

# Additional Statistical Analyses to Support Guidelines for Marine Avian Sampling



# **Additional Statistical Analyses to Support Guidelines for Marine Avian Sampling**

September 2018

Authors:

Jeffery B. Leirness and Brian P. Kinlan

Prepared under Interagency Agreement M12PG00068

By

National Oceanic and Atmospheric Administration, National Ocean Service

National Centers for Coastal Ocean Science

Marine Spatial Ecology Division, Biogeography Branch

1305 East-West Hwy, SSMC-4, N/SCI-1

Silver Spring, MD 20910

## DISCLAIMER

This study was funded, in part, by the US Department of the Interior, Bureau of Ocean Energy Management (BOEM), Office of Renewable Energy, Sterling, VA, through Interagency Agreement Number M12PG00068 with the National Oceanic and Atmospheric Administration, National Ocean Service, National Centers for Coastal Ocean Science. This report has been technically reviewed by BOEM, and it has been approved for publication. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the US Government, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

## REPORT AVAILABILITY

To download a PDF file of this report, go to the US Department of the Interior, Bureau of Ocean Energy Management Data and Information Systems webpage (<https://www.boem.gov/Environmental-Studies-EnvData/>), click on the link for the Environmental Studies Program Information System (ESPIS), and search on 2018-063. The report is also available at the National Technical Reports Library at <https://ntrl.ntis.gov/NTRL/>.

## CITATION

Leirness JB, Kinlan BP. 2018. Additional statistical analyses to support guidelines for marine avian sampling. Sterling (VA): US Department of the Interior, Bureau of Ocean Energy Management. OCS Study BOEM 2018-063. iii+43 p.

## ABOUT THE COVER

Cover photograph of Cory's Shearwater *Calonectris diomedea* with Wilson's Storm-Petrel *Oceanites oceanicus* flock provided by David M. Pereksta. Used with permission.

## ACKNOWLEDGEMENTS

We are grateful to the many scientists, seabird observers, and ship and flight crews who contributed the survey data analyzed in this study. We thank Allan O'Connell, Mark Wimer, Allison Sussman, Tim Jones, Kaycee Coleman, Kyle Dettloff, and Robert Fowler for processing and providing the data in the Northwest Atlantic Seabird Catalog (formerly Avian Compendium) database. We are also thankful to Elise Zipkin for providing computer code used in previous analyses, Jesse Cleary and Arliss Winship for technical assistance with the core abundance area calculations, and David Bigger, Mary Boatman, and Arliss Winship for comments on earlier versions of this document. Jeffery Leirness was supported by NOAA contract number EA-133C-14-NC-1384 with CSS, Inc.

## Executive Summary

Human activities and development in the ocean environment will have certain impacts on marine bird populations. A first step in evaluating these impacts is to gain a clear understanding of species distribution and abundance patterns to identify areas of high and low use. We follow a general method developed by Kinlan et al. (2012) that takes a minimum amount of input data to identify species-specific areas of high use (i.e., “hotspots”) and low use (i.e., “coldspots”) and evaluate the statistical power to detect these locations based on existing survey data. We investigated species-specific hotspots and coldspots of occurrence, non-zero abundance, and unconditional abundance at various spatial scales and the statistical power to detect them at various effect sizes.

We applied these methods to twenty species of marine birds commonly found within the US Atlantic Outer Continental Shelf (OCS) region: Common Eider *Somateria mollissima*, Surf Scoter *Melanitta perspicillata*, White-winged Scoter *M. fusca*, Long-tailed Duck *Clangula hyemalis*, Razorbill *Alca torda*, Atlantic Puffin *Fratercula arctica*, Laughing Gull *Leucophaeus atricilla*, Herring Gull *Larus argentatus*, Least Tern *Sternula antillarum*, Roseate Tern *Sterna dougallii*, Common Tern *S. hirundo*, Royal Tern *Thalasseus maximus*, Red-throated Loon *Gavia stellata*, Common Loon *G. immer*, Black-capped Petrel *Pterodroma hasitata*, Cory’s Shearwater *Calonectris diomedea*, Sooty Shearwater *Ardenna grisea*, Great Shearwater *A. gravis*, Audubon’s Shearwater *Puffinus lherminieri*, and Northern Gannet *Morus bassanus*.

Ninety standardized, science-quality datasets containing geographically referenced counts of marine birds collected between 1978 and 2015 within the Atlantic OCS were used for this study. Individual datasets were previously compiled into the Northwest Atlantic Seabird Catalog and summarized into discrete spatial units of roughly 4 km in length. Independent analyses were completed for each species and season (spring, summer, fall, winter) combination that had sufficient data.

This study provides additional guidance on the identification of species-specific hotspots and coldspots for marine birds and the statistical power to detect them in the context of wind energy development on the Atlantic OCS. However, these methods are adaptable to other geographic areas and use scenarios to aid marine spatial planning and allocation of future survey efforts. We conclude with general recommendations to consider when completing spatial power analyses and present an example interpretation of results for Common Eider *Somateria mollissima* during the winter season in Nantucket Sound.

# Contents

List of Figures . . . . .	ii
List of Tables . . . . .	ii
1 Introduction . . . . .	1
2 Methods . . . . .	1
2.1 Survey Data . . . . .	1
2.2 Spatial Grid . . . . .	6
2.3 Species Analyzed . . . . .	6
2.4 Statistical Model . . . . .	8
2.5 Hotspot/Coldspot Identification . . . . .	9
2.6 Hotspot/Coldspot Persistence . . . . .	10
2.7 Power Estimation . . . . .	11
2.8 Spatial Resolution . . . . .	12
2.9 Stratification . . . . .	12
2.10 Implementation . . . . .	13
3 Results . . . . .	13
3.1 Hotspot/Coldspot Identification . . . . .	15
3.2 Hotspot/Coldspot Persistence . . . . .	15
3.3 Power Estimation . . . . .	17
3.4 Stratification . . . . .	21
4 Discussion . . . . .	21
4.1 Model Type . . . . .	21
4.2 Type I and Type II Error Rates . . . . .	22
4.3 Adjusted vs. Unadjusted $p$ -values . . . . .	22
4.4 Effect Size and Spatial Resolution . . . . .	23
5 Case Study: Common Eider in Nantucket Sound during Winter . . . . .	24
5.1 Hotspots/Coldspots of Abundance . . . . .	24
5.2 Hotspots/Coldspots of Occurrence . . . . .	25
5.3 Hotspots/Coldspots of Non-zero Abundance . . . . .	25
5.4 Conclusions . . . . .	26
6 References . . . . .	30
Appendix A: Boxplots of Hotspot/Coldspot $p$ -values vs. Interannual Persistence . . . . .	32
Appendix B: Sample Size Requirements for 80% Power . . . . .	38

## List of Figures

Figure 1. Map of the study area . . . . .	2
Figure 2. Boxplots showing the ratio of sample sizes needed for 80% power to detect hotspots of the specified effect size combination for each model type (type I error rate = 0.05) . . . . .	20
Figure 3. Boxplots showing the ratio of sample sizes needed for 80% power to detect coldspots of the specified effect size combination for each model type (type I error rate = 0.05) . . . . .	20
Figure 4. Maps of significance test $p$ -values and power to detect hotspots/coldspots of abundance for Common Eider <i>Somateria mollissima</i> during winter in Nantucket Sound (based on the combined model with grid cell resolution of $19.2 \times 19.2$ km) . . . . .	27
Figure 5. Maps of significance test $p$ -values and power to detect hotspots/coldspots of occurrence for Common Eider <i>Somateria mollissima</i> during winter in Nantucket Sound (based on the occurrence probability model with grid cell resolution of $4.8 \times 4.8$ km) . . . . .	28
Figure 6. Maps of significance test $p$ -values and power to detect hotspots/coldspots of non-zero abundance for Common Eider <i>Somateria mollissima</i> during winter in Nantucket Sound (based on the non-zero count model with grid cell resolution of $4.8 \times 4.8$ km) . . . . .	29
Figure A1. Boxplots of hotspot $p$ -values vs. interannual persistence for the occurrence probability model (pooled across all grid cell resolutions). . . . .	32
Figure A2. Boxplots of coldspot $p$ -values vs. interannual persistence for the occurrence probability model (pooled across all grid cell resolutions). . . . .	33
Figure A3. Boxplots of hotspot $p$ -values vs. interannual persistence for the non-zero count model (pooled across all grid cell resolutions). . . . .	34
Figure A4. Boxplots of coldspot $p$ -values vs. interannual persistence for the non-zero count model (pooled across all grid cell resolutions). . . . .	35
Figure A5. Boxplots of hotspot $p$ -values vs. interannual persistence for the combined model (pooled across all grid cell resolutions). . . . .	36
Figure A6. Boxplots of coldspot $p$ -values vs. interannual persistence for the combined model (pooled across all grid cell resolutions). . . . .	37

## List of Tables

Table 1. Datasets used for analyses . . . . .	3
Table 2. Species analyzed . . . . .	7
Table 3. Number of transect segments with sightings and total counts for each species-season combination . . . . .	7
Table 4. Parameter estimates . . . . .	13

Table 5. Summary of hotspot $p$ -values . . . . .	16
Table 6. Summary of coldspot $p$ -values . . . . .	17
Table 7. Percentage of grid cells classified as hotspot, coldspot, or neither for each model type at a range of type I error rates . . . . .	18
Table B1. Sample size requirements for 80% power based on the occurrence probability model (type I error rate = 0.05) . . . . .	38
Table B2. Sample size requirements for 80% power based on the non-zero count model (type I error rate = 0.05) . . . . .	40
Table B3. Sample size requirements for 80% power based on the combined model (type I error rate = 0.05) . . . . .	42

**List of Additional Appendices (presented as separate documents)**

Appendix C: Power Analysis Results for the Occurrence Probability Model

Appendix D: Power Analysis Results for the Non-zero Count Model

Appendix E: Power Analysis Results for the Combined Model

# 1 Introduction

Interest in offshore renewable energy activities in the United States has increased dramatically in recent years. Offshore wind energy facilities, in particular, have the potential to capitalize on areas within state and federal waters that have persistent high winds and large energy production capabilities (Adams et al. 2016). However, offshore wind energy infrastructure will have certain negative impacts on marine bird populations through collision and habitat loss. Understanding these impacts is an important part of responsible marine planning and environmental stewardship, and requires a clear understanding of species distribution and abundance patterns to identify areas of low use (i.e., “coldspots”) that could serve as potential development sites and high use (i.e., “hotspots”) where development should be avoided.

A general framework for identifying species-specific hotspots and coldspots for marine birds and assessing the statistical power to detect these locations based on visual survey count data was developed by Kinlan et al. (2012). We utilize this approach and expand on examples from Kinlan et al. (2012) and Zipkin et al. (2015). Much of the motivation, assumptions, and detailed methods underlying the statistical framework will not be repeated here. Readers are highly encouraged to familiarize themselves with Kinlan et al. (2012, especially Sections 1.1–2.9) and Zipkin et al. (2015) in order to gain a full understanding.

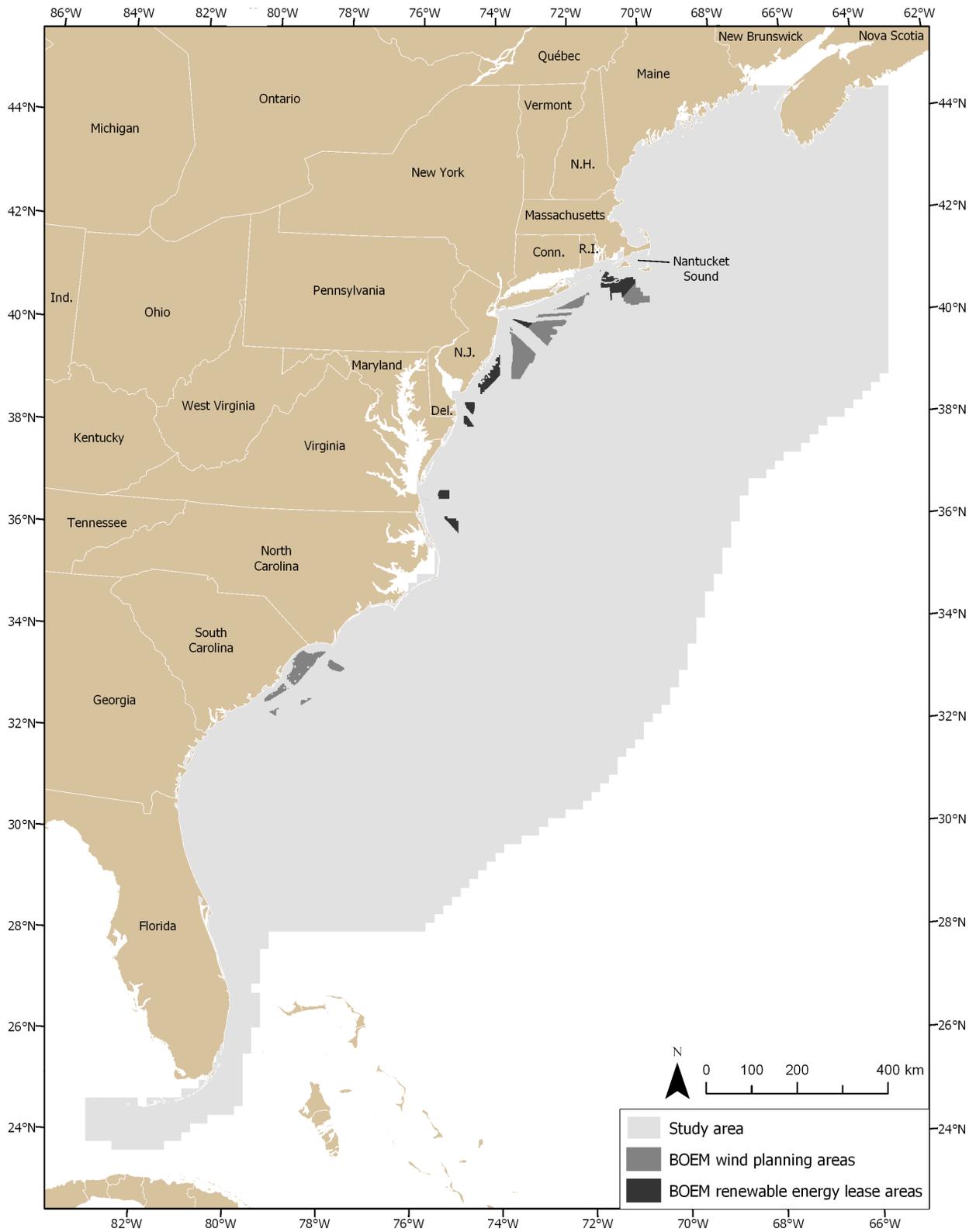
The purpose of this study is to explore refinements to the Kinlan et al. (2012) framework by investigating species-specific hotspots and coldspots of occurrence, non-zero abundance, and unconditional abundance at various spatial scales and the statistical power to detect hotspots and coldspots of various effect sizes. Results from this study provide additional guidance on the identification of species-specific hotspots and coldspots for marine birds. The methods presented in this report can be used to aid marine spatial planning and allocation of future survey efforts.

This study focuses on the Atlantic waters of the continental United States, extending inshore to the coast and offshore to the outer boundary of the Bureau of Ocean Energy Management (BOEM) Atlantic Outer Continental Shelf (OCS) region (**Figure 1**). This area contains multiple locations where renewable energy development is actively being pursued. The methods presented here, however, are not unique to this study region and could be adapted for other geographic areas of interest.

## 2 Methods

### 2.1 Survey Data

Data used in these analyses comprised 90 science-quality, geographically referenced survey datasets from the Northwest Atlantic Seabird Catalog (NWASC; formerly US Geological Survey Avian Compendium Database [O’Connell et al. 2009]), currently maintained by the US Fish and Wildlife Service (**Table 1**; additional details on individual datasets are provided in Winship et al. [2018, Appendix A]). Datasets included at-sea counts of birds collected between 1978 and 2015 using standard fixed-wing aerial and boat-based strip transect survey methodology. The width of the strip transect varied among surveys, but the vast majority of boat-based surveys used a 300 m wide strip while most aerial surveys used a 400 m wide strip. Eighty-seven survey datasets had counts that were continuously recorded and geographically referenced at the time of each bird sighting, along with geographically referenced boat/plane location data recorded at regular intervals (e.g., every five seconds). Three datasets collected between 1978 and 1988 (all boat-based) had counts that were binned into discrete transect segments (e.g., 15 minute duration) during data collection. Three aerial survey datasets (labeled as “aerial-camera” Platform in **Table 1**) used high-resolution digital video to record bird sightings, which were then processed to obtain geographically referenced counts for each species.



**Figure 1. Map of the study area**

BOEM wind planning and renewable energy lease area boundaries are current as of July 25, 2018.

**Table 1. Datasets used for analyses**

Datasets were extracted from the Northwest Atlantic Seabird Catalog (NWASC) and standardized to 4 km transect segments as described in **Section 2.1**. Seasons were defined as spring (March 1 to May 31), summer (June 1 to August 31), fall (September 1 to November 30), and winter (December 1 to February 29).

NWASC Dataset ID	Platform	Start date	End date	Number of transect segments				
				Total	Spring	Summer	Fall	Winter
AMAPPS_FWS_Aerial_Fall2012	aerial	09-29-2012	10-11-2012	3,144	0	0	3,144	0
AMAPPS_FWS_Aerial_Fall2013	aerial	09-16-2013	09-27-2013	5,015	0	0	5,015	0
AMAPPS_FWS_Aerial_Fall2014	aerial	10-06-2014	10-22-2014	3,056	0	0	3,056	0
AMAPPS_FWS_Aerial_Preliminary_Summer2010	aerial	08-03-2010	08-24-2010	1,235	0	1,235	0	0
AMAPPS_FWS_Aerial_Spring2012	aerial	03-17-2012	03-31-2012	3,129	3,129	0	0	0
AMAPPS_FWS_Aerial_Summer2011	aerial	07-30-2011	08-23-2011	3,799	0	3,799	0	0
AMAPPS_FWS_Aerial_Winter2010-2011	aerial	12-03-2010	01-17-2011	574	0	0	0	574
AMAPPS_FWS_Aerial_Winter2014	aerial	01-28-2014	02-11-2014	3,226	0	0	0	3,226
AMAPPS_NOAA/NMFS_NEFSCBoat2011	boat	06-04-2011	07-31-2011	1,363	0	1,363	0	0
AMAPPS_NOAA/NMFS_NEFSCBoat2013	boat	07-01-2013	08-18-2013	1,423	0	1,423	0	0
AMAPPS_NOAA/NMFS_NEFSCBoat2014	boat	03-12-2014	04-27-2014	952	952	0	0	0
AMAPPS_NOAA/NMFS_NEFSCBoat2015	boat	06-11-2015	06-18-2015	244	0	244	0	0
AMAPPS_NOAA/NMFS_SEFSCBoat2011	boat	06-20-2011	07-29-2011	900	0	900	0	0
AMAPPS_NOAA/NMFS_SEFSCBoat2013	boat	07-16-2013	09-09-2013	897	0	647	250	0
BarHarborWW05	boat	06-16-2005	10-19-2005	1,033	0	840	193	0
BarHarborWW06	boat	06-21-2006	10-15-2006	1,203	0	847	356	0
CapeHatteras0405	boat	08-04-2004	02-02-2005	268	0	159	0	109
CapeWindAerial	aerial	03-17-2002	02-27-2004	4,599	1,088	1,099	1,156	1,256
CapeWindBoat	boat	04-17-2002	09-12-2003	290	114	142	34	0
CDASMidAtlantic	aerial	12-19-2001	03-08-2003	1,685	247	0	0	1,438
CSAP*	boat	04-16-1980	10-12-1988	24,236	7,157	6,420	6,599	4,060
DOEBRIAerial2012	aerial-camera	03-26-2012	12-18-2012	4,433	1,434	798	1,462	739
DOEBRIAerial2013	aerial-camera	02-12-2013	12-31-2013	5,127	898	1,570	1,432	1,227
DOEBRIAerial2014	aerial-camera	01-01-2014	05-23-2014	2,310	1,004	0	0	1,306
DOEBRIBoatApr2014	boat	04-01-2014	04-05-2014	156	156	0	0	0
DOEBRIBoatApril2012	boat	04-25-2012	04-29-2012	160	160	0	0	0
DOEBRIBoatAug2012	boat	08-10-2012	08-14-2012	161	0	161	0	0
DOEBRIBoatAug2013	boat	07-30-2013	08-02-2013	162	0	162	0	0
DOEBRIBoatDec2012	boat	12-15-2012	01-03-2013	157	0	0	0	157
DOEBRIBoatDec2013	boat	12-11-2013	12-17-2013	167	0	0	0	167

continued on next page

Table 1 continued

NWASC Dataset ID	Platform	Start date	End date	Number of transect segments				
				Total	Spring	Summer	Fall	Winter
DOEBRIBoatJan2013	boat	01-28-2013	02-05-2013	158	0	0	0	158
DOEBRIBoatJan2014	boat	01-26-2014	02-01-2014	160	0	0	0	160
DOEBRIBoatJune2012	boat	06-18-2012	06-21-2012	162	0	162	0	0
DOEBRIBoatJune2013	boat	06-18-2013	06-22-2013	163	0	163	0	0
DOEBRIBoatMar2013	boat	03-20-2013	03-26-2013	160	160	0	0	0
DOEBRIBoatMay2013	boat	05-05-2013	05-09-2013	162	162	0	0	0
DOEBRIBoatNov2012	boat	11-04-2012	11-11-2012	160	0	0	160	0
DOEBRIBoatOct2013	boat	10-26-2013	10-30-2013	165	0	0	165	0
DOEBRIBoatSep2012	boat	09-06-2012	09-09-2012	160	0	0	160	0
DOEBRIBoatSep2013	boat	09-06-2013	09-10-2013	165	0	0	165	0
DominionVirginia_VOWTAP	boat	05-14-2013	04-03-2014	78	24	18	12	24
EcoMonAug08	boat	08-14-2008	08-26-2008	455	0	455	0	0
EcoMonAug09	boat	08-17-2009	08-28-2009	436	0	436	0	0
EcoMonAug10	boat	08-18-2010	09-01-2010	488	0	474	14	0
EcoMonAug2012	boat	08-07-2012	08-24-2012	626	0	626	0	0
EcoMonFeb10	boat	02-03-2010	02-17-2010	341	0	0	0	341
EcoMonFeb2012	boat	02-03-2012	02-20-2012	532	0	0	0	532
EcoMonFeb2013	boat	02-10-2013	02-25-2013	504	0	0	0	504
EcoMonJan09	boat	01-29-2009	02-12-2009	378	0	0	0	378
EcoMonJun2012	boat	05-31-2012	06-14-2012	531	31	500	0	0
EcoMonMay07	boat	05-23-2007	06-02-2007	489	423	66	0	0
EcoMonMay09	boat	05-28-2009	06-09-2009	599	189	410	0	0
EcoMonMay10	boat	05-26-2010	06-09-2010	610	260	350	0	0
EcoMonNov09	boat	11-03-2009	11-19-2009	425	0	0	425	0
EcoMonNov10	boat	11-05-2010	11-21-2010	399	0	0	399	0
EcoMonNov2011	boat	10-31-2011	11-18-2011	443	0	0	443	0
EcoMonOct2012	boat	10-27-2012	11-13-2012	480	0	0	480	0
FLPowerLongIsland_Aerial	aerial	10-13-2004	03-30-2006	272	91	42	139	0
FLPowerLongIsland_Boat	boat	04-08-2004	06-01-2006	986	404	134	274	174
FWS_MidAtlanticDetection_Spring2012	aerial	03-06-2012	03-07-2012	188	188	0	0	0
FWS_SouthernBLSC_Winter2012	aerial	02-04-2012	02-21-2012	918	0	0	0	918
FWSAtlanticWinterSeaduck2008	aerial	02-07-2008	02-17-2011	9,405	82	0	0	9,323
GeorgiaPelagic*	boat	11-15-1982	06-16-1985	1,822	538	578	508	198

continued on next page

Table 1 continued

NWASC Dataset ID	Platform	Start date	End date	Number of transect segments				
				Total	Spring	Summer	Fall	Winter
HatterasEddyCruise2004	boat	08-15-2004	08-19-2004	65	0	65	0	0
HerringAcoustic06	boat	09-19-2006	09-28-2006	278	0	0	278	0
HerringAcoustic07	boat	10-14-2007	10-25-2007	322	0	0	322	0
HerringAcoustic08	boat	09-04-2008	10-09-2008	790	0	0	790	0
HerringAcoustic09Leg1	boat	09-12-2009	09-17-2009	129	0	0	129	0
HerringAcoustic09Leg2	boat	09-21-2009	10-01-2009	279	0	0	279	0
HerringAcoustic09Leg3	boat	10-06-2009	10-15-2009	257	0	0	257	0
HerringAcoustic2010	boat	09-12-2010	10-21-2010	539	0	0	539	0
HerringAcoustic2011	boat	09-08-2011	10-13-2011	773	0	0	773	0
HerringAcoustic2012	boat	09-13-2012	10-18-2012	734	0	0	734	0
MassAudNanAerial	aerial	08-19-2002	03-29-2006	5,022	845	936	1,301	1,940
MassCEC2011-2012	aerial	01-11-2011	11-12-2012	2,535	669	503	869	494
MassCEC2013	aerial	01-21-2013	12-04-2013	2,247	498	499	920	330
MassCEC2014	aerial	01-30-2014	01-14-2015	1,513	167	501	171	674
NJDEP2009	boat	01-13-2008	12-18-2009	4,888	1,211	1,306	1,413	958
NOAA/NMFS.NEFSBoat2004	boat	06-25-2004	08-03-2004	1,101	0	1,101	0	0
NOAA/NMFS.NEFSBoat2007	boat	08-01-2007	08-28-2007	591	0	591	0	0
NOAAMBO7880*	boat	01-02-1978	11-26-1979	5,470	1,337	2,132	1,509	492
PlattsBankAerial	aerial	07-11-2005	07-29-2005	826	0	826	0	0
RISAMPAerial	aerial	12-02-2009	08-31-2010	2,485	992	774	0	719
RISAMPBoat	boat	07-10-2009	08-27-2010	329	56	183	39	51
SEFSC1992	boat	01-05-1992	02-09-1992	754	0	0	0	754
SEFSC1998	boat	07-09-1998	08-20-1998	1,286	0	1,286	0	0
SEFSC1999	boat	08-09-1999	09-25-1999	1,159	0	661	498	0
StatoilMaine	boat	05-02-2012	10-15-2013	400	40	150	140	70
WHOIJuly2010	boat	07-08-2010	07-15-2010	64	0	64	0	0
WHOISept2010	boat	09-22-2010	09-29-2010	80	0	0	80	0
<b>Totals</b>		<b>01-02-1978</b>	<b>06-18-2015</b>	<b>133,040</b>	<b>24,706</b>	<b>37,801</b>	<b>37,082</b>	<b>33,451</b>

\*Discrete-time strip transects

Prior to analysis, continuously recorded transect data were divided into spatially discrete transect segments with a target length of 4 km. Any remaining transect segment (after dividing into 4 km bins) that was greater than or equal to 2 km in length was treated as its own segment. Remaining transect segments that were less than 2 km in length were added to an adjacent segment. The placement of segments that were less than or greater than 4 km was randomized to prevent them from always occurring at the beginning or end of a transect. The distance traveled along each transect was calculated using the boat/plane location data assuming straight-line travel between recorded locations. Species-specific bird counts were summed within each transect segment and geographically referenced using the geographic midpoint of each transect segment. Counts of birds that were not identified to species were excluded from these analyses. Additionally, because our analysis methods do not directly incorporate survey effort (Kinlan et al. 2012, Sections 2.1–2.8), transect segments that were less than 2.5 km or greater than 5.5 km were excluded in order to reduce any potential bias.

Data were divided into four seasons (spring, summer, fall, winter) and analyzed separately by season to reduce the effect of the inherent temporal variability of species distributions. The spring season included transect segments that were surveyed between March 1 and May 31; summer: between June 1 and August 31; fall: between September 1 and November 30; and winter: between December 1 and February 29.

## 2.2 Spatial Grid

A spatial grid with a resolution of  $4.8 \times 4.8$  km was constructed over the study area using an oblique Mercator projected coordinate system (origin:  $35^{\circ}\text{N } 75^{\circ}\text{W}$ ; azimuth:  $40^{\circ}$ ; scale: 0.9996; geodetic datum: NAD83). The spatial resolution was chosen to match the size of a standard BOEM lease block. Segmented transect data were linked to the spatial grid by matching the midpoint of each transect segment with the grid cell in which it was geographically located. Transect segments with midpoint locations that were outside the study area were excluded.

## 2.3 Species Analyzed

Twenty species of marine birds were selected for this study based on data availability and interest to BOEM in the context of potential renewable energy development in the Atlantic OCS region (**Table 2**). Selected species included four species of sea ducks (Common Eider *Somateria mollissima*, Surf Scoter *Melanitta perspicillata*, White-winged Scoter *M. fusca*, Long-tailed Duck *Clangula hyemalis*), two alcids (Razorbill *Alca torda*, Atlantic Puffin *Fratercula arctica*), two gulls (Laughing Gull *Leucophaeus atricilla*, Herring Gull *Larus argentatus*), four terns (Least Tern *Sternula antillarum*, Roseate Tern *Sterna dougallii*, Common Tern *S. hirundo*, Royal Tern *Thalasseus maximus*), two loons (Red-throated Loon *Gavia stellata*, Common Loon *G. immer*), one petrel (Black-capped Petrel *Pterodroma hasitata*), four shearwaters (Cory's Shearwater *Calonectris diomedea*, Sooty Shearwater *Ardenna grisea*, Great Shearwater *A. gravis*, Audubon's Shearwater *Puffinus lherminieri*), and one gannet species (Northern Gannet *Morus bassanus*).

**Table 3** shows the number of transect segments with at least one sighting and the total count for each species by season. Analyses were not conducted if there were less than 50 transect segments with at least one sighting within a single season for a given species. By this criterion, 69 species-season combinations were analyzed.

**Table 2. Species analyzed**

Common name	Scientific name	Family	Order
Common Eider	<i>Somateria mollissima</i>	Anatidae	Anseriformes
Surf Scoter	<i>Melanitta perspicillata</i>	Anatidae	Anseriformes
White-winged Scoter	<i>Melanitta fusca</i>	Anatidae	Anseriformes
Long-tailed Duck	<i>Clangula hyemalis</i>	Anatidae	Anseriformes
Razorbill	<i>Alca torda</i>	Alcidae	Charadriiformes
Atlantic Puffin	<i>Fratercula arctica</i>	Alcidae	Charadriiformes
Laughing Gull	<i>Leucophaeus atricilla</i>	Laridae	Charadriiformes
Herring Gull	<i>Larus argentatus</i>	Laridae	Charadriiformes
Least Tern	<i>Sternula antillarum</i>	Laridae	Charadriiformes
Roseate Tern	<i>Sterna dougallii</i>	Laridae	Charadriiformes
Common Tern	<i>Sterna hirundo</i>	Laridae	Charadriiformes
Royal Tern	<i>Thalasseus maximus</i>	Laridae	Charadriiformes
Red-throated Loon	<i>Gavia stellata</i>	Gaviidae	Gaviiformes
Common Loon	<i>Gavia immer</i>	Gaviidae	Gaviiformes
Black-capped Petrel	<i>Pterodroma hasitata</i>	Procellariidae	Procellariiformes
Cory's Shearwater	<i>Calonectris diomedea</i>	Procellariidae	Procellariiformes
Sooty Shearwater	<i>Ardenna grisea</i>	Procellariidae	Procellariiformes
Great Shearwater	<i>Ardenna gravis</i>	Procellariidae	Procellariiformes
Audubon's Shearwater	<i>Puffinus lherminieri</i>	Procellariidae	Procellariiformes
Northern Gannet	<i>Morus bassanus</i>	Sulidae	Suliformes

**Table 3. Number of transect segments with sightings and total counts for each species-season combination**

Only combinations with  $\geq 50$  transect segments with sightings were analyzed. Seasons were defined as spring (March 1 to May 31), summer (June 1 to August 31), fall (September 1 to November 30), and winter (December 1 to February 29).

Species	Number of transect segments				Total count			
	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter
Common Eider	931	188	666	2,392	218,818	29,570	43,977	604,991
Surf Scoter	832	–	815	2,031	16,572	–	28,344	51,511
White-winged Scoter	553	–	640	1,573	20,886	–	12,011	30,329
Long-tailed Duck	1,343	–	526	3,760	87,516	–	18,438	153,386
Razorbill	998	81	183	1,934	6,719	222	1,234	16,406
Atlantic Puffin	201	262	93	309	476	573	128	541
Laughing Gull	701	1,704	1,820	137	1,600	5,663	10,647	353
Herring Gull	5,449	3,129	7,870	5,252	42,382	11,145	57,715	27,209
Least Tern	–	133	97	–	–	523	1,165	–
Roseate Tern	59	214	81	–	203	843	531	–
Common Tern	594	1,735	743	–	2,655	9,115	7,503	–
Royal Tern	268	271	380	–	732	616	1,000	–
Red-throated Loon	1,752	–	415	2,735	5,275	–	1,907	9,158

continued on next page

Table 3 continued

Species	Number of transect segments				Total count			
	Spring	Summer	Fall	Winter	Spring	Summer	Fall	Winter
Common Loon	2,791	204	1,432	3,970	6,670	272	3,129	10,171
Black-capped Petrel	118	326	77	85	222	930	157	204
Cory's Shearwater	101	2,939	1,683	–	217	13,154	7,915	–
Sooty Shearwater	740	1,566	107	–	5,801	30,481	284	–
Great Shearwater	564	6,123	6,444	122	5,178	124,466	88,740	316
Audubon's Shearwater	115	817	259	158	410	2,497	842	324
Northern Gannet	6,059	1,236	4,968	7,743	37,409	2,814	23,098	62,253
<b>Totals</b>					<b>459,741</b>	<b>232,884</b>	<b>308,765</b>	<b>967,152</b>

## 2.4 Statistical Model

We assume counts of marine birds observed within discrete spatial units (i.e., transect segments) are the outcome of a two-component hurdle model (Mullahy 1986) where the count for species  $i$  in season  $j$  is non-zero according to a Bernoulli( $\theta_{ij}$ ) distribution and the distribution of non-zero counts follows a discrete probability mass function (truncated to include only positive integer values). Kinlan et al. (2012) compared the fit of eight statistical distributions to non-zero counts of birds observed within discrete spatial units. Their findings and conclusions from Zipkin et al. (2015, 2014) suggested the discrete lognormal distribution provided the best consistent fit among the distributions considered<sup>1</sup>. Therefore, all analyses were completed assuming a discrete lognormal distribution for the non-zero counts for each species-season combination.

The probability mass function for the discrete lognormal distribution (truncated such that  $x_{ij} \in \{1, 2, 3, \dots\}$ ) is shown in **Equation 1**. Here,  $x_{ij}$  denotes the realization of random variable  $X_{ij}$  and represents the observed non-zero count for species  $i$  during season  $j$ ,  $\mu_{ij}$  is the mean and  $\sigma_{ij}$  is the standard deviation of the corresponding continuous untruncated distribution of the natural logarithm of random variable  $X_{ij}$ , and  $\Phi()$  represents the cumulative distribution function of the standard normal distribution.

$$f(x_{ij} | \mu_{ij}, \sigma_{ij}) = \frac{\Phi\left(\frac{\ln(x_{ij}+0.5)-\mu_{ij}}{\sigma_{ij}}\right) - \Phi\left(\frac{\ln(x_{ij}-0.5)-\mu_{ij}}{\sigma_{ij}}\right)}{1 - \Phi\left(\frac{\ln(0.5)-\mu_{ij}}{\sigma_{ij}}\right)} \quad (1)$$

For each species-season combination, we estimated the  $\theta_{ij}$  parameter of the Bernoulli distribution by dividing the total number of transect segments with at least one sighting by the total number of transect segments surveyed. This quantity is subsequently referred to as the *reference prevalence* for a given species-season combination. The  $\mu_{ij}$  and  $\sigma_{ij}$  parameters of the discrete lognormal distribution fit to all

<sup>1</sup>To further verify this, we fit each of the eight candidate distributions from Kinlan et al. (2012) to the non-zero counts for each species-season combination using maximum likelihood estimation. The discrete lognormal was chosen as the best fitting distribution according to Akaike's Information Criterion corrected for finite sample sizes (AICc; Burnham, Anderson 2002) in 42 of the 69 species-season combinations analyzed. However, when comparing the AICc value of the discrete lognormal to that of the best fitting distribution, all had  $\Delta\text{AICc} < 4$  (where  $\Delta\text{AICc}$  is equal to the AICc value of the discrete lognormal distribution minus the AICc value of the best fitting distribution) except five species-season combinations: Razorbill spring ( $\Delta\text{AICc} = 5.7$ ), Herring Gull summer ( $\Delta\text{AICc} = 8.4$ ), Sooty Shearwater summer ( $\Delta\text{AICc} = 4.4$ ), Black-capped Petrel summer ( $\Delta\text{AICc} = 4.4$ ), and Audubon's Shearwater summer ( $\Delta\text{AICc} = 5.1$ ). Vuong closeness tests gave little to no evidence that the best fitting distribution provided a better fit than the discrete lognormal distribution in each of these five cases ( $0.097 \leq p\text{-value} \leq 0.345$ ; Vuong 1989).

non-zero counts for each species-season combination were estimated using standard maximum likelihood estimation procedures. The resulting discrete lognormal distribution with estimated parameters  $\hat{\mu}_{ij}$  and  $\hat{\sigma}_{ij}$  is subsequently referred to as the *reference distribution* for a given species-season combination. The expected value of random variable  $X_{ij}$ , subsequently referred to as the *reference mean*, was estimated numerically by summing the products of  $x_{ij}$  and  $f(x_{ij} | \hat{\mu}_{ij}, \hat{\sigma}_{ij})$  for all integer values of  $x_{ij}$  from 1 to 10,000,000.

Three unique model types naturally arise from the two-component hurdle model, each corresponding to a distinct form of data commonly recorded during visual wildlife surveys:

1. the *occurrence probability model*, which relies only on the Bernoulli component of the hurdle model. This model is best suited for binary data recorded simply as presence or absence of a given species.
2. the *non-zero count model*, which relies only on the non-zero (i.e., the discrete lognormal) component of the hurdle model. This model is best suited for count data recorded only when a species is observed.
3. the *combined model*, which incorporates both the Bernoulli and non-zero components of the hurdle model. This model is best suited for survey data where both zero and non-zero counts are recorded during data collection.

We performed separate analyses for each model type in order to interpret results in the context of differing survey designs and data types.

## 2.5 Hotspot/Coldspot Identification

To identify species-specific hotspots and coldspots, we performed independent significance tests within each grid cell for each of the three model types. In all cases, a species-specific hotspot was defined as a location (i.e., grid cell) where the observed data value (i.e., number of transect segments with sightings for the occurrence probability model, mean non-zero count for the non-zero count model, or mean count for the combined model) was *greater* than the expected data value (i.e., reference prevalence for the occurrence probability model, mean of random draws from the reference distribution for the non-zero count model, or mean of the product of random draws from a Bernoulli( $\theta_{ij}$ ) distribution and the reference distribution for the combined model). A species-specific coldspot was defined as a location where the observed data value was *less* than the expected data value.

For the occurrence probability model, we assumed that, in each grid cell, the number of transect segments with sightings independently follow a binomial distribution with size (i.e., number of trials) equal to the number of transect segments surveyed and probability of success (i.e., probability of observing a bird of a specific species during a specific season) equal to the reference prevalence for a given species-season combination. Following this assumption, hotspot and coldspot  $p$ -values were calculated by performing one-sided exact binomial tests for each grid cell. Smaller  $p$ -values indicated greater evidence that a grid cell was a species-specific hotspot or coldspot of occurrence.

For the non-zero count model, we performed simulation-based significance tests within each grid cell. Independently for each species-season combination and for each grid cell  $k$  within the study area, we drew  $M_{ijk}$  random samples from the reference distribution, where  $M_{ijk}$  corresponds to the number of transect segments within grid cell  $k$  on which species  $i$  was observed during season  $j$ . We then calculated the arithmetic mean of the  $M_{ijk}$  random samples. This process was repeated 100,000 times in order to obtain a simulated distribution of sample means within each grid cell for each species-season combination. Hotspot  $p$ -values were calculated as the proportion of the 100,000 simulations where the simulated mean non-zero count was greater than or equal to the observed mean non-zero count for a given grid cell. Coldspot

$p$ -values were calculated as the proportion of simulations where the simulated mean non-zero count was less than or equal to the observed mean non-zero count. Smaller  $p$ -values based on the non-zero count model indicated greater evidence that a grid cell was a species-specific hotspot or coldspot of non-zero abundance.

For the combined model, we again performed simulation-based significance tests. For each grid cell,  $N_{jk}$  random samples were drawn from a Bernoulli distribution with probability equal to the reference prevalence for the given species during season  $j$ , where  $N_{jk}$  corresponds to the number of transect segments surveyed within grid cell  $k$  during season  $j$ . We also simulated  $N_{jk}$  random draws from the reference distribution for the given species-season combination. Each Bernoulli random sample value was multiplied by the corresponding random sample value from the reference distribution, yielding  $N_{jk}$  simulated counts, including both zero and non-zero values. The arithmetic mean of the  $N_{jk}$  simulated counts was calculated and the process repeated 100,000 times for each grid cell and species-season combination. Hotspot  $p$ -values were estimated by calculating the proportion of the 100,000 simulations where the simulated mean count was greater than or equal to the observed mean count for a given grid cell. Coldspot  $p$ -values were calculated as the proportion of simulations where the simulated mean count was less than or equal to the observed mean count. Smaller  $p$ -values based on the combined model indicated greater evidence that a grid cell was a species-specific hotspot or coldspot of unconditional abundance.

Because many simultaneous hotspot and coldspot  $p$ -values were calculated for each combination of species, season, and model type—one for each grid cell surveyed in both the occurrence probability and combined models and one for each grid cell where a species was observed during a given season in the non-zero count model—all  $p$ -values were adjusted for multiple testing in order to limit the number of potential false positive grid cells (i.e., grid cells with low  $p$ -values suggesting evidence of a hotspot or coldspot when in truth they are not). Many methods exist for adjusting  $p$ -values for multiple testing that control for different rates of false positives (Shaffer 1995; Wright 1992). We used the sequential method of Holm (1979) to control the family-wise error rate, which ensures that the probability of at least one false positive result is less than some user-defined threshold value (e.g., 0.05). For each species and season combination, adjusted hotspot and coldspot  $p$ -values were calculated independently, assuming the number of simultaneous tests performed was equal to either a) the number of grid cells surveyed during the specified season (occurrence probability and combined models) or b) the number of grid cells in which the specified species was observed during the specified season (non-zero count model).

## 2.6 Hotspot/Coldspot Persistence

In addition to identifying hotspots and coldspots for each species, season and model type, we also investigated the consistency of hotspot/coldspot locations through time for each species-season combination and model type. We calculated hotspot and coldspot *interannual persistence* following the same methodology described in **Section 2.5**, except that calculations/simulations were conducted separately for either a) each year in which there was survey effort during the specified season (for the occurrence probability and combined models) or b) each year in which the specified species was observed during the specified season (for the non-zero count model). For each grid cell, hotspot interannual persistence was calculated as the proportion of years with survey effort/sightings in which the single-year hotspot  $p$ -value was less than or equal to 0.05. Similarly, coldspot interannual persistence was calculated as the proportion of years in which the single-year coldspot  $p$ -value was less than or equal to 0.05. Calculations were completed using both the raw (i.e., unadjusted) and adjusted  $p$ -values separately.

## 2.7 Power Estimation

For each combination of species, season, and model type, we estimated the statistical power to detect hotspots and coldspots of effect size  $\delta$  for a range of sample sizes. For all three model types, we estimated power for hotspot effect sizes of  $\delta = 3, 10, \text{ and } 20$ , coldspot effect sizes of  $\delta = \frac{1}{3}, \frac{1}{10}, \text{ and } \frac{1}{20}$ , and for all sample sizes  $n$  from 1 to the larger of either a) 200 or b) the maximum number of transect segments surveyed within a single grid cell during each season. All power calculations assumed a type I error rate of 0.05.

For the occurrence probability model, we calculated power assuming the number of transect segments with sightings independently follow a binomial distribution with size equal to  $n$  and probability of success equal to  $\delta$  times the reference prevalence for a given species-season combination. In some instances,  $\delta$  times the reference prevalence was greater than one, which is outside the bounds of the probability of success parameter in the binomial distribution. Power to detect hotspots of occurrence was undefined at the given effect size for these species-season combinations.

For the non-zero count model, power was estimated by simulating random draws of size  $n$  from a discrete lognormal distribution with expected value equal to  $\delta$  times the reference mean and  $\sigma_{ij}$  parameter as previously estimated (see **Section 2.4**) for species  $i$  during season  $j$ . The arithmetic mean of each random sample was calculated, and the process repeated 100,000 times. The proportion of the 100,000 means that exceeded a *hotspot critical value* was used as the estimate for power to detect a hotspot of effect size  $\delta$  for a given sample size. Power to detect a coldspot was estimated by the proportion of means that were less than a *coldspot critical value* for a given sample size. Critical values were estimated by simulating random draws from the reference distribution for each sample size, calculating the arithmetic mean of each, and repeating the process 100,000 times. Hotspot and coldspot critical values were defined by the 95th and 5th percentiles, respectively, of the distribution of simulated means from the reference distribution for each sample size. For some species-season combinations, power to detect coldspots of non-zero abundance was undefined for certain effect sizes because  $\delta$  times the reference mean was less than or equal to one.

Power for the combined model was estimated by simulating random draws of size  $n$  from a Bernoulli distribution with probability equal to the reference prevalence for each species and season combination. Each Bernoulli random sample value was then multiplied by the corresponding random sample value drawn during power estimation for the non-zero count model as described above (i.e., from a discrete lognormal distribution with expected value equal to  $\delta$  times the reference mean). The arithmetic mean of each product of random samples was calculated, and the process repeated 100,000 times. Hotspot and coldspot power estimates for a given sample size were calculated as the proportion of means that were either greater than (hotspot) or less than (coldspot) the respective critical value for the combined model. Critical values were estimated by simulating random draws of size  $n$  from each component of the hurdle model (the Bernoulli component with probability equal to the reference prevalence and the reference distribution for non-zero counts), calculating the arithmetic mean of the product of both components for each sample size, repeating the process 100,000 times, and calculating the 95th (hotspot) and 5th (coldspot) percentiles of the distribution of means. For some species-season combinations power to detect coldspots of abundance was undefined for certain effect sizes because  $\delta$  times the reference mean was less than or equal to one. Note that for the combined model, the effect size  $\delta$  was only introduced via the non-zero count component (i.e., via the reference mean) of the hurdle model. We did not investigate the relationship between sample size and power when the effect size was introduced through changes in the occurrence probability (i.e., via the reference prevalence) of the combined hurdle model.

For each model type, we also calculated average power to detect hotspots and coldspots across all seasons analyzed for each species, across all species analyzed within each season, and across all species and seasons. Average power across seasons was calculated by taking the arithmetic mean of all seasonal power

estimates within each grid cell that was sampled, separately for each species. Average power across all species within each season was calculated by taking the arithmetic mean of all power estimates within each grid cell that was sampled, separately for each season. Average power across all species and all seasons was calculated by taking the arithmetic mean of the four seasonal average power values within each grid cell that was sampled. For the non-zero count model, grid cells that were sampled but had no recorded sightings of a particular species during a given season contributed a value of zero to the calculations, assuming the species-season combination was analyzed. Seasons that were either not analyzed for a given species or for which power to detect a hotspot or coldspot was undefined did not contribute to the average power calculations.

## 2.8 Spatial Resolution

In order to explore the effect of spatial resolution (i.e., grid cell size) on hotspots/coldspots and the power to detect them, we repeated all spatial analyses (**Sections 2.5–2.7**) using three additional spatial grids. Additional grids were constructed by combining the original spatial grid cells (see **Section 2.2**) into blocks of  $2 \times 2$ ,  $3 \times 3$ , and  $4 \times 4$  grid cells, corresponding to resolutions of  $9.6 \times 9.6$  km,  $14.4 \times 14.4$  km, and  $19.2 \times 19.2$  km, respectively.

## 2.9 Stratification

The methodologies described for identifying species-specific hotspots/coldspots and estimates of power assume the same reference prevalence and reference distribution across all grid cells for a given species-season combination (i.e., the estimated parameters of the Bernoulli distribution and the reference distribution remain constant throughout the study area). This assumption is important because hotspots and coldspots are necessarily defined in relative terms. In this case a hotspot of abundance, for example, defines an area of high abundance *relative* to the larger reference region (i.e., the study area). In certain instances, however, it may be of interest to understand potential hotspots/coldspots and power to detect them within unique strata of the reference region. Strata may be defined according to the values of some additional covariate or covariates (e.g., distance from shore, where  $<5$  km from shore could define stratum A,  $5\text{--}10$  km stratum B, and  $>10$  km stratum C). It is important to recognize that hotspots/coldspots would then be defined uniquely within each strata such that a hotspot of abundance within stratum A, for example, would define an area of high abundance relative to stratum A only.

To provide examples of how stratification can affect estimates of power and significance test  $p$ -values, we implemented a method to identify the core and non-core areas of abundance for four species-season combinations: Common Eider in winter, Herring Gull in fall, Roseate Tern in summer, and Red-throated Loon in winter. For each species-season combination, we identified the core area of abundance based on the predicted relative density values from Winship et al. (2018) following the core area of abundance calculation in Curtice et al. (2016, Section 3.6). This calculation resulted in each grid cell of the input data being designated as either inside or outside the core area of abundance for each species-season combination. We then spatially matched each transect segment to the resulting grid and coded each as belonging to one of the resulting two strata. Because the spatial grid used in Winship et al. (2018) did not completely cover our study area, some transect segments were excluded from this part of the analysis.

We then reran all analyses separately for each stratum. We re-estimated the  $\theta_{ij}$  parameter of the Bernoulli distribution and refit the discrete lognormal distribution to all non-zero counts within each stratum separately. Finally, we re-estimated significance test  $p$ -values and power to detect hotspots and coldspots of each effect size and model type within each stratum.

## 2.10 Implementation

Model fitting and all analyses were completed using R version 3.4.4 64-bit (R Core Team 2018). Candidate distribution model fitting relied on the following R packages: **gsl** version 1.9.10.3 (Hankin 2006), **powerLaw** version 0.70.1 (Gillespie 2015), **VGAM** version 1.0.5 (Yee 2010), and code provided in Clauset et al. (2009). Additional R packages used throughout the analyses included **raster** version 2.5.8 (Hijmans 2016), **rgdal** version 1.2.7 (Bivand et al. 2017), and **rgeos** version 0.3.23 (Bivand, Rundel 2017).

## 3 Results

The reference prevalence (i.e., the estimated  $\theta_{ij}$  parameter of the Bernoulli distribution), reference distribution parameters (i.e., the estimated  $\mu_{ij}$  and  $\sigma_{ij}$  parameters of the discrete lognormal distribution), and reference mean (i.e., the expected value of the reference distribution) are shown in **Table 4** for each species and season combination. Northern Gannet during spring had the greatest proportion of transect segments with sightings (reference prevalence = 0.245) for any analyzed species-season combination, while Black-capped Petrel during fall had the least (reference prevalence = 0.002). The expected count for a transect segment with at least one sighting was greatest for Common Eider during summer (reference mean = 374.28) and least for Common Loon during summer (reference mean = 1.33).

**Table 4. Parameter estimates**

Reference prevalence, reference mean, and maximum likelihood parameter estimates from the discrete lognormal distribution fit to non-zero counts for each species-season combination. See **Section 2.4** for more details.

Species	Season	Reference prevalence	Reference mean	$\hat{\mu}_{ij}$	$\hat{\sigma}_{ij}$
Common Eider	spring	0.038	212.45	1.626	2.655
Common Eider	summer	0.005	374.28	1.613	2.860
Common Eider	fall	0.018	67.87	1.874	2.110
Common Eider	winter	0.072	188.52	1.408	2.680
Surf Scoter	spring	0.034	17.87	1.745	1.475
Surf Scoter	fall	0.022	38.63	2.219	1.670
Surf Scoter	winter	0.061	25.44	2.029	1.529
White-winged Scoter	spring	0.022	45.04	-0.693	2.760
White-winged Scoter	fall	0.017	19.89	1.139	1.832
White-winged Scoter	winter	0.047	16.28	0.900	1.832
Long-tailed Duck	spring	0.054	38.61	1.830	1.861
Long-tailed Duck	fall	0.014	31.47	1.482	1.914
Long-tailed Duck	winter	0.112	28.96	1.875	1.688
Razorbill	spring	0.040	7.08	1.188	1.192
Razorbill	summer	0.002	2.70	0.528	0.872
Razorbill	fall	0.005	6.58	0.244	1.623
Razorbill	winter	0.058	8.30	0.794	1.514
Atlantic Puffin	spring	0.008	2.41	0.244	0.961
Atlantic Puffin	summer	0.007	2.14	-0.317	1.140
Atlantic Puffin	fall	0.003	1.37	-0.512	0.841
Atlantic Puffin	winter	0.009	1.74	-0.282	0.960
Laughing Gull	spring	0.028	2.21	-0.272	1.144
Laughing Gull	summer	0.045	3.23	-0.534	1.502

continued on next page

Table 4 continued

Species	Season	Reference prevalence	Reference mean	$\hat{\mu}_{ij}$	$\hat{\sigma}_{ij}$
Laughing Gull	fall	0.049	5.43	-0.256	1.705
Laughing Gull	winter	0.004	2.64	-0.341	1.297
Herring Gull	spring	0.221	6.77	-0.570	1.925
Herring Gull	summer	0.083	3.19	-1.260	1.724
Herring Gull	fall	0.212	6.29	-0.287	1.794
Herring Gull	winter	0.157	4.88	-1.528	2.037
Least Tern	summer	0.004	3.69	-1.348	1.834
Least Tern	fall	0.003	13.83	0.849	1.770
Roseate Tern	spring	0.002	3.36	0.244	1.207
Roseate Tern	summer	0.006	3.64	-0.073	1.401
Roseate Tern	fall	0.002	8.21	-1.408	2.259
Common Tern	spring	0.024	4.34	0.662	1.165
Common Tern	summer	0.046	4.68	-0.007	1.528
Common Tern	fall	0.020	9.29	0.179	1.830
Royal Tern	spring	0.011	2.50	-0.634	1.370
Royal Tern	summer	0.007	2.31	-0.307	1.191
Royal Tern	fall	0.010	2.56	0.368	0.934
Red-throated Loon	spring	0.071	2.83	-0.602	1.440
Red-throated Loon	fall	0.011	5.20	-1.131	1.960
Red-throated Loon	winter	0.082	3.33	-0.776	1.601
Common Loon	spring	0.113	2.36	-0.144	1.138
Common Loon	summer	0.005	1.33	-0.423	0.773
Common Loon	fall	0.039	2.19	-0.130	1.074
Common Loon	winter	0.119	2.31	-0.926	1.414
Black-capped Petrel	spring	0.005	1.91	-1.631	1.484
Black-capped Petrel	summer	0.009	2.75	-2.696	1.998
Black-capped Petrel	fall	0.002	2.08	-0.286	1.105
Black-capped Petrel	winter	0.003	2.41	0.193	0.987
Cory's Shearwater	spring	0.004	2.18	-0.361	1.171
Cory's Shearwater	summer	0.078	4.25	-1.724	2.015
Cory's Shearwater	fall	0.045	4.59	-0.855	1.810
Sooty Shearwater	spring	0.030	8.74	-3.573	2.785
Sooty Shearwater	summer	0.041	12.30	-5.917	3.359
Sooty Shearwater	fall	0.003	2.48	-15.628	3.792
Great Shearwater	spring	0.023	8.56	0.495	1.665
Great Shearwater	summer	0.162	11.42	0.027	1.985
Great Shearwater	fall	0.174	12.78	1.139	1.597
Great Shearwater	winter	0.004	2.61	-0.624	1.394
Audubon's Shearwater	spring	0.005	4.44	-5.537	2.832
Audubon's Shearwater	summer	0.022	2.80	-1.510	1.714
Audubon's Shearwater	fall	0.007	2.98	-0.337	1.378
Audubon's Shearwater	winter	0.005	2.08	-0.422	1.158
Northern Gannet	spring	0.245	5.34	-0.622	1.819
Northern Gannet	summer	0.033	2.27	-1.223	1.492

continued on next page

Table 4 continued

Species	Season	Reference prevalence	Reference mean	$\hat{\mu}_{ij}$	$\hat{\sigma}_{ij}$
Northern Gannet	fall	0.134	4.51	-0.332	1.628
Northern Gannet	winter	0.231	6.80	-0.903	2.029

### 3.1 Hotspot/Coldspot Identification

Nearly all adjusted  $p$ -values were greater than 0.2 in both hotspot and coldspot tests of significance for all model types and grid cell sizes (**Tables 5–6**): >98.14% (hotspot) and >99.38% (coldspot) for the occurrence probability model, >99.80% (hotspot) and >99.67% (coldspot) for the non-zero count model, and >99.93% (hotspot) and >99.37% (coldspot) for the combined model. By definition, unadjusted  $p$ -values were always less than or equal to  $p$ -values adjusted for multiple testing. However, unadjusted  $p$ -values showed similar patterns in that a very high percentage were greater than 0.2: >91.77% (hotspot) and >93.11% (coldspot) for the occurrence probability model, >83.89% (hotspot) and >83.84% (coldspot) for the non-zero count model, and >92.58% (hotspot) and >92.56% (coldspot) for the combined model, with percentages varying depending on grid cell size.

A large percentage of grid cells showed little evidence of being either a hotspot or coldspot (i.e., both hotspot and coldspot adjusted  $p$ -values > 0.2) for any species: 64.29–98.09%, depending on grid cell size, for the occurrence probability model, 97.46–99.81% for the non-zero count model, and 89.86–99.60% for the combined model (**Table 7**).

Geographic areas identified as potential hotspots or coldspots shifted as the spatial resolution changed. This was due to the spatial location of transect segments relative to the grid cell size and the often extreme variability in observations per transect segment, which led to large fluctuations in mean count for a given grid cell as the grid cell size changed. As a result, it was possible for a grid cell to change from being a potential hotspot/coldspot (low estimated  $p$ -value) to not being a potential hotspot/coldspot (and vice versa) as the grid cell size changed. This underscores the importance of having an *a priori* method to determine the proper spatial resolution for hotspot/coldspot identification.

### 3.2 Hotspot/Coldspot Persistence

Among grid cells that were surveyed in multiple years within the same season, the mean interannual persistence, pooled across all species, seasons, and spatial resolutions, for a hotspot with adjusted  $p$ -value  $\leq 0.05$  varied slightly depending on model type: 0.14 for the occurrence probability model, 0.20 for the non-zero count model, and 0.18 for the combined model. The mean interannual persistence of coldspots surveyed in multiple years within the same season and with adjusted  $p$ -value  $\leq 0.05$  was much lower and showed less variability: 0.06 for the occurrence probability model, 0.04 for the non-zero count model, and 0.06 for the combined model.

Across all species, seasons, and spatial resolutions, 83% of all grid cells that were surveyed in multiple years within the same season and were hotspots of occurrence in at least one year (i.e., adjusted  $p$ -value  $\leq 0.05$  in at least one single-year significance test) were hotspots of occurrence when all years were combined (adjusted  $p$ -value  $\leq 0.05$  in overall significance test). Of grid cells that were hotspots of non-zero abundance in at least one year, 80% were hotspots of non-zero abundance when all years were combined, while only 52% of grid cells that were hotspots of unconditional abundance in at least one year were overall hotspots of unconditional abundance. Similarly, 92% of all grid cells that were surveyed in multiple years within the same season and were coldspots of occurrence in at least one year were coldspots

**Table 5. Summary of hotspot  $p$ -values**

Hotspot  $p$ -values are pooled across all species and seasons and are summarized for each combination of model type (Model), adjusted/unadjusted  $p$ -value (Type), and grid cell size (Grid size). Columns 4–9 show the percentage of  $p$ -values that fall within the specified range. Percentage values sum to 100 for each row.

Model	Type	Grid size	[0, 0.001]	(0.001, 0.01]	(0.01, 0.05]	(0.05, 0.1]	(0.1, 0.2]	(0.2, 1]
occurrence probability	unadjusted	1×1	0.35	0.54	1.18	0.85	1.44	95.62
		2×2	0.96	0.85	1.55	0.93	1.50	94.20
		3×3	1.72	1.11	1.71	1.10	1.49	92.87
		4×4	2.54	1.20	1.82	1.10	1.57	91.77
	adjusted	1×1	0.08	0.02	0.02	0.01	0.02	99.85
		2×2	0.24	0.07	0.08	0.05	0.06	99.49
		3×3	0.53	0.19	0.18	0.10	0.11	98.90
		4×4	0.94	0.27	0.26	0.19	0.20	98.14
non-zero count	unadjusted	1×1	0.24	1.02	3.53	3.86	7.46	83.89
		2×2	0.25	1.14	3.42	3.64	6.71	84.86
		3×3	0.27	1.23	3.30	3.46	6.26	85.48
		4×4	0.28	1.38	3.33	3.56	5.75	85.70
	adjusted	1×1	0.01	0.00	0.01	0.01	0.02	99.95
		2×2	0.01	0.00	0.02	0.03	0.03	99.91
		3×3	0.01	0.00	0.03	0.02	0.05	99.88
		4×4	0.01	0.02	0.05	0.03	0.09	99.80
combined	unadjusted	1×1	0.13	0.67	1.59	1.25	1.49	94.87
		2×2	0.19	0.81	1.84	1.36	1.83	93.98
		3×3	0.27	0.99	2.04	1.50	1.97	93.24
		4×4	0.35	1.20	2.23	1.54	2.11	92.58
	adjusted	1×1	0.00	0.00	0.00	0.00	0.00	99.99
		2×2	0.01	0.00	0.00	0.00	0.00	99.98
		3×3	0.01	0.00	0.01	0.01	0.02	99.96
		4×4	0.01	0.00	0.02	0.02	0.02	99.93

of occurrence when all years were combined. Of grid cells that were coldspots of non-zero abundance in at least one year, only 27% were coldspots of non-zero abundance when all years were combined, while 85% of grid cells that were coldspots of unconditional abundance in at least one year were overall coldspots of unconditional abundance.

Boxplots of hotspot/coldspot  $p$ -values vs. interannual persistence for each model type are shown in **Appendix A**. Regardless of spatial resolution, all grid cells with a hotspot of occurrence interannual persistence  $\geq 0.6$  (calculated using adjusted  $p$ -values) had overall hotspot of occurrence adjusted  $p$ -values  $\leq 0.05$ , meaning there was sufficient evidence to suggest they are likely true hotspots of occurrence for at least one species-season combination (**Figure A1**). Regardless of model type or spatial resolution, all grid cells with a coldspot interannual persistence  $\geq 0.3$  showed evidence that they are likely true coldspots for at least one species-season combination (adjusted  $p$ -values  $\leq 0.05$  in overall significance tests; **Figures A2, A4, and A6**).

No formal analysis to compare hotspot/coldspot persistence to overall hotspot/coldspot identification was completed; therefore, conclusions based on the comparisons above should not be generalized. However, the results do yield some interesting hypotheses for future research. The results suggest that coldspots of occurrence, non-zero abundance, and unconditional abundance may be less persistent through time than

**Table 6. Summary of coldspot  $p$ -values**

Coldspot  $p$ -values are pooled across all species and seasons and are summarized for each combination of model type (Model), adjusted/unadjusted  $p$ -value (Type), and grid cell size (Grid size). Columns 4–9 show the percentage of  $p$ -values that fall within the specified range. Percentage values sum to 100 for each row.

Model	Type	Grid size	[0, 0.001]	(0.001, 0.01]	(0.01, 0.05]	(0.05, 0.1]	(0.1, 0.2]	(0.2, 1]
occurrence probability	unadjusted	1×1	0.02	0.03	0.09	0.11	0.30	99.45
		2×2	0.14	0.18	0.39	0.41	0.93	97.94
		3×3	0.46	0.47	0.95	0.79	1.67	95.66
		4×4	0.95	0.91	1.40	1.23	2.40	93.11
	adjusted	1×1	0.00	0.00	0.00	0.00	0.00	99.99
		2×2	0.03	0.01	0.01	0.00	0.01	99.93
		3×3	0.11	0.04	0.05	0.03	0.04	99.74
		4×4	0.26	0.11	0.11	0.05	0.08	99.38
non-zero count	unadjusted	1×1	0.04	0.12	0.58	1.03	3.63	94.60
		2×2	0.13	0.37	1.34	1.99	5.84	90.33
		3×3	0.28	0.63	2.25	2.88	7.43	86.53
		4×4	0.43	1.01	3.06	3.61	8.04	83.84
	adjusted	1×1	0.00	0.00	0.01	0.01	0.01	99.98
		2×2	0.04	0.00	0.01	0.02	0.01	99.93
		3×3	0.06	0.01	0.05	0.04	0.05	99.79
		4×4	0.09	0.02	0.10	0.06	0.05	99.67
combined	unadjusted	1×1	0.02	0.04	0.10	0.12	0.33	99.39
		2×2	0.15	0.19	0.44	0.44	1.01	97.77
		3×3	0.48	0.49	1.01	0.86	1.84	95.33
		4×4	1.00	0.93	1.50	1.37	2.64	92.56
	adjusted	1×1	0.01	0.00	0.00	0.00	0.00	99.99
		2×2	0.05	0.00	0.01	0.00	0.01	99.93
		3×3	0.17	0.00	0.04	0.03	0.03	99.73
		4×4	0.37	0.00	0.12	0.05	0.09	99.37

hotspots. The observation of a single-year hotspot of occurrence or non-zero abundance may be indicative of an overall hotspot of occurrence or non-zero abundance. Likewise, the observation of a single-year coldspot of occurrence or unconditional abundance may indicate an overall coldspot of occurrence or unconditional abundance. However, the observation of a single-year non-hotspot or non-coldspot likely gives no indication regarding the existence of a hotspot or coldspot. Yet, it may be more likely for a single-year non-coldspot to actually be an overall coldspot than for a single-year non-hotspot to actually be an overall hotspot.

### 3.3 Power Estimation

Power to detect hotspots of occurrence (i.e., based on the occurrence probability model) for effect sizes of  $\delta \geq 10$  was undefined for 11 species-season combinations: Long-tailed Duck (winter), Herring Gull (spring, fall, winter), Common Loon (spring, winter), Great Shearwater (summer, fall), and Northern Gannet (spring, fall, winter) because  $\delta$  times the reference prevalence was greater than one. Eight additional species-season combinations had undefined power to detect hotspots of occurrence for  $\delta = 20$ : Common Eider (winter), Surf Scoter (winter), Long-tailed Duck (spring), Razorbill (winter), Herring Gull

**Table 7. Percentage of grid cells classified as hotspot, coldspot, or neither for each model type at a range of type I error rates**

'Hotspot' and 'Coldspot' columns show the percentage of grid cells classified as a hotspot or coldspot (i.e., adjusted  $p$ -value  $\leq \alpha$ ) for at least one species, based on the given type I error rate ( $\alpha$ ). 'Neither' columns show the percentage of grid cells not classified as a hotspot or coldspot (i.e., adjusted hotspot  $p$ -value  $> \alpha$  and adjusted coldspot  $p$ -value  $> \alpha$ ) for any species, based on the given type I error rate.

Grid size	$\alpha$	Occurrence probability model			Non-zero count model			Combined model		
		Hotspot	Coldspot	Neither	Hotspot	Coldspot	Neither	Hotspot	Coldspot	Neither
1×1	0.01	0.98	0.08	99.02	0.03	0.02	99.96	0.11	0.17	99.72
	0.05	1.36	0.14	98.64	0.06	0.03	99.91	0.11	0.17	99.72
	0.10	1.62	0.17	98.37	0.08	0.05	99.87	0.13	0.17	99.70
	0.20	1.90	0.20	98.09	0.14	0.06	99.81	0.20	0.20	99.60
2×2	0.01	5.86	0.59	93.87	0.06	0.14	99.80	0.24	0.64	99.12
	0.05	7.56	0.71	92.12	0.14	0.15	99.71	0.35	0.73	98.94
	0.10	8.87	0.81	90.78	0.25	0.20	99.55	0.48	0.77	98.77
	0.20	10.36	0.90	89.27	0.38	0.21	99.41	0.64	0.90	98.49
3×3	0.01	14.27	2.03	85.09	0.11	0.27	99.63	0.24	2.32	97.49
	0.05	17.98	2.62	81.16	0.29	0.48	99.23	0.53	2.91	96.66
	0.10	19.98	2.97	79.13	0.43	0.61	98.96	0.88	3.29	96.10
	0.20	22.31	3.55	76.60	0.69	0.69	98.61	1.47	3.55	95.40
4×4	0.01	23.03	4.94	75.50	0.21	0.52	99.27	0.39	5.37	94.37
	0.05	27.07	6.36	71.25	0.56	0.95	98.54	1.12	6.36	92.91
	0.10	30.08	7.05	68.07	0.69	1.20	98.15	1.80	7.00	91.83
	0.20	33.30	8.25	64.29	1.25	1.33	97.46	2.62	8.34	89.86

(summer), Red-throated Loon (spring, winter), and Cory's Shearwater (summer).

Power to detect coldspots of both non-zero and unconditional abundance (i.e., based on the non-zero count and combined models, respectively) for effect size  $\delta = 1/20$  was undefined for all species except sea ducks during certain seasons: Common Eider (all seasons), Surf Scoter (fall, winter), White-winged Scoter (spring), and Long-tailed Duck (all modeled seasons). For effect size  $\delta = 1/10$ , power to detect coldspots of non-zero and unconditional abundance was defined for all sea ducks during all modeled seasons, Least Tern (fall), and Great Shearwater (summer, fall), but undefined for all other species-season combinations<sup>2</sup>. Regardless of effect size, power to detect coldspots of non-zero and unconditional abundance was undefined for 26 species-season combinations: Razorbill (summer), Atlantic Puffin (all seasons), Laughing Gull (spring, winter), Royal Tern (all modeled seasons), Red-throated Loon (spring), Common Loon (all seasons), Black-capped Petrel (all seasons), Cory's Shearwater (spring), Sooty Shearwater (fall), Great Shearwater (winter), Audubon's Shearwater (summer, fall, winter), and Northern Gannet (summer).

Power curves for each species-season combination based on the occurrence probability, non-zero count, and combined models are presented in **Appendices C–E**, respectively. Increasing effect sizes<sup>3</sup> always led to greater power to detect a hotspot or coldspot, assuming  $\delta$  times the reference prevalence was less than or equal to one and  $\delta$  times the reference mean was greater than one. Similarly, increasing the grid cell size always led to either no change or an increase in sample size within a given grid cell, which in turn

<sup>2</sup>In all cases except Sooty Shearwater (summer), power to detect coldspots of non-zero and unconditional abundance was undefined because  $\delta$  times the reference mean was less than or equal to one. Power to detect coldspots for Sooty Shearwater during summer based on the non-zero count and combined models with effect size  $\delta = 1/10$  was undefined because  $1/10$  times the reference mean combined with the relatively large  $\sigma_{ij}$  estimate of 3.359 (**Table 4**) made it infeasible to simulate an adequate number of random draws from the discrete lognormal distribution, even though  $\delta$  times the reference mean was greater than one.

<sup>3</sup>A coldspot effect size of  $1/20$  is considered a larger effect size than  $1/10$  which is considered larger than  $1/5$ , even though  $1/20$  is numerically less than  $1/10$  which is less than  $1/5$ .

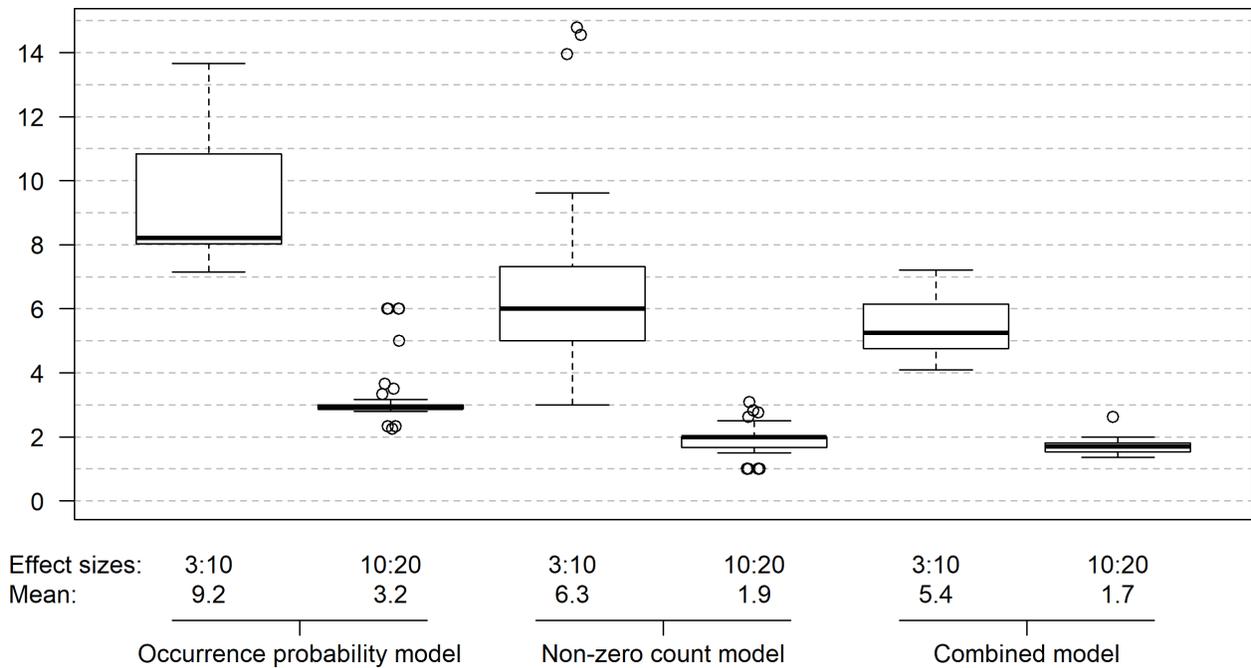
should lead to either no change or an increase in power. In general this was the case. However, power curves for the occurrence probability and combined models, which are at least partly based on the binomial distribution, exhibited a nonmonotonic “sawtooth” pattern characteristic of some discrete statistical distributions (Chernick, Liu 2002). The overall trend of power curves for these model types matched the expectation that an increase in sample size should lead to an increase in power to detect hotspots or coldspots, although small increases in sample size often led to either a) small increases in power followed by intermittent “drops” or large decreases in power to detect hotspots or b) small decreases in power followed by intermittent “jumps” or large increases in power to detect coldspots. These patterns were most evident in the power curves shown in **Appendix C**.

When estimating the sample size required to obtain a certain level of power to detect hotspots or coldspots of occurrence or unconditional abundance (i.e., based on the occurrence probability or combined models), we suggest users take a conservative approach. Using Common Eider during spring as an example, 80% power to detect a hotspot of occurrence defined as three times the reference prevalence was first reached with a sample size of 69 transect segments. Power then dropped to about 71% and did not remain above 80% until a sample size of 79 transect segments was used (**Figure C1 in Appendix C**). In this example, the suggested interpretation is that at least 79 transect segments are needed to confidently obtain 80% power to detect a hotspot of occurrence defined as three times the reference prevalence for Common Eider during spring.

The estimated sample size required to achieve 80% power to detect a hotspot or coldspot is shown in **Appendix B** for each species-season combination. Values varied drastically according to species, season, effect size, and model type. A single transect segment was the only requirement in order to achieve 80% power to detect a hotspot of occurrence defined as twenty times the reference prevalence for five species-season combinations (**Table B1**). In contrast, a sample size of 1,063 transect segments (the largest sample size considered) failed to reach 80% power to detect a hotspot of abundance of three times the reference mean for seven species-season combinations (**Table B3**). As stated above, power to detect coldspots of non-zero and unconditional abundance was undefined for many species-season combinations, but an estimated 38 transect segments were the minimum requirement in order to achieve 80% power to detect a coldspot of abundance for at least one species-season combination (Great Shearwater in fall at effect size  $\delta = 1/10$ ; **Table B3**). Achieving 80% power to detect a coldspot of occurrence required between 11 and more than 1,063 transect segments, depending on species, season, and effect size (**Table B1**).

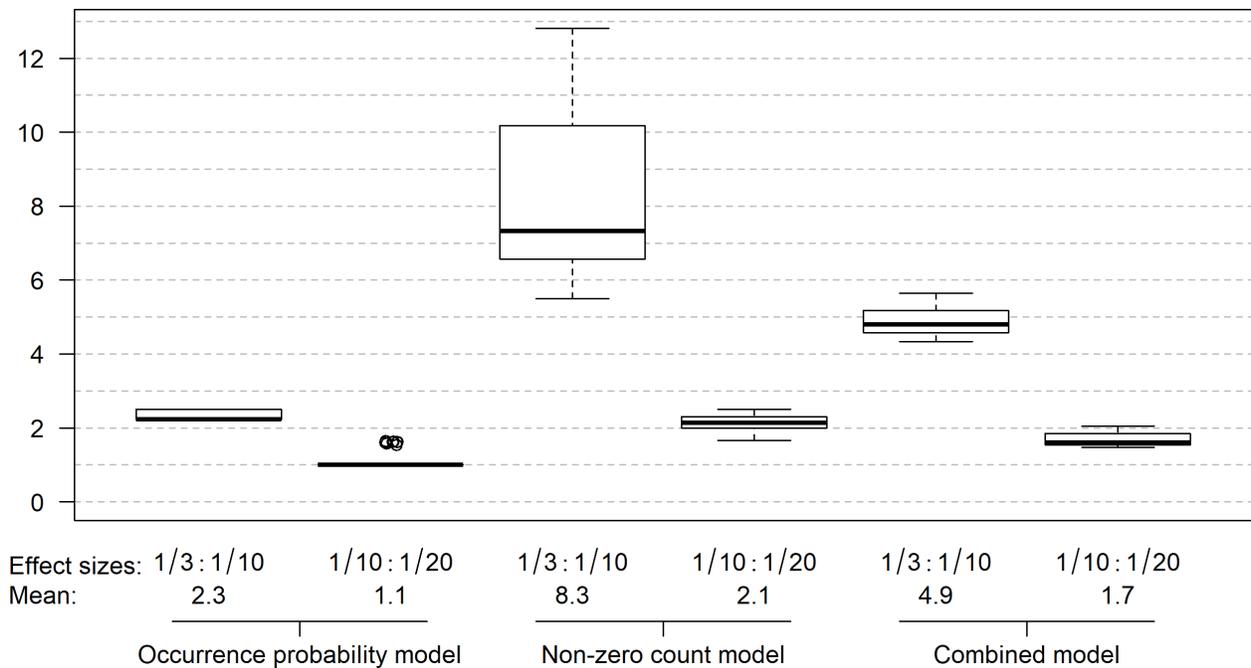
Interpreting the sample size requirements for 80% power based on the occurrence probability and combined models is straightforward since sample size corresponds directly to the number of transect segments surveyed within a single grid cell. For the non-zero count model, the interpretation is not as clear because sample size corresponds to the number of transect segments with at least one sighting of the specified species during the specified season. For a rough estimate of the number of transect segments with and without sightings (i.e., survey effort) needed to achieve 80% power to detect hotspots or coldspots of non-zero abundance, we suggest either a) simply using the sample size estimates from the combined model shown in **Table B3** or b) dividing the estimates from the non-zero count model found in **Table B2** by the corresponding reference prevalence from **Table 4**.

Summarizing across all species-season combinations, detecting a hotspot of occurrence with 80% power for effect size  $\delta = 10$  required on average three times as many transect segments within a single grid cell than for effect size  $\delta = 20$  (**Figure 2**). Achieving 80% power for effect size  $\delta = 3$  required roughly nine times as many transect segments than for effect size  $\delta = 10$ . For the non-zero count model, detecting a hotspot with 80% power required roughly twice as many transect segments with at least one sighting within a single grid cell for effect size  $\delta = 10$  compared to  $\delta = 20$  and roughly six times as many transect segments with sightings for effect size  $\delta = 3$  compared to  $\delta = 10$ . Detecting a hotspot of unconditional abundance (i.e., based on the combined model) with 80% power required roughly twice as many transect



**Figure 2. Boxplots showing the ratio of sample sizes needed for 80% power to detect hotspots of the specified effect size combination for each model type (type I error rate = 0.05)**

Species-season combinations that failed to reach 80% power within the range of sample sizes examined were excluded.



**Figure 3. Boxplots showing the ratio of sample sizes needed for 80% power to detect coldspots of the specified effect size combination for each model type (type I error rate = 0.05)**

Species-season combinations that failed to reach 80% power within the range of sample sizes examined were excluded.

segments for effect size  $\delta = 10$  compared to effect size  $\delta = 20$  and on average about five times as many transect segments within a single grid cell for effect size  $\delta = 3$  compared to  $\delta = 10$ .

Detecting a coldspot of occurrence with 80% power required roughly the same number of transect segments for effect sizes  $\delta = 1/10$  and  $\delta = 1/20$ , while effect size  $\delta = 1/3$  required on average about twice as many transect segments within a single grid cell than effect size  $\delta = 1/10$  (**Figure 3**). For the non-zero count model, detecting a coldspot with 80% power required roughly twice as many transect segments with at least one sighting for effect size  $\delta = 1/10$  compared to effect size  $\delta = 1/20$  and roughly eight times as many transect segments with sightings for effect size  $\delta = 1/3$  compared to  $\delta = 1/10$ . Detecting a coldspot of unconditional abundance with 80% power required roughly twice as many transect segments for effect size  $\delta = 1/10$  compared to  $\delta = 1/20$  and on average about five times as many transect segments within a single grid cell for effect size  $\delta = 1/3$  compared to effect size  $\delta = 1/10$ .

These power analysis results should not be interpreted to mean that larger effect sizes and coarser spatial resolutions are preferable simply because, in general, they yield greater power to detect hotspots and coldspots. Smaller effect sizes use a less stringent definition of hotspots which may be preferable for certain regulatory decisions, and finer spatial resolution allows for more precise identification of hotspots and coldspots within a spatial domain. However, smaller effect sizes and finer spatial resolution both require additional survey effort (i.e., larger sample size) to achieve the same level of statistical power, compared to larger effect sizes and coarser spatial resolutions.

### 3.4 Stratification

For all four species-season combinations, stratifying transect segments according to the core/non-core area calculation had the effect of increasing the estimated reference prevalence and reference mean values within the core area and decreasing the estimates in the non-core area. Within the core area, this led to an increase in hotspot  $p$ -values (decrease in statistical significance), a decrease in coldspot  $p$ -values (increase in statistical significance), and a decrease in power for a fixed sample size. Results showed opposite patterns within the non-core area: power estimates and coldspot  $p$ -values increased, while hotspot  $p$ -values decreased.

## 4 Discussion

This study applied methodology developed by Kinlan et al. (2012) to twenty species of marine birds and explored a number of refinements to the general framework. Useful recommendations to consider when completing spatial power analyses are presented in Kinlan et al. (2012, Sections 4.0–4.7). Additional recommendations based on this study are given below.

### 4.1 Model Type

In most cases, when both zero and non-zero counts are recorded during data collection (i.e., when data collection includes a measure of space surveyed along with species-specific counts), the combined model is most appropriate and preferable. In situations where only presence/absence is recorded or when identification of hotspots/coldspots of occurrence is sufficient, the occurrence probability model is suggested. In general, use of the non-zero count model as a basis for power analyses and hotspot/coldspot identification is discouraged, especially for management applications. The non-zero count model only considers transect segments in which the species of interest was sighted. Therefore, an area that shows evidence of being a hotspot based on the non-zero count model really means that the area is a potential

hotspot compared to all other areas the species has been sighted, as opposed to all areas sampled within the study region. Likewise and less useful in drawing meaningful conclusions, an area that shows evidence of being a coldspot based on the non-zero count model really means that the area is a potential coldspot compared to all other areas the species has been sighted. In other words, the given species could be expected to be present within the coldspot area, just in smaller numbers. In contrast, the occurrence probability and combined models use the additional information present in areas that were sampled but had no sightings of the given species. If the goal is to identify coldspots so as to avoid a given species, one would certainly want to consider areas where sampling had occurred but the species was not sighted. However, in rare situations where only positive counts are recorded, the non-zero count model may be the only option.

## 4.2 Type I and Type II Error Rates

Type I and type II error rates play an essential role in statistical hypothesis tests and power analyses. In hypothesis testing a type I error, or “false positive” result, happens when the null hypothesis is rejected in favor of the alternative hypothesis when, in fact, the null hypothesis is true. A type II error, or “false negative” result, happens when the null hypothesis is not rejected when the alternative hypothesis is actually true. When the  $p$ -value from a hypothesis test is directly used to make a yes/no decision, such as in the designation of a grid cell as a hotspot or non-hotspot, the  $p$ -value threshold at which the null hypothesis is rejected will match the specified type I error rate (i.e., the null hypothesis is rejected in favor of the alternative hypothesis when the  $p$ -value is less than or equal to the type I error rate). Power analyses use the  $p$ -value from a hypothesis test to make a yes/no decision and therefore require specification of a desired type I error rate. Furthermore, statistical power, which is defined as the probability that the null hypothesis is correctly rejected, is mathematically equivalent to one minus the type II error rate. Therefore, the outcome of any power analysis provides a direct identification of type II error rate for a specified sample size, or conversely, provides the sample size required to achieve a specified type II error rate or power level. Throughout the analyses we used a type I error rate of 0.05 and interpreted results assuming a type II error rate of 0.2 (i.e., a desired power level of 80%).

While it is impossible to entirely eliminate the probability of committing a type I or type II error, minimizing the rate of each is preferable, but this comes at a cost of increased sample size. In general, a decrease in type I error rate requires a larger sample size in order to achieve the same power level. Likewise, as discussed **Section 3.3**, an increase in power, which corresponds to a decrease in type II error rate, generally requires a larger sample size for a fixed type I error rate. These costs and benefits must be weighed when determining type I and type II error rates appropriate for the question at hand. A useful example with management implications that further details the link between type I error rate, sample size, and statistical power can be found in Taylor et al. (2007).

## 4.3 Adjusted vs. Unadjusted $p$ -values

In **Section 2.5**, we suggested and applied a method for adjusting significance test  $p$ -values by controlling for false positives. Adjusted  $p$ -values provide more certainty in identification of hotspots and coldspots since, by design, they contain fewer false positives. Therefore, areas that show evidence of being hotspots or coldspots (small adjusted  $p$ -value) are very likely true hotspots/coldspots of abundance or occurrence. However, due to the potentially large number of independent hypothesis tests performed—one test for each grid cell—adjusted  $p$ -values, especially those based on the occurrence probability or combined models, will often show no evidence of hotspots or coldspots within the study region. For any species, true hotspots and coldspots should exist, unless abundance (or occurrence probability) is uniform throughout the region, which is unlikely. These hotspots/coldspots are most likely not being detected within this analysis

framework because of insufficient survey effort.

In general, we suggest the use of adjusted  $p$ -values over unadjusted  $p$ -values. As discussed above, type I error rates can be shifted to strike the necessary balance between sample size and power. However, in certain situations where it is infeasible to obtain adequate survey data to meet even a minimal level of statistical power, unadjusted  $p$ -values can be useful in identifying possible locations of hotspots and coldspots, as long as users understand that the potential for false positives is possibly much higher than the specified type I error rate.

The use of adjusted vs. unadjusted  $p$ -values may also be situationally dependent. In situations where species protection is important, a liberal approach to identifying important areas (i.e., hotspots) for a species of interest may be preferred. In these cases, it is likely more important to ensure a high level of power (i.e., low type II error rate) than it is to have a low type I error rate. Using offshore wind energy as an example, the safest option in order to minimize risk for a given species would be to avoid development in all areas identified as possible hotspots (of a reasonable effect size) even if, in reality, some of these are false positives. While it would still be preferable to use adjusted  $p$ -values and simply increase the acceptable type I error rate to obtain adequate power, unadjusted  $p$ -values could be used to detect hotspots in this case. As long as a high level of power was achieved in order to minimize the probability of failing to identify a true hotspot, this use of unadjusted  $p$ -values could be warranted since the effect would be more areas identified as potential hotspots at the cost of a possibly unquantifiable increase in the number of false positives.

When the goal is species avoidance, a more cautious approach to detecting coldspots for a species may be preferred. In these situations, it is likely more important to ensure a low type I error rate than a high power level. Again using offshore wind energy as an example, the safest option in order to minimize risk for a given species would be to only consider for development areas identified as coldspots (of a reasonable effect size) with a high degree of certainty, even if some true coldspots failed to be identified. This could be accomplished using adjusted  $p$ -values with a low specified type I error rate. Here, power levels could safely be decreased in order to achieve reasonable sample sizes, since false negatives would be of minimal concern. Areas identified as potential coldspots based on adjusted  $p$ -values with a low type I error rate would very likely be true areas of low abundance (or occurrence) where development could be considered and the potential for impacts on the species of interest would be minimized.

#### **4.4 Effect Size and Spatial Resolution**

Greater power to detect a hotspot/coldspot can always be obtained simply by increasing the hotspot/coldspot proportional effect size. Less survey effort is required to detect a hotspot of twenty times the reference mean than a hotspot of ten times the reference mean, and less survey effort is required to detect a coldspot of  $\frac{1}{20}$  times the reference mean than a coldspot of  $\frac{1}{10}$  times the reference mean. However, smaller effect sizes, especially in the hotspot case, may be preferable in the context of identifying important areas for a species because they use a less stringent definition of hotspots. For example, a large effect size may only identify hotspots of extremely high relative abundance, whereas a smaller effect size may have the ability to identify areas with moderately high relative abundance as potential hotspots as well.

The opposite is possibly true for identifying coldspots: larger effect sizes may be preferable, assuming the effect size times the reference mean is greater than one. With regard to human activities and potential development in the offshore environment, a region of very low abundance would serve as a better development site than an area with moderately low abundance, simply because it likely carries less inherent risk of impact to the species. Assuming coldspots defined as  $\frac{1}{20}$  or  $\frac{1}{10}$  times the reference mean exist in the study region, a sufficient goal may be to obtain adequate power to detect coldspots at one of these larger effect sizes, since less survey effort would be required.

Another important consideration for power analyses is the spatial resolution (i.e., grid cell size) at which analyses are completed. Assuming complete survey coverage within the study area, larger grid cells will, in general, result in greater power to detect hotspots or coldspots simply because they contain more survey effort. However, smaller grid cells allow for more precise identification of hotspots and coldspots, as long as there is adequate power to detect them.

When determining the appropriate spatial resolution for assessing hotspots/coldspots, one must balance multiple considerations. Namely, the scale of spatial autocorrelation within the survey data, the scale at which management or regulatory decisions will be made, and the amount of existing survey data within the region of interest. Kinlan et al. (2012, Section 4.5) discussed the importance of matching the grid cell size to the scale of spatial autocorrelation. Offshore development, specifically wind energy development projects, will likely be larger in scale than the size of a single  $4.8 \times 4.8$  km grid cell. Ideally, power analyses conducted to inform such development projects will match the spatial resolution of the project. However, sufficient survey data to obtain adequate power to detect hotspots/coldspots at the scale of spatial autocorrelation and management decisions may not exist and may be cost prohibitive to gather. In these cases, it may be necessary to explore larger effect sizes and/or decreased spatial resolution.

## 5 Case Study: Common Eider in Nantucket Sound during Winter

Nantucket Sound lies off the southeastern coast of Massachusetts and is bordered by Cape Cod to the north, Nantucket to the south, and Martha's Vineyard to the west. A section of Nantucket Sound was previously considered for offshore wind power development and a comparatively large amount of marine bird survey data exists within the region. As an example of how analyses from this study could be used to inform spatial planning and management decisions, we present an interpretation of our results for Common Eider during the winter season within Nantucket Sound. We assume a maximum allowable type I error rate of 0.05 and type II error rate of 0.2 (i.e., a minimum acceptable power level of 80%) for both hotspot and coldspot identification. Since we are analyzing a large dataset containing both zero and non-zero species-specific counts of marine birds spanning many years (**Section 2.1**), analyses based on the combined model are most appropriate.

### 5.1 Hotspots/Coldspots of Abundance

All  $4.8 \times 4.8$  km grid cells within Nantucket Sound show no evidence of being either a hotspot or coldspot of Common Eider abundance in winter (all adjusted  $p$ -values  $\approx 1$ ); however, some of these grid cells may be false negatives. To understand the potential for false negatives, we need to investigate power to detect hotspots/coldspots within the region. Achieving 80% power to detect a hotspot of Common Eider abundance in winter, even at an effect size of twenty times the reference mean, requires at least 268 transect segments of about 4 km in length within a single grid cell (**Table B3**). Having 80% power to detect a coldspot of Common Eider abundance in winter also requires a very substantial amount of survey effort: at least 193 transect segments for an effect size of  $1/20$  times the reference mean. Based on currently available data, most  $4.8 \times 4.8$  km grid cells within the central part of Nantucket Sound have roughly 51–107 transect segments that have been surveyed. For each grid cell that failed to achieve 80% power (i.e., all of them), there is a  $>20\%$  chance that the grid cell was incorrectly identified as a non-hotspot and a  $>20\%$  chance that the grid cell was incorrectly identified as a non-coldspot. In other words, there is a large potential for false negatives. In order to achieve 80% power to detect a hotspot or coldspot of Common Eider abundance at a spatial resolution of  $4.8 \times 4.8$  km throughout a significant portion of Nantucket Sound, the amount of existing survey effort in the region would need to at least be tripled. This is almost certainly cost-prohibitive and infeasible within the near future, especially considering that

Nantucket Sound has been one of the most heavily surveyed regions throughout the entire study area.

At this point, there are two options that may allow us to achieve 80% power to detect hotspots and coldspots of abundance within Nantucket Sound: use more stringent definitions of hotspots/coldspots (i.e., increase the effect sizes) or decrease the spatial resolution. Increasing the effect sizes beyond twenty or  $\frac{1}{20}$  times the reference mean is not an advisable solution as these effect sizes are already quite large.

Decreasing the spatial resolution to grid cells of  $9.6 \times 9.6$  km or larger may allow us to achieve 80% power to detect hotspots/coldspots of abundance throughout Nantucket Sound without the need for additional survey effort and this is a reasonable solution. The drawback here is that the spatial scale at which conclusions can be drawn may not be relevant (i.e., too coarse) for spatial planning in the region.

Even at the coarsest spatial resolution investigated,  $19.2 \times 19.2$  km, all grid cells within Nantucket Sound show no evidence of being either a hotspot or coldspot of Common Eider abundance in winter (all adjusted  $p$ -values  $\approx 1$ ; **Figure 4b**). Adequate sample sizes exist to achieve 80% power to detect both a) hotspots of ten times the reference mean in two grid cells (**Figure 4c**) and b) coldspots of  $\frac{1}{10}$  times the reference mean in four grid cells (**Figure 4d**). We can now be fairly confident that at least two of the grid cells in Nantucket Sound are likely not hotspots of ten times the reference mean or coldspots of  $\frac{1}{10}$  times the reference mean. However, it is important to remember that even with 80% power, there is still a 20% chance that a given grid cell was incorrectly identified as a non-hotspot or non-coldspot. This probability is further compounded when multiple grid cells are considered at once, as the power estimates presented here have not been adjusted for multiple testing. Furthermore, it is entirely possible that true hotspots/coldspots of abundance exists within Nantucket Sound, but that they are hotspots/coldspots of a smaller magnitude than are detectable at effect sizes of ten or  $\frac{1}{10}$  times the reference mean. The fact that multiple grid cells in the region have unadjusted  $p$ -values that are less than 0.01 is perhaps evidence of this (**Figure 4a**).

If the spatial scale of these conclusions is too coarse for spatial planning in the region, a final option may be to consider using the occurrence probability model, which has significantly smaller sample size requirements for 80% power to detect hotspots/coldspots for Common Eider during winter (**Table B1**) with the knowledge that we will then be identifying hotspots/coldspots of occurrence instead of abundance. Before proceeding with the occurrence probability model, the zero/non-zero count data must be transformed into absence/presence data (i.e., all non-zero counts must be reduced to 1's).

## 5.2 Hotspots/Coldspots of Occurrence

At a spatial resolution of  $4.8 \times 4.8$  km there is evidence that the majority of Nantucket Sound is likely a hotspot of Common Eider occurrence in winter (adjusted  $p$ -values  $\leq 0.05$ ; **Figure 5b**). There is also a high level of power (greater than 88%) to detect hotspots of three times the reference prevalence (**Figure 5c**) and coldspots of  $\frac{1}{10}$  times the reference prevalence (**Figure 5d**). We can therefore be confident in concluding that Nantucket Sound is almost certainly an occurrence hotspot (of at least three times the reference prevalence) for Common Eider during the winter season.

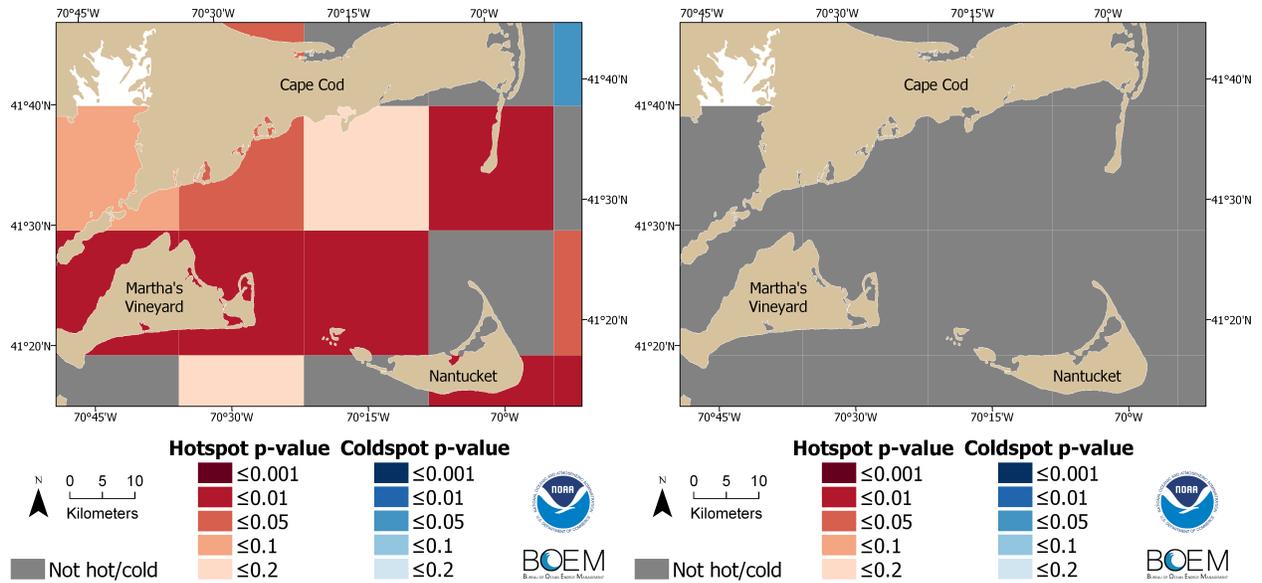
## 5.3 Hotspots/Coldspots of Non-zero Abundance

For completeness, we present conclusions based on the non-zero count model. Identifying hotspots and coldspots of non-zero abundance may be a reasonable solution for some applications as long as the implications of the non-zero count model are understood (see **Section 4.1**). Note that because our data includes both zero and non-zero counts, use of the non-zero count model excludes all transect segments with zero counts of the specified species during the specified season (Common Eider during winter in this case).

There are at least five  $4.8 \times 4.8$  km grid cells within Nantucket Sound that are likely true coldspots of non-zero abundance for Common Eider during winter (adjusted  $p$ -values  $\leq 0.1$ ; **Figure 6b**); however, the majority of grid cells show no evidence of being either a hotspot or coldspot. Moderate power levels are obtained throughout much of the region with a majority of grid cells achieving 80% power to detect hotspots of twenty times the reference mean (**Figure 6c**) and about half achieving 80% power to detect coldspots of  $\frac{1}{10}$  times the reference mean (**Figure 6d**). Unadjusted  $p$ -values based on the non-zero count model show an interesting mix of potential hotspots and coldspots within Nantucket Sound (**Figure 6a**), but at least some of these are likely false positives.

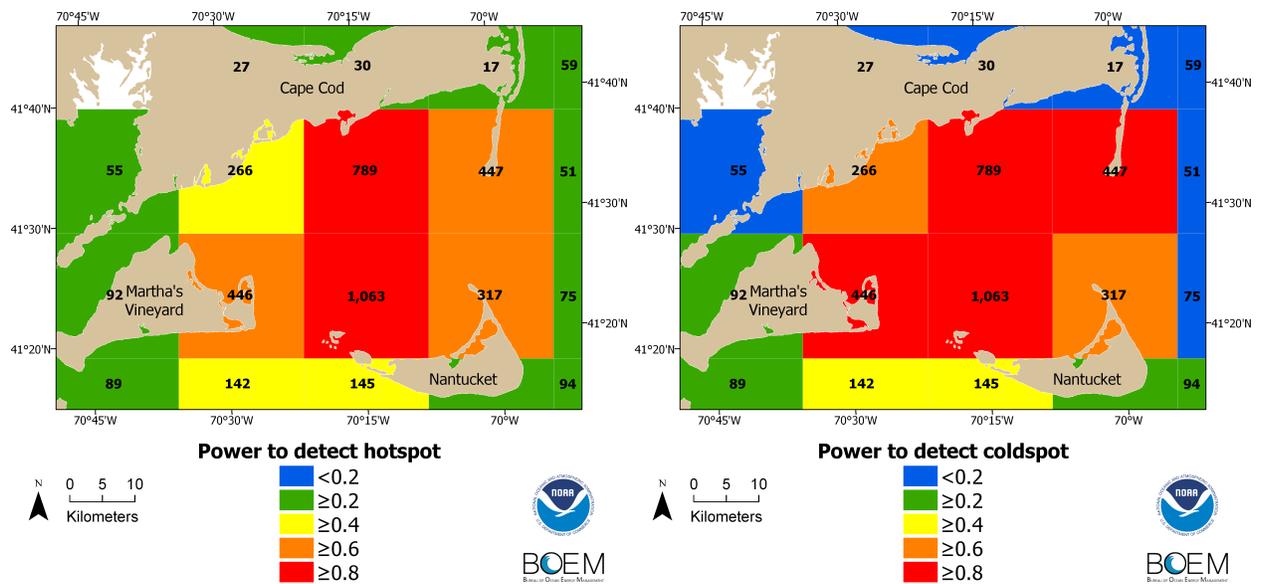
## 5.4 Conclusions

**Figures 4–6** show, rather conclusively, that the majority of Nantucket Sound is likely a true hotspot of occurrence for Common Eider during winter within the Atlantic OCS. There is also evidence for the existence of a few  $4.8 \times 4.8$  km coldspots intermixed with possible  $4.8 \times 4.8$  km hotspots of non-zero abundance for Common Eider during winter in Nantucket Sound. In terms of overall relative abundance for Common Eider during winter, evidence is limited for the existence of either hotspots or coldspots less than or equal to  $19.2 \times 19.2$  km in size. However, it is possible and rather likely that if taken as a whole the entirety of Nantucket Sound would show evidence of being a hotspot of abundance for Common Eider during the winter months when compared to the rest of the Atlantic OCS.



(a) Unadjusted  $p$ -values

(b) Adjusted  $p$ -values

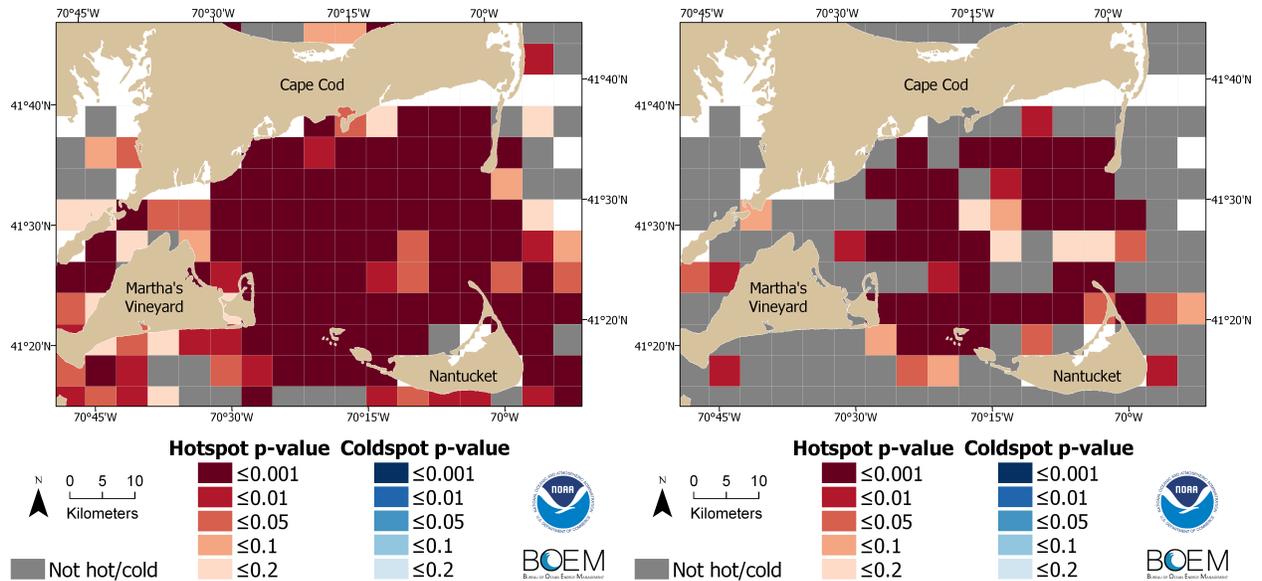


(c) Power to detect a hotspot of ten times the reference mean

(d) Power to detect a coldspot of  $\frac{1}{10}$  times the reference mean

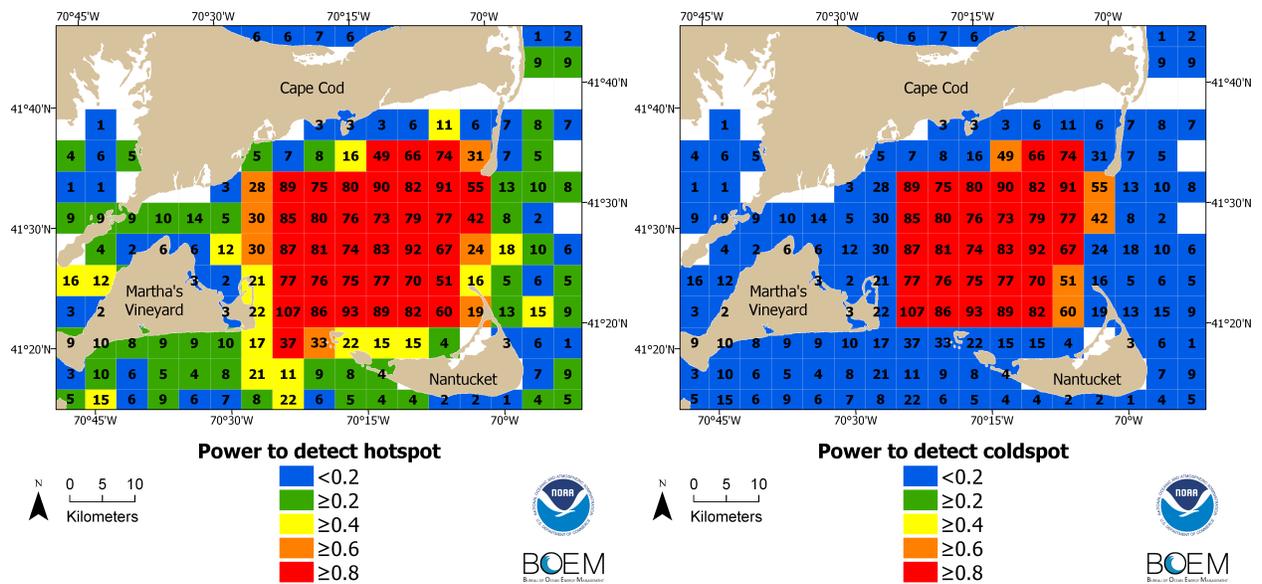
**Figure 4. Maps of significance test  $p$ -values and power to detect hotspots/coldspots of abundance for Common Eider *Somateria mollissima* during winter in Nantucket Sound (based on the combined model with grid cell resolution of  $19.2 \times 19.2$  km)**

Top panel maps show combined hotspot (red) and coldspot (blue)  $p$ -values from simulation-based significance tests of the mean count in each grid cell compared to the combined expectation from the reference prevalence and reference mean (see **Section 2.5** for details). Grid cells with  $p$ -value  $> 0.2$  are shown in gray. Bottom panel maps show power to detect hotspots/coldspots of abundance. The number of transect segments surveyed is shown within each grid cell. Blank grid cells were not surveyed during the winter season.



(a) Unadjusted  $p$ -values

(b) Adjusted  $p$ -values

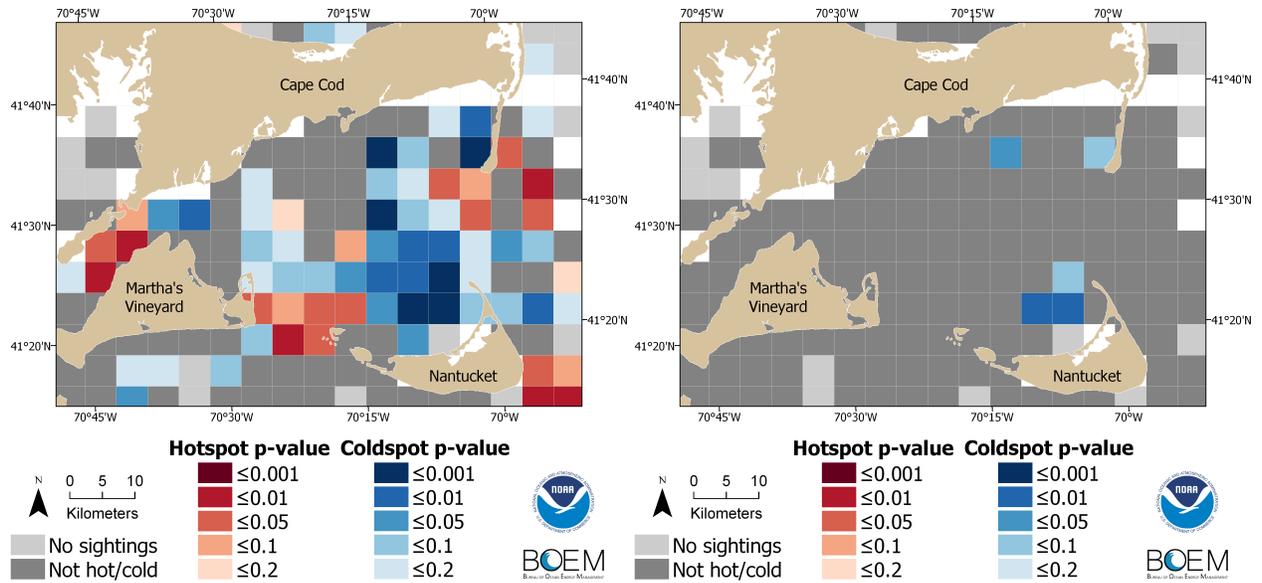


(c) Power to detect a hotspot of three times the reference prevalence

(d) Power to detect a coldspot of  $\frac{1}{10}$  times the reference prevalence

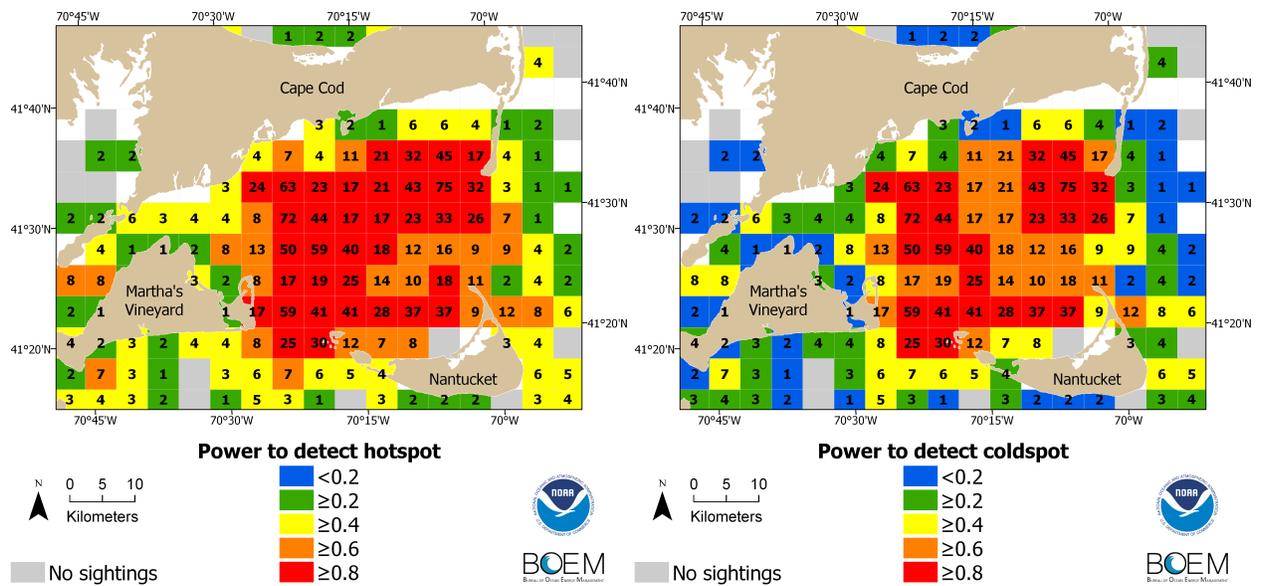
**Figure 5. Maps of significance test  $p$ -values and power to detect hotspots/coldspots of occurrence for Common Eider *Somateria mollissima* during winter in Nantucket Sound (based on the occurrence probability model with grid cell resolution of  $4.8 \times 4.8$  km)**

Top panel maps show combined hotspot (red) and coldspot (blue)  $p$ -values from significance tests of the number of transect segments with at least one sighting compared to the expectation from the reference prevalence, conditional on the number of transect segments surveyed (see Section 2.5 for details). Grid cells with  $p$ -value  $> 0.2$  are shown in gray. Bottom panel maps show power to detect hotspots/coldspots of occurrence. Blank grid cells were not surveyed during the winter season.



(a) Unadjusted  $p$ -values

(b) Adjusted  $p$ -values



(c) Power to detect a hotspot of twenty times the reference mean

(d) Power to detect a coldspot of  $\frac{1}{10}$  times the reference mean

**Figure 6. Maps of significance test  $p$ -values and power to detect hotspots/coldspots of non-zero abundance for Common Eider *Somateria mollissima* during winter in Nantucket Sound (based on the non-zero count model with grid cell resolution of  $4.8 \times 4.8$  km)**

Top panel maps show combined hotspot (red) and coldspot (blue)  $p$ -values from simulation-based significance tests of the mean non-zero count in each grid cell compared to the reference mean (see **Section 2.5** for details). Grid cells with  $p$ -value  $> 0.2$  are shown in dark gray. Bottom panel maps show power to detect hotspots/coldspots of non-zero abundance. Grid cells with survey effort, but no sightings of Common Eider are shown in light gray. Blank grid cells were not surveyed during the winter season.

## 6 References

- Adams J, Kelsey EC, Felis JJ, Pereksta DM (2016). Collision and displacement vulnerability among marine birds of the California Current System associated with offshore wind energy infrastructure. Reston (VA): US Department of the Interior, US Geological Survey. Open-File Report 2016-1154.
- Bivand RS, Keitt T, Rowlingson B (2017). `rgdal`: bindings for the Geospatial Data Abstraction Library. R package version 1.2-7.
- Bivand RS, Rundel C (2017). `rgeos`: interface to Geometry Engine - Open Source (GEOS). R package version 0.3-23.
- Burnham KP, Anderson DR (2002). Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York (NY): Springer-Verlag.
- Chernick MR, Liu CY (2002). The saw-toothed behavior of power versus sample size and software solutions: single binomial proportion using exact methods. *Am Stat.* 56(2):149–155.
- Clauset A, Shalizi CR, Newman MEJ (2009). Power-law distributions in empirical data. *SIAM Rev.* 51(4):661–703.
- Curtice C, Cleary J, Shumchenia E, Halpin PN (2016). Marine-life Data and Analysis Team (MDAT) technical report on the methods and development of marine-life data to support regional ocean planning and management. Prepared on behalf of the Marine-life Data and Analysis Team (MDAT).
- Gillespie CS (2015). Fitting heavy tailed distributions: the `powerlaw` package. *J Stat Softw.* 64(2).
- Hankin RKS (2006). Special functions in R: introducing the `gsl` package. *R News.* 6(4).
- Hijmans RJ (2016). `raster`: geographic data analysis and modeling. R package version 2.5-8.
- Holm S (1979). A simple sequentially rejective multiple test procedure. *Scand J Stat.* 6(2):65–70.
- Kinlan BP, Zipkin EF, O’Connell AF, Caldow C (2012). Statistical analyses to support guidelines for marine avian sampling: final report. Herndon (VA): US Department of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs. OCS Study BOEM 2012-101. NOAA Technical Memorandum NOS NCCOS 158.
- Mullahy J (1986). Specification and testing of some modified count data models. *J Econom.* 33(3):341–365.
- O’Connell A, Gardner B, Gilbert A, Laurent K (2009). Compendium of avian occurrence information for the continental shelf waters along the Atlantic coast of the United States, final report (database section - seabirds). Herndon (VA): US Department of the Interior, Bureau of Ocean Energy Management. OCS Study BOEM 2012-076.
- R Core Team (2018). R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.
- Shaffer JP (1995). Multiple hypothesis testing. *Annu Rev Psychol.* 46:561–584.
- Taylor BL, Martinez M, Gerrodette T, Barlow J (2007). Lessons from monitoring trends in abundance of marine mammals. *Mar Mammal Sci.* 23(1):157–175.
- Vuong QH (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica.* 57(2):307–333.
- Winship AJ, Kinlan BP, White TP, Leirness JB, Christensen J (2018). Modeling at-sea density of marine birds to support Atlantic marine renewable energy planning: final report. Sterling (VA): US Department of the Interior, Bureau of Ocean Energy Management, Office of Renewable Energy Programs. OCS Study BOEM 2018-010.

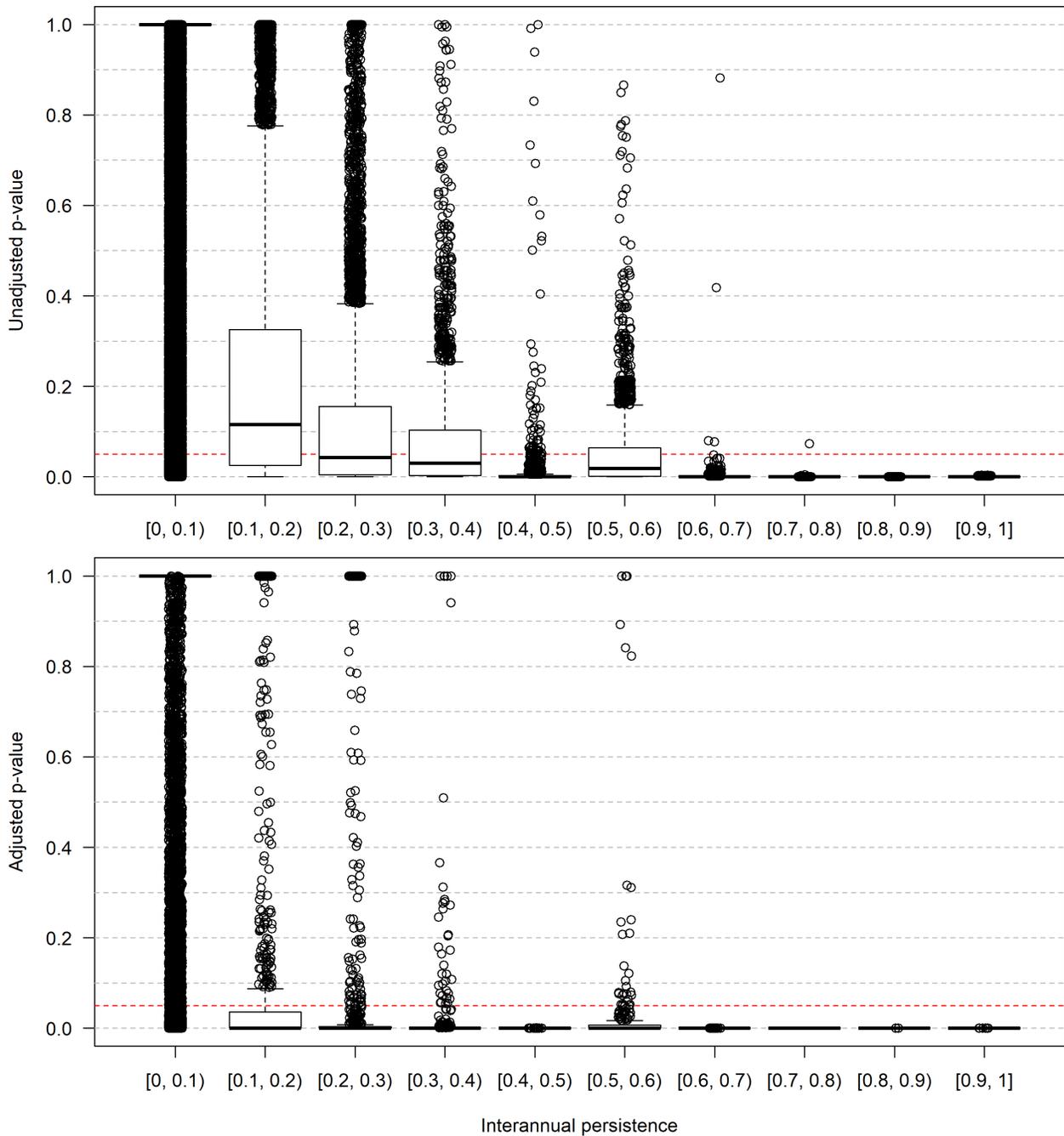
Wright SP (1992). Adjusted p-values for simultaneous inference. *Biometrics*. 48(4):1005–1013.

Yee TW (2010). The VGAM package for categorical data analysis. *J Stat Softw*. 32(10).

Zipkin EF, Kinlan BP, Sussman A, Rypkema D, Wimer M, O'Connell AF (2015). Statistical guidelines for assessing marine avian hotspots and coldspots: a case study on wind energy development in the US Atlantic Ocean. *Biol Conserv*. 191:216–223.

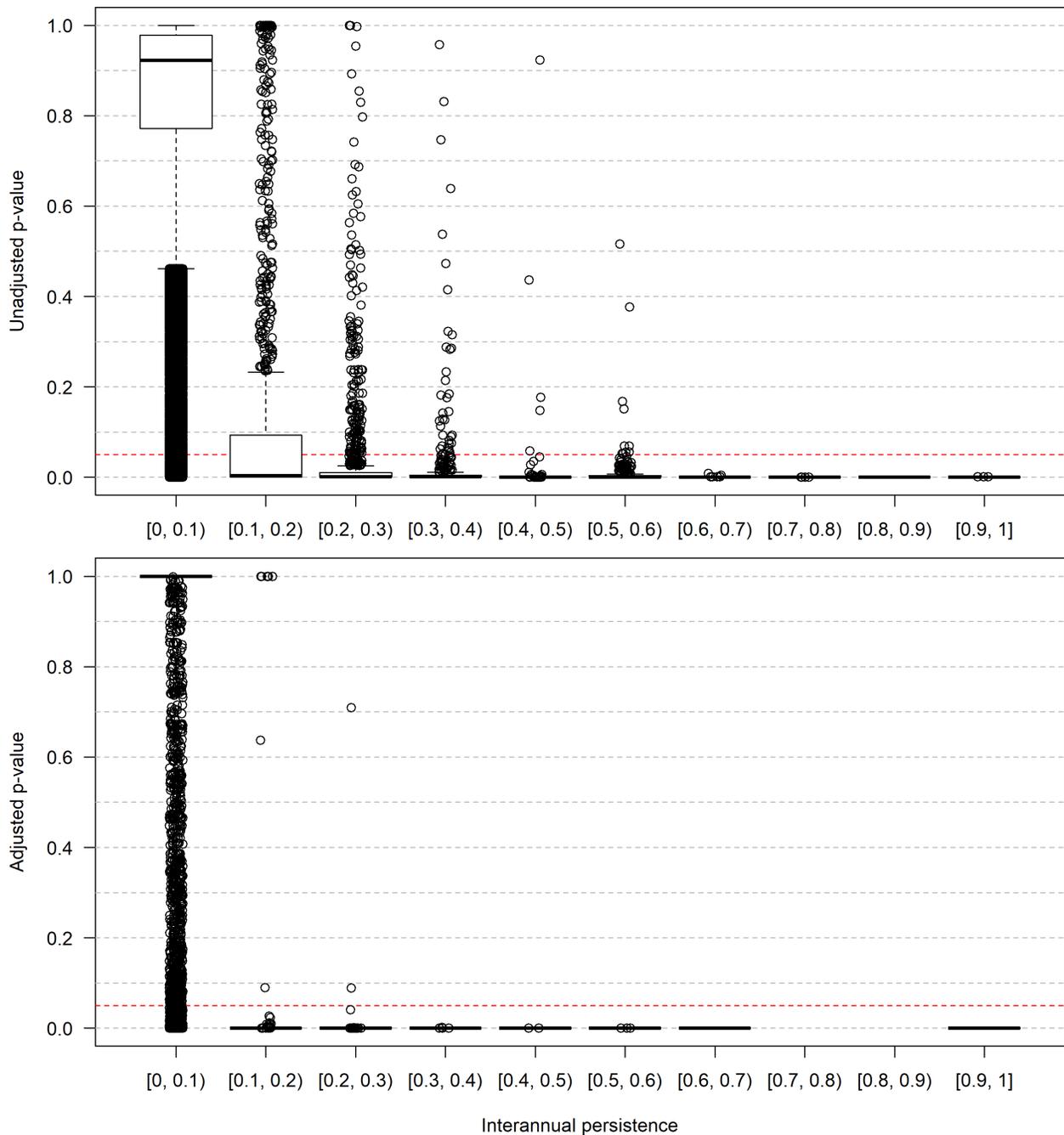
Zipkin EF, Leirness JB, Kinlan BP, O'Connell AF, Silverman ED (2014). Fitting statistical distributions to sea duck count data: implications for survey design and abundance estimation. *Stat Methodol*. 17:67–81.

## Appendix A: Boxplots of Hotspot/Coldspot $p$ -values vs. Interannual Persistence



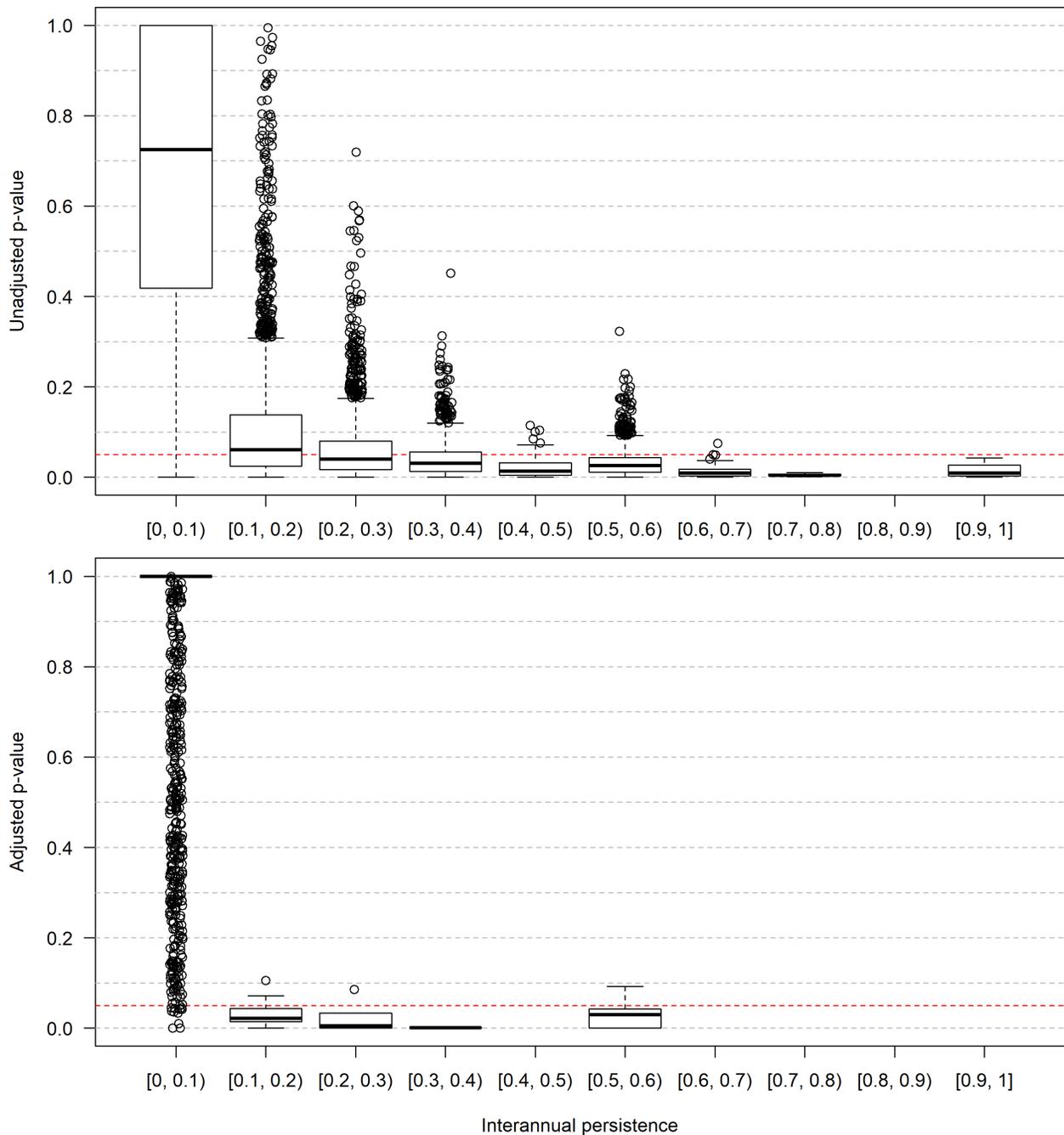
**Figure A1. Boxplots of hotspot  $p$ -values vs. interannual persistence for the occurrence probability model (pooled across all grid cell resolutions).**

Separate boxplots are drawn for each specified range of interannual persistence values. Boxplots of unadjusted  $p$ -values are shown in the top panel while adjusted  $p$ -values are shown in the bottom panel. Unadjusted/adjusted  $p$ -values are from overall significance tests (i.e., including all years of relevant data; see **Section 2.5**). The red dashed line corresponds to an unadjusted/adjusted  $p$ -value of 0.05. Interannual persistence is defined as the proportion of years with survey effort in which the single-year  $p$ -value was less than or equal to 0.05. Results are only shown for grid cells that were surveyed in multiple years during the same season.



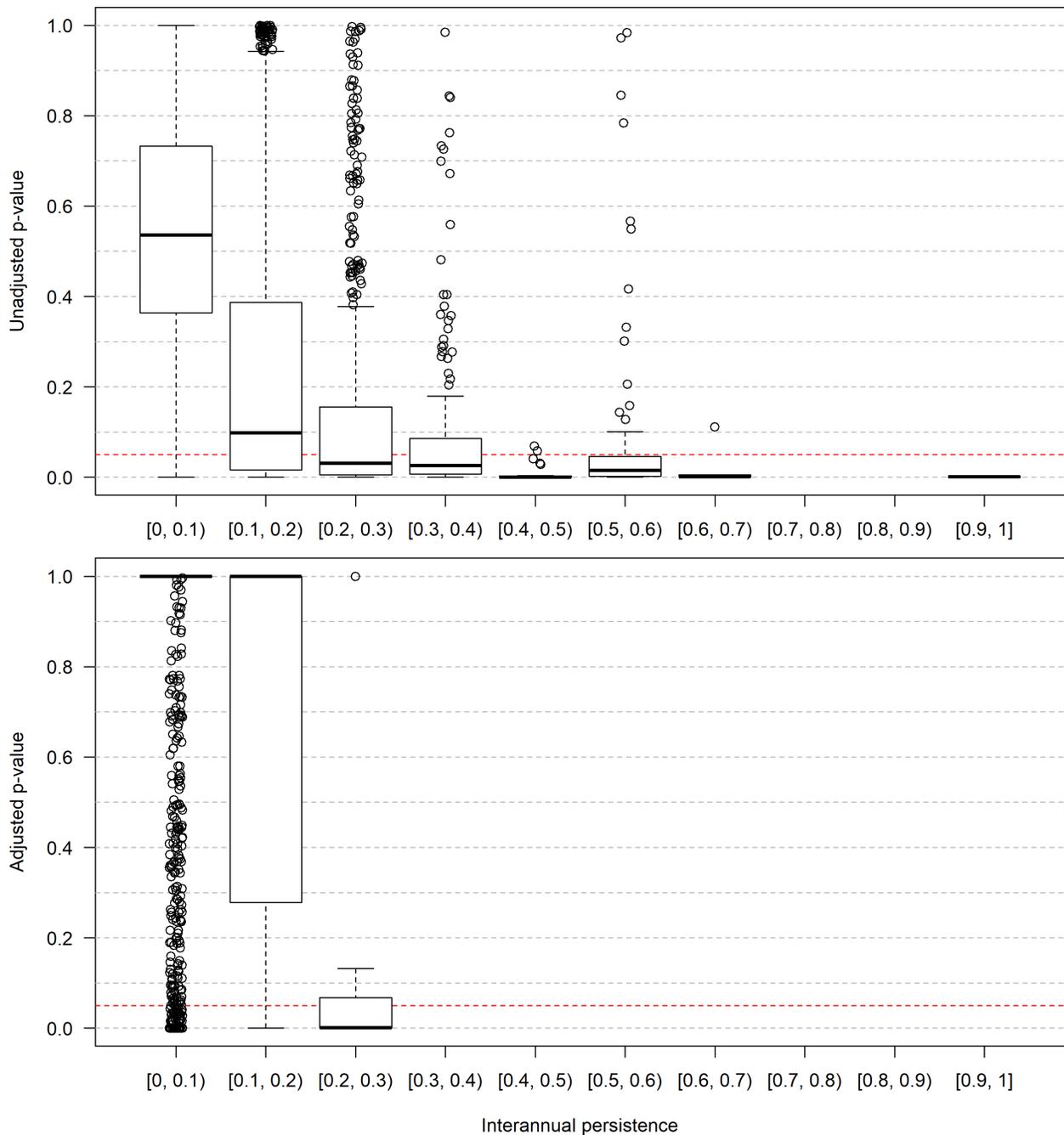
**Figure A2. Boxplots of coldspot  $p$ -values vs. interannual persistence for the occurrence probability model (pooled across all grid cell resolutions).**

Separate boxplots are drawn for each specified range of interannual persistence values. Boxplots of unadjusted  $p$ -values are shown in the top panel while adjusted  $p$ -values are shown in the bottom panel. Unadjusted/adjusted  $p$ -values are from overall significance tests (i.e., including all years of relevant data; see **Section 2.5**). The red dashed line corresponds to an unadjusted/adjusted  $p$ -value of 0.05. Interannual persistence is defined as the proportion of years with survey effort in which the single-year  $p$ -value was less than or equal to 0.05. Results are only shown for grid cells that were surveyed in multiple years during the same season.



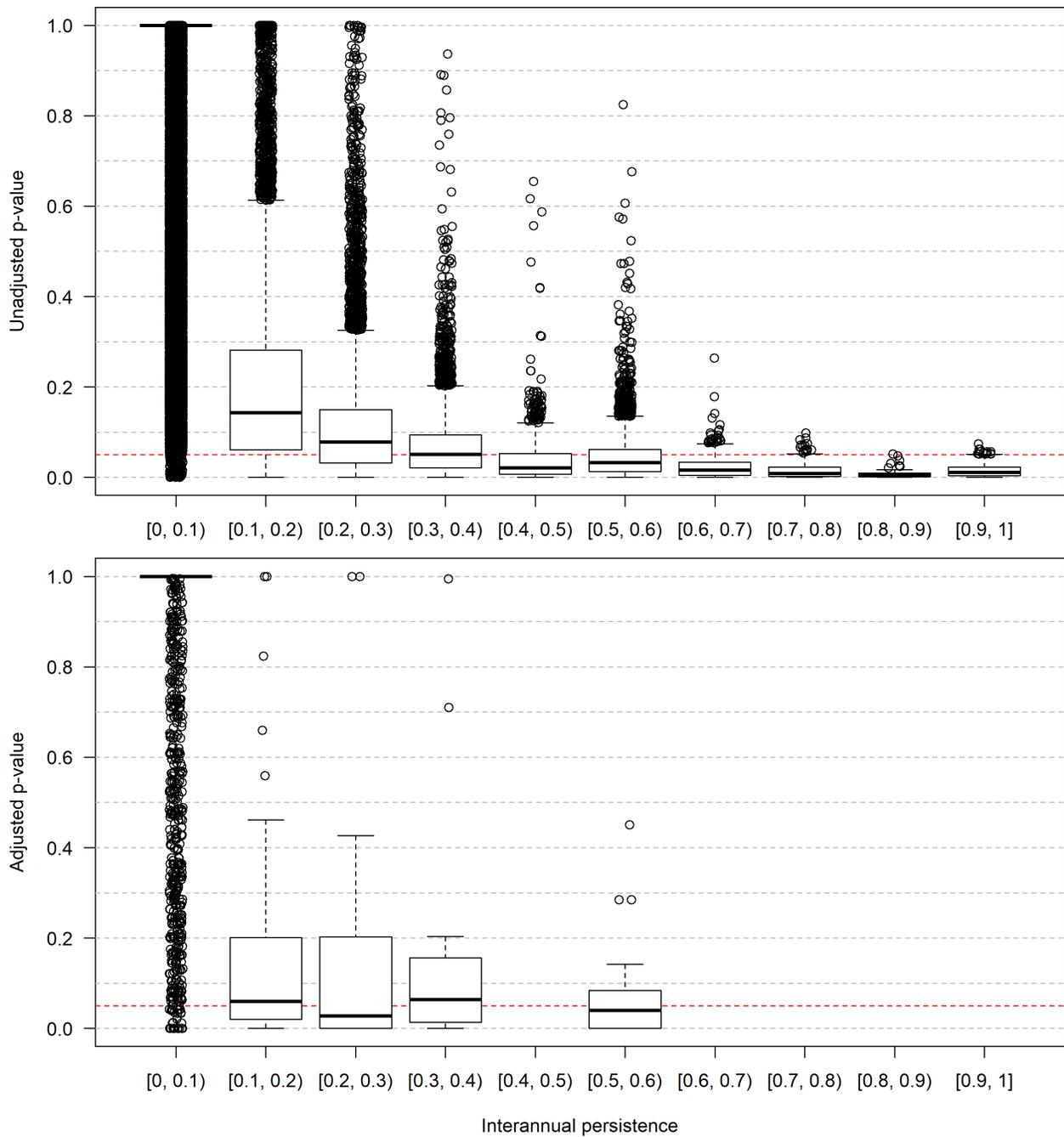
**Figure A3. Boxplots of hotspot  $p$ -values vs. interannual persistence for the non-zero count model (pooled across all grid cell resolutions).**

Separate boxplots are drawn for each specified range of interannual persistence values. Boxplots of unadjusted  $p$ -values are shown in the top panel while adjusted  $p$ -values are shown in the bottom panel. Unadjusted/adjusted  $p$ -values are from overall significance tests (i.e., including all years of relevant data; see **Section 2.5**). The red dashed line corresponds to an unadjusted/adjusted  $p$ -value of 0.05. Interannual persistence is defined as the proportion of years with sightings (of a given species during a given season) in which the single-year  $p$ -value was less than or equal to 0.05. Results are only shown for grid cells with sightings in multiple years during the same season.



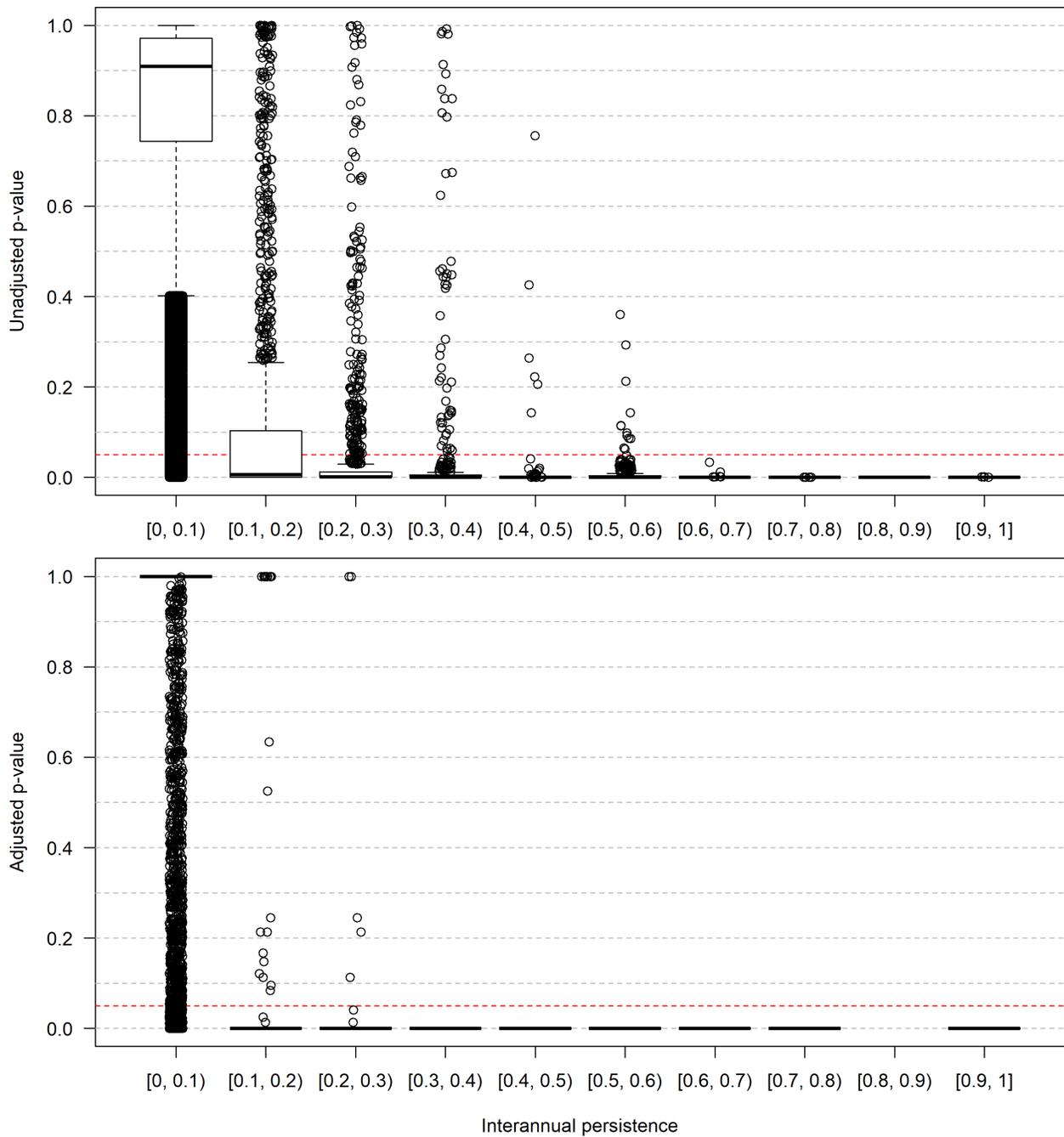
**Figure A4. Boxplots of coldspot  $p$ -values vs. interannual persistence for the non-zero count model (pooled across all grid cell resolutions).**

Separate boxplots are drawn for each specified range of interannual persistence values. Boxplots of unadjusted  $p$ -values are shown in the top panel while adjusted  $p$ -values are shown in the bottom panel. Unadjusted/adjusted  $p$ -values are from overall significance tests (i.e., including all years of relevant data; see **Section 2.5**). The red dashed line corresponds to an unadjusted/adjusted  $p$ -value of 0.05. Interannual persistence is defined as the proportion of years with sightings (of a given species during a given season) in which the single-year  $p$ -value was less than or equal to 0.05. Results are only shown for grid cells with sightings in multiple years during the same season.



**Figure A5. Boxplots of hotspot  $p$ -values vs. interannual persistence for the combined model (pooled across all grid cell resolutions).**

Separate boxplots are drawn for each specified range of interannual persistence values. Boxplots of unadjusted  $p$ -values are shown in the top panel while adjusted  $p$ -values are shown in the bottom panel. Unadjusted/adjusted  $p$ -values are from overall significance tests (i.e., including all years of relevant data; see **Section 2.5**). The red dashed line corresponds to an unadjusted/adjusted  $p$ -value of 0.05. Interannual persistence is defined as the proportion of years with survey effort in which the single-year  $p$ -value was less than or equal to 0.05. Results are only shown for grid cells that were surveyed in multiple years during the same season.



**Figure A6. Boxplots of coldspot  $p$ -values vs. interannual persistence for the combined model (pooled across all grid cell resolutions).**

Separate boxplots are drawn for each specified range of interannual persistence values. Boxplots of unadjusted  $p$ -values are shown in the top panel while adjusted  $p$ -values are shown in the bottom panel. Unadjusted/adjusted  $p$ -values are from overall significance tests (i.e., including all years of relevant data; see **Section 2.5**). The red dashed line corresponds to an unadjusted/adjusted  $p$ -value of 0.05. Interannual persistence is defined as the proportion of years with survey effort in which the single-year  $p$ -value was less than or equal to 0.05. Results are only shown for grid cells that were surveyed in multiple years during the same season.

## Appendix B: Sample Size Requirements for 80% Power

**Table B1. Sample size requirements for 80% power based on the occurrence probability model (type I error rate = 0.05)**

For the occurrence probability model, sample size corresponds to the number of transect segments surveyed.

Species	Season	Hotspot effect size			Coldspot effect size		
		3	10	20	1/3	1/10	1/20
Common Eider	spring	79	11	3	277	125	125
Common Eider	summer	607	85	30	>638	>638	601
Common Eider	fall	189	23	8	657	263	263
Common Eider	winter	41	3	–	145	65	65
Surf Scoter	spring	100	12	4	>200	139	139
Surf Scoter	fall	154	19	6	536	214	214
Surf Scoter	winter	49	4	–	171	77	77
White-winged Scoter	spring	151	18	6	527	211	211
White-winged Scoter	fall	196	24	8	684	273	273
White-winged Scoter	winter	63	6	1	221	99	99
Long-tailed Duck	spring	55	5	–	191	86	86
Long-tailed Duck	fall	239	29	10	739	333	333
Long-tailed Duck	winter	26	–	–	91	41	41
Razorbill	spring	74	7	3	>200	116	116
Razorbill	summer	>638	199	69	>638	>638	>638
Razorbill	fall	690	86	30	>771	>771	606
Razorbill	winter	51	5	–	180	81	81
Atlantic Puffin	spring	418	52	18	>728	582	582
Atlantic Puffin	summer	491	61	21	>638	>638	431
Atlantic Puffin	fall	>771	170	59	>771	>771	>771
Atlantic Puffin	winter	368	46	16	>1,063	512	512
Laughing Gull	spring	119	14	5	415	166	166
Laughing Gull	summer	66	6	1	231	104	104
Laughing Gull	fall	61	5	1	212	95	95
Laughing Gull	winter	832	104	36	>1,063	>1,063	730
Herring Gull	spring	9	–	–	45	20	13
Herring Gull	summer	35	3	–	125	56	56
Herring Gull	fall	9	–	–	47	21	13
Herring Gull	winter	16	–	–	65	29	18
Least Tern	summer	>638	121	42	>638	>638	>638
Least Tern	fall	>771	163	57	>771	>771	>771
Roseate Tern	spring	>728	179	62	>728	>728	>728
Roseate Tern	summer	>200	75	26	>200	>200	>200
Roseate Tern	fall	>771	195	68	>771	>771	>771
Common Tern	spring	141	17	6	490	196	196
Common Tern	summer	65	6	1	227	102	102
Common Tern	fall	169	21	7	589	235	235
Royal Tern	spring	313	39	13	>728	436	436

continued on next page

Table B1 continued

Species	Season	Hotspot effect size			Coldspot effect size		
		3	10	20	1/3	1/10	1/20
Royal Tern	summer	475	59	20	>638	>638	>638
Royal Tern	fall	332	41	14	>771	462	462
Red-throated Loon	spring	42	4	–	146	66	66
Red-throated Loon	fall	304	38	13	>771	423	423
Red-throated Loon	winter	36	3	–	126	57	57
Common Loon	spring	26	–	–	91	41	41
Common Loon	summer	631	79	27	>638	>638	554
Common Loon	fall	77	7	3	>200	121	121
Common Loon	winter	21	–	–	86	39	39
Black-capped Petrel	spring	713	89	31	>728	>728	626
Black-capped Petrel	summer	394	49	17	>638	549	549
Black-capped Petrel	fall	>771	205	72	>771	>771	>771
Black-capped Petrel	winter	>1,063	168	58	>1,063	>1,063	>1,063
Cory's Shearwater	spring	>728	104	36	>728	>728	>728
Cory's Shearwater	summer	38	3	–	133	60	60
Cory's Shearwater	fall	66	6	1	229	103	103
Sooty Shearwater	spring	113	14	4	393	157	157
Sooty Shearwater	summer	72	10	3	251	113	113
Sooty Shearwater	fall	>771	148	51	>771	>771	>771
Great Shearwater	spring	148	18	6	516	206	206
Great Shearwater	summer	15	–	–	63	28	17
Great Shearwater	fall	14	–	–	58	26	16
Great Shearwater	winter	934	117	41	>1,063	>1,063	820
Audubon's Shearwater	spring	>728	91	32	>728	>728	643
Audubon's Shearwater	summer	157	19	6	545	218	218
Audubon's Shearwater	fall	487	61	21	>771	678	678
Audubon's Shearwater	winter	721	90	31	>1,063	1,003	1,003
Northern Gannet	spring	8	–	–	40	18	11
Northern Gannet	summer	103	9	4	360	144	144
Northern Gannet	fall	19	–	–	76	34	21
Northern Gannet	winter	9	–	–	43	19	12
Minimum	–	8	3	1	40	18	11
Maximum	–	>1,063	205	72	>1,063	>1,063	>1,063

**Table B2. Sample size requirements for 80% power based on the non-zero count model (type I error rate = 0.05)**

For the non-zero count model, sample size corresponds to the number of transect segments with sightings.

Species	Season	Hotspot effect size			Coldspot effect size		
		3	10	20	1/3	1/10	1/20
Common Eider	spring	684	47	17	261	23	10
Common Eider	summer	>638	68	22	410	32	14
Common Eider	fall	141	15	7	75	9	5
Common Eider	winter	710	48	17	263	23	10
Surf Scoter	spring	26	4	2	19	3	–
Surf Scoter	fall	45	6	3	30	5	3
Surf Scoter	winter	30	5	2	22	4	2
White-winged Scoter	spring	698	50	19	205	17	7
White-winged Scoter	fall	62	8	4	36	5	–
White-winged Scoter	winter	60	8	4	34	5	–
Long-tailed Duck	spring	71	9	4	44	6	3
Long-tailed Duck	fall	80	10	5	46	6	3
Long-tailed Duck	winter	45	6	3	31	5	2
Razorbill	spring	12	2	1	9	–	–
Razorbill	summer	5	1	1	–	–	–
Razorbill	fall	32	5	3	16	–	–
Razorbill	winter	27	4	2	16	–	–
Atlantic Puffin	spring	6	2	1	–	–	–
Atlantic Puffin	summer	8	2	1	–	–	–
Atlantic Puffin	fall	4	1	1	–	–	–
Atlantic Puffin	winter	5	1	1	–	–	–
Laughing Gull	spring	9	2	1	–	–	–
Laughing Gull	summer	20	4	2	7	–	–
Laughing Gull	fall	36	6	3	16	–	–
Laughing Gull	winter	13	3	2	–	–	–
Herring Gull	spring	60	8	4	24	–	–
Herring Gull	summer	28	5	3	7	–	–
Herring Gull	fall	45	7	3	19	–	–
Herring Gull	winter	62	9	5	19	–	–
Least Tern	summer	38	6	3	11	–	–
Least Tern	fall	52	7	4	29	4	–
Roseate Tern	spring	11	2	1	5	–	–
Roseate Tern	summer	17	3	2	7	–	–
Roseate Tern	fall	122	15	7	39	–	–
Common Tern	spring	11	2	1	7	–	–
Common Tern	summer	24	4	2	11	–	–
Common Tern	fall	54	7	4	26	–	–
Royal Tern	spring	13	3	2	–	–	–
Royal Tern	summer	10	2	1	–	–	–
Royal Tern	fall	6	1	1	–	–	–

continued on next page

Table B2 continued

Species	Season	Hotspot effect size			Coldspot effect size		
		3	10	20	1/3	1/10	1/20
Red-throated Loon	spring	17	3	2	–	–	–
Red-throated Loon	fall	56	8	4	18	–	–
Red-throated Loon	winter	24	4	2	7	–	–
Common Loon	spring	9	2	1	–	–	–
Common Loon	summer	4	1	1	–	–	–
Common Loon	fall	7	2	1	–	–	–
Common Loon	winter	14	3	2	–	–	–
Black-capped Petrel	spring	13	3	2	–	–	–
Black-capped Petrel	summer	38	7	4	–	–	–
Black-capped Petrel	fall	8	2	1	–	–	–
Black-capped Petrel	winter	7	2	1	–	–	–
Cory's Shearwater	spring	9	2	1	–	–	–
Cory's Shearwater	summer	54	8	4	14	–	–
Cory's Shearwater	fall	40	6	3	15	–	–
Sooty Shearwater	spring	327	34	14	65	–	–
Sooty Shearwater	summer	>638	82	33	152	–	–
Sooty Shearwater	fall	158	29	16	–	–	–
Great Shearwater	spring	38	6	3	20	–	–
Great Shearwater	summer	80	10	5	36	4	–
Great Shearwater	fall	35	5	3	22	3	–
Great Shearwater	winter	15	3	2	–	–	–
Audubon's Shearwater	spring	184	24	12	25	–	–
Audubon's Shearwater	summer	25	5	3	–	–	–
Audubon's Shearwater	fall	15	3	2	–	–	–
Audubon's Shearwater	winter	8	2	1	–	–	–
Northern Gannet	spring	45	7	4	16	–	–
Northern Gannet	summer	16	3	2	–	–	–
Northern Gannet	fall	29	5	3	12	–	–
Northern Gannet	winter	72	10	5	27	–	–
Minimum	–	4	1	1	5	3	2
Maximum	–	>710	82	33	410	32	14

**Table B3. Sample size requirements for 80% power based on the combined model (type I error rate = 0.05)**

For the combined model, sample size corresponds to the number of transect segments surveyed.

Species	Season	Hotspot effect size			Coldspot effect size		
		3	10	20	1/3	1/10	1/20
Common Eider	spring	>728	>728	488	>728	726	370
Common Eider	summer	>638	>638	>638	>638	>638	>638
Common Eider	fall	>771	>771	429	>771	765	443
Common Eider	winter	>1,063	705	268	>1,063	395	193
Surf Scoter	spring	>200	168	99	>200	199	–
Surf Scoter	fall	>771	348	192	>771	388	253
Surf Scoter	winter	594	100	58	516	119	81
White-winged Scoter	spring	>728	>728	>728	>728	>728	523
White-winged Scoter	fall	>771	567	297	>771	526	–
White-winged Scoter	winter	>1,063	209	110	964	186	–
Long-tailed Duck	spring	>728	191	98	>728	184	114
Long-tailed Duck	fall	>771	>771	414	>771	732	457
Long-tailed Duck	winter	448	69	38	351	73	47
Razorbill	spring	>200	93	61	>200	–	–
Razorbill	summer	>638	>638	>638	–	–	–
Razorbill	fall	>771	>771	>771	>771	–	–
Razorbill	winter	555	102	58	443	–	–
Atlantic Puffin	spring	>728	348	243	–	–	–
Atlantic Puffin	summer	>638	480	327	–	–	–
Atlantic Puffin	fall	>771	>771	721	–	–	–
Atlantic Puffin	winter	>1,063	297	215	–	–	–
Laughing Gull	spring	483	118	79	–	–	–
Laughing Gull	summer	556	115	72	389	–	–
Laughing Gull	fall	>771	146	84	526	–	–
Laughing Gull	winter	>1,063	995	637	–	–	–
Herring Gull	spring	292	45	24	150	–	–
Herring Gull	summer	407	80	47	227	–	–
Herring Gull	fall	235	38	21	134	–	–
Herring Gull	winter	433	69	37	187	–	–
Least Tern	summer	>638	>638	>638	>638	–	–
Least Tern	fall	>771	>771	>771	>771	>771	–
Roseate Tern	spring	>728	>728	>728	>728	–	–
Roseate Tern	summer	>200	>200	>200	>200	–	–
Roseate Tern	fall	>771	>771	>771	>771	–	–
Common Tern	spring	665	149	98	682	–	–
Common Tern	summer	638	122	73	475	–	–
Common Tern	fall	>771	467	245	>771	–	–
Royal Tern	spring	>728	402	260	–	–	–
Royal Tern	summer	>638	497	329	–	–	–
Royal Tern	fall	>771	274	191	–	–	–

continued on next page

Table B3 continued

Species	Season	Hotspot effect size			Coldspot effect size		
		3	10	20	1/3	1/10	1/20
Red-throated Loon	spring	308	66	42	–	–	–
Red-throated Loon	fall	>771	>771	491	>771	–	–
Red-throated Loon	winter	354	72	43	221	–	–
Common Loon	spring	118	28	20	–	–	–
Common Loon	summer	>638	438	321	–	–	–
Common Loon	fall	>200	79	55	–	–	–
Common Loon	winter	162	37	23	–	–	–
Black-capped Petrel	spring	>728	>728	618	–	–	–
Black-capped Petrel	summer	>638	>638	594	–	–	–
Black-capped Petrel	fall	>771	>771	>771	–	–	–
Black-capped Petrel	winter	>1,063	>1,063	812	–	–	–
Cory's Shearwater	spring	>728	>728	571	–	–	–
Cory's Shearwater	summer	>638	130	72	348	–	–
Cory's Shearwater	fall	>771	178	101	571	–	–
Sooty Shearwater	spring	>728	>728	540	>728	–	–
Sooty Shearwater	summer	>638	>638	>638	>638	–	–
Sooty Shearwater	fall	>771	>771	>771	–	–	–
Great Shearwater	spring	>728	318	180	>728	–	–
Great Shearwater	summer	519	72	36	288	51	–
Great Shearwater	fall	226	37	21	174	38	–
Great Shearwater	winter	>1,063	>1,063	787	–	–	–
Audubon's Shearwater	spring	>728	>728	>728	>728	–	–
Audubon's Shearwater	summer	>638	299	180	–	–	–
Audubon's Shearwater	fall	>771	652	408	–	–	–
Audubon's Shearwater	winter	>1,063	723	490	–	–	–
Northern Gannet	spring	197	33	18	107	–	–
Northern Gannet	summer	>638	147	93	–	–	–
Northern Gannet	fall	247	47	27	160	–	–
Northern Gannet	winter	>200	49	25	154	–	–
Minimum	–	118	28	18	107	38	47
Maximum	–	>1,063	>1,063	>812	>1,063	>771	>638



### **Department of the Interior (DOI)**

The Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors the Nation's trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities.



### **Bureau of Ocean Energy Management (BOEM)**

The mission of the Bureau of Ocean Energy Management is to manage development of U.S. Outer Continental Shelf energy and mineral resources in an environmentally and economically responsible way.

### **BOEM Environmental Studies Program**

The mission of the Environmental Studies Program is to provide the information needed to predict, assess, and manage impacts from offshore energy and marine mineral exploration, development, and production activities on human, marine, and coastal environments. The proposal, selection, research, review, collaboration, production, and dissemination of each of BOEM's Environmental Studies follows the DOI Code of Scientific and Scholarly Conduct, in support of a culture of scientific and professional integrity, as set out in the DOI Departmental Manual (305 DM 3).