pp.

Limpert, E., W.A. Stahel, and M. Abbt. 2001. Log-normal distributions across the sciences: keys and clues. BioScience 51:341-352.

Link, W.A. and J.R. Sauer. 2007. A hierarchical analysis of population change with application to cerulean warblers. Ecol. 83:2832-2840.

Ma, Q., A. Johansson, and D.J.T. Sumpter. 2011. A first principles derivation of animal group size distributions. J. Theo. Biol. 283:35-43.

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. Ecol. Let. 8:1235-1246.

Mitzenmacher, M. 2004. A brief history of generative models for power law and lognormal distributions. Internet Math. 1:226-251.

Mullahy. J. 1986. Specification and testing of some modified count data models. J. Econometrics 33:341–365.

Murphy, K.R., B. Myors and A. Wolach. 2008. Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests, Third Edition. Routledge Academic. New York, NY. 224 pp.

Niwa, H.S. 2003. Power-law versus exponential distributions of animal group sizes. J. Theor. Biol. 224:451–457.

O'Connell, Jr., A.F., B. Gardner, A.T. Gilbert, and K. Laurent. 2009. Compendium of avian occurrence information for the continental shelf waters along the Atlantic coast of the United States (database section: seabirds). A final report for the U.S. Department of the Interior, Minerals Management Service, Atlantic OCS Region, Herndon, VA. 50 pp. Contract No. M08PG20033.

Okubo, A. 1986. Dynamical aspects of animal grouping. Advances in Biophysics 22:1-94.

Pearse, A.T., P.D. Gerard, S.J. Dinsmore, R.M. Kaminski, and K.J. Reinecke. 2008. Estimation and correction of visibility bias in aerial surveys of wintering ducks. J. Wild. Man. 72:808–813.

Pollock, K.H. and W.L. Kendall. 1987. Visibility bias in aerial surveys: a review of estimation procedures. J. Wild. Man. 51:502–510.

Preston, F.W. 1948. The commonness and rarity of species. Ecology 29:254-283.

Quenouille, M.H. 1949. A relationship between the logarithmic, Poisson, and negative binomial series. Biometrics 5:162-164.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Silverman, E.D., M. Kot, and E. Thompson. 2001. Testing a simple stochasitic model for the dynamics of waterfowl aggregation. Oecol. 128:608-617.

Sjoberg, M., B. Albrectsen, and J. Hjalten. 2000. Truncated power law: a tool for understanding aggregation patterns in animals? Ecol. Lett. 3:90–94.

Sokal, R.R. and F.J. Rohlf. 2012. Biometry: the principles and practice of statistics in biological research. 4th edition.

Tremblay, Y., S. Bertrand, R.W. Henry, M.A. Kappes, D.P. Costa and S.A. Shaffer. 2009. Analytical approaches to investigating seabird–environment interactions: a review. Mar. Ecol. Prog. Ser. 391:153–163.

Veit, R.R. and W.A. Montevecchi. 2006. The influences of climate change on marine birds. Acta Zool Sin 52: S165−S168. Freeman and Co. New York, NY. 937 pp.

Vuong, Q.H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57:307-333.

Wood, C.C. 1985. Aggregative response of common mergansers (*Mergus merganser*): predicting flock size and abundance on Vancouver Island salmon streams. Can. J. Fish. Aqua. Sci. 42:1259-1271.

Yee, T.W. 2010. The VGAM package for categorical data analysis. J. Stat. Soft. 32:1-34.

Zipkin, E.F., B. Gardner, A. Gilbert, A.F. O'Connell, 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. Oecol. 163:893-902.

Zipkin, E.F., J.B. Leirness, B.P. Kinlan, A.F. O'Connell, and E.D. Silverman. 2012. Fitting statistical distributions to sea duck count data: implications for survey design and abundance estimation. Stat. Methodol. doi:10.1016/j.stamet.2012.10.002

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS

Additional figures and tables are provided in seven Digital Supplements to this report, lettered from Digital Supplement A through Digital Supplement G. Each Digital Supplement is provided as an Adobe PDF file, and is available on request. Requests for Digital Supplements may be directed to:

Chris Caldow
Chief, Biogeography Branch
Center for Coastal Monitoring and Assessment
National Centers for Coastal Ocean Science
1305 East-West Hwy, SSMC-4, N/SCI-1
Silver Spring, MD 20910
Chris.Caldow@noaa.gov

Digital Supplement A (7 pages). Power vs. sample size curves for hotspot/coldspot tests of non-zero mean counts.

Figures A1-A6. Power vs. sample size curves for hotspot/coldspot tests of non-zero mean counts (i.e., case (1) described in section 1.2). Curves are presented for the six distributions in Table 1 for which a finite mean exists for realistic parameter values (Poisson, Negative Binomial, Geometric, Logarithmic, Discretized Lognormal, and Zeta with exponential cutoff). For each distribution, six panels show curves for different values of the reference mean, and within each panel lines of different colors show curves for different effect sizes, represented as multiples of the reference mean (e.g., an effect size of 0.33 for a reference mean of 10 corresponds to power to detect a coldspot with a mean of 3.3 or smaller). Note that the number of curves per panel varies, because some combinations of the reference mean and effect size do not make sense (for example, with a reference mean of 2, a 0.33 effect size would correspond to a mean of 0.66, which is not possible given that non-zero counts must be greater than or equal to 1). For distributions with more than one parameter, the first parameter is adjusted to produce the desired reference mean, and additional ("nuisance") parameters are held constant. Curves shown are examples for the value(s) of the nuisance parameter(s) given in the figure heading.

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS (cont'd)

Digital Supplement B (3 pages). Power vs. sample size curves for hotspot/coldspot tests of occurrence probability.

Figure B1. Power vs. sample size curves for hotspot/coldspot tests of occurrence probability (i.e., case (2) described in section 1.2). Curves are presented based on the binomial distribution, assuming that the probability of occurrence remains constant for a given species in a given place over the study period, and that the statistical test used is Fisher's Two-Proportion Exact Test (one-tailed, α=0.05). Points show where tests were evaluated; curves are linearly interpolated in between points. Each panel shows curves for a different value of the reference (e.g., regional) prevalence. Each color represents a different multiplicative effect size. For example, the red curve in the lower left panel is for the test of the alternative hypothesis (Ha): probability of occurrence = 0.6 versus the null hypothesis (Ho): probability of occurrence = 0.2, i.e. a 300% higher prevalence than the reference value.

Figure B2. Relationship of sampling effort to expected number of presences observed for different prevalence values. Curves show number of surveys (y axis) need to have a specified probability (color), of observing at least k presences (x axis) under a binomial distribution Binomial(N,p), where N is the total number of surveys and p is the species' prevalence (probability of occurrence). For example, the red curve in the lower left panel shows that when a species' prevalence is 33%, one would have to conduct 44 surveys to have a 95% chance of observing 10 presences. Analysis assumes surveys are independent and prevalences do not change over the time period studied. Each panel shows curves for a different value of prevalence.

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS (cont'd)

Digital Supplement C (7 pages). Power vs. sample size curves for hotspot/coldspot tests of unconditional mean counts (full hurdle model, including zero and non-zero components).

Figures C1-C6. Power vs. sample size curves for hotspot/coldspot tests of unconditional mean counts (i.e., case (3) described in section 1.2). Curves are presented for the six distributions in Table 1 for which a finite mean exists for realistic parameter values (Poisson, Negative Binomial, Geometric, Logarithmic, Discretized Lognormal, and Zeta with exponential cutoff). For each distribution, nine panels show curves for different combinations of the reference mean value and the prevalence value. Within each panel lines of different colors show curves for different effect sizes, represented as multiples of the non-zero reference mean (e.g., an effect size of 0.33 for a reference mean of 10 corresponds to power to detect a coldspot with a non-zero mean of 3.3 or smaller). Note that the number of curves per panel varies, because some combinations of the reference mean and effect size do not make sense (for example, with a reference mean of 2, a 0.33 effect size would correspond to a mean of 0.66, which is not possible given that non-zero counts must be greater than or equal to 1). For distributions with more than one parameter, the first parameter is adjusted to produce the desired reference mean, and additional ("nuisance") parameters are held constant. Curves shown are examples for the value(s) of the nuisance parameter(s) given in the figure heading. The prevalence is assumed to remain unchanged regardless of effect size (that is, changes in the mean abundance are assumed to occur through changes in the non-zero counts, rather than through changes in prevalence).

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS (cont'd)

Digital Supplement D (9 pages). Additional information about datasets extracted from the USGS Avian Compendium Database.

Table D1. List of science-quality datasets in the USGS Avian Compendium Database as of August 2012.

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS (cont'd)

Digital Supplement E (88 pages). Model fit and model selection information.

Table E1. Model fit and selection statistics for non-zero count data in (a) Spring, (b) Summer, (c) Fall, (d) Winter. Maximum likelihood estimates of the best-fitting parameters for each of the top three candidate distributions are shown for each species. Model selection statistics (AICc and log-likelihood values) are also given. For each species, the top three models are shown ranked from lowest to highest AICc. The top-ranked model (lowest AIC) was used for subsequent analyses (see Tables 4 and 5 in main document). Species appear in the same order within each season as in Table 4 of the main document.

Figures E1-E74. Model fit plots. Maximum likelihood model fits (lines) and observed probabilities (black dots) for non-zero count data for all modeled species. Fits are shown for the top four models, ranked in the legend from lowest to highest AICc. Plots are presented grouped by season, with species appearing in the same order within each season as in Table 4 of the main document:

Figures E1-E19. Spring

Figures E20-E37. Summer

Figures E38-E59. Fall

Figures E60-E74. Winter

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS (cont'd)

Digital Supplement F (286 pages). Maps and figures for **conditional (non-zero count)** power analyses and significance tests.

Maps depict results in BOEM Atlantic OCS lease blocks.

The user should keep in mind that the spatial distribution of information in maps is dependent on the input data used. There are a variety of reasons that some datasets may not be reflected in these maps: some datasets existed but were not available to us, others were excluded because they were not of a consistent high scientific quality, and others may not yet been collected or made available at the time of this analysis. These maps are intended as a demonstration of the methods described in OCS Study BOEM 2012-101.

SECTION I. Summary Statistic Maps Calculated for All Species

Summary statistics (number of occurrences and average, maximum, and minimum hotspot and coldspot power) were calculated across all species in all seasons combined and for each season.

Figures F1-F7. All Seasons Combined

Number of occurrences summed over all species in all seasons
Average, maximum, and minimum power to detect 3x hotspots of non-zero abundance
Average, maximum, and minimum power to detect 1/3x coldspots of non-zero abundance

Figures F8-F14. Spring

Number of occurrences summed over all species in spring
Average, maximum, and minimum power to detect 3x hotspots of non-zero abundance
Average, maximum, and minimum power to detect 1/3x coldspots of non-zero abundance

Figures F15-F21. Summer

Number of occurrences summed over all species in summer
Average, maximum, and minimum power to detect 3x hotspots of non-zero abundance
Average, maximum, and minimum power to detect 1/3x coldspots of non-zero abundance

Figures F22-F28. Fall

Number of occurrences summed over all species in fall
Average, maximum, and minimum power to detect 3x hotspots of non-zero abundance
Average, maximum, and minimum power to detect 1/3x coldspots of non-zero abundance

Figures F29-F35. Winter

Number of occurrences summed over all species in winter
Average, maximum, and minimum power to detect 3x hotspots of non-zero abundance
Average, maximum, and minimum power to detect 1/3x coldspots of non-zero abundance

SECTION II. Species-specific Power Analysis Maps and Figures

Results of the non-zero conditional model are presented as a set of 6 figures for each included species in each season. Within each season, species are presented in the same order as in Table 4 of the main document, except that only species for which maps were created ("Maps created?" = "Yes" in 3rd column of Table 4) are included.

Figures F36-F101. Spring power analysis maps and figures (11 species x 6 figures per species).
Figures F102-F143. Summer power analysis maps and figures (7 species x 6 figs. per species).
Figures F144-F215. Fall power analysis maps and figures (12 species x 6 figs. per species).
Figures F216-F275. Winter power analysis maps and figures (10 species x 6 figs. per species).

1st Figure for each Species: Map of number of occurrences of this species in this season in BOEM Atlantic OCS lease blocks.

2nd Figure for each Species: Map of the mean non-zero count in for this species in this season in BOEM Atlantic OCS lease blocks.

3rd Figure for each Species: Power vs. sample size curves for 3x hotspot and 1/3x coldspot detection for this species, given the selected model fit and reference mean.

4th Figure for each Species: Map of power to detect 3x hotspots of non-zero abundance.

5th Figure for each Species: Map of power to detect 1/3x coldspots of non-zero abundance.

6th Figure for each Species: Combined map of hotspot (red) and coldspot (blue) significance test p-values, based on one-sample, one-tailed (hotspot) Monte Carlo significance tests of the mean non-zero count in each lease block compared to the reference mean. Darker shading indicates greater statistical significance. Lease blocks that did not approach statistical significance ($p>0.2$) are shown in grey, with the intensity of the shading proportional to the average of 3x hotspot and 1/3x coldspot power values for that cell. That is, 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. In contrast, light grey shading indicates lease blocks not identified as significant hotspots or coldspots, but for which there was little or no power to detect a hotspot or coldspot, had it existed. The darkest blue lease blocks can therefore 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. Hotspot (coldspot) significance does not consider whether high (low) abundances persisted across years or occurred in the same year; if interannual persistence is of concern, the temporal distribution of the data should be examined. P-values are not corrected for the large number of simultaneous tests performed (two tests per lease block in which the species occurred), so many of the lighter red and blue lease blocks are likely false positives. The most significant values (darkest red and blue) are more reliable, but will still contain some false positives. Similarly, the lightest grey cells have the highest chance of being false negatives, whereas the darkest grey cells have the lowest chance of being false negatives.

APPENDIX A. LIST OF DIGITAL SUPPLEMENTS (cont'd)

Digital Supplement G (246 pages). Maps and figures for **full hurdle model (zero and non-zero count)** power analyses and significance tests.

Maps depict results in BOEM Atlantic OCS lease blocks.

The user should keep in mind that the spatial distribution of information in maps is dependent on the input data used. There are a variety of reasons that some datasets may not be reflected in these maps: some datasets existed but were not available to us, others were excluded because they were not of a consistent high scientific quality, and others may not yet been collected or made available at the time of this analysis. These maps are intended as a demonstration of the methods described in OCS Study BOEM 2012-101.

SECTION I. Summary Statistic Maps Calculated for All Species

Summary statistics (number of times each lease block was surveyed and average, maximum, and minimum hotspot and coldspot power) were calculated across all species in all seasons combined and for each season individually.

Figures G1-G7. All Seasons Combined

Number of times each lease block was surveyed, summed over all seasons
Average, maximum, and minimum power to detect 3x hotspots of abundance
Average, maximum, and minimum power to detect 1/3x coldspots of abundance

Figures G8-G14. Spring

Number of times each lease block was surveyed in spring
Average, maximum, and minimum power to detect 3x hotspots of abundance
Average, maximum, and minimum power to detect 1/3x coldspots of abundance

Figures G15-G21. Summer

Number of times each lease block was surveyed in summer
Average, maximum, and minimum power to detect 3x hotspots of abundance
Average, maximum, and minimum power to detect 1/3x coldspots of abundance

Figures G22-G28. Fall

Number of times each lease block was surveyed in fall
Average, maximum, and minimum power to detect 3x hotspots of abundance
Average, maximum, and minimum power to detect 1/3x coldspots of abundance

Figures G29-G35. Winter

Number of times each lease block was surveyed in winter
Average, maximum, and minimum power to detect 3x hotspots of abundance
Average, maximum, and minimum power to detect 1/3x coldspots of abundance

SECTION II. Species-specific Power Analysis Maps and Figures

Results of the full hurdle model (for zero and non-zero counts) are presented as a set of 5 figures for each included species in each season. Within each season, species are presented in the same order as in Table 4 of the main document, except that only species for which maps were created ("Maps created?" = "Yes" in 3rd column of Table 4) are included.

Figures G36-G90. Spring power analysis maps and figures (11 species x 5 figures per species).
Figures G91-G125. Summer power analysis maps and figures (7 species x 5 figs. per species).
Figures G126-G185. Fall power analysis maps and figures (12 species x 5 figs. per species).
Figures G186-G235. Winter power analysis maps and figures (10 species x 5 figs. per species).

1st Figure for each Species: Map of the mean count (including any zeros) for this species in this season in BOEM Atlantic OCS lease blocks.

2nd Figure for each Species: Power vs. sample size curves for 3x hotspot and 1/3x coldspot detection for this species, given the selected model fit, reference mean, and reference prevalence.

3rd Figure for each Species: Map of power to detect 3x hotspots of abundance.

4th Figure for each Species: Map of power to detect 1/3x coldspots of abundance.

5th Figure for each Species: Combined map of hotspot (red) and coldspot (blue) significance test p-values, based on one-sample, one-tailed (hotspot) Monte Carlo significance tests of the mean count in each lease block compared to the expectation from the reference mean/prevalence. Darker shading indicates greater statistical significance. Lease blocks that did not approach statistical significance ($p>0.2$) are shown in grey, with the intensity of the shading proportional to the average of 3x hotspot and 1/3x coldspot power values for that cell. That is, 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. In contrast, light grey shading indicates lease blocks not identified as significant hotspots or coldspots, but for which there was little or no power to detect a hotspot or coldspot, had it existed. The darkest blue lease blocks can therefore 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 that were not surveyed in this season. Hotspot (coldspot) significance does not consider whether high (low) abundances persisted across years or occurred in the same year; if interannual persistence is of concern, the temporal distribution of the data should be examined. P-values are not corrected for the large number of simultaneous tests performed (two tests for each lease block that was surveyed in this season), so many of the lighter red and blue lease blocks are likely false positives. Note that there are many more tests performed in these maps than in the corresponding maps presented in Digital Supplement F, because of the larger number of lease blocks considered; the number of false-positives will be correspondingly higher. The most significant values (darkest red and blue) are more reliable, but will still contain some false positives. Similarly, the lightest grey cells have the highest chance of being false negatives, whereas the darkest grey cells have the lowest chance of being false negatives.

The Department of the Interior Mission

As the Nation's principal conservation agency, the Department of the Interior has responsibility for most of our nationally owned public lands and natural resources. This includes fostering the sound use of our land and water resources, protecting our fish, wildlife and biological diversity; preserving the environmental and cultural values of our national parks and historical places; and providing for the enjoyment of life through outdoor recreation. The Department assesses our energy and mineral resources and works to ensure that their development is in the best interests of all our people by encouraging stewardship and citizen participation in their care. The Department also has a major responsibility for American Indian reservation communities and for people who live in island communities.

The Bureau of Ocean Energy Management

The Bureau of Ocean Energy Management (BOEM) works to manage the exploration and development of the nation's offshore resources in a way that appropriately balances economic development, energy independence, and environmental protection through oil and gas leases, renewable energy development and environmental reviews and studies.

www.boem.gov

U.S. Department of Commerce

Dr. Rebecca M. Blank, Acting Secretary

National Oceanic and Atmospheric Administration

Dr. Jane Lubchenco, Under Secretary for Oceans and Atmosphere

National Ocean Service

David M. Kennedy, Assistant Administrator for Ocean Service and Coastal Zone Management

The National Centers for Coastal Ocean Science

Dr. W. Russell Callender, Director

The National Centers for Coastal Ocean Science provides research, scientific information and tools to help balance the nation's ecological, social and economic goals. Our partnerships with local and national coastal managers are essential in providing science and services to benefit communities around the nation.

coastalscience.noaa.gov