BOEM Report 2016-077

Abundance Estimates of Ice-Associated Seals: Bering Sea Populations that Inhabit the Chukchi Sea During the Open-Water Period

FINAL REPORT

By
Peter L. Boveng, Michael F. Cameron, Paul B. Conn, and Erin E. Moreland
Marine Mammal Laboratory
Alaska Fisheries Science Center
7600 Sand Point Way NE
Seattle, WA 98115

Prepared for
U.S. Department of Interior
Bureau of Ocean Energy Management
Alaska Outer Continental Shelf Region
Environmental Studies Section
3801 Centerpoint Drive, Suite 500
Anchorage, AK 99503-5823

This study was funded by the U.S. Department of the Interior, Bureau of Ocean Energy Management (BOEM), through Interagency Agreement M12PG00017 with the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, as part of the MMS Alaska Environmental Studies Program.

DISCLAIMER

This report has been reviewed by the Bureau of Ocean Energy Management and is approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Service, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

REPORT AVAILABILITY

This document is available to the public through:
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161
FAX: (703) 605-6900

And at

http://www.boem.gov/Environmental-Stewardship/Environmental-Studies/Alaska-Region/Index.aspx

CITE REPORT AS

Boveng, P.L., M.F. Cameron, Paul B. Conn, and E.E. Moreland. 2017. Abundance Estimates of Ice-Associated Seals: Bering Sea Populations that Inhabit the Chukchi Sea During the Open-Water Period. Final Report. BOEM Report 2016-077. Bureau of Ocean Energy Management, Alaska Outer Continental Shelf Region, Anchorage, Alaska, USA. ix + 119 pp with appendices.

PROJECT ORGANIZATION

Peter Boveng; NMFS, Seattle, Washington:
Principal Investigator and report author

Michael Cameron; NMFS, Seattle, Washington:
Co-Principal Investigator and report author

Erin Moreland; NMFS, Seattle, Washington:
Project Coordinator and report author

Josh London; NMFS, Seattle, Washington:
Data Manager

Paul Conn, NMFS, Seattle, Washington:
Statistician and report author

Brett McClintock, NMFS, Seattle, Washington:
Statistician

Jay Ver Hoef; NMFS, Seattle, Washington:
Statistician

ABSTRACT

Bearded, spotted, ribbon and ringed seals are key components of Arctic marine ecosystems and they are important subsistence resources for northern coastal Alaska Native communities. Although these seals are protected under the Marine Mammal Protection Act (MMPA) and bearded and ringed seals have been listed as threatened[1] under the Endangered Species Act (ESA), no reliable, comprehensive abundance estimates are available for any of the species. Obtaining reliable abundance estimates for ice-associated seals is vital for developing sound plans for management, conservation, and responses to potential environmental impacts of oil and gas activities and climate change. The Bering Okhotsk Seal Surveys (BOSS) project addressed the most critical need for fundamental assessment data on ice-associated seals (also known as ice seals) in the Bering and Okhotsk Seas. Improved monitoring of ice seals is fundamental for the National Marine Fisheries Service (NMFS) to meet its management and regulatory mandates for stock assessments under the MMPA and extinction-risk assessments under the ESA.

The best way to estimate the abundances of ice-associated seals is to conduct aerial photographic or sightings surveys during the reproductive and molting period when the geographic structure of the population reflects the breeding structure and the greatest proportions of the populations are hauled out on the ice and are available to be seen. The distributions of these seals are broad and patchy and so surveys must cover large areas. Similarly, the extent, locations, and conditions of the sea ice habitat change so rapidly that surveys must be conducted in a relatively short period of time. The expense and logistic complexity of these surveys have been the primary impediments to acquisition of comprehensive and reliable estimates, though the complexity of the seals' behavior is also a factor.

Scientists at the Polar Ecosystems Program of NOAA's National Marine Mammal Laboratory, Alaska Fisheries Science Center, collaborated with colleagues from the State Research and Design Institute for the Fishing Fleet ("Giprorybflot") in Saint Petersburg, Russia, to conduct synoptic aerial surveys of ice-associated seals in the Bering and Okhotsk Seas. Conducting spring-time surveys in those areas will yield abundance estimates for the entire population of ribbon seals, and all but a small fraction of the spotted seal population. For bearded seals, the surveys included the large and important fraction of the population that overwinters and breeds in the Bering and Okhotsk Seas. The U.S. Bureau of Ocean

[1] Pending litigation.

Energy Management (BOEM) provided critical financial support in 2012 and 2013 to complete the U.S. surveys of the central and eastern Bering Sea. Surveys for the portions of the bearded and ringed seal populations that breed in the Chukchi and Beaufort seas will require separate and subsequent surveys, possibly with different seasonal timing.

Two years of survey effort were required to achieve adequate precision for abundance estimates and to ensure that sufficient periods of suitable weather occurred during survey periods. Aerial surveys were conducted in spring 2012 and 2013. In the United States and Russia combined, the teams flew more than 47,000 nautical miles (nmi) (87,000 km) of survey track. The completion of this project marks the largest survey of ice-associated seals ever completed and will provide the first comprehensive estimates of abundance for bearded, spotted, ribbon, and ringed seals in the Bering Sea and Sea of Okhotsk.

Analysis of full data sets from both years indicate substantial annual variation in numbers of seals in the U.S. portion of the Bering Sea during April and early May. Model-averaged estimates in 2012 were 240,000 spotted seals, 117,000 ribbon seals, 170,000 bearded seals, and 186,000 ringed seals. In contrast, the estimates for 2013 were lower: 163,000 spotted seals, 38,000 ribbon seals, 125,000 bearded seals, and 119,000 ringed seals. Seals may have been distributed farther to the west in 2013 (i.e. more in Russian waters), but there is substantial uncertainty about ribbon and spotted seal numbers in 2012 because weather constraints prohibited us from conducting many flights over the southwest portion of our study (at the ice edge) where densities were the highest. Based on the proportions of 165 seals instrumented with satellite tags in separate studies, we estimate that 69,000-101,000 (42%) of spotted seals and 6,000-25,000 (21%) of ribbon seals that occupy the eastern (U.S.) Bering Sea in spring used the Chukchi Sea during the summer, open-water period in 2013 and 2012, respectively.

KEY WORDS: bearded seal, *Erignathus barbatus*, ribbon seal, *Histriophoca fasciata*, ringed seal, *Phoca hispida*, spotted seal, *Phoca largha*, ice-associated seals, abundance, distribution, aerial survey, Bering Sea, Sea of Okhotsk, Chukchi Sea, Arctic

CONTENTS

PROJECT ORGANIZATION i

ABSTRACT ii

CONTENTS iv

LIST OF FIGURES vii

LIST OF TABLES ix

INTRODUCTION 1

Species Background 1

Spotted Seal (*Phoca largha*) 1

Bearded Seal (*Erignathus barbatus nauticus*) 2

Ribbon Seal (*Histriophoca fasciata*) 4

Ringed Seal (*Phoca hispida hispida*) 5

Previous Research 7

Using recent abundance surveys to estimate statistical power and effort required for Bering-Okhotsk seal surveys (BOSS) 7

Using seasonal movement studies to determine proportions of seals from the Bering Sea that use the Chukchi Sea during summer 10

PROJECT BACKGROUND 12

GOAL AND OBJECTIVES 13

Goal: 13

Objectives: 13

METHODS 13

Objective 1: Contract for a NOAA Twin Otter and a second long-range aircraft 14

Objective 2: Conduct surveys of ice-associated seals. 16

Objective 3: Effectively retrieve, manage, and process all imagery for analyses. 20

Thermal detection 22

Species Identification 23

Abundance estimation 24

Objective 4: For ribbon and spotted seals, which breed in the Bering Sea but are present seasonally in the Chukchi Sea, produce estimates of the numbers of individuals that use the Chukchi Sea during the open-water season. 25

RESULTS AND DISCUSSION 25

Objective 1: Contract for a NOAA Twin Otter and a second long-range aircraft 25

Objective 2: Conduct surveys of ice-associated seals. 26

Objective 3: Effectively retrieve, manage, and process all imagery for analyses. 27

Thermal Detection 27

Species Identification 28

Abundance Estimation 28

Objective 4: For ribbon and spotted seals, which breed in the Bering Sea but are present seasonally in the Chukchi Sea, produce estimates of the numbers of individuals that use the Chukchi Sea during the open-water season. 32

ACKNOWLEDGMENTS 32

PRESENTATIONS, SIGNIFICANT MEETINGS, AND PUBLICATIONS 33

Presentations and significant meetings 33

Publications 35

Agency reports 35

Peer-reviewed journals 35

LITERATURE CITED 36

APPENDIX 1. - Conn, P. B., J. M. Ver Hoef, B. T. McClintock, E. E. Moreland, J. M. London, M. F. Cameron, S. P. Dahle, and P. L. Boveng. 2014. Estimating multispecies abundance using automated detection

systems: ice-associated seals in the Bering Sea. Methods in Ecology and Evolution DOI 10.1111/2041-210X.12127 39

APPENDIX 2. - McClintock, B. T., E. E. Moreland, J. M. London, S. P. Dahle, G. M. Brady, E. L. Richmond, K. M. Yano, and P. L. Boveng. 2015. Quantitative assessment of species identification in aerial transect surveys for ice-associated seals. Marine Mammal Science 21:1057-1076. 54

APPENDIX 3. – Conn, P. B., D. S. Johnson, J. M. V. Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs 85:235-252. 75

APPENDIX 4. - Conn, P. B., D. S. Johnson, and P. L. Boveng. 2015. On extrapolating past the range of observed data when making statistical predictions in ecology. PLoS ONE 10(10): e0141416. doi:10.1371/journal.pone.0141416 77

APPENDIX 5. - Conn, P. B., J. T. Thorson, and D. S. Johnson. *In review*. Confronting preferential sampling in wildlife surveys: diagnosis and model-based triage. Methods in Ecology and Evolution 94

LIST OF FIGURES

Figure 1 Map of the approximate distribution of spotted seals, from Boveng et al. (2009). 2

Figure 2. Map of the approximate distribution of bearded seals, from Cameron et al. (2010). 3

Figure 3. Map of the approximate distribution of ribbon seals, from Boveng et al. (2013). 5

Figure 4. Map of the approximate distribution of ringed seals, from Kelly et al. (2010). 7

Figure 5. Map showing the abundance and distribution of seals observed during the Healy surveys, from Cameron and Boveng (2007). Counts of animals were summed over 5 nmi of track line and are represented by a pie chart. The diameter of the pie represents the total number of animals in the 5 nmi of track line, and the relative proportions of species seen are shown with different colored pie wedges. 8

Figure 6. Sea ice extent in the Bering Sea on April 21, 2007 (left) and May 16, 2007 (right). 9

Figure 7. Use of the Bering and Chukchi seas by ribbon (green) and spotted (orange) seals during the open water period. Ice extent is indicated by the dashed line. 11

Figure 8. Use of the Bering and Chukchi seas by ribbon (green) and spotted (orange) seals during the pupping and molt period. Ice extent is indicated by the dashed line. 12

Figure 9. NOAA Twin Otter (N56RF). This U.S. aircraft mostly surveyed the southern portion of the Bering Sea. 15

Figure 10. Aero Commander (N222ME). This U.S. aircraft mostly surveyed the northern portion of the Bering Sea. 15

Figure 11. Antonov AN-38-100. Russian survey aircraft that covered the Sea of Okhotsk and the Russian (western) portion of the Bering Sea. 16

Figure 12. Example of two adult bearded seals detected using thermal imagery. 17

Figure 13. NOAA Twin Otter belly-port camera setup: three FLIR SC645 thermal imagers (top), paired with three Canon 1Ds Mark III digital SLR cameras (bottom). 18

Figure 14. Aero Commander instrument setup. A Nikon D3X digital SLR camera paired with a FLIR SC645 thermal imagers, in each of two belly ports. 18

Figure 15. Russian survey aircraft camera setup showing a downward facing Nikon D3X and two oblique Nikon D300s. The cooled thermal imager, Malahit-M, is in a separate compartment of the aircraft. 19

Figure 16. Example of an animal (seal pup) likely to be missed during a manual review of SLR imagery, but easily detected using thermal imagery. 21

Figure 17. The initial US hot spot detection method utilized a temperature threshold applied to a plot of maximum pixel temperature per frame to identify which thermal frames to evaluate. Digital SLR images were matched using the timestamps and ice features to locate the source of the thermal signature. 22

Figure 18. The characteristic bands on the coats of ribbon seals are not necessarily clearly visible in an aerial image. The images on the top right and bottom right were taken with a Canon 1Ds Mark III fitted with a Zeiss 100 mm lens from 1000 ft during BOSS 2012. In the top right image, a species identification expert would likely rely on the clearly visible bands to conclude that the seal is certainly a ribbon seal. In the bottom right image, a species identification expert would rely on a combination of body shape, head size, flipper size and shape, and what could be one or more faint bands to conclude that the seal is likely a ribbon seal. 24

Figure 19. BOSS 2012 (pink) and 2013 (green) survey track lines in the Bering and Okhotsk seas covering over 90,000 km (56,000 miles) completed during the joint US-Russian survey effort. Water depths shallower than 500m are shown in light blue and the April, 2013 ice extent is in white. 26

Figure 20. Estimates of bearded, ribbon, ringed, and spotted seal abundance for three dates in the spring of 2012 in the eastern (U.S.) Bering Sea. Each map is bounded by the Bering Strait to the north, Alaska to the east, the U.S.-Russia Exclusive Economic Zone to the west, and maximum spring sea ice extent to the south. 30

Figure 21. Estimates of bearded, ribbon, ringed, and spotted seal abundance for three dates in the spring of 2013 in the eastern (U.S.) Bering Sea. Each map is bounded by the Bering Strait to the north, Alaska to the east, the U.S.-Russia Exclusive Economic Zone to the west, and maximum spring sea ice extent to the south. Note the scale is the same as Figure 20 to permit proper comparison. 31

LIST OF TABLES

Table 1. Instrument and camera resolution of US and Russian BOSS 2012 and 2013 survey efforts. 19

Table 2. Estimates of ice-associated seal abundance (thousands) in the eastern Bering Sea (U.S. airspace) in different years by different analysis methods. We present model-based standard errors for each model run, as well as model-averaged standard errors for combined estimates. The abbreviation Sp-T is used to indicate a spatio-temporal model; RE indicates random effects. We suggest basing inference on the combined, model-averaged estimates rather than results from individual model runs. 29

INTRODUCTION

Bearded, spotted and ribbon seals are key components of Arctic marine ecosystems and they have long been essential resources in the subsistence economies of coastal Arctic communities. Although these seals are protected under the Marine Mammal Protection Act (MMPA) and bearded and ringed seals have been listed as threatened under the Endangered Species Act (ESA), no reliable, comprehensive abundance estimates are available for any of the species. Obtaining reliable abundance estimates for ice-associated seals is vital for developing sound plans for management, conservation, and responses to potential environmental impacts of oil and gas activities and climate change.

Species Background

Spotted Seal (*Phoca largha*)

Spotted seals are distributed along the continental shelves of the Bering, Chukchi, and Beaufort Seas, and from the Sea of Okhotsk south to the western Sea of Japan and northern Yellow Sea (Shaughnessy and Fay 1977) Figure 1).

Satellite-tagging studies showed that seals tagged during summer in the northeastern Chukchi Sea moved south in October and passed through the Bering Strait in November. Seals overwintered in the Bering Sea along the ice edge and made east-west movements along the edge (Lowry et al. 1998). During spring, spotted seals tend to prefer small floes (i.e., < 20 m in diameter), and inhabit mainly the southern margin of the ice in areas where the water depth does not exceed 200 m, and move to coastal habitats after the retreat of the sea ice (Fay 1974, Shaughnessy and Fay 1977, Lowry et al. 2000, Simpkins et al. 2003). In summer and fall, spotted seals use coastal haul-out sites regularly (Frost et al. 1977, Lowry et al. 1998), and may be found as far north as 69-72°N in the Chukchi and Beaufort Seas (Porsild 1945, Shaughnessy and Fay 1977). To the south, along the west coast of Alaska, spotted seals are known to occur around the Pribilof Islands, Bristol Bay, and the eastern Aleutian Islands.

In 2007, the National Marine Mammal Laboratory (NMML) conducted aerial surveys of the central and eastern Bering Sea pack ice using helicopters based aboard the U.S. Coast Guard icebreaker, Healy (Ver Hoef et al. 2014). Frequencies of sightings data and information on ice distribution and the timings of seal haul-out behavior were analyzed to develop a population estimate of 141,479 (95% CI = 92,769-321,882) spotted seals in the areas surveyed within the eastern and central Bering Sea.

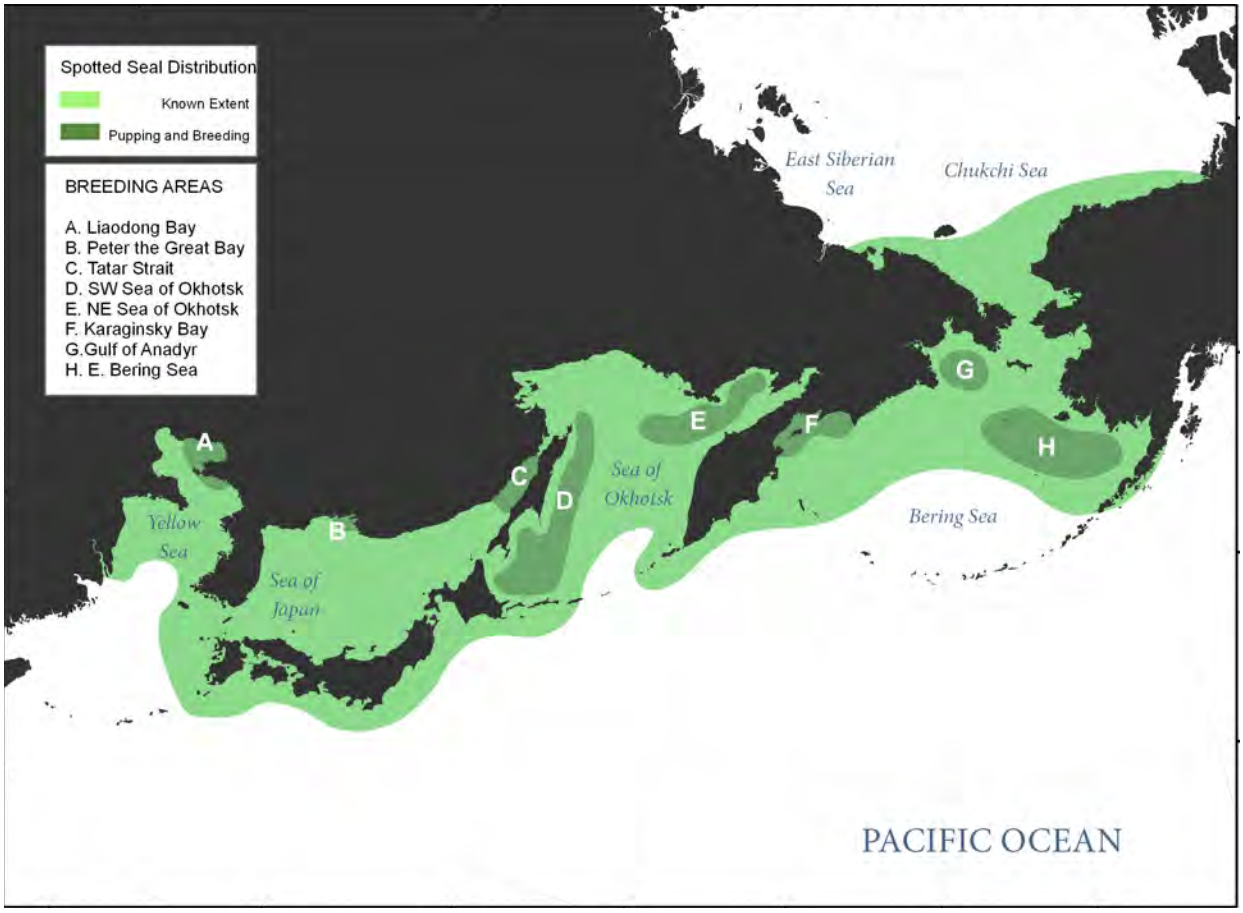

Figure 1 Map of the approximate distribution of spotted seals, from Boveng et al. (2009).

Bearded Seal (*Erignathus barbatus nauticus*)

Bearded seals are a boreo-arctic species with a circumpolar distribution (Fedoseev 1965, Johnson et al. 1966, Burns 1967, Burns and Frost 1979, Burns 1981a, Smith 1981, Kelly 1988a). Their normal range (Figure 2), extends from the Arctic Ocean (85°N) south to Sakhalin Island (45°N) in the Pacific, and south to Hudson Bay (55°N) in the Atlantic (Allen 1880, Ognev 1935, King 1983). Bearded seals inhabit the seasonally ice-covered seas of the Northern Hemisphere where they whelp and rear their pups, and molt their coats on the ice in the spring and early summer. Bearded seals feed primarily on benthic organisms, including epifaunal and infaunal invertebrates, and demersal fishes; therefore, they are closely linked to areas where the seafloor is shallow (i.e., less than 200 m). Spring surveys conducted in 1999 and 2000 along the Alaskan coast, and 2001 in the central Bering Sea, indicated that bearded seals in Alaska tend to prefer areas of between 70% and 90% sea-ice coverage (Simpkins et al. 2003), and were typically more abundant 20-100 nmi from shore than within 20 nmi of shore (Bengtson et al.

2005), with the exception of high concentrations nearshore to the south of Kivalina. Many of the seals that winter in the Bering Sea move north through the Bering Strait from late April through June, and

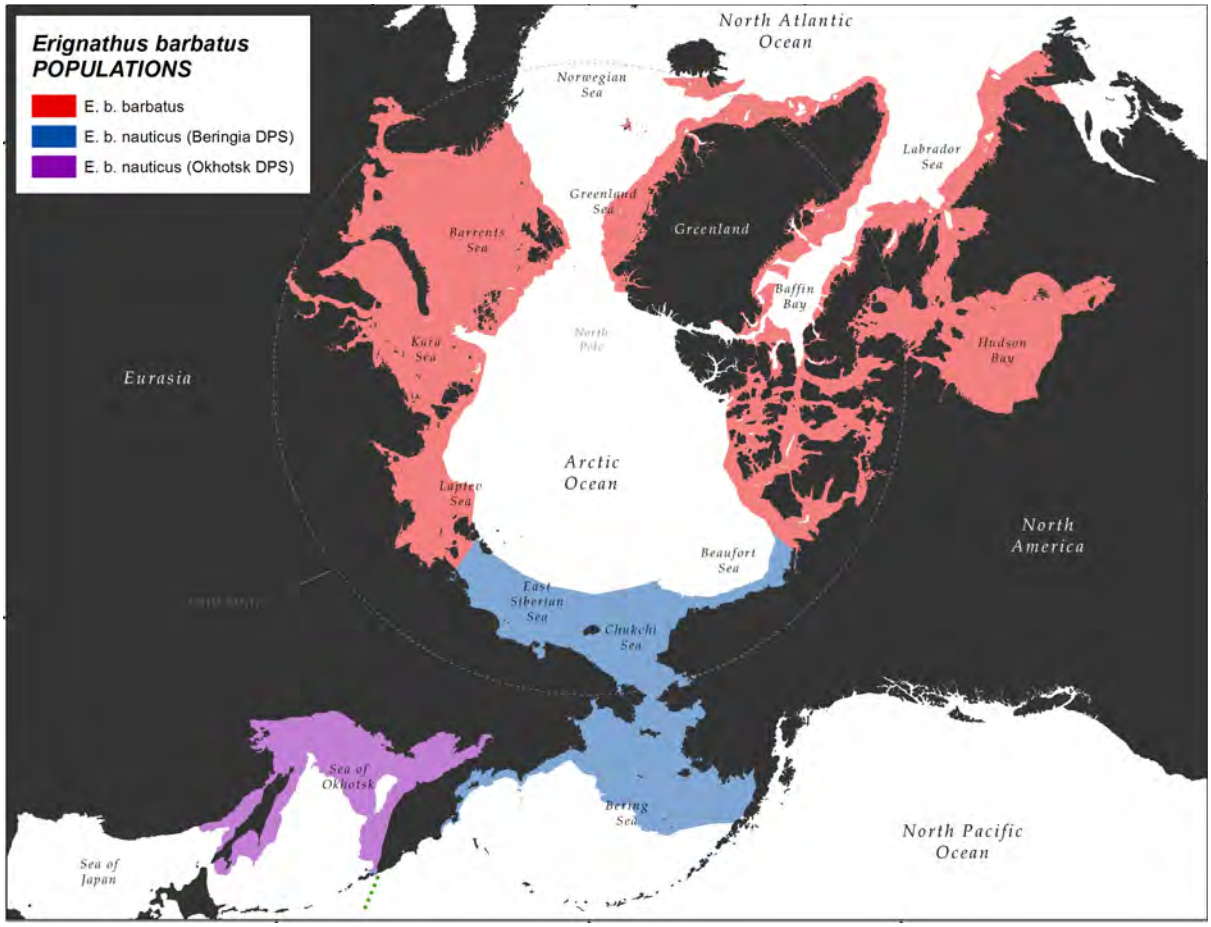

Figure 2. Map of the approximate distribution of bearded seals, from Cameron et al. (2010).

spend the summer along the ice edge in the Chukchi Sea (Burns 1967, Burns 1981a). The overall summer distribution is quite broad, with seals rarely hauled out on land, and some seals may not follow the ice northward but remain in open-water areas of the Bering and Chukchi Seas (Burns 1981a, Nelson 1981, Smith 1981). An unknown proportion of the population moves southward from the Chukchi Sea in late fall and winter, and Burns (1967) noted a movement of bearded seals away from shore during that season as well.

A reliable population estimate for this stock is currently considered not available. However, based on studies by Ver Hoef et al. (2010), Fedoseev (2000), and Bengtson et al. (2005), Cameron et al. (2010) estimated about 125,000 bearded seals in the Bering Sea and 27,000 bearded seals in the Chukchi Sea.

Cameron et al. (2010) did not determine population estimates for the East Siberian and Beaufort Seas, but they did estimate that the Alaska stock contained approximately 155,000 bearded seals.

Ribbon Seal (*Histriophoca fasciata*)

Ribbon seals inhabit the North Pacific Ocean and adjacent parts of the Arctic Ocean. In Alaska waters, ribbon seals are found in the open sea, on the pack ice, and only rarely on shore-fast ice (Kelly 1988b). They range northward from Bristol Bay in the Bering Sea into the Chukchi and western Beaufort Seas (Figure 3). From late March to early May, ribbon seals inhabit the Bering Sea ice front (Burns 1970, Burns 1981b, Braham et al. 1984). They are most abundant in the northern part of the ice front in the central and western parts of the Bering Sea (Burns 1970, Burns et al. 1981).

As the ice recedes in May to mid-July the seals move farther to the north in the Bering Sea, where they haul out on the receding ice edge and remnant ice (Burns 1970, Burns 1981b, Burns et al. 1981). There is little known about the range of ribbon seals during the rest of the year. Kelly (1988b), suggested that many ribbon seals migrate into the Chukchi Sea for the summer. Satellite tag data from 2005 and 2007 indicate that ribbon seals disperse widely. Ten seals tagged in 2005 near the eastern coast of Kamchatka spent the summer and fall throughout the Bering Sea and Aleutian Islands; eight of the 26 seals tagged in 2007 in the central Bering Sea moved to the Bering Strait, Chukchi Sea, or Arctic Basin as the seasonal ice retreated (Boveng et al. 2008).

At present, reliable data on trends in population abundance for the Alaska stock of ribbon seals are unavailable. Although the current population trend is unknown, results from recent aerial surveys and reports from Alaska Native hunters suggest that there has not been a dramatic decline in the recent past. (Boveng et al. 2013). This stock is thought to occupy its entire historically-observed range (Boveng et al. 2013).

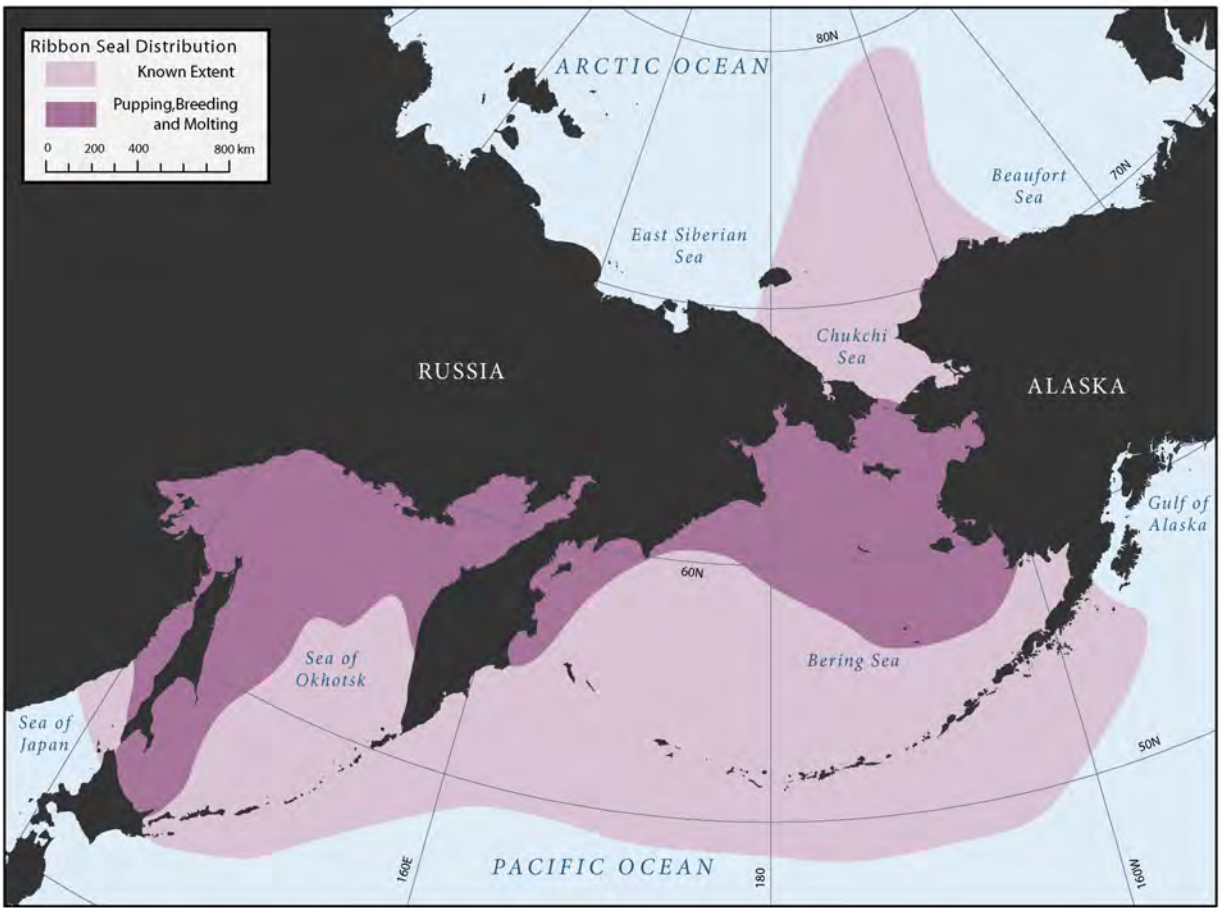

Figure 3. Map of the approximate distribution of ribbon seals, from Boveng et al. (2013).

Ringed Seal (*Phoca hispida hispida*)

Ringed seals have a circumpolar distribution and are found in all seasonally ice-covered seas of the Northern Hemisphere. The Alaska stock of ringed seals is considered the portion of the subspecies *Phoca hispida hispida* that occurs within the U.S. Exclusive Economic Zone (EEZ) of the Beaufort, Chukchi, and Bering Seas (Figure 4).

Throughout their range, ringed seals have an affinity for ice-covered waters and are well adapted to occupying both shore-fast and pack ice (Kelly 1988c). They remain in contact with ice most of the year and use it as a platform for pupping and nursing in under-snow lairs in late winter to early spring, for molting in late spring to early summer, and for resting at other times of the year. In Alaskan waters, during winter and early spring when sea ice is at its maximal extent, ringed seals are abundant in the northern Bering Sea, Norton and Kotzebue Sounds, and throughout the Chukchi and Beaufort Seas. They

occur as far south as Bristol Bay in years of extensive ice coverage but generally are not abundant south of Norton Sound except in nearshore areas (Frost 1985). Although details of their seasonal movements have not been adequately documented, it is thought that most ringed seals that winter in the Bering and Chukchi Seas migrate north in spring as the seasonal ice melts and retreats (Burns 1970) and historically spend summer in the pack ice of the northern Chukchi and Beaufort seas, as well as in nearshore ice remnants in the Beaufort Sea (Frost 1985).

Ringed seal population surveys in Alaska have used various methods and assumptions, had incomplete coverage of their habitats and range, and were conducted more than a decade ago; therefore, current, comprehensive, and reliable abundance estimates or trends for the Alaska stock are not available. Frost et al. (2004) conducted surveys within 40 km of shore in the Alaskan Beaufort Sea during May-June 1996-1999, and observed ringed seal densities ranging from 0.81 seals/km^2 in 1996 to 1.17 seals/km^2 in 1999. Bengtson et al. (2005) conducted surveys in the Alaskan Chukchi Sea during May-June 1999 and 2000. Population estimates were derived from observed densities corrected for availability bias using a haul-out model from 6 tagged seals. Ringed seal abundance estimates for the entire survey area were 252,488 (SE = 47,204) in 1999 and 208,857 (SE = 25,502) in 2000. Using the estimates from Frost et al. (2004) and Bengtson et al. (2005), Kelly et al. (2010) calculated a total population estimate for in the Alaskan Chukchi and Beaufort Seas regions of at least 300,000 ringed seals, which is likely an underestimate.

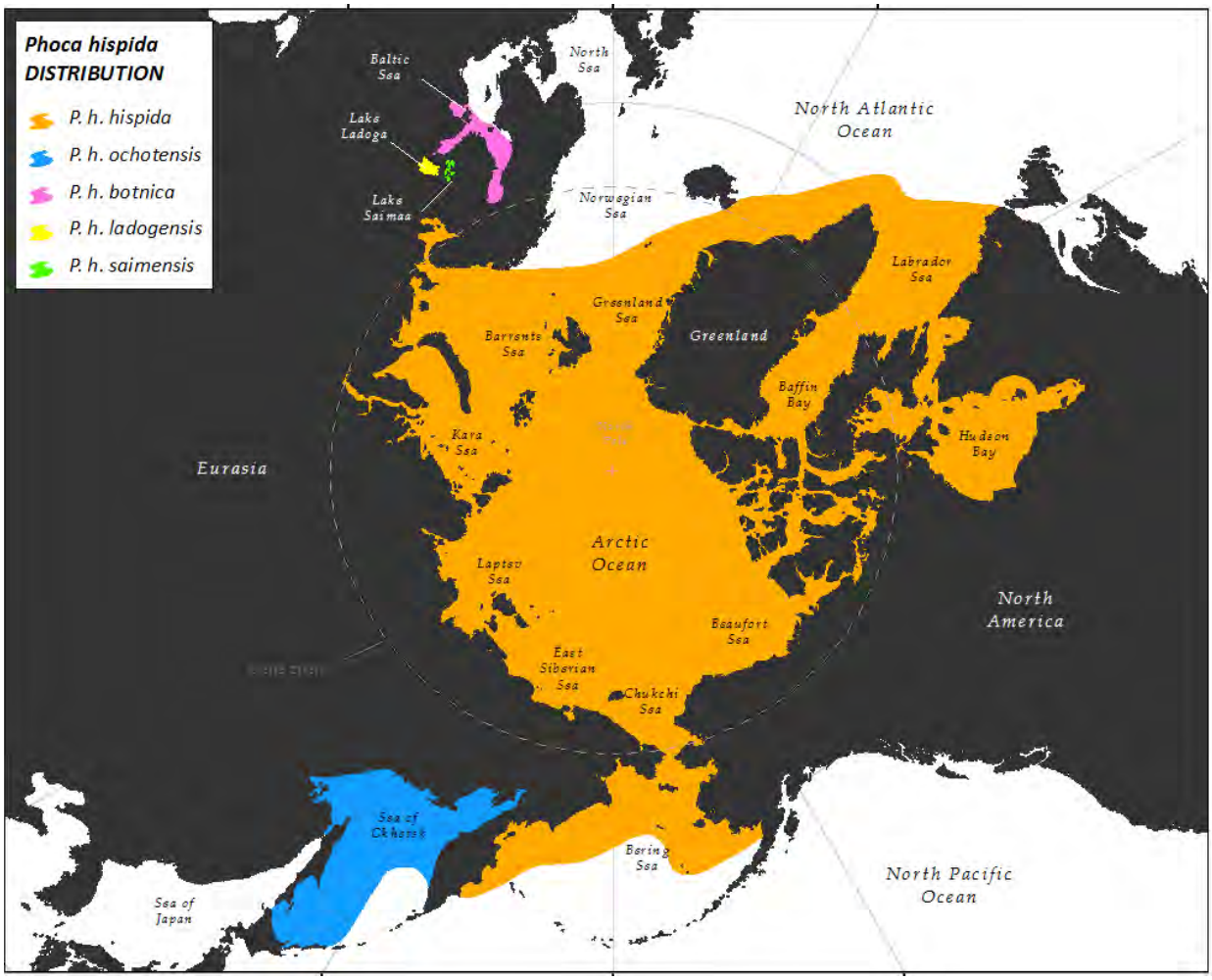

Figure 4. Map of the approximate distribution of ringed seals, from Kelly et al. (2010).

Previous Research

Using recent abundance surveys to estimate statistical power and effort required for Bering-Okhotsk seal surveys (BOSS)

Between 13 April and 26 May 2007, NMML conducted population surveys of spotted, ribbon, and bearded seals in an area of 81,600 km^2 in the central Bering Sea. Researchers surveyed 2,748 km^2 with line transect methods from helicopters deployed from the U.S. Coast Guard icebreaker *Healy* (Figure 5; (Cameron and Boveng 2007). During that period, the sea ice conditions and the spatial distributions of seals changed dramatically (Figure 6), which greatly complicated abundance estimation.

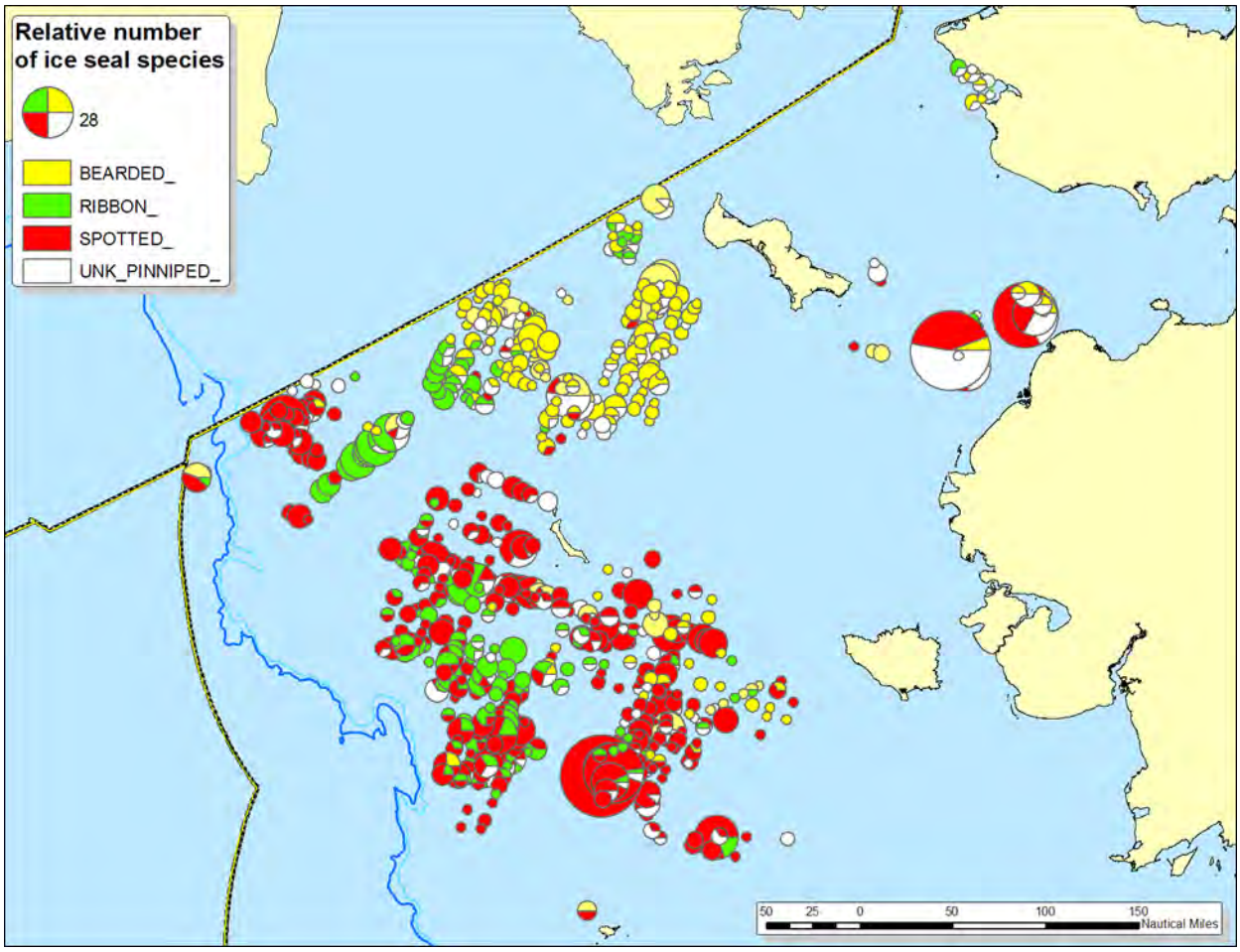

Figure 5. Map showing the abundance and distribution of seals observed during the Healy surveys, from Cameron and Boveng (2007). Counts of animals were summed over 5 nmi of track line and are represented by a pie chart. The diameter of the pie represents the total number of animals in the 5 nmi of track line, and the relative proportions of species seen are shown with different colored pie wedges.

Using methods described by Ver Hoef et al. (2014), we accounted for: 1) the incomplete availability of seals, with a generalized linear mixed model fitted to seal haul-out timelines obtained by satellite telemetry, 2) the incomplete detection, through standard distance-sampling methods along with a double-observer model, and 3) the shifting ice and seal movements, with a spatially-autocorrelated regression model using remotely sensed sea ice concentrations to predict abundance for each survey date; all were combined in a hierarchical model to obtain stable estimates of population abundances for this region, even though the populations were spatially dynamic. Within this study area, spotted seals were most abundant (233,700, 95% CI 137,300–793,100), followed by bearded seals (61,800, 95% CI 34,900–171,600) and ribbon seals (61,100, 95% CI 35,200–189,300).

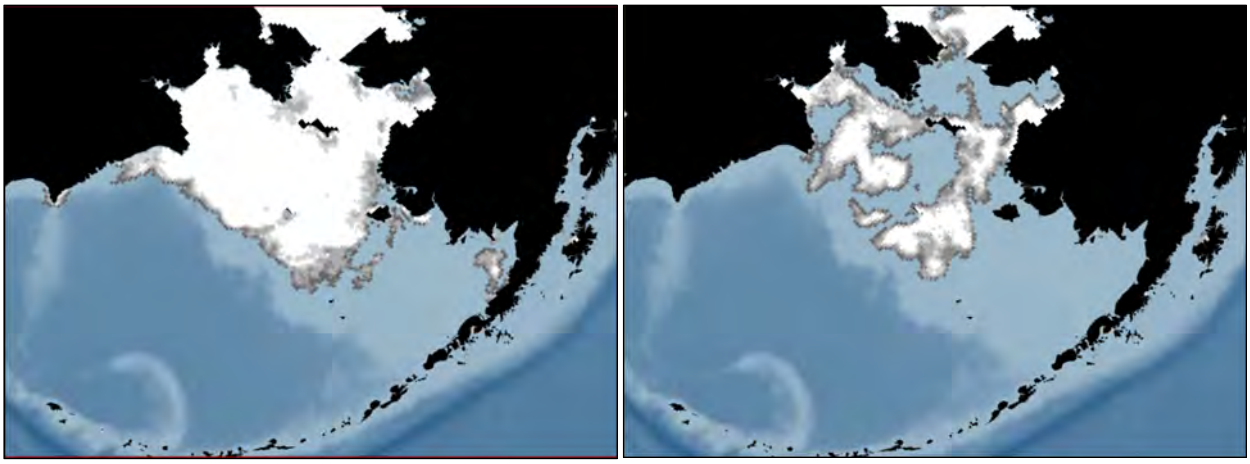

Figure 6. Sea ice extent in the Bering Sea on April 21, 2007 (left) and May 16, 2007 (right).

These surveys and estimates were for only a portion of the ranges of these species, and were conducted in part to provide data necessary to design and plan the 2012 and 2013 synoptic surveys in the present project. For example, we assumed that seals are spatially distributed as an overdispersed Poisson process with variance = $\theta\lambda\mathrm{B}$, where θ is the overdispersion factor (estimated from the 2007 surveys as: bearded=18; spotted=79; ribbon=99) λ is density (bearded=0.2; spotted=0.5; ribbon=0.2); and B is the area to be surveyed. Given the strip width of our survey sensors W (about 360 m), we calculated the survey track length *L*, required to achieve a coefficient of variation (CV) of 0.2 in our abundance estimates for U.S. waters using:

$$L = \frac{\theta}{W(CV)^2 \lambda}$$

For bearded seals, we estimated required track lengths of 6,250 km; for spotted seals, 10,972 km; and for ribbon seals, 34,375 km. Averaging these three estimates yields 16,667 km, or about 17000 km of survey track length as our goal for adequate precision in each of the U.S. and Russian surveys.

We recognized however, that this goal for survey transect length was fairly conservative. The density and overdispersion factors from previous experience were based on distance sampling, a haul-out correction and spatial modeling, and included all of the imprecision in combining them for estimating abundance from those data. The 2012 and 2013 surveys were to be based on direct counts from aerial photos, which are more precise. Together with the increased precision in the haul-out correction factors from additional satellite-tag records and the reduction in uncertainty of the overall population estimate from the combined and simultaneous effort of the U.S. and Russian surveys, we estimated a CV of about 0.1 for the total estimate of spotted seals.

We chose a CV of 0.1 in order to detect a 35% change in population abundance with 90% confidence at $\alpha = 0.1$. Surveys of this type are expensive and logistically and technically challenging. As such, one could expect that they will only be conducted every decade or more. A population change of 35% over 10 years would require an annual decline of only 4-5% per year. Detecting such a change is critically important in the context of habitat changes resulting from climate change, oil and gas development, and other impacts.

Using seasonal movement studies to determine proportions of seals from the Bering Sea that use the Chukchi Sea during summer

NMML deployed 156 satellite tags on ribbon and spotted seals (67 adults, 38 sub-adults and 51 young of the year) in the Bering Sea between 2007 and 2010. Results from these tagging efforts provided a basis for quantifying the use of habitats north of the Bering Strait by both species. Chukchi Sea habitat use occurs mostly during the open water period from July through October (Figure 7). We estimated that during this time, spotted and ribbon seals of the Bering Sea population spend 26.1% and 9.5% of their time, respectively, north of the Bering Strait. Based on proportions of seals in our tagged sample, about 42% of Bering Sea spotted seals and 21% of Bering Sea ribbon seals use the Chukchi Sea during the summer, open-water period. The consistent use of habitats within the Chukchi Sea by both of these

species suggests they are likely to encounter activities related to oil and gas extraction and exploration. Because ribbon and spotted seals concentrate in the Bering Sea in the spring (March-May), using sea ice as a platform for giving birth, nursing and breeding (Figure 8), springtime aerial surveys in this region are the most efficient means for estimating abundance. Therefore, abundance estimates obtained during spring in the Bering Sea, and proportions of the Bering Sea populations that move north through Bering Strait for the open-water season, form a basis for computing more informative estimates of the numbers of ribbon and spotted seals that rely on the Chukchi Sea during the open water season, and that may interact with oil and gas activities in that region By combining abundance estimates from the BOSS surveys with results from ongoing analysis of seal movements and behavior, we wished to provide more informative estimates of the numbers of ribbon and spotted seals that rely on the Chukchi Sea during the open water season, and that may interact with oil and gas activities in that region.

For bearded and ringed seals, the satellite-telemetry data are not currently sufficient for estimating the proportions of the Bering Sea populations that summer in the Chukchi Sea. However, these proportions are typically thought to be high (Cameron et al. 2010, Kelly et al. 2010), so as a first approximation the entire Bering Sea populations could be considered present in the Chukchi Sea during part or all of the open-water season.

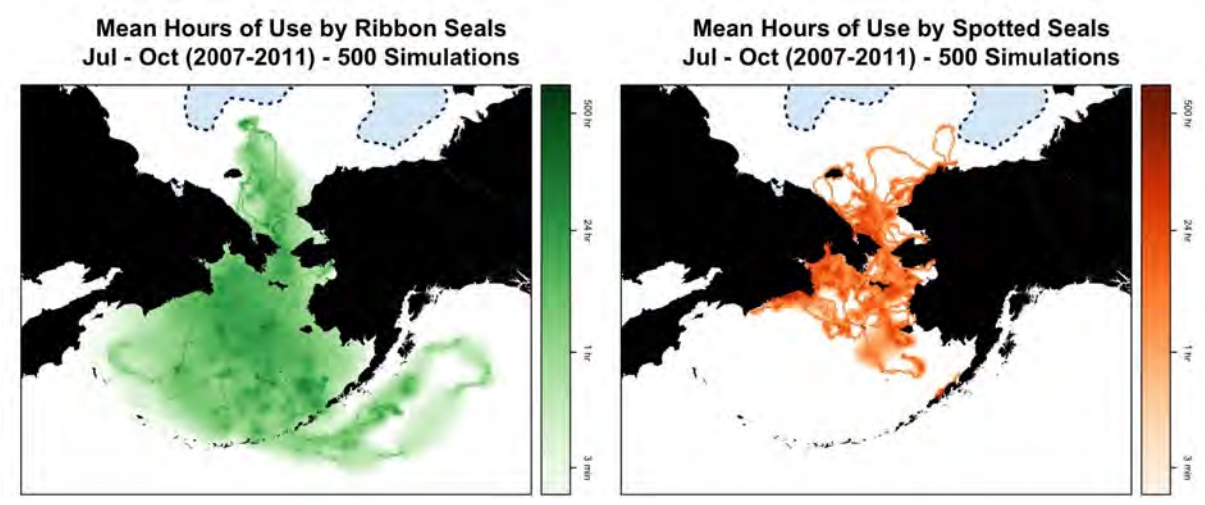

Figure 7. Use of the Bering and Chukchi seas by ribbon (green) and spotted (orange) seals during the open water period. Ice extent is indicated by the dashed line.

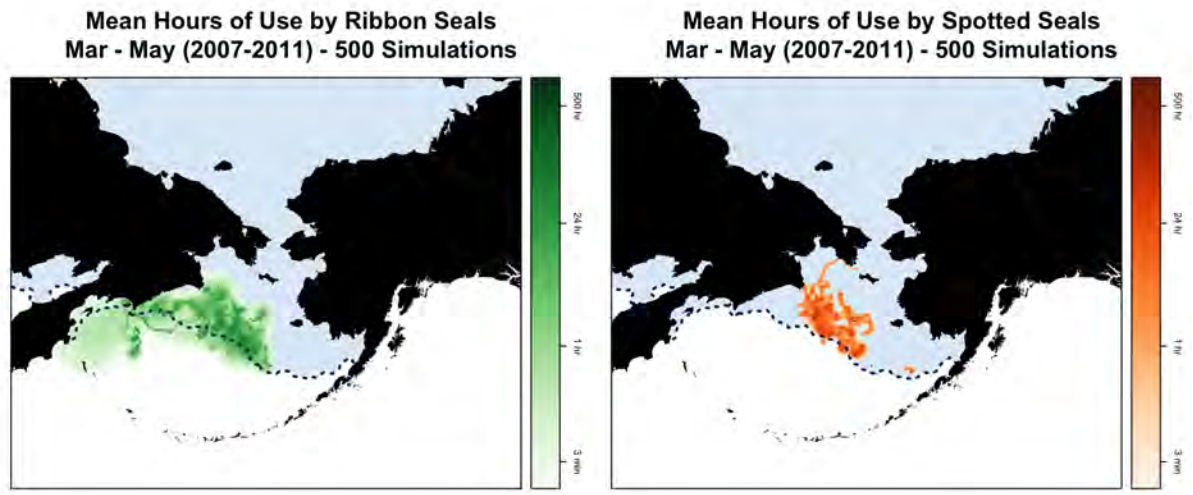

Figure 8. Use of the Bering and Chukchi seas by ribbon (green) and spotted (orange) seals during the pupping and molt period. Ice extent is indicated by the dashed line.

PROJECT BACKGROUND

Interagency Agreement (IAA) M12PG00017 is part of an overall project supported by NOAA and BOEM funds. The BOEM funds have been provided as charter aircraft support under a separate IAA between BOEM and the U.S. Department of Interior, Office of Aircraft Services, Aviation Management Division, and as reimbursable funds to NOAA under this IAA. This report includes activities conducted and supported under all three funding sources, to provide sufficient context for the more limited scope of this IAA.

Ice-associated seal surveys were conducted by the NMML in the spring of 2012 and 2013 (Moreland et al. 2013). The surveys are a component of the Alaska Fisheries Science Center's Loss of Sea Ice (LOSI) Program and are part of a joint US-Russian effort (BOSS: Bering-Okhotsk Seal Surveys) to estimate the abundance of bearded, ringed, spotted, and ribbon seals in the Bering Sea and the Sea of Okhotsk. In 2010-2011, LOSI funding was used to support planning and preparation for joint U.S.-Russia surveys, including support for aircraft charters in Russia. Government-wide budget reductions in 2012 derailed plans to use LOSI funds in 2012-2013 to procure aircraft charters for the U.S. survey area of this international effort. NOAA had already committed significant funds to the U.S. surveys by hiring personnel and purchasing necessary supplies and equipment (e.g., cameras and thermal sensors). In consideration of overlapping goals and needs for data between NOAA and BOEM, NMML requested

assistance from BOEM to support aircraft contracts and enable completion of the U.S. portion of this program.

The NMML obtained all necessary research permits. The NMML conducted the aerial surveys under Marine Mammal Protection Act (MMPA) Scientific Research Permits permit 15126, issued in 2010 and valid until 2015. The satellite telemetry studies used for estimating availability were conducted under MMPA Permits 15126, 782-1765, 782-1676 and 358-1787. NMML research is subject to review and oversight by an animal care and use committee (IACUC) to ensure compliance with the Animal Welfare Act. This project was reviewed and authorized under IACUC Number A/NW 2010-3 through 2013.

GOAL AND OBJECTIVES

Goal:

The long-term goal of the BOSS program is to calculate abundance estimates for bearded, spotted, ribbon, and ringed seals in the Bering and Okhotsk Seas. The specific objectives of the BOEM-funded portion of the program are below.

Objectives:

1. Contract for a NOAA Twin Otter and a second longer-range aircraft to conduct surveys of the central and eastern Bering Sea shelf in April and May of 2012 and 2013.
2. Conduct surveys of ice-associated seals using high-resolution digital photographic and thermal imaging sensors, with the coverage required to obtain annual seal abundance estimates with adequate precision (CV=0.2).
3. Effectively retrieve, manage, and process the sensors' imagery for analyses.
4. For ribbon and spotted seals, which breed in the Bering Sea but are present seasonally in the Chukchi Sea, produce estimates of the numbers of individuals that use the Chukchi Sea during the open-water season.

METHODS

The best way to estimate the abundances of ice-associated seals is to conduct aerial photographic and sightings surveys during the reproductive and molting period when the geographic structure of the population reflects the breeding structure and the greatest proportions of the populations are hauled out on the ice and are available to be seen. The distributions of these seals are broad and patchy and so

surveys must cover large areas. Similarly, the extent, locations, and conditions of the sea ice habitat change so rapidly that any surveys must be conducted in a relatively short period of time.

Scientists at the Polar Ecosystems Program of NOAA's National Marine Mammal Laboratory (NMML), Alaska Fisheries Science Center (AFSC), collaborated with colleagues from the State Research and Design Institute for Fishing Fleet ("Giprorybflot") in Saint Petersburg, Russia, to design and conduct synoptic aerial surveys of ice-associated seals in the Bering and Okhotsk Seas. Conducting spring-time surveys in those areas yield abundance estimates for the entire population of ribbon seals, and all but a small fraction of the spotted seal population. For bearded seals, the surveys included the large and important fraction of the population that overwinters and breeds in the Bering and Okhotsk Seas. Though some ringed seals were expected to be observed outside of their snow lairs, it was assumed that the visible proportion of the actual population would be low, so that without appropriate correction factors, estimates of their abundance would be negatively biased. Two years of survey effort were required to ensure adequate precision for abundance estimates and to ensure that sufficient periods of suitable weather occurred during survey periods.

Objective 1: Contract for a NOAA Twin Otter and a second long-range aircraft

The NMML's Bering Sea pack ice surveys for ice-associated seals have progressed from ship-based helicopter flights reliant on observer-collected data in 2007 and 2008 to instrument-only surveys on long-range, fixed-wing aircraft. Using long-range, fixed-wing aircraft makes it possible to achieve greater coverage of the survey area in a shorter period of time, improving abundance estimates by minimizing the change in sea ice habitat during the survey window. Two aircraft operating simultaneously would enable us to complete surveys of the US portion of the Bering Sea in less time. We determined that a NOAA-owned Twin Otter DHC-6 aircraft (Figure 9) and a longer-range Aero Commander AC-690 owned and operated by Clearwater Air (Figure 10), would be the best platforms for our surveys. Both aircraft have a long and safe history of outstanding support for science operations in Alaska. They have long ranges and existing belly-mounted open-air ports required for our imaging sensors. For comparison, the Russian team used a Russian-made survey aircraft, the Antonov AN-38-100 (Figure 11).

Figure 9. NOAA Twin Otter (N56RF). This U.S. aircraft mostly surveyed the southern portion of the Bering Sea.

Figure 10. Aero Commander (N222ME). This U.S. aircraft mostly surveyed the northern portion of the Bering Sea.

Figure 11. Antonov AN-38-100. Russian survey aircraft that covered the Sea of Okhotsk and the Russian (western) portion of the Bering Sea.

Objective 2: Conduct surveys of ice-associated seals.

U.S. flights were planned for 4-8 hours per day at an altitude of 1,000 ft (300 m) to maximize the area surveyed while maintaining the required minimum ground resolution requirements (i.e., ~20 cm/pixel for thermal detection and ~2cm/pixel for species identification), and minimizing the chance of disturbance to seals and other wildlife. Flights originated from airports in Nome, Bethel, and St. Paul Island, Alaska, while also using an airstrip in Gambell, on St. Lawrence Island as a refueling point to reach the most remote areas of sea ice in the central Bering Sea. The Russian team's Sea of Okhotsk surveys began from Khabarovsk in Tatar Strait and worked their way through Shelikhov Bay and into Karaginsky Bay using airports at Yuzhno-Sakhalinsk, Nikolaevsk, Nogliki, Okha, Tukchi, Okhotsk and Magadan. Surveys of the western Bering Sea used airstrips at Ossora, Tilichiki, Beringovsky, Anadyr and Provideniya on Russia's Kamchatka Peninsula.

Advanced thermal-imaging technology was used on both the U.S. and Russian survey aircraft to detect the warm bodies of seals against the background of the cold sea ice (Figure 12). High-resolution digital images are used to identify the species of seals detected by the thermal imagers.

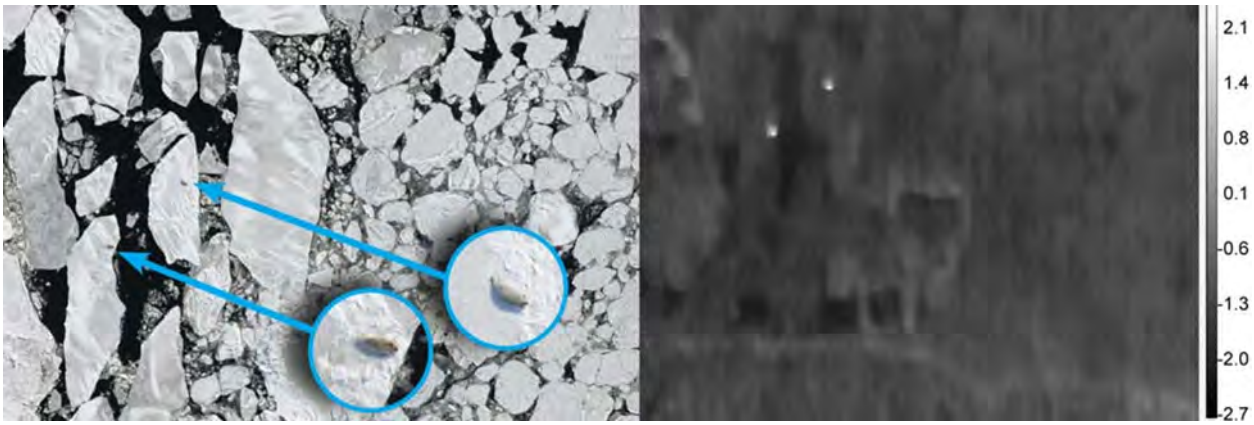

Figure 12. Example of two adult bearded seals detected using thermal imagery.

A NOAA Twin Otter (N56RF) aircraft housed three FLIR SC645 thermal imagers that continuously recorded data in the 7.5-13.0 µm wavelength. Each thermal imager was paired with a Canon Mark III 1Ds digital single-lens reflex (SLR) camera fitted with a 100-mm Zeiss lens. All six instruments were mounted in an open-air belly port (Figure 13). The center pair was set to a 0° angle while the port and starboard pairs were set to +/- 25.5° to avoid left and right overlap. The combined thermal swath width was approximately 1,500 ft (470 m), at an altitude of 1,000 ft. A contracted Aero Commander aircraft (Figure 14) carried two sets of paired thermal imagers (SC645) and digital SLR cameras (Nikon D3X) set to 12.5° angles outward and surveyed a swath width of approximately 900 ft (280 m). The thermal imagers were controlled by FLIR software (ExaminIR) and set to record at a rate of 4 frames per second. The SLR cameras were set to fire at their maximum write speed of approximately 1 image per second. The Russian aircraft carried a large, cooled thermal imager, Malakhit-M, which was paired with three fixed, digital SLR cameras fitted with 50-mm lenses (Figure 15). On board observers also collected images with hand-held SLR cameras with zoom lenses. Aircraft and instrument details are provided in Table 1.

Figure 13. NOAA Twin Otter belly-port camera setup: three FLIR SC645 thermal imagers (top), paired with three Canon 1Ds Mark III digital SLR cameras (bottom).

Figure 14. Aero Commander instrument setup. A Nikon D3X digital SLR camera paired with a FLIR SC645 thermal imagers, in each of two belly ports.

Figure 15. Russian survey aircraft camera setup showing a downward facing Nikon D3X and two oblique Nikon D300s. The cooled thermal imager, Malahit-M, is in a separate compartment of the aircraft.

Table 1. Instrument and camera resolution of US and Russian BOSS 2012 and 2013 survey efforts.

	Russian Surveys	**US Surveys**	
Aircraft	Antonov AH-38-100	NOAA Twin Otter DHC-	Aero Commander AC-
Thermal Imager	Malahit-M	FLIR SC645	FLIR SC645
Digital SLR Cameras	Nikon D800, D300,	Canon 1Ds Mark III	Nikon D3X
SLR Lens	50mm	100mm	100mm
Survey Altitude	200-250m	300m	300m
Thermal Swath	500m	470m	280m
SLR Swath	500m	390m	237m
SLR Resolution	2-7 cm/pixel	1.9-2.1 cm/pixel	2.0-2.5 cm/pixel

Objective 3: Effectively retrieve, manage, and process all imagery for analyses.

The BOSS project capitalized on recent advances in technology by pairing thermal and high-resolution digital SLR imagery to detect warm seal bodies hauled out on cold sea ice. Compared to observer-based surveys, thermal-detection surveys require fewer personnel and less post-survey processing time, can be flown at a higher altitude (reducing disturbance of seals), and yield higher rates of seal detection. Thermal detection is particularly useful in detecting animals that are well camouflaged (Figure 16). High-resolution digital imagery is being used to identify seal species and differentiate hot spots generated by seals from anomalous thermal signals (false positives caused by melt pools, dirty ice, etc.). Using imagery also allows greater flexibility to explore potential sources of variability such as detection probability and examine species misclassification rates.

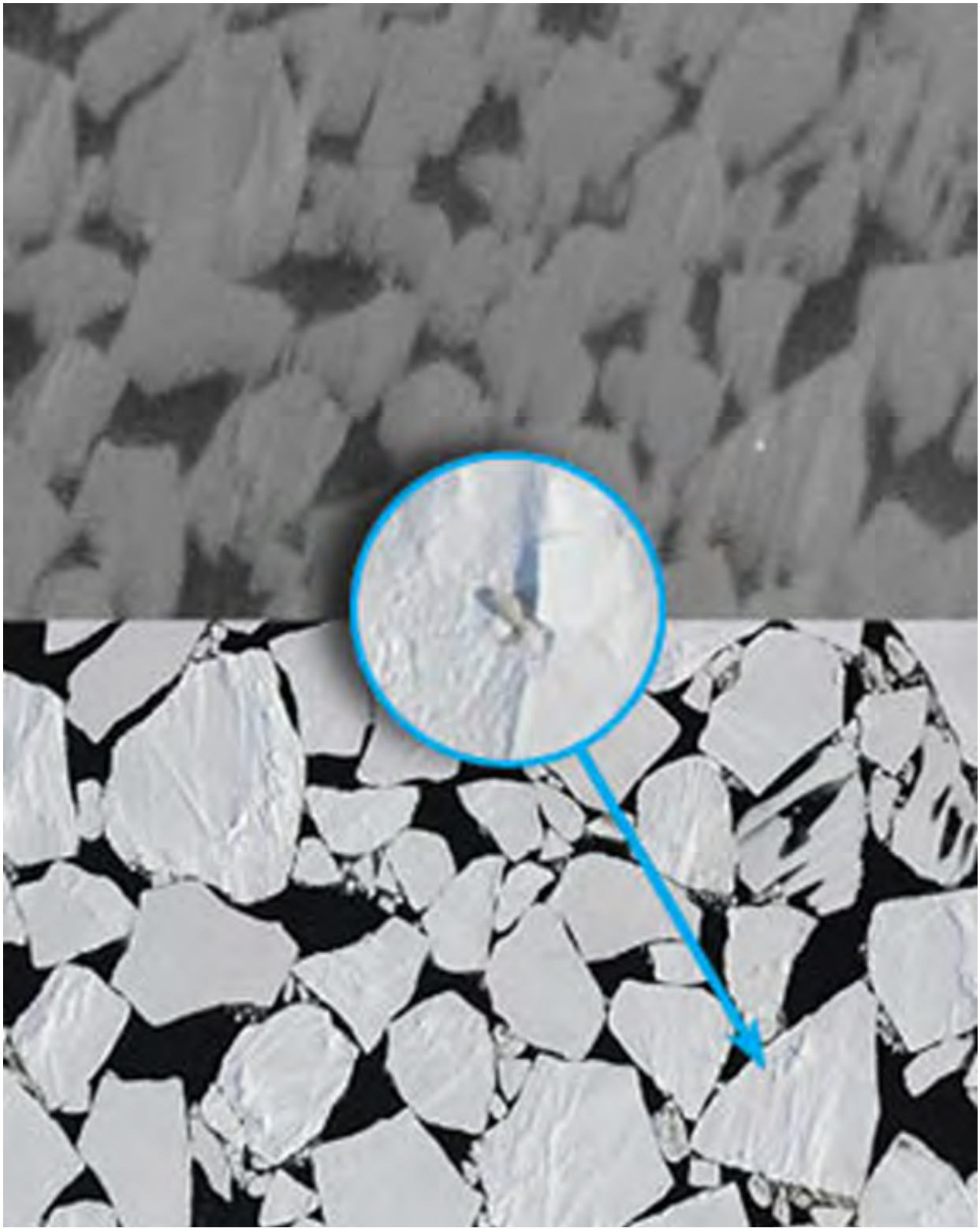

Figure 16. Example of an animal (seal pup) likely to be missed during a manual review of SLR imagery, but easily detected using thermal imagery.

Thermal detection

Manual hot spot detection

Preliminary analysis of 2012 data for both the U.S. and Russian surveys was based on setting a temperature threshold and manually reviewing potential seal hot spots. Russian technique relied on software to identify hot spots, which were then manually reviewed and matched to SLR imagery. The U.S. thermal video recordings were first analyzed using temperature profiles. A temperature threshold was applied to a graph of maximum pixel temperature per frame and video frames corresponding to peaks that exceeded the threshold ('hot spots') were reviewed (Figure 17). Distinct peaks falling below the threshold were also evaluated. Valid hot spots were matched to corresponding SLR imagery. Seal species, thermal anomalies (e.g., melt pool or dirty ice) and additional image characteristics were recorded. Unfortunately, it was not possible to identify a consistent threshold between the temperature of the ice and the temperature of the seals. Background temperature noise was variable throughout each long flight and ringed seals disproportionately fell below the threshold.

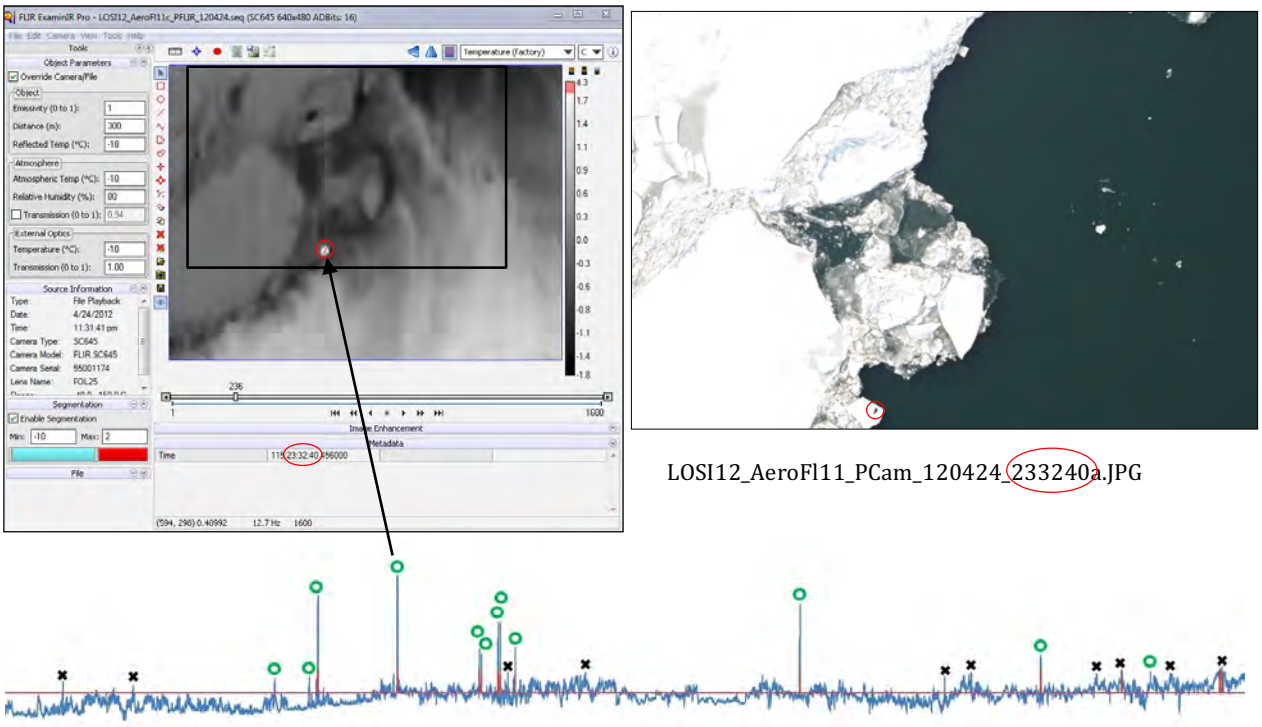

Figure 17. The initial US hot spot detection method utilized a temperature threshold applied to a plot of maximum pixel temperature per frame to identify which thermal frames to evaluate. Digital SLR images were matched using the timestamps and ice features to locate the source of the thermal signature.

Automated hot spot detection (Skeyes)

We then developed automated hot spot detection software (Skeyes) that relies on anomalous temperature shifts rather than specific, absolute, temperature thresholds. This "outlier" method reduces the number of seals missed by eliminating the need to set (subjective) temperature thresholds that require frequent adjustment in response to changes in atmospheric conditions (e.g., humidity, air temperature, sea ice temperature, and seal temperature, which changes with time out of the water). Skeyes is designed to operate inflight as well as post process FLIR video files, filtering out frames identified by the algorithm. Filtered frames were reviewed manually and valid hot spots were matched to color imagery using patterns in the ice and the source of the hot spot was determined. One thermal camera video file per 2012 flight was analyzed manually, all remaining thermal data were processed by Skeyes.

Species Identification

The different characteristics that distinguish these ice-associated seal species can sometimes be difficult to discern from imagery taken at survey altitude. For example, the characteristic bands on the coats of ribbon seals will not always be visible in a photo, depending on the orientation of the seal and angle of the image (Figure 18). The identifying characteristics of spotted and ringed seals can be even more difficult to discern from aerial photos. Although typically ignored in population estimates, errors can be common when attempting to identify similar-looking seal species from aerial photographs.

We accounted for species misidentification in our abundance model by estimating misclassification probabilities for species identified in the images. Several ice seal experts with NMML's Polar Ecosystems Program identified the species of more than 600 seals detected by thermal imagers and photographed during the 2012 surveys. To learn more about the factors influencing the species identification process, our experts also recorded the specific morphological characteristics that are visible in each image. In addition, experts ranked their confidence in each species identification as "positive," "likely," or "guess." By replicating the species-identification process with multiple observers for each seal, and assuming that a positive species-identification is the correct species, the probabilities of correct (and incorrect) species identification were estimated and accounted for in our final estimates of population abundance for each species (McClintock et al. 2015).

Figure 18. The characteristic bands on the coats of ribbon seals are not necessarily clearly visible in an aerial image. The images on the top right and bottom right were taken with a Canon 1Ds Mark III fitted with a Zeiss 100 mm lens from 1000 ft during BOSS 2012. In the top right image, a species identification expert would likely rely on the clearly visible bands to conclude that the seal is certainly a ribbon seal. In the bottom right image, a species identification expert would rely on a combination of body shape, head size, flipper size and shape, and what could be one or more faint bands to conclude that the seal is likely a ribbon seal.

Abundance estimation

Analyzing abundance from thermal video and digital photography presents several statistical challenges due to incomplete detection, false positives, and species misidentification. We developed several novel statistical approaches to estimate abundance in face of such nuisance factors. The first approach we developed expressed abundance of each species using a spatial regression model that related expected density to covariates (e.g., sea ice concentration, distance from land, distance from 1000m shelf break, instance from the southern ice edge); observed counts of each species were then assumed to arise from an overdispersed Poisson process subject to thinning and a multinomial species misclassification process (Conn et al. 2015). Thinning included both incomplete detection (cameras do not detect 100% of seals) and availability less than 1.0 (not all animals are on ice while sampling is being conducted). In all analyses reported in this paper, we used data from previously deployed satellite tags to estimate availability for bearded, ribbon, and spotted seals. For ringed seals, we simply used a point estimate of 0.6 (taken from Bengtson et al. 2005).

The previous described model requires that species distributions be constant in space, so was only applied to data we collected during a one week window (April 20-27) in 2012. However, we developed a second modeling approach that was a generalization of the previous method that allows abundance to change in space while sampling is being conducted. This is accomplished with a spatio-temporal

statistical modeling framework where total abundance is assumed constant but allowed to redistribute itself each day in response to changing habitat (e.g., sea ice; Conn et al. 2016). We fit several variants of this model to full data from the U.S. portion of the Bering Sea from 2012 and 2013 (separate analyses were conducted for each year). Once again, expected density was written as a function of habitat and landscape covariates.

Objective 4: For ribbon and spotted seals, which breed in the Bering Sea but are present seasonally in the Chukchi Sea, produce estimates of the numbers of individuals that use the Chukchi Sea during the open-water season.

By combining abundance estimates from the BOSS surveys with results from ongoing analysis of seal movements and behavior, we wished to provide more informative estimates of the numbers of ribbon and spotted seals that rely on the Chukchi Sea during the open water season, and that may interact with oil and gas activities in that region. We used the proportions of eastern Bering Sea ribbon and spotted seals using the Chukchi Sea during summer (presented in the Previous Research section above), and multiplied those proportions by our abundance estimates from the BOSS surveys in 2012 and 2013.

RESULTS AND DISCUSSION

Objective 1: Contract for a NOAA Twin Otter and a second long-range aircraft

The DHC-6 Twin Otter N56RF, operated by NOAA's Office of Marine and Aircraft Operations, was retained for Twin Otter to fly our abundance surveys for the required time periods. We were also able to negotiate a contract with Clearwater Air to fly our required survey routes. The 2012 surveys took place over a 6-week period from early April to late May. The Twin Otter flew 139 hours on 28 days and completed 21 survey flights. The Aero Commander was chartered for one month and flew 97 hours on 20 days to complete 18 surveys. In 2013, the Otter flew 94 hours on 20 days and completed 16 surveys, while the Aero Commander flew 107 hours on 17 days to complete 18 surveys. The 2013 survey effort was completed within a 4-week window of good weather from mid-April to early May.

Objective 2: Conduct surveys of ice-associated seals.

Aerial surveys for bearded, spotted, ribbon, and ringed seals were conducted from 6 April to 23 May 2012 and from 4 April to 9 May 2013. The two US aircraft flew a total of 73 surveys covering more than 31,000 nmi (57,400 km) of track line and collected 1.8 million images. The Russian team completed 32 flights from 13 airports and flew more than 16,000 nmi (29,600 km). Combined, the teams flew more than 47,000 nmi (87,000 km) of survey track (Figure 19) The completion of this project marks the largest survey of ice-associated seals ever completed and will provide the first comprehensive estimates of abundance for bearded, spotted, ribbon, and ringed seals in the Bering Sea and Sea of Okhotsk.

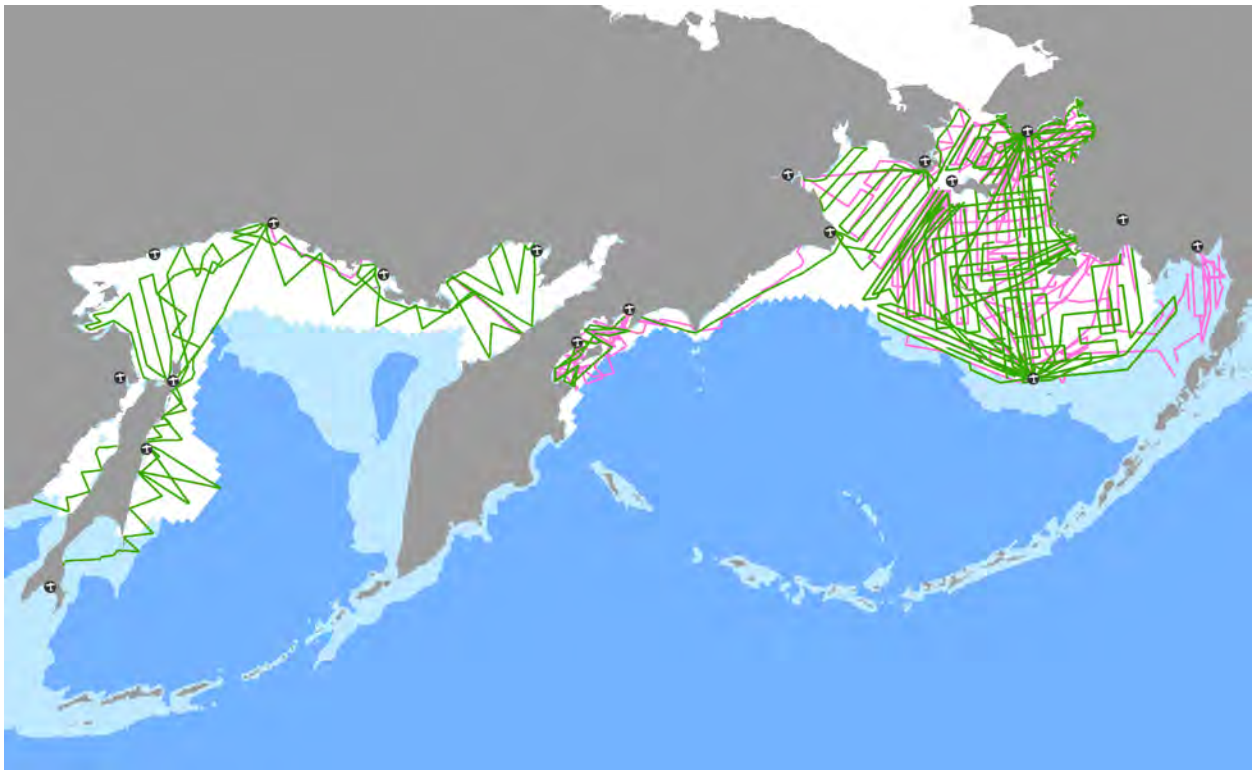

Figure 19. BOSS 2012 (pink) and 2013 (green) survey track lines in the Bering and Okhotsk seas covering over 90,000 km (56,000 miles) completed during the joint US-Russian survey effort. Water depths shallower than 500m are shown in light blue and the April, 2013 ice extent is in white.

Objective 3: Effectively retrieve, manage, and process all imagery for analyses.

Thermal Detection

Combined, the two aircraft collected a total of 1.8 million color images and over 5.4 TB of thermal video. Manual and automated detection approaches were established and tested using a subset of survey data reviewed by a technician for the presences of seals.

Automated hot spot detection (Skeyes)

To test Skeyes effectiveness for finding seals, every tenth SLR photo from a 10 flight subsample of the 2012 Bering Sea ice-seal survey was examined for seals by an experienced observer without the aid of thermal data. Of the 70 seal groups found in 11,724 SLR images examined, Skeyes found 75 seal groups, including 68 (97%) seal groups found by the experienced technician, using only about 1/13th the time.

We have completed analyzing all of the thermal video for hot spot detection. The impact of incorporating thermal imaging into our aerial survey methodology has been significant. The Bering Sea surveys required half of the field personnel needed for traditional observer-based surveys and data processing has occurred in parallel to detection advances. This has allowed rapid production of preliminary data analysis, including examples of reliable population estimates, with human error minimized and detection probability substantially increased.

Planned improvements for future surveys

Development of this system is iterative and ongoing. Our current focus is on algorithm refinement to reduce false positives (without sacrificing detection performance), to reduce the collection of extraneous imagery, and to improve efficiency of image processing. Towards this end we have begin developing an automated thermal detection system called “Snowflake.” This system triggers the collection of thermal and SLR images when a seal-like thermal signature is detected. The system can be used in flight or as a post-processing module to replace the threshold detection approach described above. The current version of Snowflake detects 94% of the seals found with the threshold detection and 97% of the seal groups found by manual review of SLR imagery. Future improvements to Snowflake focus on reducing the false-positive trigger rate, tracking and projecting GPS data for each hot spot, and improved SLR camera control for in-flight triggering. We are also exploring machine vision cameras as an alternative to professional off-the-shelf SLR cameras. This would allow greater camera control (i.e., one computer can control both the thermal and visual instruments), provide access to additional data to

improve the filtering of anomalous hot spots, and improved efficiency of data download, processing, and management.

Species Identification

To date, we have completed analyzing the thermal data and SLR images in both years for seals. We now have a better understanding of the frequency of ice seal species misidentification errors from aerial photos and are able to properly adjust our population estimate and variance for each species accordingly. In addition, we have gained a better understanding of the specific morphological characteristics that are most effective for distinguishing each species in photos from aerial transect surveys. It is also noteworthy that we had originally assumed that the number of ringed seals still occupying snow lairs in the Bering Sea in April would preclude us from obtaining a reliable population estimate for that species. Examination of the photographs, however, has indicated that snow cover was relatively light during our surveys and large numbers of ringed seals were basking in the open on top of the ice or snow. Furthermore, our photographs of the pack ice zone had greater numbers of ringed seals than we expected from our previous, visual observer line-transect surveys, perhaps because of greater rates of disturbance and misclassification of ringed seals in the former surveys. Therefore, we now believe that we will be able to provide a reliable estimate for ringed seals in the Bering Sea, though that will require a separate analytical effort for the shore-fast ice zone and collaboration with other researchers to obtain a haul-out adjustment for ringed seals. The estimates presented below assume a deterministic correction factor of 0.6 for ringed seals.

Abundance Estimation

Estimates and standard errors from each year and analysis type performed are reported in Table 2. Estimates from the static spatial regression model applied to April 20-27, 2012 were much higher than previous estimates of density in this area (e.g., Ver Hoef et al. 2014). We suspect that static results may have been anomalously high because a greater proportion of seals were hauling out on ice during this one week window. The dates selected for the one week, static analysis happened to correspond to a time when temperatures were high and weather was better than normal. We are currently gathering data from satellite tags and weather reanalysis products to test the hypothesis that seal haul-out probabilities are related to weather conditions.

Results from spatio-temporal statistical analyses using the full data sets in 2012 and 2013 yielded smaller abundance estimates (Table 2). There was also considerable variation in estimates among years,

was which likely attributable (at least in part) to shifts in spatial distributions (Figures 20 & 21). In particular, the quantity and location of sea ice was very different for the two years, with 2012 having record levels of spring sea ice. There was also very limited coverage of the southwestern corner of our study area in 2012 (where concentrations of ribbon and spotted seals were at their highest), so estimates of total abundance are somewhat uncertain for 2012.

Table 2. Estimates of ice-associated seal abundance (thousands) in the eastern Bering Sea (U.S. airspace) in different years by different analysis methods. We present model-based standard errors for each model run, as well as model-averaged standard errors for combined estimates. The abbreviation Sp-T is used to indicate a spatio-temporal model; RE indicates random effects. We suggest basing inference on the combined, model-averaged estimates rather than results from individual model runs.

Model	**Year**	**Bearded**	**Ribbon**	**Ringed**	**Spotted**
Static, Apr 20-27	2012	301 (33)	183 (25)	286 (15)[1]	458 (44)
Sp-T, no REs	2012	170 (6)	133 (11)	180 (3)[1]	234 (9)
Sp-T, REs (1)	2012	167 (7)	86 (9)	187 (3)[1]	273 (10)
Sp-T, REs (2)	2012	174 (6)	133 (12)	190 (3)[1]	220 (9)
Sp-T, Combined[2]	2012	170 (7)	117 (25)	186 (3)	240 (24)
Sp-T, no REs	2013	122 (5)	38 (4)	121 (4)	163 (5)
Sp-T, REs	2013	127 (5)	38 (3)	116 (2)	162 (6)
Sp-T, Combined[2]	2013	125 (5)	38 (4)	119 (4)	163 (6)

[1]Ringed seal estimates do not include specific efforts to target fast ice or extra uncertainty due to incomplete knowledge of haul-out correction factors

[2]Combined estimates generated using simple model-averaging formula (i.e., Burnham and Anderson 2002) and equal model weights

Although we produced an estimate of ringed seals, this estimate should be viewed with caution. First, we used a simple point estimate of availability (0.6) from a study in the Chukchi Sea (Bengtson et al. 2005), which does not properly account for extra variance attributable to uncertainty about the availability process. Also, it does not include survey counts from the shore-fast ice zone. The shore-fast zone is relatively small region of high ringed seal density that required a different approach for

allocation of survey effort (narrow, along-shore sawtooth). Therefore, it will require a separate analysis to avoid bias from preferential sampling (Conn et al. Submitted).

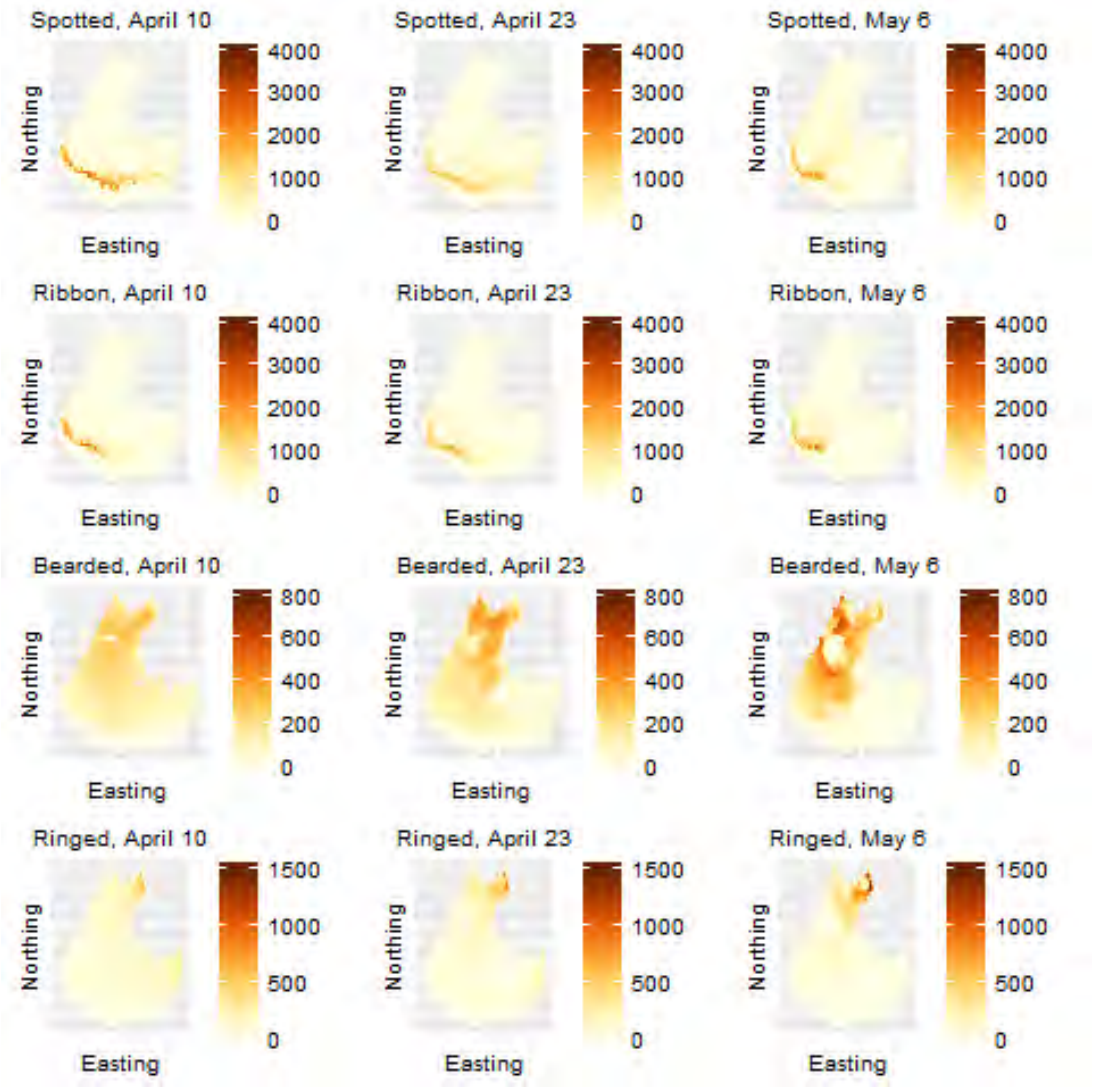

Figure 20. Estimates of bearded, ribbon, ringed, and spotted seal abundance for three dates in the spring of 2012 in the eastern (U.S.) Bering Sea. Each map is bounded by the Bering Strait to the north, Alaska to the east, the U.S.-Russia Exclusive Economic Zone to the west, and maximum spring sea ice extent to the south.

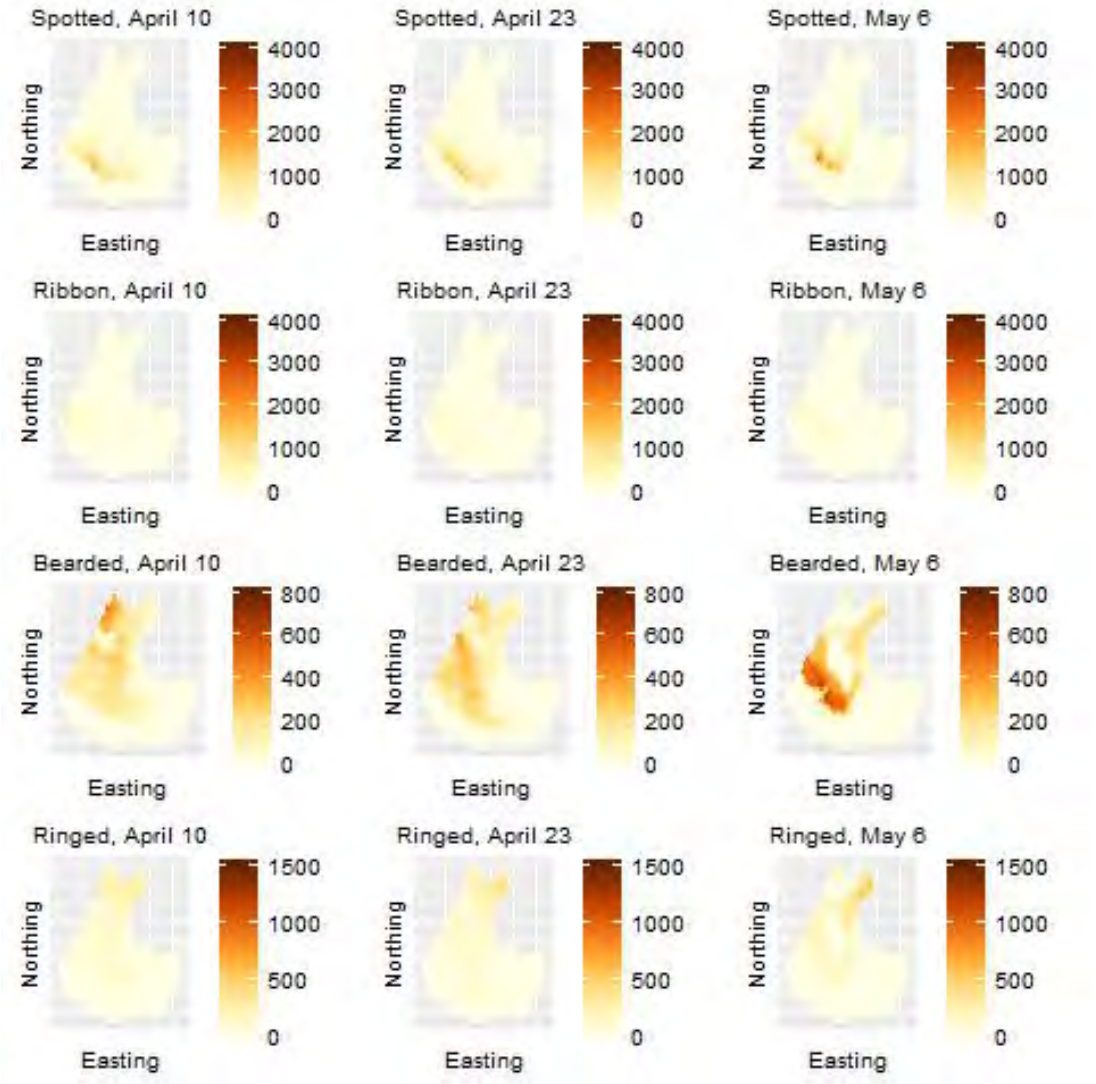

Figure 21. Estimates of bearded, ribbon, ringed, and spotted seal abundance for three dates in the spring of 2013 in the eastern (U.S.) Bering Sea. Each map is bounded by the Bering Strait to the north, Alaska to the east, the U.S.-Russia Exclusive Economic Zone to the west, and maximum spring sea ice extent to the south. Note the scale is the same as Figure 20 to permit proper comparison.

Although we have made substantial progress in developing estimation methodology and software for conducting demanding spatiotemporal analyses, we intend to analyze satellite tagging data as a function of weather data to better account for variation in seal availability (haul-out) probabilities. Also, the variation in seal densities between year (especially ribbon and spotted seals) suggests that there may be substantial differences in the numbers of seals using Russian versus U.S. waters between years; the

maritime boundary crosses through a region of typically high seal densities, so slight shifts in distribution between years could have substantial influence on separate U.S. and Russian estimates. If so, the eventual combination of the U.S. and Russian results might be expected to smooth out some of this interannual variability. To account for this possibility, a final step will be to incorporate data collected by our Russian collaborators to provide the most comprehensive estimates of abundance for bearded, spotted, ribbon, and ringed seals in the Bering Sea and the Sea of Okhotsk.

Objective 4: For ribbon and spotted seals, which breed in the Bering Sea but are present seasonally in the Chukchi Sea, produce estimates of the numbers of individuals that use the Chukchi Sea during the open-water season.

Based on movement records from ribbon and spotted seals instrumented with satellite tags in separate studies, we estimated that 42% of spotted seals and 21% of ribbon seals from the eastern Bering Sea population spend at least part of the open-water season (July – October) in the Chukchi Sea. Multiplying those proportions by our abundance estimates for these species, we estimated that 69,000-101,000 spotted seals and 6,000-25,000 of ribbon seals that occupy the eastern (U.S.) Bering Sea in spring used the Chukchi Sea during the summer, open-water period in 2013 and 2012, respectively.

ACKNOWLEDGMENTS

This study was partially funded by the U.S. Department of the Interior, Bureau of Ocean Energy Management, through Interagency Agreement M12PG00017 with the U.S. Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, as part of the MMS Alaska Environmental Studies Program. We thank all NOAA personnel and contractors that helped collect and process seal data. Most of the bearded seal haul-out data were collected and made available by the Native Village of Kotzebue and the Alaska Department of Fish and Game, with support from the Tribal Wildlife Grants Program of the US Fish and Wildlife Service (Grant Number U-4-IT). Views expressed are those of the authors and do not necessarily represent findings or policy of any government agency. Use of trade or brand names does not indicate endorsement by the U.S. government.

PRESENTATIONS, SIGNIFICANT MEETINGS, AND PUBLICATIONS

Presentations and significant meetings

September 28, 2012: Presentation "Bering Okhotsk Seal Surveys (BOSS): Synoptic surveys for ice-associated seals", at the Marine Mammals of the Holarctic Conference, Suzdal, Russia

October 2, 2012: CAFF Workshop on International Ringed Seal Monitoring; summary presentation of BOSS project by NOAA scientists

January 21–25, 2013: Two presentations; "Bering-Okhotsk Seal Surveys (BOSS): Joint US-Russian Aerial Surveys for Ice-Associated Seals" and " Ice-associated seal detection and identification from the 2012 Bering-Okhotsk Seal Surveys (BOSS)" at the Alaska Marine Science Symposium, Anchorage, AK

January 24–25, 2013: Alaska Native Ice Seal Committee, Anchorage, AK: summary presentations of BOSS project to Committee members by NOAA scientists

February 12-13, 2013: Presentations and discussions of survey plans with hunters and the St. Lawrence Island Marine Mammal Advisory Council in Savoonga and Gambell.

March 20, 2013: Presentation, "How many seals are in the Bering Sea?", to Kawerak, the Bering Straits Region an Alaska Native, non-profit corporation.

September 4, 2013: A seminar, "How many seals are in the Bering Sea?", describing the BOSS surveys and preliminary results, was presented at the Sea Mammal Research Unit, Scottish Oceans Institute, University of St. Andrews, Scotland.

November 22, 2013: A seminar, "Estimating animal abundance using an automated detection system: Ice-associated seals in the Bering Sea", describing some statistical methods developed for the BOSS program, was presented to the University of Washington, School of Fisheries and Aquatic Sciences Quantitative Seminar series, Seattle, Washington.

December 12, 2013: A presentation, "How many seals are in the Bering Sea?", describing the BOSS surveys and preliminary results, was presented to Sam Rauch III (Assistant Administrator for Fisheries, Acting) at NOAA's National Marine Mammal Laboratory, Seattle, Washington USA.

January 22-23, 2014: A presentation, “Bering-Okhotsk Seal Surveys (BOSS): Joint US-Russian Aerial Surveys for Ice-Associated Seals, Spring 2012 and 2013” an overview of the BOSS program, including preliminary results, was presented at the Alaska Marine Science Symposium and at the Alaska Native Ice Seal Committee and Co-management meeting, Anchorage, Alaska, USA.

May 15, 2014: Powerpoint slides for “Bering-Okhotsk Seal Surveys (BOSS): Joint U.S.-Russian aerial surveys for ice-associated seals, Spring 2012 and 2013” provided to the Bristol Bay Marine Mammal Commission meeting, Dillingham, Alaska

June 2, 2014: A web presentation, “How many seals are in the Bering Sea?”, describing the BOSS surveys and preliminary results, was presented to the 5th Annual Meeting of the Scientific Working Group under the U.S. - Russia Polar Bear Commission, National Conservation Training Center, Shepherdstown, West Virginia.

July 4, 2014. A presentation “Using spatio-temporal statistical models to estimate animal abundance from transect counts” was presented at the International Statistical Ecology Conference in Montpelier, France.

September 22-27, 2014: Two presentations: “Bering-Okhotsk Seal Surveys (BOSS) Program progress toward comprehensive estimates of abundance” an overview of the BOSS program including preliminary results; and “Automated thermal detection of seals on ice, Bering-Okhotsk Seal Surveys (BOSS): Finding a needle in a haystack” an overview of the BOSS survey methodology, were presented at the Marine Mammals of the Holarctic Conference, St. Petersburg, Russia.

August 10-11, 2015. Two presentations, “using automated aerial imaging to estimate Arctic seal abundance: the devil is in the details” and “Spatio-temporal models for aerial survey counts: An application to ice-associated seals in the Bering Sea” were presented at the Ecological Society of America annual meeting in Baltimore, MD.

Multiple dates: Meetings with Russian scientists, aircraft pilots, software developers and thermal-systems contractors to prepare for, and share the results of, the 2012 and 2013 fieldwork and preliminary analyses.

Publications

The following publications made use of data from this project. Copies of the peer-reviewed publications are provided in Appendices 1-5.

Agency reports

Moreland, E., M. Cameron, and P. Boveng. 2013. Bering Okhotsk Seal Surveys (BOSS), joint U.S.-Russian aerial surveys for ice-associated seals, 2012-13. Alaska Fisheries Science Center Quarterly Report July-August-September 2013:1-6. http://www.afsc.noaa.gov/Quarterly/jas2013/JAS13-Feature.pdf

Peer-reviewed journals

Conn, P. B., J. M. Ver Hoef, B. T. McClintock, E. E. Moreland, J. M. London, M. F. Cameron, S. P. Dahle, and P. L. Boveng. 2014. Estimating multispecies abundance using automated detection systems: ice-associated seals in the Bering Sea. Methods in Ecology and Evolution DOI 10.1111/2041-210X.12127.

McClintock, B. T., E. E. Moreland, J. M. London, S. P. Dahle, G. M. Brady, E. L. Richmond, K. M. Yano, and P. L. Boveng. 2015. Quantitative assessment of species identification in aerial transect surveys for ice-associated seals. Marine Mammal Science 21:1057-1076.

Conn, P. B., D. S. Johnson, J. M. V. Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs 85:235-252.

Conn, P. B., D. S. Johnson, and P. L. Boveng. 2015. On extrapolating past the range of observed data when making statistical predictions in ecology. PLoS ONE 10(10): e0141416.

Conn, P. B., J. T. Thorson, and D. S. Johnson. *In review*. Confronting preferential sampling in wildlife surveys: diagnosis and model-based triage. Methods in Ecology and Evolution.

LITERATURE CITED

Allen, J. A. 1880. History of North American pinnipeds: a monograph of the walruses, sea-lions, sea-bears and seals of North America. U.S. Department of the Interior, U.S. Government Printing Office, Washington, D.C. 785 p.

Bengtson, J. L., L. M. Hiruki-Raring, M. A. Simpkins, and P. L. Boveng. 2005. Ringed and bearded seal densities in the eastern Chukchi Sea, 1999-2000. Polar Biology 28:833-845.

Boveng, P. L., J. L. Bengtson, T. W. Buckley, M. F. Cameron, S. P. Dahle, B. P. Kelly, B. A. Megrey, J. E. Overland, and N. J. Williamson. 2009. Status review of the spotted seal (*Phoca largha*). U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-200. 153 p.

Boveng, P. L., J. L. Bengtson, T. W. Buckley, M. F. Cameron, S. P. Dahle, B. A. Megrey, J. E. Overland, and N. J. Williamson. 2008. Status review of the ribbon seal (*Histriophoca fasciata*). U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-191. 115 p.

Boveng, P. L., J. L. Bengtson, M. F. Cameron, S. P. Dahle, E. A. Logerwell, J. M. London, J. E. Overland, J. T. Sterling, D. E. Stevenson, B. L. Taylor, and H. L. Ziel. 2013. Status review of the ribbon seal (*Histriophoca fasciata)*. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-255. 175 p.

Braham, H. W., J. J. Burns, G. A. Fedoseev, and B. D. Krogman. 1984. Habitat partitioning by ice-associated pinnipeds: distribution and density of seals and walruses in the Bering Sea, April 1976 Pages 25-47 *in* F. H. Fay and G. A. Fedoseev, editors. Soviet-American Cooperative Research on Marine Mammals. Volume 1 - Pinnipeds. NOAA Technical Report NMFS 12. U.S. Department of Commerce, NOAA, Washington, DC.

Burns, J. J. 1967. The Pacific bearded seal. Alaska Department of Fish and Game, Pittman-Robertson Project Report W-6-R and W-14-R. 66 p.

Burns, J. J. 1970. Remarks on the distribution and natural history of pagophilic pinnipeds in the Bering and Chukchi Seas. Journal of Mammalogy 51:445-454.

Burns, J. J. 1981a. Bearded seal *Erignatus barbatus* Erxleben, 1777. Pages 145-170 *in* S. H. Ridgway and R. J. Harrison, editors. Handbook of Marine Mammals Volume 2: Seals. Academic Press, New York, NY.

Burns, J. J. 1981b. Ribbon seal *Phoca fasciata* Zimmermann, 1783. Pages 89-109 *in* S. H. Ridgway and R. J. Harrison, editors. Handbook of Marine Mammals. Volume 2: Seals. Academic Press, New York, NY.

Burns, J. J., and K. J. Frost. 1979. The natural history and ecology of the bearded seal, *Erignathus barbatus*. Alaska Department of Fish and Game. 77 p.

Burns, J. J., L. F. Lowry, and K. J. Frost. 1981. Trophic relationships, habitat use, and winter ecology of ice-inhabiting phocid seals and functionally related marine mammals in the Arctic: Part 1 - Trophic relationships and Part 2 - Habitat use and winter ecology. Division of Game, Alaska Department of Fish and Game, Annual Report. 114 p.

Cameron, M., and P. Boveng. 2007. Abundance and distribution surveys for ice seals aboard USCG *Healy* and the *Oscar Dyson,* April 10-June 18, 2007. Alaska Fisheries Science Center Quarterly Report, April-May-June 2007:12-14.

Cameron, M. F., J. L. Bengtson, P. L. Boveng, J. K. Jansen, B. P. Kelly, S. P. Dahle, E. A. Logerwell, J. E. Overland, C. L. Sabine, G. T. Waring, and J. M. Wilder. 2010. Status review of the bearded seal (*Erignathus barbatus)*. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-211. 246 p.

Conn, P. B., J. T. Thorson, and D. S. Johnson. Submitted. Confronting preferential sampling in wildlife surveys: diagnosis and model-based triage. Methods in Ecology and Evolution

Fay, F. H. 1974. The role of ice in the ecology of marine mammals of the Bering Sea. Pages 383-399 *in* D. W. Hood and E. J. Kelley, editors. Oceanography of the Bering Sea. Institute of Marine Science, Hakodate, Japan.

Fedoseev, G. A. 1965. The ecology of the reproduction of seals on the northern part of the Sea of Okhotsk. Izvestiya TINRO 65:212-216. (Translated from Russian by the Fisheries and Marine Service, Quebec, Canada, Translation Series No. 3369, 8 p.).

Fedoseev, G. A. 2000. Population biology of ice-associated forms of seals and their role in the northern Pacific ecosystems. Center for Russian Environmental Policy, Russian Marine Mammal Council, Moscow, Russia. 271 p. (Translated from Russian by I. E. Sidorova, 271 p.).

Frost, K. J. 1985. The ringed seal (*Phoca hispida*). Pages 79-87 *in* J. J. Burns, K. J. Frost, and L. F. Lowry, editors. Marine Mammals Species Accounts. Alaska Department Fish and Game, Juneau, AK.

Frost, K. J., L. F. Lowry, and J. J. Burns. 1977. Spring foods and feeding of phocid seals in the Bering Sea ice front (Abstract). Page 20 *in* The Second Conference on the Biology of Marine Mammals, San Diego, CA. Society for Marine Mammalogy.

Frost, K. J., L. F. Lowry, G. Pendleton, and H. R. Nute. 2004. Factors affecting the observed densities of ringed seals, *Phoca hispida*, in the Alaskan Beaufort Sea, 1996-99. Arctic 57:115-128.

Johnson, M. L., C. H. Fiscus, B. T. Ostenson, and M. L. Barbour. 1966. Marine mammals. Pages 877-924 *in* N. J. Wilimovsky and J. N. Wolfe, editors. Environment of the Cape Thompson Region, Alaska. U.S. Atomic Energy Commission, Oak Ridge, TN.

Kelly, B. P. 1988a. Bearded seal, *Erignathus barbatus*. Pages 77-94 *in* J. W. Lentifer, editor. Selected Marine Mammal Species of Alaska: Species Accounts with Research and Managment Recommendations. Marine Mammal Commission, Washington, D.C.

Kelly, B. P. 1988b. Ribbon seal, *Phoca fasciata*. Pages 95-106 *in* J. W. Lentfer, editor. Selected Marine Mammal Species of Alaska: Species Accounts with Research and Managment Recommendations. Marine Mammal Commission, Washington, D.C.

Kelly, B. P. 1988c. Ringed seal, *Phoca hispida*. Pages 57-75 *in* J. W. Lentifer, editor. Selected Marine Mammal Species of Alaska: Species Accounts with Research and Managment Recommendations. Marine Mammal Commission, Washington, D.C.

Kelly, B. P., J. L. Bengtson, P. L. Boveng, M. F. Cameron, S. P. Dahle, J. K. Jansen, E. A. Logerwell, J. E. Overland, C. L. Sabine, G. T. Waring, and J. M. Wilder. 2010. Status review of the ringed seal (*Phoca hispida*). U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-212. 250 p.

King, J. E. 1983. Seals of the world. 2nd edition. British Museum (Natural History) and Oxford University Press, London, UK. 240 p.

Lowry, L. F., V. N. Burkanov, K. J. Frost, M. A. Simpkins, R. Davis, D. P. DeMaster, R. Suydam, and A. Springer. 2000. Habitat use and habitat selection by spotted seals (*Phoca largha*) in the Bering Sea. Canadian Journal of Zoology 78:1959-1971.

Lowry, L. F., K. J. Frost, R. Davis, D. P. DeMaster, and R. S. Suydam. 1998. Movements and behavior of satellite-tagged spotted seals (*Phoca largha*) in the Bering and Chukchi Seas. Polar Biology 19:221-230.

McClintock, B. T., E. E. Moreland, J. M. London, S. P. Dahle, G. M. Brady, E. L. Richmond, K. M. Yano, and P. L. Boveng. 2015. Quantitative assessment of species identification in aerial transect surveys for ice-associated seals. Marine Mammal Science 21:1057-1076.

Moreland, E., M. Cameron, and P. Boveng. 2013. Bering Okhotsk Seal Surveys (BOSS), joint U.S.-Russian aerial surveys for ice-associated seals, 2012-13. Alaska Fisheries Science Center Quarterly Report July-August-September 2013:1-6.

Nelson, R. K. 1981. Harvest of the sea: coastal subsistence in modern Wainwright. North Slope Borough, Coastal Management Program, A report for the North Slope Borough's Coastal Management Program. 112 p.

Ognev, S. I. 1935. Mammals of U.S.S.R. and adjacent countries. Volume 3. Carnivora. Glavpushnina NKVT, Moscow, Russia. 641 p. (Translated from Russian by the Israel Program for Scientific Translations, Jerusalem 1962, 741 p.).

Porsild, A. E. 1945. Mammals of the Mackenzie delta. Canadian Field-Naturalist 59:4-22.

Shaughnessy, P. D., and F. H. Fay. 1977. A review of taxonomy and nomenclature of North Pacific harbor seals. Journal of Zoology 182:385-419.

Simpkins, M. A., L. M. Hiruki-Raring, G. Sheffield, J. M. Grebmeier, and J. L. Bengtson. 2003. Habitat selection by ice-associated pinnipeds near St. Lawrence Island, Alaska in March 2001. Polar Biology 26:577-586.

Smith, T. G. 1981. Notes on the bearded seal, *Erignathus barbatus*, in the Canadian Arctic. Department of Fisheries and Oceans, Arctic Biological Station, Canadian Technical Report of Fisheries and Aquatic Sciences No. 1042. 49 p.

Ver Hoef, J. M., M. F. Cameron, P. L. Boveng, J. M. London, and E. E. Moreland. 2014. A spatial hierarchical model for abundance of three ice-associated seal species in the eastern Bering Sea. Statistical Methodology 17:46-66. (Published online April 8, 2013).

Ver Hoef, J. M., J. M. London, and P. L. Boveng. 2010. Fast computing of some generalized linear mixed pseudo-models with temporal autocorrelation. Computational Statistics 25:39-55.

APPENDIX 1. - Conn, P. B., J. M. Ver Hoef, B. T. McClintock, E. E. Moreland, J. M. London, M. F. Cameron, S. P. Dahle, and P. L. Boveng. 2014. Estimating multispecies abundance using automated detection systems: ice-associated seals in the Bering Sea. Methods in Ecology and Evolution DOI 10.1111/2041-210X.12127

This paper established a basis for the analytical method that will be used to estimate the abundance and distribution of bearded, spotted, ringed, and ribbon seals in the Bering Sea. Significantly, it included consideration of the effects and variance from incomplete detection, incomplete availability (haul-out proportion), and species miss-classification rates. Preliminary estimates of abundance, from just 10 flights conducted in a 1-week period of the 2012 survey, indicated that the final analysis will produce higher and more precise abundance estimates than have been obtained from previous surveys. The higher abundance may be due to better detection of seals by thermal video than by traditional, visual observation, and lower rates of seal disturbance that were obtained by flying at higher altitudes. The greater precision is likely due to the extensive and relatively dense coverage that we were able to achieve.

Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2014, **5,** 1280–1293 doi: 10.1111/2041-210X.12127

SPECIAL ISSUE. MODELLING DEMOGRAPHIC PROCESSES IN MARKED POPULATIONS: PROCEEDINGS OF THE EURING 2013 ANALYTICAL MEETING

Estimating multispecies abundance using automated detection systems: ice-associated seals in the Bering Sea

Paul B. Conn*, Jay M. Ver Hoef, Brett T. McClintock, Erin E. Moreland, Josh M. London, Michael F. Cameron, Shawn P. Dahle and Peter L. Boveng

National Marine Mammal Laboratory, NOAA-NMFS, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, USA

Summary

1. Automated detection systems employing advanced technology (e.g. infrared imagery, auditory recording systems, pattern recognition software) are compelling tools for gathering animal abundance and distribution data since investigators can often collect data more efficiently and reduce animal disturbance relative to surveys using human observers.

2. Even with these improvements, analysing animal abundance with advanced technology can be challenging because of potential for incomplete detection, false positives and species misidentification. We argue that double sampling with an independent sampling method can provide the critical information needed to account for such errors.

3. We present a hierarchical modelling framework for jointly analysing automated detection and double sampling data obtained during animal population surveys. Under our framework, observed counts in different sampling units are conceptualized as having arisen from a thinned log-Gaussian Cox process subject to spatial autocorrelation (where thinning accounts for incomplete detection). For multispecies surveys, our approach handles incomplete species observations owing to (i) structural uncertainties (e.g. in cases where the automatic detection data do not provide species observations) and (ii) species misclassification; the latter requires auxiliary information on the misclassification process.

4. As an example of combining an automated detection system and a double sampling procedure, we consider the problem of estimating animal abundance from aerial surveys that use infrared imagery to detect animals, and independent, high-resolution digital photography to provide information on species composition and thermal detection accuracy. We illustrate our approach by analysing simulated data and data from a survey of four ice-associated seal species in the eastern Bering Sea.

5. Our analysis indicated reasonable performance of our hierarchical modelling approach, but suggested a need to balance model complexity with the richness of the data set. For example, highly parameterized models can lead to spuriously high predictions of abundance in areas that are not sampled, especially when there are large gaps in spatial coverage.

6. We recommend that ecologists employ double sampling when enumerating animal populations with automated detection systems to estimate and correct for detection errors. Combining multiple data sets within a hierarchical modelling framework provides a powerful approach for analysing animal abundance over large spatial domains.

Key-words: abundance estimation, aerial survey, automated detection, data augmentation, hierarchical models, pattern recognition, spatially restricted regression, species misidentification, thermal imagery

Introduction

Several promising approaches have been developed to monitor animal populations using advanced animal detection technology. Pattern recognition algorithms (e.g. Kogan & Margoliash 1998) applied to automated auditory collection systems (cf. Blumstein *et al.* 2011) are capable of discriminating different species, sexes and groups of animals. Ecologists have deployed acoustic arrays to study a range of taxa including terrestrial (Blumstein *et al.* 2011), marine (Moretti *et al.* 2010; Ward

*Correspondence author. E-mail: paul.conn@noaa.gov

et al. 2012) and amphibian (Waddle, Thigpen & Glorioso 2009) species. Another active area of research is application of object-based image analysis to automate animal counts from remotely sensed high-resolution images (see e.g. Groom *et al*. 2013). In this case, a computer algorithm is trained to automatically count animals on a sequence of images. Lastly, when animals give heat signatures different from their surrounding environment, infrared imagery can be used to enumerate animal populations. This approach is often combined with digital photography to provide information about species identity and has been used to monitor big horn sheep (Bernatas & Nelson 2004), pinnipeds (Chernook, Kuznetsov & Yakovenko 1999; Speckman *et al*. 2011), polar bears (Amstrup *et al*. 2004) and, most frequently, ungulates (see e.g. Kissell & Nimmo 2011; Franke *et al*. 2012, and references therein).

Historically, researchers employed human observers to conduct large-scale animal population surveys, and a variety of sampling designs and statistical models are available to cope with imperfect detection when estimating density and abundance from such records (see e.g. Williams, Nichols & Conroy 2002, for a review). Advanced technologies (e.g. infrared imagery, automated acoustic detectors, pattern recognition software) are a promising alternative for increasing survey coverage and reducing detection error, but are far from perfect. For instance, advanced technologies may still miss animals and may also pick up non-target signatures (resulting in false positives). In multispecies surveys, species misidentification errors may also be present. To accurately estimate abundance from automated detection data, it is thus often necessary to collect sufficient auxiliary information to estimate and correct for multiple error types. However, few statistical methods have been developed to incorporate these error rates into abundance estimates (but see Marques *et al*. 2013).

In this paper, we develop a hierarchical modelling framework to estimate animal abundance on landscapes surveyed using an automated detection system. Our approach assumes that the investigator collects independent data using a different sampling approach (hereafter, 'double sampling') over a subset of the survey area to help estimate error rates. In particular, we require that double sampling data be collected in such a manner that it can be used to estimate the probability of false negatives (missed animals), false positives (erroneous detections) and species misidentification, if applicable. Further, we assume that double sampling data can be used to accurately measure individual covariates (e.g. group sizes for clusters of animals).

We demonstrate our approach on simulated data and also on aerial survey data of ice-associated seals. In both cases, automatic detection data consisted of thermal imagery and double sampling consisted of automated high-resolution digital photography. Under our approach, thermal imagery is used to find 'hot spots' – points in the infrared video that have more extreme heat signatures when compared to the surrounding environmental matrix. Digital photographs with matched time stamps can then be searched to get information on the species composition of each hot spot, as well as the number of animals present. Further, independent searches of photographs can be conducted to estimate the proportion of animals missed.

Our manuscript is organized as follows. First, we describe the data necessary to conduct a joint analysis of automatic detection and double sampling data. Next, we describe a model-based framework for estimating animal abundance from such records. After describing a simple simulation study, we analyse a test data set of flights conducted over the eastern Bering Sea in the spring of 2012. In this case, we wish to make inference about the abundance of four ice-associated seal species from data that are contaminated by species misclassifications and anomalous thermal readings.

Methods

DATA REQUIREMENTS

We suppose that the investigator partitions their survey area into J (possibly irregular) sampling units, each of which has area A_j (see Table 1 for a complete list of notation). In practice, the size of the sampling unit will likely be constrained by the resolution of available habitat covariates (e.g. remote sensing data). We assume that transects through each sampling unit occur more or less randomly with respect to available habitat so that the investigator is not making fine-scale adjustments within units to target areas of higher habitat quality. We suppose that L ($L \leq J$) sampling units are surveyed using an automatic detection system and R_j gives the proportion of unit j that is surveyed. We suppose that the spatial domain surveyed by the double sampling method in unit j is a subset of that surveyed by the automatic detection system. As such, we allow for the possibility that some (potentially a large fraction) of automatic detections are not double-sampled.

For each unit that is sampled, we suppose that an automatic detection algorithm is employed on remotely sensed data (e.g. thermal imagery, audio recordings) to compile a list of detections of focal taxa. In practice, some tuning of this algorithm may be needed to balance the resulting sensitivity and specificity; making the algorithm too sensitive can markedly increase the number of false positives, while making it too specific can result in a large number of missed animals. Note that we allow for both false positives (anomalies) and false negatives (non-detections) in subsequent modelling.

Data notation

Let Y_j denote the total number of automatic detections that are recorded in survey unit j. We assign an indicator $I_{ij} = 1$ to automatic detections for which a species observation can be made, and set $I_{ij} = 0$ otherwise (here, $i \in \{1, 2, \ldots, Y_j\}$ identifies the ith automatic detection in surveys of sampling unit j). Note that for some automatic detection data (e.g. thermal imagery), double sampling data may actually be necessary to make species determinations, while for others (e.g. auditory detections using speech recognition algorithms), I_{ij} may equal one for every record. The investigator assigns each automatic detection with $I_{ij} = 1$ an observation type, O_{ij}. There is considerable latitude in selecting species classification schemes (see e.g. Species misclassification model and subsequent examples). The investigator

Table 1. Definitions of parameters and data used in the hierarchical model for automatic detection and double sampling data. Symbols appearing in boldface represent vectors or matrices

	Definition
Parameter	
N_s	Total abundance of species s in the study area ($= \sum_j N_{js}$)
N_{js}	Abundance of species s in sampling unit j ($N_{js} = N_{js}^{\text{obs}} + N_{js}^{*} + N_{js}^{**}$)
N_{js}^{obs}	Number of observed animals in sampling unit j that are truly of species s
N_{js}^{*}	Number of undetected animals in surveyed regions of sampling unit j that are of species s
N_{js}^{**}	Abundance of species s in unsurveyed regions of sampling unit j
G_{js}	Number of groups of animals of species s located in sampling unit j
G_{js}^{obs}	Number of groups of animals of species s located in the surveyed region of sampling unit j detected by the automatic detection system
ν_{js}	The log of abundance intensity for species s in sampling unit j
$\tau_{\nu s}$	Precision of the log of abundance intensity for species s; possibly used to impart overdispersion relative to the Poisson distribution
$\tau_{\eta s}$	Precision parameter for spatial random effects associated with species s
λ_{js}	Abundance intensity for species s in sampling unit j ($\lambda_{js} = A_j R_j p_{js} \exp(\nu_{js})$)
$\boldsymbol{\beta}_s$	Parameters of the linear predictor describing variation in the log of abundance intensity as a function of landscape and habitat covariates for species s
$\boldsymbol{\eta}_s$	Vector of spatial random effects for species s
$\boldsymbol{\alpha}_s$	Vector of reduced-dimension random effects for species s [when restricted spatial regression (RSR) is employed]
$\boldsymbol{\theta}_s$	Parameters describing the distribution of individual covariates at the population level for species s
S_{ij}	True species associated with the ith automatic detection obtained while surveying sampling unit j
$\pi_{ij}^{O\|s}$	Probability that the ith group of animals encountered while surveying sampling unit j are assigned observation type O given that they are truly of species s.
p_{js}	Probability that a member of species s associated with the area surveyed in sampling unit j is detected ($p_{js} = p_s a_{js}$)
a_{js}	Probability that an animal of species s is available to be detected at the time(s) when surveys are conducted in sampling unit j (for seals, this is their haul-out probability)
p_s	Probability that a member of species s will be detected by the automatic detection system given that it is available to be detected
Data	
Y_j	Total count of automatic detections recorded during surveys of sampling unit j
Z_{ijk}	The value of the kth individual covariate associated with automatic detection i in sampling unit j
I_{ij}	Indicator for whether the ith automatic detection recorded in the jth sampling unit was also subject to double sampling
$\mathbf{X}_s$	Design matrix associated with abundance intensity model for species s
A_j	The area of sampling unit j (perhaps scaled to its mean)
R_j	Proportion of sampling unit j that is sampled via the automatic detection method during the survey
O_{ij}	Observation type for the ith automatic detection in sampling unit j (e.g. observed species)
J	Total number of sampling units in the study area
L	Total number of sampling units in the study area that are actually sampled
$\mathbf{W}$	Association matrix describing spatial neighbourhood structure of sampling units
$\mathbf{Q}$	Structure matrix for spatial random effects (note the precision matrix for random effects is given by $\tau_{\eta s}\mathbf{Q}$)
$\mathbf{K}_s$	Design matrix for spatial random effects when dimension reduction (RSR) is employed

also records any individual covariates, Z_{ijk} (e.g. group size), where k identifies the kth covariate (Table 1). For the present development, we require that covariates are available for each record where $I_{ij} = 1$.

MODEL

The observed data include a set of species classifications for each sampling unit, a count of unclassified automatic detections for each sampling unit (i.e. those for which $I_{ij} = 0$), together with individual covariates such as group size. We also allow for the possibility that the investigator has auxiliary data (through double sampling or some other mechanism) to estimate components of the detection process (e.g. detection probability, species misclassification probabilities) and has gathered habitat covariates to help explain variation in abundance. Our next task shall be to devise a way to conduct inference on animal abundance and species–habitat relationships from such a seemingly disparate data amalgam.

When conceptualizing how the observed data arise, we find it intuitive to break the problem down into several components within a hierarchical modelling framework (e.g. Fig. 1). First, we consider the way in which expected abundance for each species varies over the landscape. When space is discretized into individual sampling units (as we have done here), a common way to relate counts to habitat covariates is through a spatial regression model. In our case, we do not know the actual abundance in each sampling unit, but we can still borrow this framework to describe variation in expected abundance in each cell. Secondly, realized animal counts in a given sampling unit will typically be different than the expected abundance for several reasons, including random variation, incomplete coverage of the sampling unit and detection probabilities that are <1. We refer to the model describing the relationship between true species counts and expected abundance as a 'Local abundance model'. Finally, the type of observations that are spawned when a group of animals is detected depends on (i) an observation process relating the true species to different observations classifications and (ii) a process relating the true species to individual covariate values. We refer to models for these

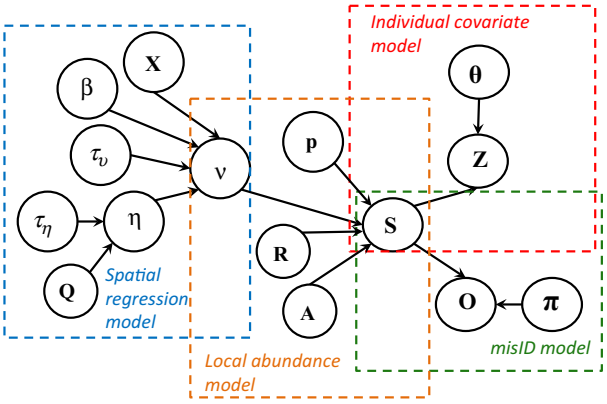

Fig. 1. Directed, acyclic graph for the model proposed for multispecies abundance estimation from thermal imagery and digital photography (adapted from Conn *et al.* 2013, fig. A1). Notation is defined in Table (subscripts and superscripts omitted for clarity).

processes as the 'Species misclassification model' and 'Individual covariate model', respectively.

We now describe each of these four components (spatial regression, local abundance, species misidentification and individual covariate models) in turn. In doing so, we make a number of distributional choices that may require refinement in certain sampling scenarios (see Discussion). We then describe Markov chain Monte Carlo (MCMC) methods and approaches for generating posterior predictions of abundance across the study area. Throughout, we use bold symbols to denote vectors and matrices. For a fuller mathematical treatment, see Appendix S1.

Spatial regression model

For each species s, we write the log of expected abundance in each sampling unit as a function of habitat covariates, spatially autocorrelated random effects and unstructured random effects. For the moment, we treat all sampling units as if they were the same size (adjustments for unequal area are made in the following section). In particular, we express the log of expected abundance (ν_{js}) across the collection of sampling units as

$$\boldsymbol{\nu}_s = \mathbf{X}_s\boldsymbol{\beta}_s + \boldsymbol{\eta}_s + \boldsymbol{\epsilon}_s, \qquad \text{eqn 1}$$

where $\mathbf{X}_s$ denotes a design matrix relating environmental and habitat covariates to expected abundance, $\boldsymbol{\beta}_s$ gives related regression parameters, the $\boldsymbol{\eta}_s$ specify random effects with spatially autocorrelated, Gaussian errors, and $\boldsymbol{\epsilon}_s$ represents mean zero Gaussian error with precision parameter $\tau_{\nu s}$.

There are several common choices for inducing spatial autocorrelation in hierarchical spatial regression models (see e.g. Banerjee, Carlin & Gelfand 2004). In the following, we specify an intrinsic conditionally autoregressive prior distribution (ICAR; Besag & Kooperberg 1995; Rue & Held 2005) for $\boldsymbol{\eta}_s$ such that

$$\boldsymbol{\eta}_s \sim \mathcal{N}\left(\mathbf{0}, (\tau_{\eta s}\mathbf{Q})^{-1}\right),$$

where $\mathcal{N}()$ denotes a multivariate normal (Gaussian) distribution and $\tau_{\eta s}\mathbf{Q}$ gives precision of the Gaussian spatial process. Here, $\tau_{\eta s}$ is a precision parameter to be estimated, and $\mathbf{Q}$ is defined as $\mathbf{Q} = \mathbf{D} - \mathbf{W}$, where $\mathbf{W}$ is an association matrix describing the spatial neighbourhood structure of sampling units and $\mathbf{D}$ is a diagonal matrix with elements $-\mathbf{W}\mathbf{1}$ ($\mathbf{1}$ being a column vector of ones). For purposes of this paper, we use a formulation for $\mathbf{W}$ that approximates thin-plate splines (Rue & Held 2005, section 3.4.2). This approach implies a greater degree of smoothing than first-order formulations for $\mathbf{Q}$, a potentially useful feature when analysing sparse data from abundance surveys (see Discussion). In an effort to eliminate parameter redundancy and confounding between spatial regression parameters and spatial random effects, we also implemented a restricted spatial regression (RSR; Reich, Hodges & Zadnik 2006; Hodges & Reich 2010; Hughes & Haran 2013) version of eqn 1 (see Appendix S1 for further details).

Table 2. Species classification probabilities used in the hierarchical seal abundance model. True species appear in columns, while observation types occur on rows. The column (and row) for 'Other' indicate non-seals (e.g. thermal anomalies, non-target taxa)

			True species				
Obs Index	Obs species	Confidence	Bearded	Ribbon	Ringed	Spotted	Other
1	Bearded	Certain	$\pi^{1\|1}$	0	0	0	0
2	Bearded	Likely	$\pi^{2\|1}$	$\pi^{2\|2}$	$\pi^{2\|3}$	$\pi^{2\|4}$	0
3	Bearded	Guess	$\pi^{3\|1}$	$\pi^{3\|2}$	$\pi^{3\|3}$	$\pi^{3\|4}$	0
4	Ribbon	Certain	0	$\pi^{4\|2}$	0	0	0
5	Ribbon	Likely	$\pi^{5\|1}$	$\pi^{5\|2}$	$\pi^{5\|3}$	$\pi^{5\|4}$	0
6	Ribbon	Guess	$\pi^{6\|1}$	$\pi^{6\|2}$	$\pi^{6\|3}$	$\pi^{6\|4}$	0
7	Ringed	Certain	0	0	$\pi^{7\|3}$	0	0
8	Ringed	Likely	$\pi^{8\|1}$	$\pi^{8\|2}$	$\pi^{8\|3}$	$\pi^{8\|4}$	0
9	Ringed	Guess	$\pi^{9\|1}$	$\pi^{9\|2}$	$\pi^{9\|3}$	$\pi^{9\|4}$	0
10	Spotted	Certain	0	0	0	$\pi^{10\|4}$	0
11	Spotted	Likely	$\pi^{11\|1}$	$\pi^{11\|2}$	$\pi^{11\|3}$	$\pi^{11\|4}$	0
12	Spotted	Guess	$\pi^{12\|1}$	$\pi^{12\|2}$	$\pi^{12\|3}$	$\pi^{12\|4}$	0
13	Other	NA	0	0	0	0	1
14	Unknown	NA	$\pi^{14\|1}$	$\pi^{14\|2}$	$\pi^{14\|3}$	$\pi^{14\|4}$	0

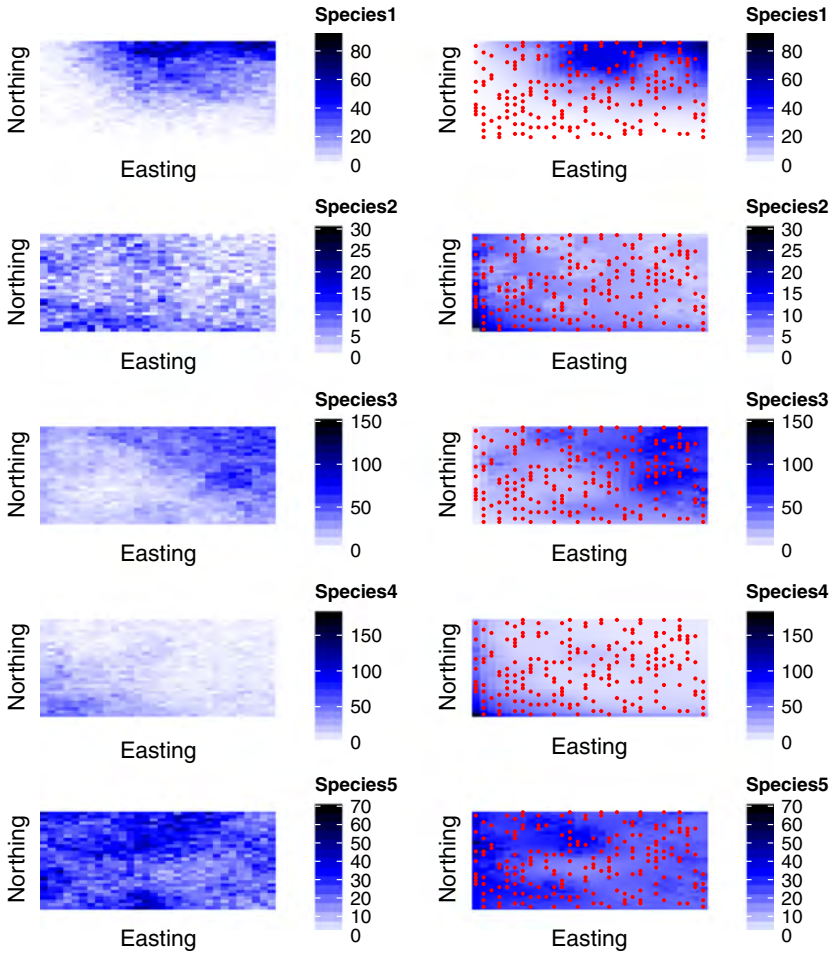

Fig. 2. Simulated (left panels) and estimated (right panels) abundance across a landscape for five hypothetical species. Red circles on estimated abundance panels indicate sampled cells.

Local abundance model

The preceding formulation describes variation in the log of abundance intensity, but does not include other factors affecting the expected number of animals encountered by surveys in a given cell. For instance, sampling units may vary in size, the proportion of area surveyed may vary across sampled units, automatic detections may miss animals, and not all animals associated with sampling unit j may be present while surveys are being conducted. Also, we expect random fluctuations in the number of animals present relative to the expected abundance intensity. For these reasons, we model the number of automatic detections of species s in sampling unit j ($G_{js}^{\text{obs}} = \sum_i I_{[S_{ij}=s]}$) as $G_{js}^{\text{obs}} \sim \text{Poisson}(\lambda_{js})$ where $\lambda_{js} = A_j R_j p_{js} \exp(\nu_{js})$ and $p_{js} = a_{js} p_s$ (recall that notation is defined in Table 1). In conjunction with our choice of a Gaussian distribution for ν_{js}, this formulation implies that the actual number of detections for each species is a realization of a thinned version of the log-Gaussian Cox process (see e.g. Rathbun & Cressie 1994; Møller, Syversveen & Waagepetersen 1998).

The data collected on aerial surveys do not provide sufficient information to estimate availability, (a_{js}), so auxiliary data or strong priors are needed on these parameters (see e.g. Example: Ice-Associated Seals). One approach for getting information on p_s is to conduct an unaided search of double sampling data, and to treat animals found in the unaided search as trials to test the false-negative rate of the automatic detection algorithm; in this case, the number of successful automatic detections can be treated as binomial with success probability p_s.

Species misclassification model

The preceding sections describe how animals (or animal groups) of each species are detected. However, in order to allow imperfect species observations, we need to specify a model relating the true species to actual observations. For observations where species can be assigned ($I_{ij} = 1$), we suppose that observations O_{ij} arise according to a multinomial process conditional on the true species S_{ij} and classification probabilities $\pi_{ij}^{O_{ij}|S_{ij}}$. In practice, this specification requires that we treat the true species as a latent parameter (i.e. that we admit uncertainty about its value).

Automatic detection data are typically not sufficient to estimate the misclassification parameters, $\boldsymbol{\pi}$, so strong priors or

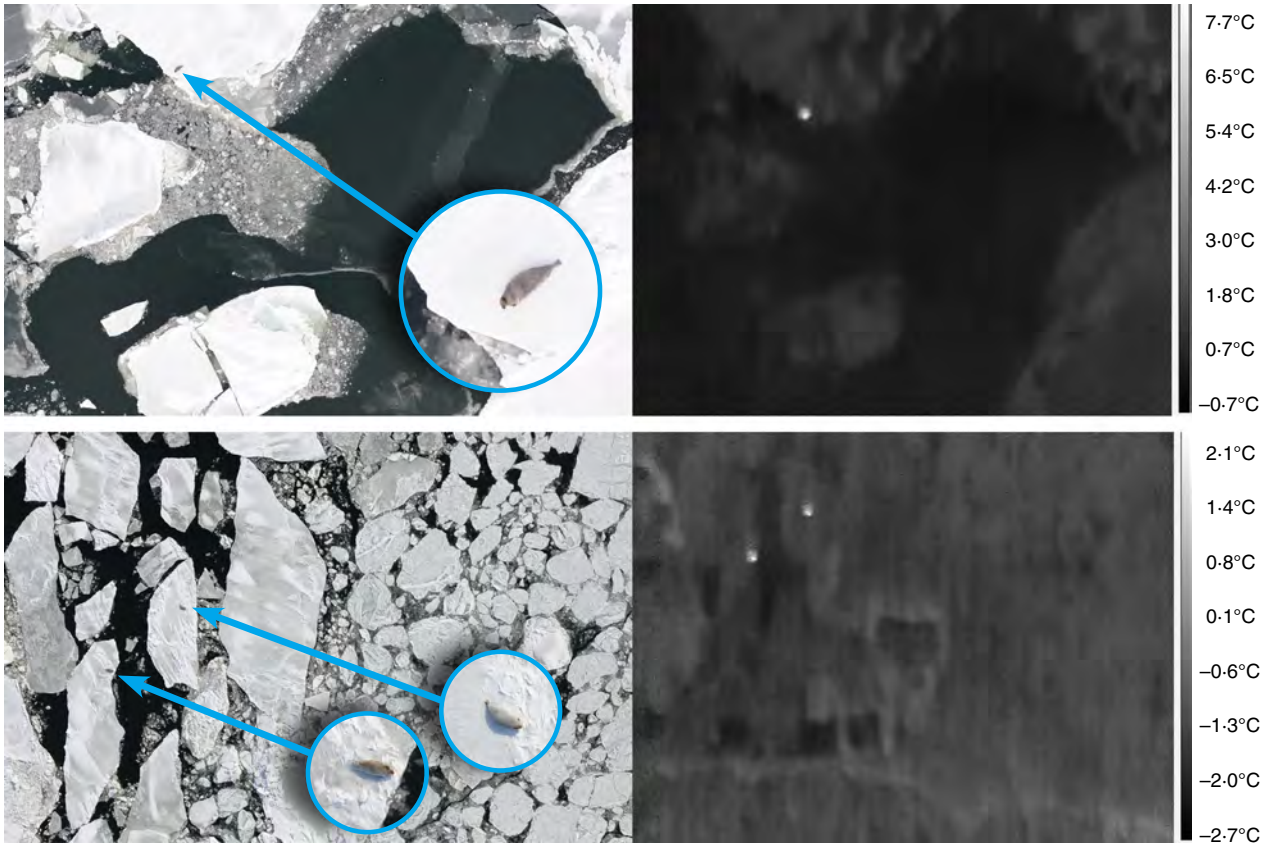

Fig. 3. A composite image showing two high-resolution digital photographs (left) with a matched thermal hot spots (right). Thermal videos are screened for such hot spots, and corresponding photographs are searched (when available) to provide information on species identity.

auxiliary data are needed to provide structure on these (see subsequent examples). In the following sections, we suppress dependence on individual and transect (i.e. we set $\pi_{ij}^{O_{ij}|S_{ij}} = \pi^{O_{ij}|S_{ij}}$). However, we suspect that expressing classification parameters as a function of covariates using a multinomial logit link (Agresti 2002) will be useful in future applications (see e.g. Conn *et al.* 2013, for further discussion). Although there is considerable flexibility for structuring the species classification matrix, we use a formulation specifically tailored to our seal study in subsequent applications. This formulation requires observers to classify observations by both species and certainty, and also permits them to record species as 'unknown' or 'other' (Table 2). The 'other' category accounts for false positives.

Individual covariate model

We allow for the possibility that automatic detection i in sample unit j has k associated individual covariates, which are only assumed to be observed if $I_{ij} = 1$ (see Discussion). The most important of these is likely group size, the number of animals that are associated with a specific automatic detection. In cases where automatic detections can consist of more than one animal, one must include the group size distribution when generating an overall abundance estimate. Our approach is to parametrically model covariates as

$$Z_{ijk} \sim f_{S_{ij}}(\boldsymbol{\theta}_{S_{ij}}),$$

where $f_s(\boldsymbol{\theta}_s)$ gives a probability mass or density function with parameters specific to species s. In the applications that follow, we only use one individual covariate (group size), which we give a zero-truncated Poisson distribution with a species-specific intensity parameter to be estimated. We also require that each automatic detection be composed of like species.

MCMC sampler

Writing our model hierarchically, we are able to envision how broad landscape-scale processes can ultimately be translated into observed data in a probabilistic fashion. Provided that we believe our model is a reasonable approximation to reality and are willing to assign prior distributions for model parameters, Bayesian calculus provides a convenient way of making inference about the data-generating process (including parameters describing species–habitat relationships and animal abundance). We used a hybrid Gibbs–Metropolis sampler to draw samples from the joint posterior distribution symbolically specified by Fig. 1 (see Appendix S1 for a mathematical specification). This involves iteratively sampling model parameters from their full conditional distributions. Owing to our judicious choice of Gaussian error on log-scale abundance intensity (i.e. ν_{js}), many of the parame-

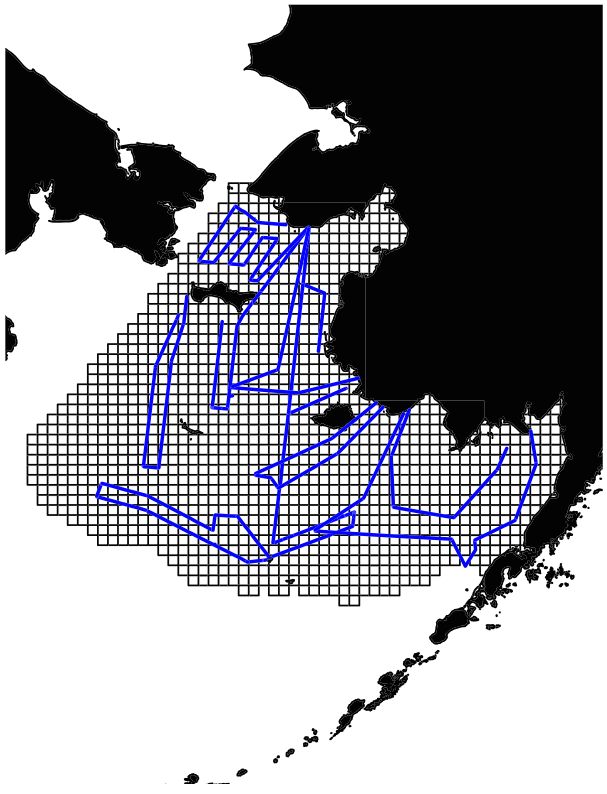

Fig. 4. Map of eastern Bering Sea study area showing 25 × 25 km sampling units and survey lines for flights that were included in the analysis. The western boundary of the study area was determined by the U.S. Exclusive Economic Zone (EEZ); the southern boundary was determined by limiting analysis to cells that had ≥1% sea ice for at least one day from 1 April 2012 to 20 May 2012. Cells comprised of >99% land were excluded from analysis.

ter groups can be sampled directly using the same strategies commonly used in Bayesian analysis of linear models (see e.g. Gelman *et al.* 2004, Chapter 14). Our strategy for updating parameters shares many features with some of our past work on distance sampling (see e.g. Conn, Laake & Johnson 2012; Conn *et al.* 2013) and is presented in Appendix S1.

Posterior prediction and model comparison

Our models provide estimates of parameters explaining variation in animal abundance (i.e. spatial regression parameters) as well as species-specific abundance estimates for animals detected in sampled areas. To extend inference to the abundance of species s over the entire landscape, we use posterior predictive distributions. For sampling units that we do not sample, posterior predictions can be simulated as

$$N_{js} \sim \text{Poisson}(\exp(\nu_{js}) R_j A_j (1 + \theta_s)),$$

where θ_s gives the zero-truncated Poisson intensity parameter for group size.

For units that are sampled, we generate posterior samples of abundance that are a combination of (i) animals detected during surveys ($N_{js}^{\text{obs}} = \sum_i I_{[S_{ij}=s]} Z_{ij1}$; note that the total number of such animals is fixed, but species can vary), (ii) animals that are associated with sampled areas, but were not detected (N_{js}^{*}), and (iii) animals in portions of sampled cells that were not surveyed (N_{js}^{**}), such that $N_{js} = N_{js}^{\text{obs}} + N_{js}^{*} + N_{js}^{**}$. This approach is attractive in that it implicitly includes a finite population correction. For instance, if all animals are detected in a given sampling unit and there is no species misclassification, then abundance in that unit is known with certainty (Ver Hoef 2008; Johnson, Laake & Ver Hoef 2010). Specifically, we can generate abundance predictions as

$$N_{js}^{*} \sim \text{Poisson}(\exp(\nu_{js}) R_j A_j (1 - p_{js})(1 + \theta_s)), \text{and}$$

$$N_{js}^{**} \sim \text{Poisson}(\exp(\nu_{js})(1 - R_j) A_j (1 + \theta_s)).$$

Predictions of total abundance across the study area can then be calculated as $N_s = \sum_j N_{js}$.

We also compute a posterior predictive loss statistic to compare the performance of alternative models with different sources of variation in modelled abundance (e.g. different combinations of covariates, presence/absence of spatial autocorrelation). Suggested by Gelfand & Ghosh (1998), this approach measures the ability of a given model to generate data sets similar to the one collected. In particular, a loss statistic is computed for each model m as $\mathcal{D}_m = \mathcal{G}_m + \mathcal{P}_m$, where $\mathcal{G}_m$ is a measure of posterior loss and $\mathcal{P}_m$ is a penalty for variance. Models with a smaller overall $\mathcal{D}_m$ are favoured in this context. Our implementation largely follows that of Conn *et al.* (2013); see Appendix S1 for further details.

EXAMPLE: SIMULATED DATA

To assess the ability of our proposed model to accurately estimate abundance, we simulated a survey of four species over a 30 × 30 grid ($J = 900$). We generated true abundance using the same general model structure as used in estimation. Log abundance for each species was generated as a function of several covariates (easting, northing and a Matern-distributed hypothetical covariate) as well as spatially correlated error where spatial random effects were generated for each species assuming an ICAR ($\tau = 20$) distribution. Covariate relationships were configured such that for species one, abundance intensity increased linearly in both eastern and northern directions; species two exhibited a low but constant abundance across the landscape; species three exhibited a high abundance on the western edge of the landscape which declined slightly towards the east; and species four had a strong relationship with the hypothetical covariate. We also included a fifth 'species' in an attempt to mimic anomalous readings (false positives), where expected abundance intensity was set to be constant across the landscape. In some cases, the ICAR random effects obscured the covariate relationships (Fig. 2).

We simulated a survey over 200 randomly selected grid cells, assuming that each survey covered 10% of its target sampling unit (Fig. 2) and that double sampling was conducted for 80% of automatic detections. Total sample coverage was thus ≈2·2% of the population. The observation model was built to resemble our seal example (see Example: Ice-Associated Seals);

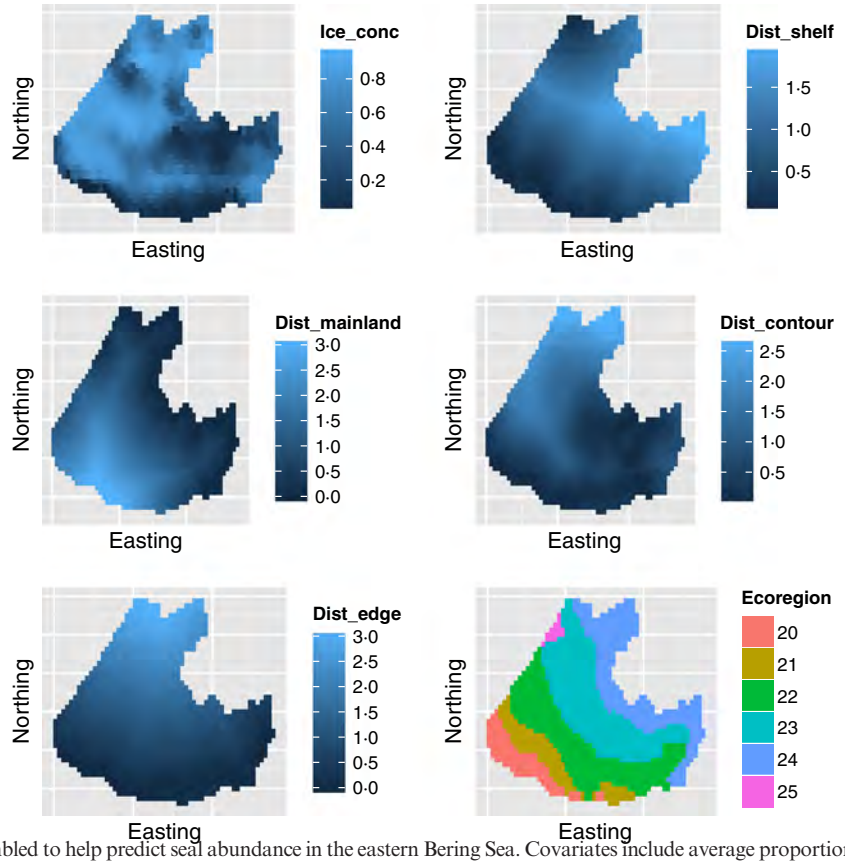

Fig. 5. Covariates assembled to help predict seal abundance in the eastern Bering Sea. Covariates include average proportion of sea ice while surveys were conducted ('ice_conc'), distance from 1000-m depth contour ('dist_shelf'), distance from mainland ('dist_mainland'), distance from 10% sea ice contour ('dist_contour'), distance from southern sea ice edge ('dist_edge') and ecoregion (see Piatt & Springer 2007). All covariates other than ice_conc and ecoregion were standardized to have a mean of 1·0 prior to plotting and analysis. Unsampled ecoregions were combined with the closest sampled ecoregion for estimation.

double-sampled animals could be classified as belonging to any of the four target 'species' or could be recorded as 'unknown' or 'other'. In addition, there were three classes of target species classification certainty: 'certain', 'likely' and 'guess' (Table 2). Observations were determined according to a multinomial distribution, with probabilities given in Appendix S2.

We supplied our hierarchical model with the same covariates that were used to generate the data (thus utilizing the 'correct' functional form and assuming no covariate measurement error), and permitted estimation of RSR ICAR random effects. We fixed overdispersion relative to the Poisson distribution to be small ($\tau_v = 100$) to stabilize estimation (see Discussion). We summarized the posterior distribution by running the Markov chain for 600 000 iterations, discarding 100 000 iterations as a burn-in and recording values from every 100th iteration to save disk space. This procedure took ≈ 2·5 days on a 2·93-GHz Dell Precision T1500 desktop with 8·0 GB of RAM.

EXAMPLE: ICE-ASSOCIATED SEALS

We conducted aerial surveys of four ice-associated seal species (bearded seals *Erignathus barbatus*, ribbon seals *Histriophoca fasciata*, ringed seals *Phoca hispida* and spotted seals *Phoca largha*) over the eastern Bering Sea between 10 April and 22 May 2012. Our strategy was to use infrared cameras as an automatic detection procedure and to use a set of independent, automated digital photographs as a form of double sampling (Fig. 3). Two aircraft were used in surveys, a NOAA DeHavilland DHC-6 Twin Otter and an AC-690 Aero Commander. The Twin Otter was configured with three FLIR SC645 far-IR infrared cameras with 25 mm lenses measuring data in the 7·5- to- 13-μm wavelength, each of which was paired with a 21 megapixel high-resolution digital single-lens reflex (SLR) camera fitted with a 100 mm lens. All six cameras were mounted in the belly port of the airplane. To avoid counting the same animal twice, the infrared cameras were mounted such that their thermal swaths abutted each other but did not overlap. Flying at a target altitude of 300 m, this configuration produced a thermal swath width of *c.* 470 m. The Aero Commander was similarly configured with two sets of infrared and SLR cameras, resulting in a thermal swath width of *c.* 280 m. SLRs were automated to take pictures approximately every 1–1·4 s; flying at a target speed of 130 kts, photographs covered ≈84% of the thermal swath.

As the quantity and distribution of sea ice varied considerably over the course of the surveys, we selected 10 flights that provided good spatial coverage within a 1-week period (20–27 April) for analysis (Fig. 4), assuming that abundance was con-

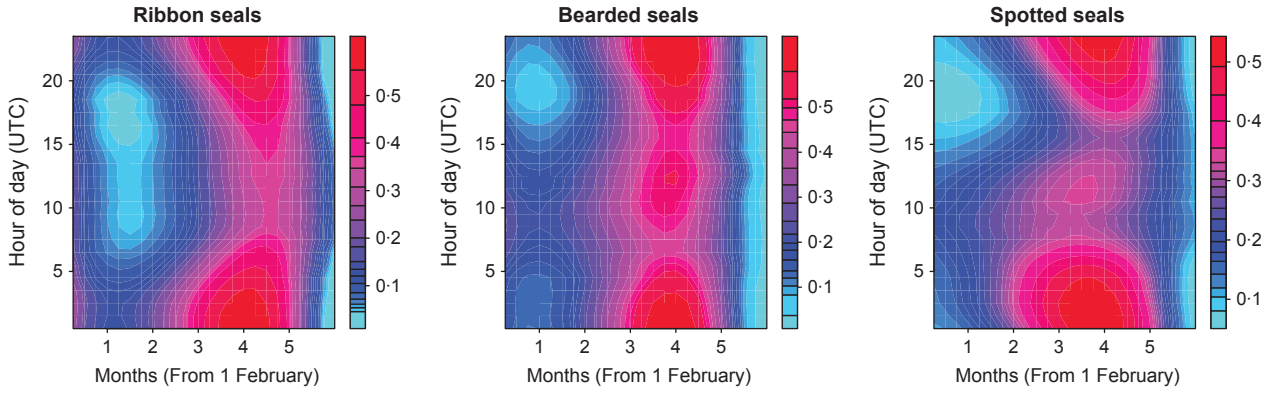

Fig. 6. Predicted haul-out (availability) probability as a function of day of year and time of day for each species. Note that estimates are currently unavailable for ringed seals.

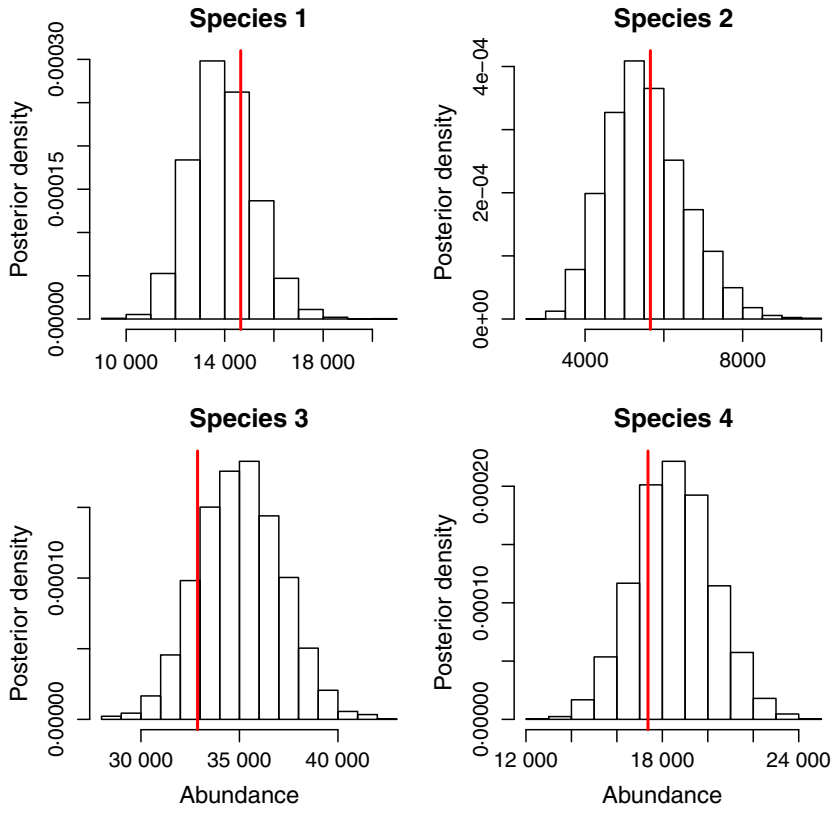

Fig. 7. Posterior predictive distributions for species abundance as estimated from simulated data. True values are indicated in red.

stant over the study area during this period. Analysis was also limited to one set of cameras from each plane. In total, our analysis included 9076 km of survey effort (40·7 h of flying time). We limited effort to times and locations when altitude was 228·6–335·3 m and roll was <2·5° from centre. Aircraft yaw could not be calculated reliably and was not included in area calculations.

We compiled several covariates we thought might be useful in predicting seal abundance in our study area. These included marine ecoregion (cf. Piatt & Springer 2007), distance from mainland, distance from 1000-m depth contour, sea ice concentration, distance from southern ice edge and distance from 10% sea ice contour (Fig. 5). Remotely sensed sea ice data were obtained at a 25 × 25 km resolution from the National Snow and Ice Data Center, Boulder, CO, USA, on an EASE Grid 2.0 projection. We used this projection and same resolution to define sampling units (Fig. 4). Calculations of covariates were made relative to the centroid of each sampling unit.

To estimate the probability of detection associated with infrared detections (p_s), a technician manually searched an independent, systematic random sample of 11 724 digital photographs (out of a total of 117 225 images) for the presence of seals. The technician spent *c.* 120 h searching photographs and found a total of 70 seal groups. We then examined whether

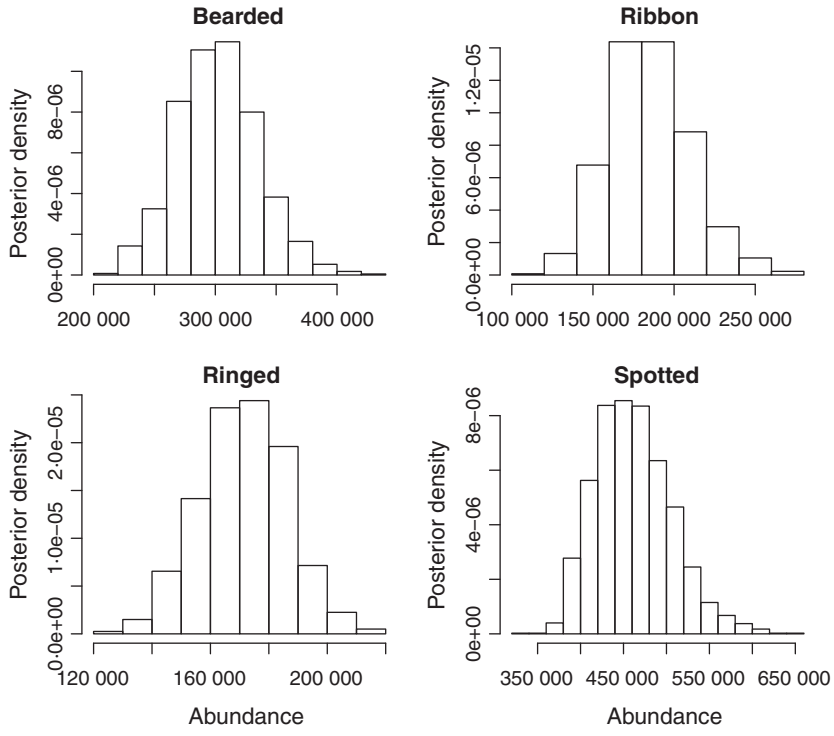

Fig. 8. Posterior predictive distributions of seal abundance in the eastern Bering Sea from Model 1. Estimates of ringed seal abundance are uncorrected for haul-out (availability) probability.

these seal groups were also detected as hot spots using our infrared hot spot detection method, finding that 66 (94·3%) of them were detected. As species could not always be identified, we set $p_s = p$ for all species and used these data to help estimate the overall probability of detection (see below). For reference, the conditional probability of detection for our technician (calculated using seals detected by infrared) was lower at 66/82 = 80·5%.

We obtained data on availability probability (a_{js}) from ARGOS-linked satellite transmitters affixed to spotted, bearded and ribbon seals in the Bering Sea from 2004 through 2012. A conductivity sensor placed on each transmitter provided hourly data on the proportion of time each tag was dry. As in previous analyses (e.g. Bengtson *et al.* 2005; Ver Hoef, London & Boveng 2010), dry-time percentages were converted into Bernoulli responses to analyse seal haul-out behaviour, where a success was recorded whenever tags were mostly (≥50%) dry in a given hour (seals could only be detected by thermal imagery when they were out of water). Because we were only interested in explaining variation in haul-out behaviour during spring, we limited analysis to records between 1 February and 31 July of each year, treating each individual-year combination as an independent replicate (i.e. data for individuals obtained in two separate years were treated as if they were statistically independently). This approach resulted in a total of 19 individual-year combinations for bearded seals, 92 for ribbon seals and 55 for spotted seals. These data were analysed within a generalized linear mixed modelling framework that explicitly acknowledges temporal autocorrelation in responses (see Ver Hoef, London & Boveng 2010). For our purposes, the linear predictor was written as a function of hour of day and day of year. Hour of day was treated as a categorical variable with 24 levels, while day of year was calculated as proportion of year since 1 February. We modelled linear, quadratic and cubic effects for day of year and included all interactions between day of year and hour of day. After separate models were fitted to data for each species, predictions in logit space (Fig. 6) and an associated variance–covariance matrix could be computed for any set of availability probabilities (a_{js}) of interest using standard mixed model theory (see e.g. Littell *et al.* 1996; Ver Hoef *et al.* 2013).

We used the following procedure to produce prior samples of detection probabilities (p_{js}) for surveyed sampling units:

1. Determine an average time of day and day of year when sampling was conducted on each sample unit.

2. For rep ∈ {1, 2, …, 1000}, sample $\text{logit}(\mathbf{a}_s^{\text{rep}}) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ gives mixed model haul-out (availability) predictions in logit space and $\boldsymbol{\Sigma}$ gives the prediction variance–covariance matrix.

3. For rep ∈ {1, 2, …, 1000}, sample infrared detection probability as $p_s^{\text{rep}} \sim \text{Beta}(67, 5)$. This formulation implies a conjugate Beta(1,1) prior on p_s.

4. Compute samples of detection probability (availability × infrared detection) by sampling unit as $\mathbf{p}_s^{\text{rep}} = \mathbf{a}_s^{\text{rep}} p_s^{\text{rep}}$.

Samples of $\mathbf{p}_s^{\text{rep}}$ could then be used as a prior distribution within a Metropolis–Hastings step to account for detection probabilities that varied by hour, day of year and species (see Appendix S1 for further details). Note that there were no availability data for ringed seals, so a_{js} was set to 1·0. As such, ringed seal abundance estimates are uncorrected for availability.

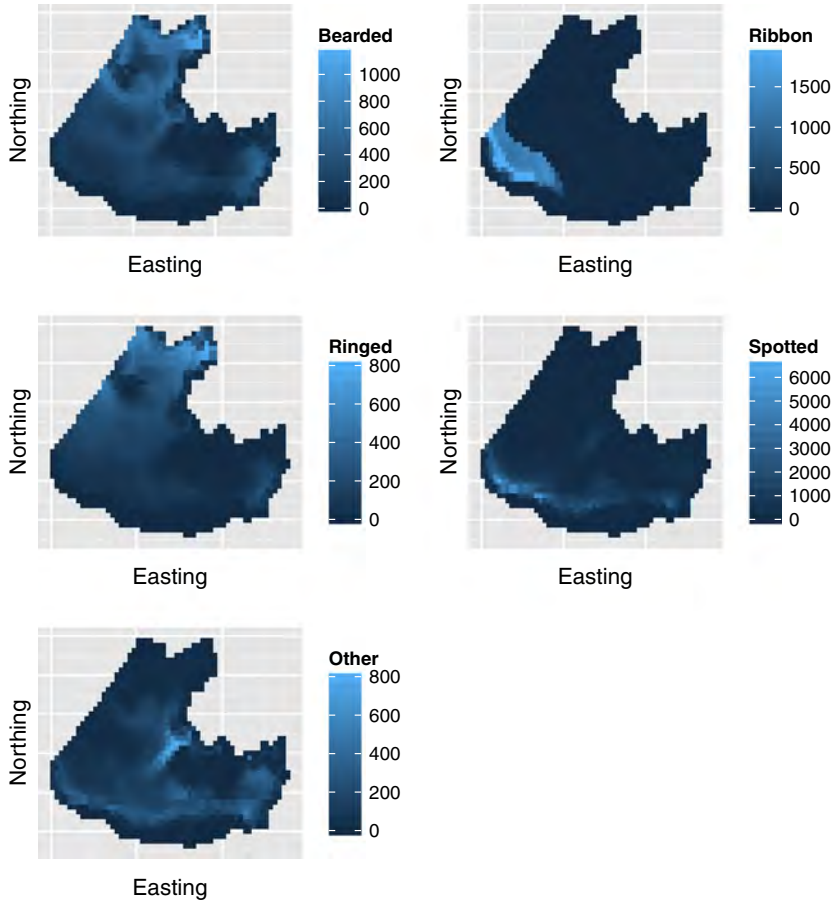

Fig. 9. Mean posterior predictions of seal abundance across our study area in the eastern Bering Sea. Legends indicate posterior predictions of abundance in 25 × 25 km grid cells. Abundance for 'other' indicates abundance of other heat signatures that were not seals (e.g. sea lions, walrus, melt pools, birds).

An independent experiment was performed to generate a prior distribution of species classification probabilities (B. McClintock, unpublished data). This analysis used readings by multiple observers and certainty categories (certain, likely, guess) to produce posterior predictions of classification probabilities, with a constraint that observations recorded as 'certain' were 100% accurate. These predictions were used directly as a joint prior distribution for the species classification matrix (see Appendix S1). The classification matrix specified by the posterior mean of these predictions is provided in Appendix S2.

We considered several model formulations for each species. Based on prior surveys in the region (see e.g. Conn *et al.* 2013; Ver Hoef *et al.* 2013), our *a priori* expectation was that ribbon and spotted seals would be concentrated in the southern portions of our study area, whereas bearded and ringed seals would be primarily located farther north. We also expected that abundance would be nonlinearly related to sea ice concentration, where zero seals would be detected in cells with no ice and few seals (possibly with the exception of ringed seals) would be detected in cells with 100% ice. Ideally, a model for ribbon seal abundance would be written as a function of the distance from the continental shelf, where nutrient upwelling supports an abundant prey base. However, models with continuous predictors proved problematic for ribbon seals, as covariates (and combinations of covariates) were often maximized in the south-west corner of our study area, producing estimates of abundance that were unbelievably high (note that there were considerable gaps in sampling in this region). To avoid extrapolation past the range of observed data, we thus wrote all models for ribbon seals as a function of ecoregion and sea ice only. For the remaining species, we fit two possible models to the data. In the first, the log of abundance intensity was written as an additive function of *ice_conc*, ice_conc^2, *dist_mainland*, *dist_shelf*, *dist_contour* and *dist_edge*. In the second model, the log of abundance intensity was the same as ribbon seals, namely an additive function of *ice_conc*, ice_conc^2 and *ecoregion*.

We initially tried to fit models that included RSR ICAR random effects, but these often produced overinflated estimates of abundance in areas where there were large gaps in spatial coverage, even when the spatial neighbourhood defining the **Q** structure matrix was a relatively smooth RW2 structure (as in Rue & Held 2005, section 3.4.2). As such, we limit estimation to pure trend surface models that do not include spatial autocorrelation (i.e. $\mathbf{v}_s = \mathbf{X}_s\boldsymbol{\beta}_s + \boldsymbol{\varepsilon}_s$), acknowledging that posterior predictions of abundance likely overstate precision (see Discussion). Initial runs also produced positive predictions of seal abundance in cells without ice, likely because we only surveyed

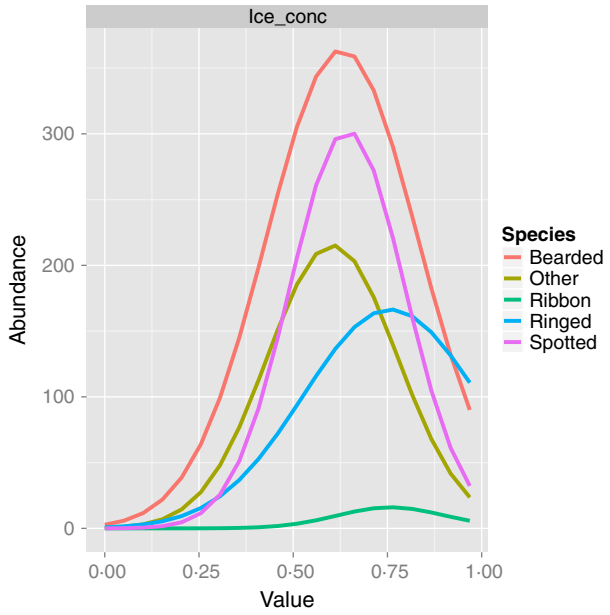

Fig. 10. Mean posterior prediction of abundance for each seal species as a function of sea ice concentration. Predictions for each species were made by setting all other modelled covariates to their means, so that they are best interpreted as the relative effect of sea ice on a species-specific basis (absolute values of predictions are not necessarily biologically meaningful). For instance, predicted ribbon seal abundance was calculated by averaging predicted abundance over all ecoregions, some of which had a predicted abundance near zero.

cells that had ice. To anchor this intercept at zero, we introduced dummy data into estimation that indicated we encountered zero seals in cells with <0·1% sea ice. As with the simulated data example, we set $\tau_{vs} = 100$ and summarized the posterior distribution by running the Markov chain for 600 000 iterations, discarding 100 000 iterations as a burn-in and recording values from every 100th iteration to save disk space. This procedure took ≈3·5 days on a 2·93-GHz Dell Precision T1500 desktop with 8·0 GB of RAM.

Results

SIMULATED DATA

Posterior predictive distributions for estimated abundance reasonably approximated the spatial distribution for each species (Fig. 2), and posterior predictive distributions of total abundance captured true abundance in all cases (Fig. 7). This suggests that our estimation scheme produces reasonable estimates, at least for the sample coverage (≈2·2%) and high frequency of double sampling (80%) assumed here.

ICE-ASSOCIATED SEALS

Our posterior loss statistic favoured model one (with continuous covariates for all species other than ribbon seals; $\mathcal{D}_1 = 4066$) over model two (where ecoregion was used for all species; $\mathcal{D}_2 = 4118$), although estimated seal abundance was similar for each. Patterns in seal abundance conformed to our *a priori* expectations regarding species distributions for each model; for brevity, we present overall abundance estimates (Fig. 8) and mean posterior prediction abundance maps (Fig. 9) from model 1 only. Posterior mean density estimates, calculated using an effective study area of 767 114 km^2, were 0·39 bearded seals km^{-2} (95% CI 0·32–0·47), 0·24 ribbon seals km^{-2} (95% CI 0·19–0·30) and 0·60 spotted seals km^{-2} (95% CI 0·51–0·73). We also were able to estimate the relationship between seal abundance and ice concentration, finding that for most species abundance was maximized when the proportion of sea ice in a sampling unit was in the 0·6–0·8 range (Fig. 10).

Discussion

Automated detection systems offer several potential advantages over human observer surveys. For example, infrared survey flights can be flown faster and at higher altitudes than conventional (human observer) surveys, increasing the effective area that can be surveyed, decreasing the likelihood of animal disturbance and making surveys safer for pilots and crew. Surveys using automated detection devices have the added advantage of providing a physical, archivable record of animal detections. However, such surveys can still miss animals or pick up non-target signatures. Here, we have shown that double sampling (in the seal example, digital photography) is a viable avenue for allowing species-specific inferences about abundance from automated detection data. However, this approach requires rather sophisticated hardware and software, as well as modelling techniques to account for the vagaries of the detection process, including imperfect detection, availability <1, anomalies (false positives) and species misclassification (note that these factors also occur in studies with human observers, even if they are usually ignored!). Despite the complexity, the simulation study suggested that our approach is capable of estimating maps of species distributions that capture large-scale trends in abundance, with posterior predictive distributions of total abundance including true values.

The subset of seal data we used was quite sparse, with survey tracks covering about 0·4% of the study area. Nevertheless, we were able to fit trend surface models to these data and generate posterior predictions for abundance that largely reflected our *a priori* expectations. For instance, our seal density estimates compared favourably to results from 2006 helicopter transect surveys over a 279 880-km^2 subset of our study area (Ver Hoef *et al.* 2013), where densities were estimated as 0·22 bearded seals km^{-2} (95% CI 0·12–0·61), 0·22 ribbon seals km^{-2} (95% CI 0·13–0·68) and 0·84 spotted seals km^{-2} (95% CI 0·49–2·83). In addition to the actual numbers, the relationships between abundance and underlying landscape and environmental covariates may also be of interest. For instance, we were able to relate seal abundance to landscape features (e.g. distance from land), remotely-sensed sea ice data and ecoregion, and to compare alternative models via a posterior loss statistic. Seal density appeared to peak at slightly higher values of sea ice concentration than previously observed (cf. Ver Hoef *et al.* 2013), possibly due to the uncharacteristically high levels

of ice in the Bering Sea in 2012. We plan to build upon this modelling framework to arrive at more definitive estimates of seal abundance and covariate relationships in the near future. This effort will likely include adding a temporal dimension in the process model to account for changing sea ice conditions (Ver Hoef *et al*. 2013) and expanding the survey grid to include data from concurrent Russian surveys in the western Bering Sea. Owing to current CPU run-times (i.e. 3·5 days for the seal analysis), this effort will also likely require improvements to computer code (e.g. by using parallel processing).

Although not presented here, our experience with fitting models to both simulated and real data is that there needs to be relatively intense spatial coverage to support estimation of overdispersion (i.e. τ_{vs}) and/or spatial random effects using our modelling approach. Since modelling occurs on the log of abundance intensity, the tendency with overparameterized models is for positive bias, particularly in unsampled cells. The robustness of our approach is likely viewed along a continuum. With low spatial coverage, trend surface models (i.e. those without spatial autocorrelation) may still do a reliable job of predicting abundance at the expense of overstated precision. However, even with trend surface models, investigators should take care to avoid situations where the linear predictor for abundance has maximum values in unsampled areas. With higher levels of spatial coverage (and low species misclassification rates), estimation of spatial random effects and overdispersion may be more reliable, particularly when considering reduced rank spatial models like the RSR approach outlined in Appendix S1.

The methods developed in this paper were largely motivated by our seal data, and we recognize that further developments and refinements may be needed when different automatic detection systems and double sampling strategies are employed. For example, our use of double sampling data to estimate detection probability implicitly relies on the assumption that animal detections in each data set (automatic detection, double sampling) are independent. We think this assumption is reasonable in our seal example, but would likely fail in terrestrial applications where habitat cover affects thermal and visual detections similarly (Franke *et al*. 2012). For many terrestrial applications, as well as surveys using automated image processing, an alternative double sampling data set would likely be needed (e.g. using surveys of known animals). For auditory surveys, assessment of error rates could be conducted using test data sets where true species is known. However, auditory surveys would likely need to account for additional factors such as cue rate and variation in auditory detection distances (Marques *et al*. 2013), as well as availability probability (Diefenbach *et al*. 2007).

An additional consideration is the amount of double sampling that needs to be conducted. Required coverage largely depends on the amount of information provided by double sampling, as well as the propensity for spatial variation in detection errors. In our seal example, double sampling (i.e. automated digital photography) was used in at least three ways: (i) to provide species observations (including false positives), (ii) to estimate detectability and (iii) to examine species identification errors via an experiment with multiple image readers. Since species distributions and false-positive rates varied considerably across the landscape, it was necessary to have considerable coverage in double sampling data (e.g. 79% of detected hot spots had associated photographs). By contrast, we only searched ≈10% of available images to estimate detection probability and used 716 photographs for our species identification experiment. We did not expect these error rates to vary spatially, and target proportions were largely informed by power analysis (J. Ver Hoef, unpublished data). We do not expect to increase these latter sample sizes in future work (i.e. even when increasing the number of flights), since we expect to use the same technology and transect protocols.

Our approach was to account for missed animals by including detection probability as a thinning parameter relative to the log-Gaussian Cox process, which is likely appropriate for many populations. However, when automated detection of animals is a function of individual-level covariates (e.g. size, distinctiveness), an alternative approach such as data augmentation (Royle 2009; Conn, Laake & Johnson 2012) would likely be necessary since detection probability must then be modelled at the level of the individual animal. Additional approaches to account for overdispersion (e.g. zero inflation, variance inflation factors) would also be useful and are a subject of current research.

It is important to contrast our approach in this paper, which uses double sampling to estimate detection probability, to approaches that rely on temporal replication or distance data. For instance, *N*-mixture models (Royle, Dawson & Bates 2004) also specify a hierarchical framework for spatially replicated animal count data; in this case, a population closure assumption and temporal replication render detection probability estimable. However, since detection probability includes a number of processes (e.g. detectability, availability; cf. Nichols, Thomas & Conn 2009), it is usually not possible to scale up to absolute abundance with *N*-mixture models. Another related approach is hierarchical models for distance sampling data, which rely on the assumption of declining detection with increased distance from the transect line to help estimate detection probability (cf. Schmidt *et al*. 2011; Conn, Laake & Johnson 2012; Ver Hoef *et al*. 2013).

Despite the complexities associated with modelling the detection process, we are optimistic about the future of automated detection systems as a tool for estimating animal abundance over large spatial domains. These tools provide the means to markedly increase survey coverage and reduce data processing times. Hierarchical models, like the one we have developed in this paper, provide a natural framework to combine multiple data sets that can be used to estimate different components of the detection process, and to correctly propagate uncertainties associated with each component into final estimates.

Acknowledgements

We thank all NOAA personnel and contractors that helped collect and process seal data and J. Jansen, D. Johnson, J. Laake and an anonymous reviewer for providing helpful comments on a previous version of this manuscript. Funding

for aerial surveys was provided by the U.S. National Oceanic and Atmospheric Administration and by the U.S. Bureau of Ocean Energy Management (Interagency Agreement M12PG00017). Most of the bearded seal haul-out data were collected and made available by the Native Village of Kotzebue and the Alaska Department of Fish and Game, with support from the Tribal Wildlife Grants Program of the US Fish and Wildlife Service (Grant Number U-4-IT). The satellite telemetry studies were conducted under the authority of Marine Mammal Protection Act Scientific Research Permits 15126, 782-1765, 782-1676 and 358-1787. Views expressed are those of the authors and do not necessarily represent findings or policy of any government agency. Use of trade or brand names does not indicate endorsement by the U.S. government.

Data accessibility

R scripts: uploaded as online supporting information.

Seal survey data: uploaded as online supporting information.

References

Agresti, A. (2002) *Categorical Data Analysis*, 2nd edn. Wiley-Interscience, Hoboken, New Jersey.

Amstrup, S.C., York, G., McDonald, T.L., Nielson, R. & Simac, K. (2004) Detecting denning polar bears with forward-looking infrared (FLIR) imagery. *BioScience*, **54**, 337–344.

Banerjee, S., Carlin, B. & Gelfand, A.E. (2004) *Hierarchical Modeling and Analysis of Spatial Data*. Chapman & Hall/CRC, Boca Raton, FL.

Bengtson, J.L., Hiruki-Raring, L.M., Simpkins, M.A. & Boveng, P.L. (2005) Ringed and bearded seal densities in the eastern Chukchi Sea, 1999–2000. *Polar Biology*, **28**, 833–845.

Bernatas, S. & Nelson, L. (2004) Sightability model for California bighorn sheep in canyonlands using forward-looking infrared (FLIR). *Wildlife Society Bulletin*, **32**, 638–647.

Besag, J. & Kooperberg, C. (1995) On conditional and intrinsic autoregressions. *Biometrika*, **82**, 733–746.

Blumstein, D.T., Mennill, D.J., Clemins, P., Girod, L., Yao, K., Patricelli, G. *et al.* (2011) Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *The Journal of Applied Ecology*, **48**, 758–767.

Chernook, V.I., Kuznetsov, N.B. & Yakovenko, M.Y. (1999) *Multispectral Aerial Survey of Haulouts of Ice Seals*. PINRO Publications, Murmansk, Russia (in Russian).

Conn, P.B., Laake, J.L. & Johnson, D.S. (2012) A hierarchical modeling framework for multiple observer transect surveys. *PLoS ONE*, **7**, e42294.

Conn, P.B., McClintock, B.T., Cameron, M.F., Johnson, D.S., Moreland, E.E. & Boveng, P.L. (2013) Accommodating species identification errors in transect surveys. *Ecology*, in press.

Diefenbach, D.R., Marshall, M., Mattice, J. & Brauning, D. (2007) Incorporating availability for detection in estimates of bird abundance. *The Auk*, **124**, 96–106.

Franke, U., Goll, B., Hohmann, U. & Heurich, M. (2012) Aerial ungulate surveys with a combination of infrared and high-resolution natural colour images. *Animal Biodiversity and Conservation*, **35**, 285–293.

Gelfand, A.E. & Ghosh, S. (1998) Model choice: a minimum posterior predictive loss approach. *Biometrika*, **85**, 1–11.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004) *Bayesian Data Analysis*, 2nd edn. Chapman and Hall, Boca Raton, Florida.

Groom, G., Stjernholm, M., Nielsen, R.D., Fleetwood, A. & Petersen, I.K. (2013) Remote sensing image data and automated analysis to describe marine bird distributions and abundance. *Ecological Informatics*, **14**, 2–8.

Hodges, J. & Reich, B. (2010) Adding spatially-correlated errors can mess up the fixed effects you love. *The American Statistician*, **64**, 325–334.

Hughes, J. & Haran, M. (2013) Dimension reduction and alleviation of confounding for spatial generalized mixed models. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, **75**, 139–159.

Johnson, D., Laake, J. & Ver Hoef, J. (2010) A model-based approach for making ecological inference from distance sampling data. *Biometrics*, **66**, 310–318.

Kissell, R.E. Jr & Nimmo, S.K. (2011) A technique to estimate white-tailed deer *Odocoileus virginanus* density using vertical-looking infrared imagery. *Wildlife Biology*, **17**, 85–92.

Kogan, J.A. & Margoliash, D. (1998) Automated recognition of bird song elements from continuous recordings using dynamic time warping and hidden Markov models: a comparative study. *The Journal of the Acoustical Society of America*, **103**, 2185–2196.

Littell, R.C., Milliken, R.C., Stroup, W.W. & Wolfinger, R. (1996) *SAS System for Mixed Models*. SAS Publishing, Cary, North Carolina.

Møller, J., Syversveen, A.R. & Waagepetersen, R.P. (1998) Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, **25**, 451–482.

Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., Harris, D. & Tyack, P.L. (2013) Estimating animal population density using passive acoustics. *Biological Reviews*, **88**, 287–309.

Moretti, D., Marques, T.A., Thomas, L., DiMarzio, N., Dilley, A., Morrissey, R., McCarthy, E., Ward, J. & Jarvis, S. (2010) A dive counting density estimation method for Blainville's beaked whale (*Mesoplodon densirostris*) using a bottom-mounted hydrophone field as applied to a Mid-Frequency Active (MFA) sonar operation. *Applied Acoustics*, **71**, 1036–1042.

Nichols, J., Thomas, L. & Conn, P. (2009) Inferences about landbird abundance from count data: recent advances and future directions. *Modeling Demographic Processes in Marked Populations*, Vol. 3, Environmental and Ecological Statistics Series (eds D. Thomson, E. Cooch & M. Conroy), pp. 201–236. Springer, New York.

Piatt, J.F. & Springer, A.M. (2007) Marine ecoregions of Alaska. *Long-Term Ecological Change in the Northern Gulf of Alaska*, pp. 522–526. Elsevier, Amsterdam.

Rathbun, S.L. & Cressie, N. (1994) A space-time survival point process for a longleaf pine forest in Southern Georgia. *Journal of the American Statistical Association*, **89**, 1164–1174.

Reich, B., Hodges, J. & Zadnik, V. (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, **62**, 1197–1206.

Royle, J. (2009) Analysis of capture-recapture models with individual covariates using data augmentation. *Biometrics*, **65**, 267–274.

Royle, J., Dawson, D. & Bates, S. (2004) Modeling abundance effects in distance sampling. *Ecology*, **85**, 1591–1597.

Rue, H. & Held, L. (2005) *Gaussian Markov Random Fields*. Chapman & Hall/CR, Boca Raton, Florida.

Schmidt, J., Rattenbury, K., Lawler, J. & Maccluskie, M. (2011) Using distance sampling and hierarchical models to improve estimates of Dall's sheep abundance. *The Journal of Wildlife Management*, **76**, 317–327.

Speckman, S.G., Chernook, V.I., Burn, D.M., Udevitz, M.S., Kochnev, A.A., Vasilev, A. *et al.* (2011) Results and evaluation of a survey to estimate Pacific walrus population size, 2006. *Marine Mammal Science*, **27**, 514–553.

Ver Hoef, J.M. (2008) Spatial methods for plot-based sampling of wildlife populations. *Environmental and Ecological Statistics*, **15**, 3–13.

Ver Hoef, J., London, J. & Boveng, P. (2010) Fast computing of some generalized linear mixed pseudo-models with temporal autocorrelation. *Computational Statistics*, **25**, 39–55.

Ver Hoef, J.M., Cameron, M.F., Boveng, P.L., London, J.M. & Moreland, E.E. (2013) A hierarchical model for abundance of three ice-associated seal species in the eastern Bering Sea. *Statistical Methodology*, doi:10.1016/j.stamet.2013.03.001. Early online.

Waddle, J.H., Thigpen, T.F. & Glorioso, B.M. (2009) Efficacy of automatic vocalization recognition software for anuran monitoring. *Herpetological Conservation and Biology*, **4**, 384–388.

Ward, J.A., Thomas, L., Jarvis, S., Baggenstoss, P., DiMarzio, N., Moretti, D. *et al.* (2012) Passive acoustic density estimation of sperm whales in the Tongue of the Ocean, Bahamas. *Marine Mammal Science*, **28**, E444–E455.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego, California.

Received 16 May 2013; accepted 26 September 2013
Handling Editor: Charles Francis

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. Further mathematical details for estimating animal abundance from a combination of automatic detection and double sampling data.

Appendix S2. Prior means for misclassification parameters used in simulated data and seal analyses.

APPENDIX 2. - McClintock, B. T., E. E. Moreland, J. M. London, S. P. Dahle, G. M. Brady, E. L. Richmond, K. M. Yano, and P. L. Boveng. 2015. Quantitative assessment of species identification in aerial transect surveys for ice-associated seals. Marine Mammal Science 21:1057-1076.

This article describes the method used to quantify the rate of errors in misclassifying species from the digital color photos and also the diagnostic characters that observers used to classify the species.

Marine Mammal Science

MARINE MAMMAL SCIENCE, 31(3): 1057–1076 (July 2015)

DOI: 10.1111/mms.12206

Quantitative assessment of species identification in aerial transect surveys for ice-associated seals

BRETT T. MCCLINTOCK,[1] ERIN E. MORELAND, JOSHUA M. LONDON, SHAWN P. DAHLE, GAVIN M. BRADY, ERIN L. RICHMOND, KYMBERLY M. YANO, and PETER L. BOVENG, National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA-NMFS, 7600 Sand Point Way NE, Seattle, Washington 98115, U.S.A.

ABSTRACT

Technological advances have facilitated collection of vast quantities of photographic data from aerial surveys of marine mammals. However, when it is difficult to distinguish species from a distance, reliable identification from aerial images can often be challenging. This is the case for ice-associated seals, species for which global climate change has motivated intensive monitoring efforts in recent years. We assess species and age class identification from aerial images of four ice seal species (bearded seals, *Erignathus barbatus*; ribbon seals, *Histriophoca fasciata*; ringed seals, *Pusa hispida*; spotted seals, *Phoca largha*) in the Bering Sea. We also investigate the specific phenomenological and behavioral traits commonly associated with species identification and observer confidence. We generally found species and age class misidentification occurred at relatively low levels, but only 83% of spotted seals tended to be correctly identified (with 11% mistaken as ribbon seals). We also found certain traits were strong predictors for observed species, age class, or observer confidence. Our findings add to the growing body of evidence that species misidentification is pervasive in passive sampling of animal populations. Even low levels of misidentification have been demonstrated to induce substantial biases in estimators of species distribution and abundance, and it is important that statistical models account for such errors.

Key words: abundance, aerial survey, latent state model, partial state observation, Phocidae, ice-associated seals, photo-identification, satellite imagery, species misclassification, species misidentification, species occurrence.

Several species of seals (family Phocidae) require sea ice as a platform for whelping and rearing pups during spring and for molting during spring to early summer. These "ice-associated" seals, or "ice seals" have become a focus of conservation concern, primarily stemming from expected losses or modifications to their sea-ice habitat due to the effects of global climate change (Tynan and DeMaster 1997, Moore and Huntington 2008). Until recently, relatively little was known about the distribution and abundance of many of these species (*e.g.*, Conn *et al*. 2013*a*, *b*; Ver Hoef *et al*. 2014).

Distribution and abundance data for both marine and terrestrial mammals are difficult, time-consuming, and expensive to collect, often involving aerial transect surveys over large geographic areas (*e.g.*, O'Brien and Lindzey 1998, Ver Hoef *et al*. 2014). Historically, surveys required observers to visually identify and count target

[1]Corresponding author (e-mail: brett.mcclintock@noaa.gov).

species during flight, but technological advances in digital imaging have facilitated the collection of large quantities of photographic data from aerial (*e.g.*, Conn *et al*. 2013*b*) and satellite (*e.g.*, LaRue *et al.* 2011, Fretwell *et al*. 2012) surveys, which have the benefit of documenting all observations for postprocessing, verification, and archiving. However, species and age class identification from aerial or satellite imagery can often be difficult (*e.g.*, O'Brien and Lindzey 1998, Fretwell *et al*. 2014), and if not accounted for, misidentification can result in unreliable inference about species distribution and abundance (*e.g.*, McClintock *et al*. 2010*a*, Miller *et al*. 2011, Caillat *et al*. 2013, Conn *et al*. 2013*a*).

Species identification of ice-associated seals from aerial imagery is particularly challenging. These species can be difficult to distinguish from a distance, occupy similar habitats, and have overlapping distributions. In addition, safety and potential disturbance of seals necessitate relatively high survey altitudes, which, even with modern digital image sensors, still limit the potential resolution of imagery to examine species-specific characteristics. It is therefore challenging to balance the conflicting goals of maximizing spatial coverage and minimizing disturbance (improved with higher survey altitudes and wider-angle lenses) while retaining sufficient pixel resolution for accurate species identification (improved with lower survey altitudes, higher zoom lenses, and higher resolution cameras). Image quality is also highly variable due to environmental factors such as weather and light conditions. Even under ideal environmental conditions, animal behavior such as body position, orientation, or movement can further limit our ability to reliably determine species from aerial images. Because many of the characteristics commonly used to identify ice-associated seal species (*e.g.*, chapter 5 in Jefferson *et al*. 2008) are often not readily visible in aerial photographs, coarser characteristics (*e.g.*, body shape, pelage pattern, or associations with other seals) must often be used. Therefore, unlike species identifications based on physical or close visual inspection, aerial survey photographs will often produce species identifications of variable degrees of certainty depending on which characteristics are visible in any particular image.

Focusing on the observation process, we examine the ability of observers to identify four ice-associated seal species (bearded seals, *Erignathus barbatus*; ribbon seals, *Histriophoca fasciata*; ringed seals, *Pusa hispida*; and spotted seals, *Phoca largha*) from aerial images collected during transect surveys of the Bering Sea. These four species are easily distinguishable from physical or close visual inspection, but can be challenging to identify from aerial imagery (see Fig. 1, 2). For example, body position, animal movement, or light conditions can potentially obscure the unmistakable white ribbons of an adult ribbon seal. When the characteristic "torpedo-like" body shape and "dog-like" muzzle of a spotted seal are not visible, it could be more likely to be mistaken for a ringed seal. Juveniles and pups of all of the species tend to be the most difficult to positively identify as they have not developed many of the distinguishing traits of the adults, and therefore tend to look alike from a distance. Furthermore, pups of ribbon, spotted, and ringed seals are born with a white lanugo coat that is generally present until weaning. For pups and nonpups of each species, we quantify both correct identification probabilities and misidentification probabilities resulting from false-positive species or age class identifications. We also identify the specific characteristics most commonly associated with identifications of each species and age class, as well as which traits are associated with greater degrees of observer confidence. As inference about species distribution and abundance is generally the ultimate goal of aerial transect and satellite surveys, we conclude with a discussion of the implications of our findings for ecological inference when species are difficult to distinguish.

Figure 1. The characteristic bands on the coats of ribbon seals are not necessarily clearly visible in an aerial image. The images on the top right and bottom right were taken with a Canon 1Ds Mark III fitted with a Zeiss 100 mm lens from 300 m during a 2012 line transect survey in the Bering Sea. In the top right image, an observer would likely rely on the clearly visible bands to conclude that the seal is certainly a ribbon seal. In the bottom right image, an observer would likely rely on a combination of body shape, head size, flipper size and shape, and what could be one or more bands to conclude that the seal is probably a ribbon seal.

Figure 2. A red face, which is one of the characteristics associated mostly with bearded seals, is not always present, nor is it necessarily visible in an aerial image. The image on the right was taken with a Canon 1Ds Mark III fitted with a Zeiss 100 mm lens from 300 m during a 2012 line transect survey in the Bering Sea. In this image, an observer would likely rely on the combination of body shape, head size, front-flipper size and shape, and position on the floe to conclude that the seal is probably or certainly a bearded seal.

Methods

Data Collection

Between April and May in 2012 and 2013, we conducted aerial transect surveys of ice-associated seal species in the eastern Bering Sea. Flying at a target altitude of 300 m, images were captured using automated 21 or 24 megapixel high-resolution digital single-lens reflex cameras (Canon 1Ds Mark III and Nikon D3X) fitted with a 100 mm Zeiss lens. These specifications were selected to produce images of

previously established minimum ground resolution requirements for species identification (2 cm per pixel) while maximizing survey area and minimizing animal disturbance. Images were collected continuously at a rate of approximately one frame per second with minimal or no overlap; hence, some images contained partial seals at the edge of the frame (*e.g.*, torso only, head and partial torso, *etc.*). Using systematic random sampling, 716 images were chosen from 10 flights during a one week period from 20 to 27 April 2012. This set of flights provided representative spatial coverage of the entire study area (see Conn *et al.* 2013*b* for full details). The 716 images included $R=759$ distinct seals for species and age class identification.

Four seal biologists from the Polar Ecosystem Program of the National Marine Mammal Laboratory participated in our species identification trials. All four observers assigned species to 600 photographed individuals, and only one of two observers assigned species to the remaining 159 individuals. For each trial, observers assigned a species (bearded = BD, ribbon = RN, ringed = RD, spotted = SD, unknown = UK), species identification confidence level for BD, RN, RD, or SD classifications (guess: <50%, likely: 51%–99%, or positive: 100%), age class (pup, nonpup, unknown), and age class confidence for pup or nonpup classifications (guess, likely, positive). Here we define pup as any young animal with a white lanugo coat or, in the case of bearded seals and those that have already molted their lanugo, within a few body lengths of an adult seal. Nonpups are defined as all other age classes (*e.g.,* young-of-the-year or weaned pup, subadult, adult). Observers also recorded whether the torso, head, fore flippers, and rear flippers were visible.

To gain a better understanding of the traits and behaviors commonly used to identify ice-associated seal species and age classes from aerial photographs, we asked our observers to record specific characteristics associated with each seal. Prior to commencing the trials, a comprehensive list of potential characteristics was compiled from extensive discussions with ice seal biologists (Table 1). This included traits as seen specifically in aerial imagery, which were not necessarily consistent with traits typically seen from the ground (*e.g.*, white band around neck and serpentine body shape). Based on this list, observers recorded specific visible traits for each trial seal image, including those associated with pelage (light coat, mottled coat, dark coat, distinct ribbon, 1 faint ribbon, 2–3 faint ribbons), body shape ("American football," "torpedo-like," serpentine, tubular, other), fore flipper shape (long and slender, short broad square, other), muzzle shape ("cat-like," "dog-like," other), the color of any lanugo (white, off-white), and the presence of long rear flippers. The presence of a long neck, slender posterior, small head (relative to body size), red face, beard-like vibrissae, or "T"-shaped pattern on the forehead were also recorded. Some behavioral characteristics were recorded, including track path (straight, serpentine), track pattern (alternating, paired), associations with other seals (triad, nonpup group, single adult), the presence of a maintained hole in the ice, maximum ice floe dimension (>3 body lengths, <2 body lengths), and proximity to ice edge (1 body length, >1 body length). For convenience, we will henceforth refer to these traits using the abbreviations defined in Table 1.

Species and Age Classification Analysis

We are able to estimate misidentification probabilities by repeated sampling of multiple observers and treating observations with positive confidence levels as the "gold standard." Formally, let $s \in \{1,2,3,4,5\}$ denote respective observed species classifications of spotted, ribbon, bearded, ringed, and unknown, $a \in \{1,2,3\}$ denote

Table 1. Phenomenological and behavioral traits used to identify four species of ice-associated seal (spotted = SD, ribbon = RN, bearded = BD, ringed = RD) from aerial survey images.

Trait	Description	Category	Affiliation
alt_track	Alternating flipper pattern in tracks	Track	RN
assoc	"Associated" with another seal within 6 body lengths	Behavior	
assoc_23len	Approximately 2/3 the length of an associated seal	Body	BD pups
beard	Beard-like vibrissae	Head	BD
catlike	"Cat-like" face; compact features, short muzzle	Head	RD
distinctrbn	1 or more distinct ribbons	Pelage	RN
dkcoat	Dark coat with no spots	Pelage	RN
doglike	"Dog-like" face; wide skull, long muzzle	Head	SD
edge_1body	Within 1 body length of edge on non-small ice floe	Behavior	BD
edge_gt1body	>1 body length from edge on non-small ice floe	Behavior	
faintrbn1	1 faint ribbon	Pelage	RN
faintrbn23	2–3 faint ribbons	Pelage	RN
fball	"American football" or "comma" shape	Body	RD
group	3 or more associated nonpups	Behavior	SD
gt3body	On non-small ice floe (>3 body lengths)	Behavior	
torpedo	"Torpedo-like" or "elongated teardrop" shape	Body	SD
hole	Close proximity to a maintained hole in ice floe	Behavior	RD
lanugo_offwh	Off-white lanugo	Pelage	SD pups
lanugo_white	White lanugo	Pelage	RN pups, RD pups
long_neck	Long, slender neck	Body	RN
long_rearflip	Long hindflippers	Body	RN
lsff	Long slender fore flippers	Body	RN
lt2body	On small ice floe (<2 body lengths)	Behavior	BD
ltcoat	Light, uniform coat	Pelage	BD
mot	Mottled coat; spots or rings	Pelage	SD, RD
neckband	White band around neck	Pelage	RD
nocatdog	Other face type (not *catlike* or *doglike*)	Head	RN, BD
other_ffchar	Other fore flipper characteristics (not *lsff* or *sbsff*)	Body	SD, RD
othershape	Other body shape	Body	
pair_track	Paired flipper pattern in tracks	Track	SD
red_face	Reddish coloration on face	Head	BD
sbsff	Short broad square fore flippers	Body	BD
serp_track	Serpentine track pattern	Track	RN
serpentine	Serpentine body position	Body	RN
slender_post	Slender posterior	Body	RN
small_head	Small blunt head relative to body size	Head	BD
str_track	Straight track pattern	Track	SD
Thead	"T"-shaped pattern on forehead	Head	BD
triad	Two nonpups associated with one pup	Behavior	SD
tubular	Tubular or "cigar-like" body shape	Body	BD

respective observed age classifications of pup, nonpup, and unknown, and $s^*, a^* \in \{1,2,3\}$ denote respective observed species and age confidence levels of guess, likely, and positive. Our species, age, and confidence classifications therefore yield $(4 \times 3 + 1) \times (2 \times 3 + 1) = 91$ possible observation types. Letting $y_{r,i} \in \{1,2, \ldots, 91\}$ denote the observation type of observer i for seal r with indices

$$y_{r,i} = \begin{cases} 21(s_{r,i} - 1) + 7(s^*_{r,i} - 1) + 3(a_{r,i} - 1) + a^*_{r,i} & \text{if } s_{r,i} < 5 \text{and } a_{r,i} < 3 \\ 21(s_{r,i} - 1) + 7(s^*_{r,i} - 1) + 7 & \text{if } s_{r,i} < 5 \text{and } a_{r,i} = 3 \\ 84 + 3(a_{r,i} - 1) + a^*_{r,i} & \text{if } s_{r,i} = 5 \text{and } a_{r,i} < 3 \\ 91 & \text{if } s_{r,i} = 5 \text{and } a_{r,i} = 3, \end{cases}$$

we can model the species, age, and confidence level data using a categorical likelihood:

$$[\boldsymbol{y}|\boldsymbol{p}, \boldsymbol{z}] = \prod_{r=1}^{R} \prod_{i=1}^{n} \text{Categorical}(y_{r,i}; p_{z_r,1,i}, p_{z_r,2,i}, \ldots, p_{z_r,91,i}),$$

where $p_{z_r,y,i}$ is the conditional probability of observation $y_{r,i} = y$ for observer i given the true species and age classification for seal image $r = 1,\ldots,R$ is $z_r \in \{1,2,3,4,5,6,7,8\}$, corresponding to spotted pup (SDP), spotted nonpup (SDN), ribbon pup (RNP), ribbon nonpup (RNN), bearded pup (BDP), bearded nonpup (BDN), ringed pup (RDP), and ringed nonpup (RDN).

We assume any positive species or age observation (*i.e.*, $s^*_{r,i} = 3$ or $a^*_{r,i} = 3$) is the true species or age class for seal r. For ease of interpretation, we parameterized $p_{z_r,y,i}$ in terms of the probability that true species $l \in \{1,2,3,4\}$ (or {SD, RN, BD, RD}) is identified as species s $(\alpha_{l,s,i})$, the probability that true species l identified as species s is assigned to species confidence level $s^*(\beta_{l,s,s^*,i})$, the probability that true age class $j \in \{1,2\}$ (or {pup, nonpup}) is identified as age class a $(\delta_{j,a,i})$, and the probability that true age class j identified as age class a is assigned to age class confidence level a^* $(\gamma_{j,a,a^*,i})$:

$$p_{z_r,y,i} = \begin{cases} \alpha_{l,s,i}\beta_{l,s,s^*,i}\delta_{j,a,i}\gamma_{j,a,a^*,i} & \text{if } s < 5 \text{ and } a < 3 \\ \alpha_{l,s,i}\beta_{l,s,s^*,i}\delta_{j,a,i} & \text{if } s < 5 \text{ and } a = 3 \\ \alpha_{l,s,i}\delta_{j,a,i}\gamma_{j,a,a^*,i} & \text{if } s = 5 \text{ and } a < 3 \\ \alpha_{l,s,i}\delta_{j,a,i} & \text{if } s = 5 \text{ and } a = 3, \end{cases}$$

where

$$l = \begin{cases} 1 & \text{if } z_r \in \{1,2\} \\ 2 & \text{if } z_r \in \{3,4\} \\ 3 & \text{if } z_r \in \{5,6\} \\ 4 & \text{if } z_r \in \{7,8\} \end{cases}$$

and

$$j = \begin{cases} 1 & \text{if } z_r \in \{1,3,5,7\} \\ 2 & \text{if } z_r \in \{2,4,6,8\}. \end{cases}$$

Note that $\sum_{s=1}^{5} \alpha_{l,s,i} = 1$, $\sum_{s^*=1}^{3} \beta_{l,s,s^*,i} = 1$ for $s = 1, \ldots, 4$, and $\sum_{a=1}^{3} \delta_{j,a,i} = 1$, $\sum_{a^*=1}^{3} \gamma_{j,a,a^*,i} = 1$ for $a = 1, 2$.

To illustrate, suppose image r is of a bearded seal pup (*i.e.*, $z_r = 5$). For $s_{r,i} = 1$ (observed spotted seal), $s^*_{r,i} = 3$ (species positive), $a_{r,i} = 2$ (observed nonpup) and $a^*_{r,i} = 1$ (age guess), we have $\Pr(y_{r,i} = 18|z_r = 5) = p_{5,18,i} = \alpha_{3,1,i}\beta_{3,1,3,i}\delta_{1,2,i}\gamma_{1,2,1,i} = 0$

because we assume a bearded seal cannot be positively misidentified as a spotted seal (*i.e.*, $\beta_{3,1,3,i} = 0$). Now suppose image r is of a spotted seal nonpup (*i.e.*, $z_r = 2$). For $s_{r,i} = 3$ (observed bearded seal), $s^*_{r,i} = 1$ (species guess), $a_{r,i} = 1$ (observed pup), and $a^*_{r,i} = 3$ (age positive), then $\Pr(y_{r,i} = 45|z_r = 2) = p_{2,45,i} = \alpha_{1,3,i}\beta_{1,3,1,i}\delta_{2,1,i}\gamma_{2,1,3,i} = 0$ because we assume nonpups cannot be positively misclassified as pups (*i.e.*, $\gamma_{2,1,3,i} = 0$). Note that while these parameters are conditional on the true species and age class, Bayes' rule allows calculation of the probability that any given observation corresponds to the true species and age class:

$$\Pr(z_r = z|y_{r,i} = y) = \frac{\psi_z p_{z,y,i}}{\sum_k \psi_k p_{k,y,i}}.$$

For each seal image, the latent species identity and age class (z_r) is assumed known with certainty only if $\sum_{i=1}^{n} I(s^*_{r,i} = 3) > 0$ and $\sum_{i=1}^{n} I(a^*_{r,i} = 3) > 0$, where $I(x)$ is an indicator function taking the value 1 when x is true and 0 otherwise. We assume $z_r|\boldsymbol{\psi} \sim \text{Categorial}(\psi_1,\psi_2,\ldots,\psi_8)$ and use Bayesian analysis methods to estimate the joint posterior distribution of the model parameters and latent variables:

$$[\boldsymbol{\psi}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}, z|\boldsymbol{y}] \propto [\boldsymbol{y}|\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma}, z][z|\boldsymbol{\psi}][\boldsymbol{\psi}][\boldsymbol{\alpha}][\boldsymbol{\beta}][\boldsymbol{\delta}][\boldsymbol{\gamma}],$$

where ψ_z is the probability that any given seal image belongs to true species and age class z. We assigned uninformative Dirichlet priors on $\boldsymbol{\psi}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\gamma}$ (with appropriate constraints $\beta_{l,s,3,i} = 0$ for $l \neq s$ and $\gamma_{j,a,3,i} = 0$ for $j \neq a$).

Our model assumes positive species or age class confidence levels $s^* = 3$ or $a^* = 3$ are correct, and conflicting positive classifications are therefore not permitted. Out of 2,559 observer trials, there were initially five conflicting positive species identifications and nine conflicting positive age classifications. Observers were asked to revisit these trials, and all but one of these conflicts was deemed a data-entry error and corrected. The single remaining discrepancy was resolved through a civilized discussion between the observers. We performed our analysis in R (R Development Core Team 2013) using the *rjags* package (Plummer 2013; see Appendix S1 for JAGS model code). Starting from overdispersed initial values, we obtained three chains of 200,000 iterations for posterior summaries (after pilot tuning and burn-in of 20,000 iterations). To compare observer effects and obtain posterior summaries for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\gamma}$ across all observers, we performed an additional analysis assuming no observer effects; that is, $p_{z_r,y,i} = p_{z_r,y}$ for $i = 1,\ldots,n$. Standard diagnostics provided no evidence for lack of convergence (*e.g.*, Gelman-Brooks-Rubin diagnostic < 1.01 for all parameters).

Species and Age Classification Characteristics Analysis

For each observed species and age class (SDP, SDN, RNP, RNN, BDP, BDN, RDP, and RDN), we performed logistic regressions to identify the traits that best predicted the observed species and species confidence levels. To identify the traits that best explained the observed species, we first ignored observer confidence levels and treated the response as binary (*e.g.*, SDP or not SDP). To identify the traits that best predicted positive species identifications, we performed logistic regressions using only the observations in which each particular species and age class was identified (*i.e.*, we used only those observations with a guess, likely, or positive confidence level for both the observed species and age class; observations with an unknown species

and/or age class were not used). For this second set of analyses, we ignored age class confidence levels and again treated the response as binary (*e.g.*, positive SDP or non-positive SDP). For both sets of analyses, the predictors were binary indicators for the presence or absence of each trait (see Table 1).

With 40 different predictor traits, examining all possible models is impractical, and we had no *a priori* basis for establishing a manageable subset of candidate models. For both sets of logistic regression analyses, we therefore started from a null model with no trait effects and performed bidirectional stepwise selection based on Akaike's Information Criterion (AIC) using the *stepAIC*()and *glm*() functions in R (R Development Core Team 2013). When the minimum-AIC model exhibited (quasi-) complete separation due to sparseness or the dichotomous nature of the data, we used Firth logistic regression, which uses a penalized likelihood, to obtain estimates of effect sizes, standard errors, and profile-likelihood confidence intervals (implemented in the R package *logistf*; Heinze *et al*. 2013).

Results

Species and Age Classification Analysis

Based on estimated species and age classifications, the majority of images consisted of spotted or ringed seal nonpups. Estimates of true species and age class proportions in our sample were nearly identical between the analyses with and without observer effects (Table 2). For the model excluding observer effects, posterior medians were $\psi_1 = 0.08$ (95% CI: 0.07–0.11) for spotted pups, $\psi_2 = 0.29$ (95% CI: 0.26–0.32) for

Table 2. Comparison of estimated proportions (and standard errors) of each ice seal species and age class from our misidentification model (ψ) and naïve estimates assuming no misidentification (ψ^i_{naive}) with and without observer effects ($i = 1,\ldots,4$). Seal species and age classes are spotted pup (SDP), spotted nonpup (SDN), ribbon pup (RNP), ribbon nonpup (RNN), bearded pup (BDP), bearded nonpup (BDN), ringed pup (RDP), and ringed nonpup (RDN).

	Observer effects					No observer effects	
Species/age	ψ	ψ^1_{naive}	ψ^2_{naive}	ψ^3_{naive}	ψ^4_{naive}	ψ	ψ_{naive}
SDP	0.08	0.07	0.08	0.07	0.06	0.08	0.07
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
SDN	0.29	0.28	0.27	0.24	0.23	0.29	0.26
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
RNP	0.02	0.03	0.02	0.03	0.04	0.02	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
RNN	0.06	0.08	0.07	0.12	0.11	0.06	0.10
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
BDP	0.03	0.02	0.03	0.02	0.04	0.03	0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
BDN	0.18	0.19	0.18	0.16	0.17	0.18	0.17
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
RDP	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
RDN	0.34	0.32	0.35	0.35	0.35	0.34	0.34
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.00)

spotted nonpups, $\psi_3 = 0.02$ (95% CI: 0.01–0.03) for ribbon pups, $\psi_4 = 0.06$ (95% CI: 0.04–0.07) for ribbon nonpups, $\psi_5 = 0.03$ (95% CI: 0.02–0.04) for bearded pups, $\psi_6 = 0.18$ (95% CI: 0.15–0.21) for bearded nonpups, $\psi_7 = 0.00$ (95% CI: 0.00–0.01) for ringed pups, and $\psi_8 = 0.34$ (95% CI: 0.31–0.38) for ringed nonpups. Ignoring unknown species or age classifications (*i.e.*, $s = 5$ or $a = 3$), naïve proportions assuming no misidentification were 0.07 (SE = 0.01) for spotted pups, 0.26 (SE = 0.01) for spotted nonpups, 0.03 (SE = 0.01) for ribbon pups, 0.10 (SE = 0.00) for ribbon nonpups, 0.03 (SE = 0.00) for bearded pups, 0.17 (SE = 0.01) for bearded nonpups, 0.00 (SE = 0.01) for ringed pups, and 0.34 (SE = 0.00) for ringed nonpups (Table 2).

We found the type and magnitude of misidentification to vary between species (Fig. 3). Across observers, correct species identification probabilities were relatively high for ribbon ($\alpha_{2,2} = 0.97$; 95% CI: 0.93–0.99) and bearded seals ($\alpha_{3,3} = 0.96$; 95% CI: 0.93–0.98). Spotted seals were correctly identified with the lowest probability ($\alpha_{1,1} = 0.83$; 95% CI: 0.80–0.86) and tended to be misidentified as ribbon seals ($\alpha_{1,2} = 0.11$; 95% CI: 0.09–0.14). Ringed seals tended to be identified correctly ($\alpha_{4,4} = 0.94$; 95% CI: 0.92–0.96), but were more often misidentified as spotted seals ($\alpha_{4,1} = 0.03$; 95% CI: 0.02–0.04) than the other species. With $\alpha_{l,5}$ ranging between 0.01 and 0.02, all species had similarly low probabilities of being assigned to the unknown species observation category (see Appendix S2 for more detailed posterior summaries). We found some variability in species identification probabilities between observers (Fig. 3). For example, one observer was significantly more likely to misidentify spotted seals as ribbon seals ($\alpha_{1,2,i} = 0.18$; 95% CI: 0.14–0.24), while another observer was more likely to misidentify spotted seals as bearded seals ($\alpha_{1,3,i} = 0.04$; 95% CI: 0.02–0.07) and ringed seals as spotted seals ($\alpha_{4,1,i} = 0.07$; 95% CI: 0.04–0.10).

Within species, we found similar levels of misidentification for pups and nonpups (Fig. 3). Pups and nonpups of all four species had similar age class identification probabilities, with respective overall correct age class identification probabilities of $\delta_{1,1} = 0.96$ (95% CI: 0.93–0.98) and $\delta_{2,2} = 0.98$ (95% CI: 0.97–0.99). We found marginal evidence of slightly higher probabilities for pups being misclassified as nonpups, with $\delta_{1,2} = 0.02$ (95% CI: 0.01–0.04) and $\delta_{2,1} = 0.01$ (95% CI: 0.01–0.02), and pups being assigned to the unknown age class observation category, with $\delta_{1,3} = 0.02$ (95% CI: 0.01–0.04) and $\delta_{2,3} = 0.01$ (95% CI: 0.00–0.01). In the analysis including observer effects, age class identification probabilities were similar for all four observers.

Observers overall tended to be most confident in correct identifications of ribbon seals ($\beta_{2,2,3} = 0.91$; 95% CI: 0.84–0.95), followed by bearded seals ($\beta_{3,3,3} = 0.72$; 95% CI: 0.68–0.76), ringed seals ($\beta_{4,4,3} = 0.60$; 95% CI: 0.56–0.63), and spotted seals ($\beta_{1,1,3} = 0.39$; 95% CI: 0.35–0.42), but we found confidence to vary among observers (Fig. 4). For example, one observer tended to be more confident when identifying spotted seals ($\beta_{1,1,3,i} = 0.63$; 95% CI: 0.57–0.69), and another observer tended to be less confident for all four species ($\beta_{1,1,3,i} = 0.17$, 95% CI: 0.12–0.23; $\beta_{2,2,3,i} = 0.77$, 95% CI: 0.61–0.88; $\beta_{3,3,3,i} = 0.50$; 95% CI: 0.41–0.59; $\beta_{4,4,3,i} = 0.12$, 95% CI: 0.08–0.17). With the exception of ribbon seals, we generally found correct species identifications were assigned to the "guess" confidence level with the lowest probability (see Appendix S2).

For age class confidence, observers overall tended to be more confident in correct classifications of pups ($\gamma_{1,1,3} = 0.88$; 95% CI: 0.84–0.92) than nonpups ($\gamma_{2,2,3} = 0.76$; 95% CI: 0.74–0.78). However, age class confidence also varied among observers

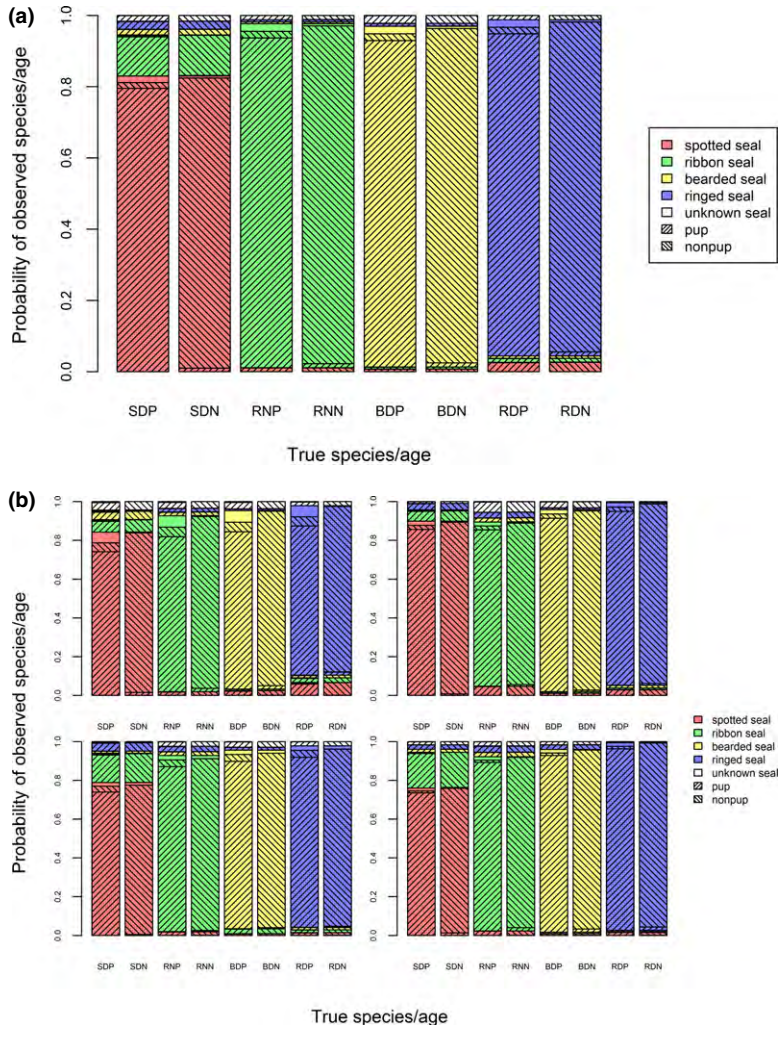

Figure 3. Observed species and age class identification probabilities for four species of ice-associated seals in the Bering Sea. True species and age classes include spotted seal pup (SDP), spotted seal nonpup (SDN), ribbon seal pup (RNP), ribbon seal nonpup (RNN), bearded seal pup (BDP), bearded seal nonpup (BDN), ringed seal pup (RDP), and ringed seal nonpup (RDN). Observed species classifications include spotted seal (red), ribbon seal (green), bearded seal (yellow), ringed seal (blue), and unknown seal (white). Observed age classes include pup, nonpup, and unknown. Solid colors with no hashing indicate unknown age classification. Top panel (a) includes results from an analysis with no observer effects on model parameters. Bottom four panels (b) correspond to four different observers from an analysis including observer effects.

(Fig. 4). For example, the same observer that generally exhibited lower species confidence also exhibited a substantially lower age class confidence for nonpups ($\gamma_{2,2,3,i}$ = 0.06; 95% CI: 0.05–0.09).

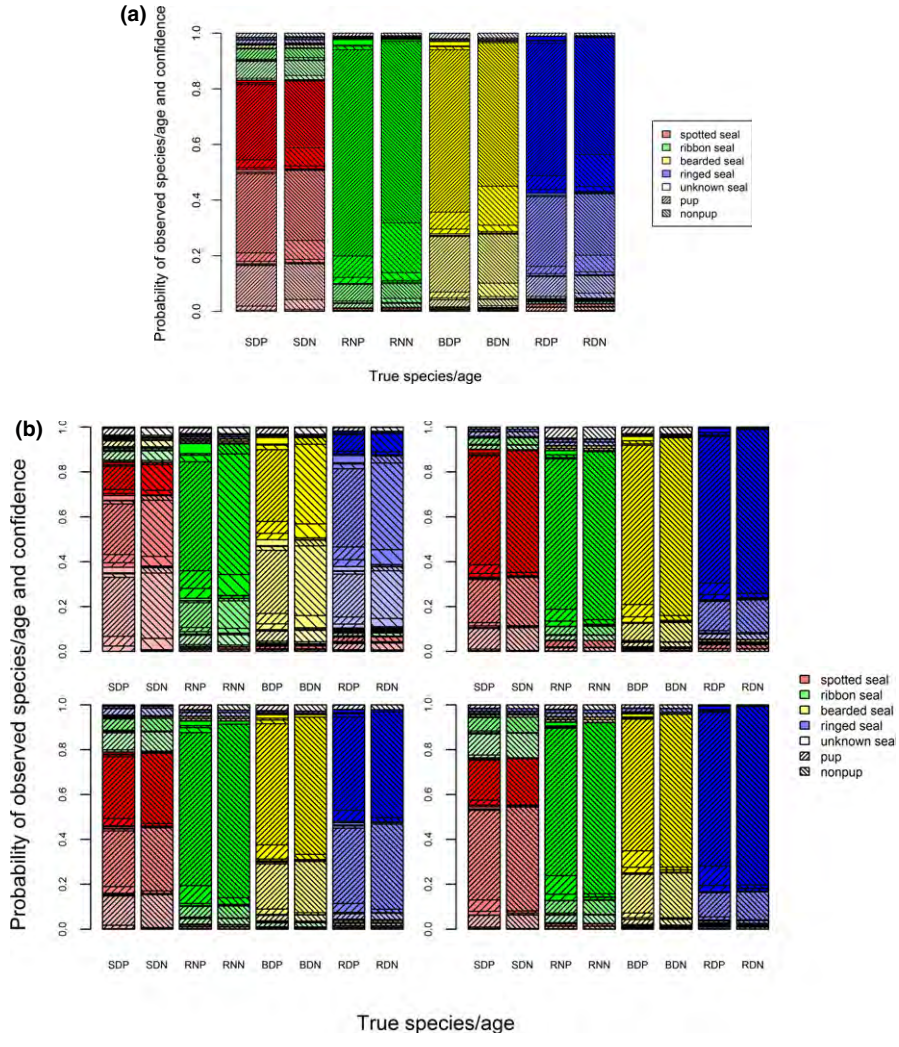

Figure 4. Observed species, age class, and confidence level probabilities for four species of ice-associated seals in the Bering Sea. True species and age classes include spotted seal pup (SDP), spotted seal nonpup (SDN), ribbon seal pup (RNP), ribbon seal nonpup (RNN), bearded seal pup (BDP), bearded seal nonpup (BDN), ringed seal pup (RDP), and ringed seal nonpup (RDN). Observed species classifications include spotted seal (red), ribbon seal (green), bearded seal (yellow), ringed seal (blue), and unknown seal (white). Observed age classes include pup, nonpup, and unknown. For observed species classifications, darker shades indicate greater confidence (*e.g.*, light red = spotted seal guess, red = spotted seal likely, and dark red = spotted seal positive). For observed age classes, the relative density of hash lines indicate greater confidence (*e.g.*, low density = guess, medium density = likely, high density = positive). Solid colors with no hashing indicate unknown age classification. Top panel (a) includes results from an analysis with no observer effects on model parameters. Bottom four panels (b) correspond to four different observers from an analysis including observer effects.

Species and Age Classification Characteristics Analysis

The vast majority of seal images included the torso (99%), head (98%), and rear flippers (98%), while fore flippers (61%) and tracks (50%) had a greater tendency to be obstructed, absent, or otherwise indiscernible. A "complete" seal (torso, head, fore flippers, and rear flippers) with tracks was visible in 35% of images. Many of the traits included in our list of potential characteristics (see Table 1) were commonly associated with observed species and age class assignments, but with varying frequencies and confidence levels (Fig. 5). Our aerial survey images did not contain sufficient numbers of ringed seal pup observations for our logistic regression analysis, but there were sufficient sample sizes for all other species and age classes. Although presence on an ice floe with a maximum dimension greater than three body lengths (*gt3body*) was the trait most commonly associated with each species and age class, it was not generally found to be a useful predictor of observed species and age classifications (Tables 3–6).

For the first set of logistic regression analyses ignoring species confidence levels, we found several traits to be useful predictors for observations of spotted, ribbon, and bearded seal pups. Based on the minimum-AIC model for spotted seal pups, the most important predictors with positive associations were *triad*, *lanugo_offwh*, and *group* (Table 3). Observations of ribbon seal pups were positively associated with both *lanugo_white* and *lanugo_offwh*. Observed bearded seal pups were positively associated with *small_head*, *mot*, *assoc_23len*, *slender_post*, *tubular*, *nocatdog*, *edge_1body*, and *ltcoat*.

For spotted seal nonpups, the minimum-AIC model included positive effects for *triad*, *lanugo_white, doglike*, *lanugo_offwh, lsff*, *torpedo*, *assoc*, *other _ffchar*, *pair_track*, *mot*, and *dkcoat* (Table 4). For ribbon seal nonpups, positive associations were found for *distinctrbn*, *faintrbn23*, *serpentine*, *othershape*, *faintrbn1*, *torpedo*, *fball*, *catlike*, *long_neck*, *edge_gt1body*, *dkcoat*, and *neckband*. Bearded seal nonpups were positively associated with *edge_1body*, *red_face*, *small_head*, *tubular*, *ltcoat*, *nocatdog*, and *str_track*. For ringed seal nonpups, we found positive effects for *hole*, *fball*, and *neckband*.

In our second set of logistic regression analyses accounting for observed species confidence levels, we found certainty in identification (*i.e.*, positive confidence) within each species of pup to be strongly related to the presence of an associated nonpup (Table 5). Based on the minimum-AIC model for spotted seal pups, we found that *triad*, *assoc*, and *long_rearflip* were positively associated with species certainty, while *ltcoat* tended to be associated with the lower confidence levels (*i.e.*, guess and likely). For ribbon seal pups, we found positive species confidence to be strongly associated with *assoc*. Certainty for bearded seal pups was positively associated with *assoc_23len* and *tubular*, while the lower confidence levels were associated with *slender_post* and *ltcoat*.

For observations of spotted seal nonpups, the minimum-AIC model suggested species identification certainty was positively associated with *triad*, *torpedo*, *tubular*, *mot*, *sbsff*, *assoc*, *doglike*, *other_ffchar*, *lsff*, and *pair_track* (Table 6). Ribbon seal nonpups tended to be identified with positive species confidence when *distinctrbn*, *faintrbn23*, *alt_track*, and *other_ffchar* were observed. Certainty for bearded seal nonpups was positively associated with *red_face*, *assoc*, *nocatdog*, *doglike*, *ltcoat*, *small_head*, *edge_1body*, *catlike*, *dkcoat*, *str_track*, and *tubular*. For ringed seal nonpups, we found positive species confidence to be strongly associated with the presence of *hole*, *fball*, and *mot*. For each species, additional traits were negatively associated with species confidence for nonpups (Table 6).

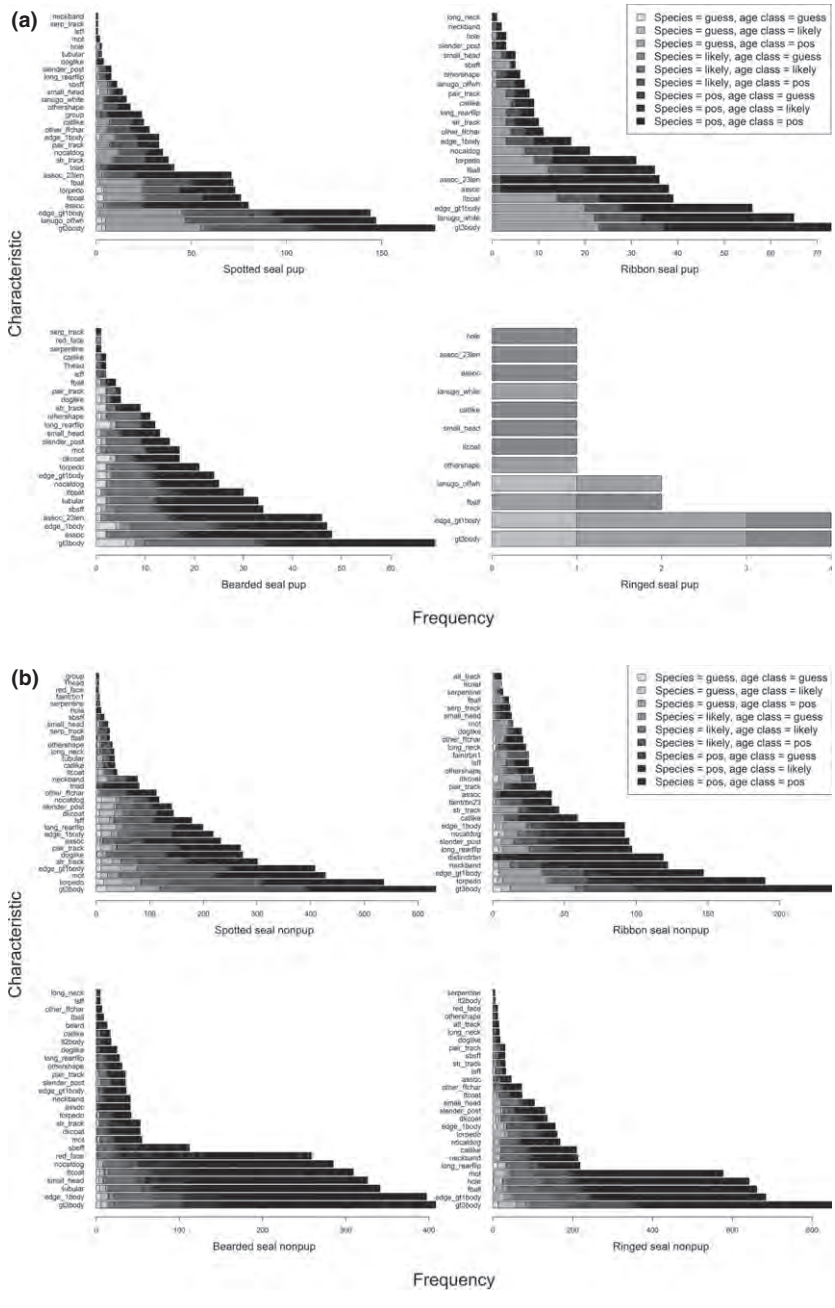

Figure 5. Frequencies of observed characteristics from images identified as pups (a) and nonpups (b) of four ice-associated seal species in the Bering Sea. Bars are stacked according to the frequencies of nine observed species and age class confidence categories. For pups, only those traits with at least one observation are included for each species. For nonpups, only traits with ≥ 5 observations are included. Trait definitions are provided in Table 1.

Table 3. Characteristics that best predicted observed species identifications across four species of ice-associated seal pups.[a] Coefficients and 95% confidence intervals (LCI, UCI) for each species are from the minimum-AIC logistic regression model. Trait definitions are provided in Table 1.

Observed species	Positive effects				Negative effects			
	Trait	Coef.	LCI	UCI	Trait	Coef.	LCI	UCI
spotted	*triad*	4.18	1.95	9.09	*tubular*	−2.05	−4.01	−0.49
	lanugo_offwh	3.63	2.67	4.70	*dkcoat*	−1.67	−6.56	0.55
	group	3.15	−0.36	8.29	*othershape*	−1.45	−2.76	−0.17
					lanugo_white	−0.97	−1.99	0.04
					edge_1body	−0.79	−1.75	0.14
					(intercept)	−0.65	−1.52	0.15
ribbon	*lanugo_white*	5.80	4.36	8.05	*triad*	−4.34	−9.23	−2.13
	lanugo_offwh	1.29	−0.26	3.55	(intercept)	−4.07	−6.24	−2.78
bearded	*small_head*	8.96	−1.01	36.06	*lanugo_offwh*	−19.03	−64.10	−4.24
	mot	7.55	2.35	31.49	*triad*	−14.05	−47.21	−2.90
	assoc_23len	7.00	2.70	28.41	*lanugo_white*	−10.07	−40.22	−3.68
	slender_post	3.80	0.33	17.18	*fball*	−7.01	−37.13	−1.11
	tubular	3.47	0.33	13.31	*assoc*	−5.85	−21.98	−1.47
	nocatdog	2.68	−0.53	16.89	*catlike*	−4.85	−28.24	0.86
	edge_1body	2.53	0.36	8.68	(intercept)	−1.78	−6.76	0.12
	ltcoat	2.10	−0.39	13.08	*othershape*	−1.44	−8.84	2.16

[a]Sample sizes for ringed seal pups were insufficient for inclusion in this analysis.

Discussion

We generally found species and age class misidentification to occur across all species and observers. While species misidentification rates appear to be relatively low for ribbon, bearded, and ringed seals, we found spotted seals were frequently mistaken for other species, and ribbon seals in particular. This is most noticeable in the differences between estimated species proportions based on our misidentification model and naïve estimates ignoring misidentification (see Table 2). We attribute this to observers generally being less confident about spotted seal observations (*e.g.*, only 39% of correct spotted seal observations were assigned a positive confidence level) and a tendency for spotted seals to resemble ribbon seals when their distinctive pelage patterns are obscured or absent. We had expected ringed seals to be more frequently misidentified as spotted seals, as suggested during past observer-based helicopter surveys, but we suspect the high resolution imagery and additional time observers had to review imagery reduced this error. Age class misidentification rates were similarly low across species, although we found evidence that pups may be slightly more likely to be mistaken for nonpups.

The subset of flights for this analysis were chosen because they provided the best coverage of the survey area within a limited amount of time (one week), ensuring minimal ice movement. For the four species of ice-associated seals included in our study, we found ringed and spotted seals to be the most prevalent in our sample, while ribbon seals were encountered with the lowest frequency. These proportions were all within the ranges of our expectations given natural history observations and previous surveys of these species. For example, nonpups of spotted, ringed, and bearded seals were more prevalent than nonpups of ribbon seals because the 10 survey

Table 4. Characteristics that best predicted observed species identifications across four species of ice-associated seal nonpups. Coefficients and 95% confidence intervals (LCI, UCI) for each species are from the minimum-AIC logistic regression model. Trait definitions are provided in Table 1.

Observed species	Positive effects				Negative effects			
	Trait	Coef.	LCI	UCI	Trait	Coef.	LCI	UCI
spotted	*triad*	4.55	2.32	9.45	*distinctrbn*	−6.20	−11.09	−4.00
	lanugo_white	2.50	−0.59	7.54	*hole*	−4.89	−5.99	−3.99
	doglike	1.82	1.27	2.40	*faintrbn23*	−4.63	−7.17	−2.74
	lanugo_offwh	1.80	−0.59	6.72	*red_face*	−1.96	−3.27	−0.91
	lsff	1.72	1.12	2.35	*faintrbn1*	−1.75	−3.00	−0.59
	torpedo	1.54	1.04	2.05	(intercept)	−1.60	−2.41	−0.84
	assoc	1.47	0.87	2.10	*small_head*	−1.56	−2.37	−0.81
	other_ffchar	1.32	0.67	2.01	*fball*	−1.52	−2.25	−0.84
	pair_track	1.25	0.73	1.78	*lt2body*	−1.31	−2.80	−0.04
	mot	1.22	0.47	1.99	*catlike*	−0.81	−1.50	−0.15
	dkcoat	0.64	2.32	9.45	*edge_1body*	−0.69	−1.09	−0.28
					ltcoat	−0.68	−1.54	0.18
					neckband	−0.43	−0.99	0.13
ribbon	*distinctrbn*	8.87	6.44	13.94	*hole*	−7.12	−12.55	−3.84
	faintrbn23	6.13	4.44	8.54	(intercept)	−5.41	−8.01	−3.78
	serpentine	5.76	3.44	8.72	*sbsff*	−2.24	−5.14	0.19
	othershape	3.81	1.96	6.48	*assoc*	−2.03	−3.38	−0.92
	faintrbn1	3.48	2.43	4.64	*red_face*	−1.99	−6.86	0.09
	torpedo	2.94	1.29	5.51	*pair_track*	−1.23	−2.12	−0.43
	fball	1.81	0.00	4.46	*doglike*	−0.93	−1.85	−0.12
	catlike	1.13	0.36	1.87	*mot*	−0.92	−1.75	−0.07
	long_neck	1.11	−0.04	2.13				
	edge_gt1body	0.68	0.10	1.28				
	dkcoat	0.65	−0.08	1.43				
	neckband	0.58	−0.18	1.29				
bearded	*edge_1body*	3.25	2.18	4.52	*triad*	−4.29	−9.50	−1.38
	red_face	2.96	1.82	4.37	*lsff*	−4.25	−6.61	−1.52
	small_head	2.38	1.39	3.49	*long_neck*	−3.70	−6.71	−0.45
	tubular	2.30	1.12	3.60	*hole*	−3.31	−7.66	−0.07
	ltcoat	2.29	1.48	3.18	*fball*	−3.00	−4.49	−1.68
	nocatdog	1.39	0.54	2.28	*other_ffchar*	−2.82	−5.74	−0.60
	str_track	1.35	0.13	2.55	*distinctrbn*	−2.72	−7.70	−0.27
					(intercept)	−2.71	−4.63	−0.96
					faintrbn23	−2.48	−7.58	0.28
					long_rearflip	−1.80	−3.08	−0.63
					torpedo	−1.59	−2.71	−0.50
					gt3body	−1.39	−2.86	0.03
					mot	−1.25	−2.38	−0.23
					neckband	−1.17	−2.45	−0.01
					pair_track	−1.03	−2.45	0.40
					doglike	−0.89	−2.21	0.33
					slender_post	−0.86	−2.14	0.33
ringed	*hole*	5.49	4.63	6.53	*distinctrbn*	−5.61	−10.77	−3.22
	fball	3.30	2.79	3.83	*faintrbn1*	−3.85	−8.79	−1.51

(Continued)

Table 4. (Continued)

Observed species	Positive effects				Negative effects			
	Trait	Coef.	LCI	UCI	Trait	Coef.	LCI	UCI
	neckband	0.86	0.24	1.49	*faintrbn23*	−3.59	−8.45	−1.53
	(intercept)	0.10	−0.85	1.01	*lanugo_white*	−2.87	−7.90	0.20
					triad	−2.74	−7.66	−0.49
					tubular	−2.63	−4.11	−1.48
					serpentine	−2.49	−4.46	−0.15
					doglike	−2.47	−3.47	−1.59
					red_face	−1.77	−3.07	−0.62
					str_track	−1.33	−2.21	−0.50
					assoc	−1.10	−1.89	−0.37
					lsff	−1.00	−1.95	−0.10
					dkcoat	−0.99	−1.58	−0.42
					othershape	−0.95	−2.19	0.09
					gt3body	−0.88	−1.81	0.10
					other_ffchar	−0.75	−1.60	0.05
					pair_track	−0.72	−1.77	0.30
					ltcoat	−0.66	−1.30	−0.03
					nocatdog	−0.50	−1.07	0.05

Table 5. Characteristics that best predicted positive observer confidence levels within observations of four species of ice-associated seal pups.[a] Coefficients and 95% confidence intervals (LCI, UCI) for each species are from the minimum-AIC logistic regression model. Trait definitions are provided in Table 1.

Observed species	Positive effects				Negative effects			
	Trait	Coef.	LCI	UCI	Trait	Coef.	LCI	UCI
spotted	*triad*	2.83	1.43	4.62	*ltcoat*	−2.61	−4.27	−1.37
	assoc	2.64	1.62	3.76	(intercept)	−1.85	−2.64	−1.18
	long_rearflip	2.05	0.04	4.23				
ribbon	*assoc*	6.94	4.59	11.90	(intercept)	−4.26	−9.10	−2.31
bearded	*assoc_23len*	3.50	1.97	5.44	(intercept)	−2.40	−4.14	−1.04
	tubular	1.60	0.35	3.08	*slender_post*	−2.00	−3.71	−0.51
					ltcoat	−0.89	−2.29	0.34

[a]Sample sizes for ringed seal pups were insufficient for inclusion in this analysis.

flights did not include much effort in the southwest portion of the study area where ribbon seal densities have previously been observed to be the highest (Conn *et al*. 2013*b*, Ver Hoef *et al*. 2014). There were no positive ringed seal pup observations, and we believe this is attributable to reproductive behavior; Arctic ringed seal pups are whelped and nursed in subnivean lairs during spring (Smith and Stirling 1975) and thus would not be visible during our surveys. Our ability to make inference about ringed seal pups was therefore limited.

Strict sampling protocols or observer training are often proposed as means for reducing or eliminating species misidentification (*e.g.*, Miller *et al*. 2012). Despite

Table 6. Characteristics that best predicted positive observer confidence levels within observations of four species of ice-associated seal nonpups. Coefficients and 95% confidence intervals (LCI, UCI) for each species are from the minimum-AIC logistic regression model. Trait definitions are provided in Table 1.

Observed species	Positive effects				Negative effects			
	Trait	Coef.	LCI	UCI	Trait	Coef.	LCI	UCI
spotted	*triad*	3.18	2.23	4.25	(intercept)	−4.84	−6.06	−3.76
	torpedo	2.36	1.46	3.40	*serp_track*	−2.18	−3.76	−0.80
	tubular	1.41	−0.06	2.90	*long_neck*	−1.73	−3.10	−0.57
	mot	1.37	0.89	1.87	*long_rearflip*	−0.93	−1.46	−0.42
	sbsff	1.32	−0.25	2.84	*nocatdog*	−0.86	−1.66	−0.13
	assoc	1.24	0.76	1.73	*neckband*	−0.80	−1.54	−0.10
	doglike	1.20	0.74	1.68	*slender_post*	−0.60	−1.19	−0.02
	other_ffchar	0.97	0.39	1.56				
	lsff	0.68	0.14	1.23				
	pair_track	0.65	0.21	1.11				
ribbon	*distinctrbn*	10.87	7.86	16.18	(intercept)	−3.74	−6.12	−2.30
	faintrbn23	4.49	2.80	7.01	*long_rearflip*	−2.49	−4.44	−0.96
	alt_track	2.93	−0.46	11.46	*doglike*	−1.44	−3.96	0.57
	other_ffchar	1.81	−0.67	4.18				
bearded	*red_face*	2.01	1.42	2.65	(intercept)	−3.92	−5.49	−2.41
	assoc	1.62	0.36	3.24	*pair_track*	−1.64	−2.92	−0.40
	nocatdog	1.51	0.86	2.19	*othershape*	−1.63	−2.90	−0.40
	doglike	1.46	0.18	2.96	*slender_post*	−1.61	−2.61	−0.64
	ltcoat	1.26	0.51	2.02	*long_rearflip*	−1.17	−2.27	−0.06
	small_head	1.15	0.46	1.85				
	edge_1body	1.13	0.10	2.11				
	catlike	1.07	−0.34	2.53				
	dkcoat	1.00	0.00	2.03				
	str_track	0.98	−0.04	2.07				
	tubular	0.61	−0.17	1.39				
ringed	*hole*	2.22	1.82	2.63	(intercept)	−2.93	−3.48	−2.39
	fball	2.09	1.68	2.51	*small_head*	−1.88	−2.41	−1.36
	mot	0.87	0.51	1.23	*pair_track*	−1.50	−2.61	−0.45
					lsff	−1.02	−2.00	−0.12
					long_rearflip	−0.59	−0.99	−0.20

our observers being professional ice seal biologists, they all exhibited tendencies for misidentification to varying degrees. This is consistent with other studies assessing species misidentification based on expert observers (*e.g.*, Simons *et al*. 2007, McClintock *et al*. 2010*b*, Miller *et al*. 2012), and it therefore seems unlikely that observer training or experience will eliminate misidentification errors entirely. Furthermore, variation in both misidentification rates and confidence levels among our observers suggest it may be important to include observer effects in species distribution and abundance models that account for these sources of uncertainty.

Our observers had differing levels of field and photo-identification experience, and while experience is expected to affect observer performance, we did not attempt to incorporate individual covariates as predictors in our model. We instead included generic effects for each observer and found strong evidence of differences among our observers. These differences could be attributable to many factors, including (but not

limited to) experience, personality, age, vision, and health. This remains an interesting avenue for future research.

We used actual survey data in our study; the true species and age classes in our seal images were therefore unknown and needed to be estimated. This constitutes a key difference between our analysis and others that examined species misidentification based on experimentally generated field data (*e.g.*, Simons *et al*. 2007, McClintock *et al*. 2010*b*, Miller *et al*. 2012). The critical assumption of our approach is that positive species or age class identifications are correct, and the true species or age classes are observed in these instances. If incorrect positive identifications were pervasive, then our inferences about misidentification and the relative proportions of each species in our sample would clearly be erroneous. Our observers were informed of this critical assumption and instructed to assign positive confidences only when they were absolutely certain of species or age class. We cannot say with certainty that there were no "positive misidentifications," but the lack of conflicting positive identifications among our observers provides at least some indication that this assumption was reasonable.

Our findings add to the growing body of evidence that misidentification is pervasive in passive sampling of multiple, related animal species (*e.g.*, Lukacs and Burnham 2005, Simons *et al*. 2007, McClintock *et al*. 2010*b*). Although we found misidentification rates to be relatively low for most of our ice seal species, even low levels of misidentification have been demonstrated to induce substantial biases in estimators of species distribution and abundance (*e.g.*, McClintock *et al*. 2010*a*, Miller *et al*. 2011, Caillat *et al*. 2013, Conn *et al*. 2013*a*). Thus, even when these events are rare, it is important that analytical methods account for such errors. Fortunately, statistical models for species distribution and abundance data that account for misidentification probabilities have begun to appear in the literature, often with accompanying software (*e.g.*, Marques *et al*. 2009; Miller *et al*. 2011; Conn *et al*. 2013*a*, *b*). Similar to Miller *et al*. (2011), our misclassification model is based on repeated sampling (which can be across time, space, or observers), and it is relatively straightforward to integrate this observation model with a system process model (*e.g.*, species distribution or abundance) in a hierarchical framework (*e.g.*, Royle and Dorazio 2008).

The phenomenological and behavioral traits found to be associated with species identification and observer confidence were largely consistent with our expectations. For example, we expected observed bearded seals to be associated with a close vicinity to the ice-floe edge, red face, relatively small head, tubular body shape, and light coat. We also expected the presence of an associated nonpup (presumably the mother) would be a key predictor of observer confidence for each species of pup. However, we also found some surprising results. For example, we expected the presence of distinct ribbons and a serpentine body position to be strongly related to ribbon seal observations, but we also found that traits typically associated with other species (*e.g.*, "torpedo-like" shape, "American football" shape, and "cat-like" face) were useful predictors for observations of ribbon seal nonpups. However, these traits were not found to be important predictors for observer confidence of ribbon seals. We suspect this reflects the subjectivity of many of the phenomenological traits used to identify these species and that the perceived presence of a single trait is not necessarily indicative of one particular species. Rather, the species identification process is complicated and often requires numerous variables to be simultaneously evaluated and weighed as a whole. As such, the species identification process is very difficult to fully characterize, and our quantitative assessment represents an initial attempt to do so for these species.

For large databases, such as those arising from aerial transect and satellite surveys, a great deal of time and effort is required for observers to manually identify species from digital images. Automated pattern recognition software (*e.g.*, Arzoumanian *et al*. 2005) may have the potential to help reduce this burden, while possibly reducing misidentification and eliminating observer effects as well. In addition to learning more about the general process of species and age class identification from aerial or satellite images, we hope our characteristics analysis can serve as a baseline for initial attempts to improve identification (and account for misidentification rates) of these and similar species. However, our findings demonstrate this is a very complicated process, and any automated classification algorithm based on imagery will necessarily be very complex (but perhaps not insurmountably so). In the meantime, our methodology can be used to assess the identification process for a wide variety of species from aerial or satellite imagery and provides a mechanism for accounting for misidentification in models of species distribution and abundance.

Acknowledgments

The findings and conclusions in the paper are those of the author(s) and do not necessarily represent the views of the National Marine Fisheries Service, NOAA. Any use of trade, product, or firm names does not imply an endorsement by the U.S. Government. Funding for this study was provided by the U.S. National Oceanic and Atmospheric Administration and by the U.S. Bureau of Ocean Energy Management (Interagency Agreement M12PG00017). Aerial surveys were conducted under the authority of NMFS research permit 15126-02.

Literature Cited

Arzoumanian, Z., J. Holmberg and B. Norman. 2005. An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*. Journal of Applied Ecology 42:999–1011.

Caillat, M., L. Thomas and D. Gillespie. 2013. The effects of acoustic misclassification on cetacean species abundance estimation. Journal of the Acoustical Society of America 134:2469–2476.

Conn, P. B., B. T. McClintock, M. F. Cameron, D. S. Johnson, E. E. Moreland and P. L. Boveng. 2013*a*. Accommodating species identification errors in transect surveys. Ecology 94:2607–2618.

Conn, P. B., J. M. Ver Hoef, B. T. McClintock, *et al.* 2013*b*. Estimating multispecies abundance using automated detection systems: Ice-associated seals in the Bering Sea. Methods in Ecology and Evolution, doi:10.1111/2041-210X.12127.

Fretwell, P. T., M. A. LaRue, P. J. Morin, *et al.* 2012. An emperor penguin population estimate: The first global, synoptic survey of a species from space. PLOS ONE 7:e33751.

Fretwell, P. T., I. J. Staniland and J. Forcada. 2014. Whales from space: Counting southern right whales by satellite. PLOS ONE 9:e88655.

Heinze, G., M. Ploner, D. Dunkler and H. Southworth. 2013. logistf: Firth's bias reduced logistic regression. R package version 1.21. Available at http://CRAN.R-project.org/package=logistf.

Jefferson, T. A., M. A. Webber and R. L. Pitman. 2008. Marine mammals of the world. Academic Press, New York, NY.

LaRue, M. A., J. J. Rotella, R. A. Garrott, *et al.* 2011. Satellite imagery can be used to detect variation in abundance of Weddell seals (*Leptonychotes weddellii*) in Erebus Bay, Antarctica. Polar Biology 34:1727–1737.

Lukacs, P. M., and K. P. Burnham. 2005. Estimating population size from DNA-based closed capture-recapture data incorporating genotyping error. Journal of Wildlife Management 69:396–403.

Marques, T. A., L. Thomas, J. Ward, N. DiMarzio and P. L. Tyack. 2009. Estimating cetacean population density using fixed passive acoustic sensors: An example with Blainville's beaked whales. Journal of the Acoustical Society of America 125:1982–1994.

McClintock, B. T., L. L. Bailey, K. H. Pollock and T. R. Simons. 2010*a*. Unmodeled observation error induces bias when inferring patterns and dynamics of species occurrence via aural detections. Ecology 91:2446–2454.

McClintock, B. T., L. L. Bailey, K. H. Pollock and T. R. Simons. 2010*b*. Experimental investigation of observation error in anuran call surveys. Journal of Wildlife Management 74:1882–1993.

Miller, D. A. W., J. D. Nichols, B. T. McClintock, E. H. C. Grant, L. L. Bailey and L. A. Weir. 2011. Improving occupancy estimation when two types of observational error occur: Nondetection and species misidentification. Ecology 92:1422–1428.

Miller, D. A. W., L. A. Weir, B. T. McClintock, E. H. Campbell Grant, L. L. Bailey and T. R. Simons. 2012. Experimental investigation of false positive errors in auditory species occurrence surveys. Ecological Applications 22:1665–1674.

Moore, S. E., and H. P. Huntington. 2008. Arctic marine mammals and climate change: Impacts and resilience. Ecological Applications 18:S157–S165.

O'Brien, S. L., and F. G. Lindzey. 1998. Aerial sightability and classification of grizzly bears at moth aggregation sites in the Absaroka Mountains, Wyoming. Ursus 10:427–435.

Plummer, M. 2013. rjags: Bayesian graphical models using MCMC. R package version 3-11. Available at http://CRAN.R-project.org/package=rjags.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Royle, J. A., and R. M. Dorazio. 2008. Hierarchical modeling and inference in ecology. Academic Press, New York, NY.

Simons, T. R., M. W. Alldredge, K. H. Pollock and J. M. Wettroth. 2007. Experimental analysis of the auditory detection process on avian point counts. Auk 124:986–999.

Smith, T. G., and I. Stirling. 1975. The breeding habitat of the ringed seal (*Phoca hispida*). The birth lair and associated structures. Canadian Journal of Zoology 53:1297–1305.

Tynan, C. T., and D. P. DeMaster. 1997. Observations and predictions of Arctic climatic change: Potential effects on marine mammals. Arctic 50:308–322.

Ver Hoef, J. M., M. F. Cameron, P. L. Boveng, J. M. London and E. E. Moreland. 2014. A spatial hierarchical model for abundance of three ice-associated seal species in the eastern Bering Sea. Statistical Methodology 17:46–66.

Received: 6 October 2014
Accepted: 6 December 2014

Supporting Information

The following supporting information is available for this article online at http://onlinelibrary.wiley.com/doi/10.1111/mms.12206/suppinfo.

Appendix S1. JAGS code for the species and age classification model with observer effects.

Appendix S2. Posterior summaries for the species and age classification model without observer effects.

APPENDIX 3. – Conn, P. B., D. S. Johnson, J. M. V. Hoef, M. B. Hooten, J. M. London, and P. L. Boveng. 2015. Using spatiotemporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs 85:235-252.

This article presents a statistical advance in the area of space-time models for animal abundance and provides a description of our proposed method for incorporating the temporal dimension into our hierarchical model for estimating abundance to account for changing sea ice conditions that occurred within our survey period. Because the article is not an Open Access publication, only the abstract is provided here. The full article may be obtained from http://onlinelibrary.wiley.com/wol1/doi/10.1890/14-0959.1/full

Abstract--Ecologists often fit models to survey data to estimate and explain variation in animal abundance. Such models typically require that animal density remains constant across the landscape where sampling is being conducted, a potentially problematic assumption for animals inhabiting dynamic landscapes or otherwise exhibiting considerable spatiotemporal variation in density. We review several concepts from the burgeoning literature on spatiotemporal statistical models, including the nature of the temporal structure (i.e., descriptive or dynamical) and strategies for dimension reduction to promote computational tractability. We also review several features as they specifically relate to abundance estimation, including boundary conditions, population closure, choice of link function, and extrapolation of predicted relationships to unsampled areas. We then compare a suite of novel and existing spatiotemporal hierarchical models for animal count data that permit animal density to vary over space and time, including formulations motivated by resource selection and allowing for closed populations. We gauge the relative performance (bias, precision, computational demands) of alternative spatiotemporal models when confronted with simulated and real data sets from dynamic animal populations. For the latter, we analyze spotted seal (*Phoca largha*) counts from an aerial survey of the Bering Sea where the quantity and quality of suitable habitat (sea ice) changed dramatically while surveys were being conducted. Simulation analyses suggested that multiple types of spatiotemporal models provide reasonable inference (low positive bias, high precision) about animal abundance, but

have potential for overestimating precision. Analysis of spotted seal data indicated that several model formulations, including those based on a log-Gaussian Cox process, had a tendency to overestimate abundance. By contrast, a model that included a population closure assumption and a scale prior on total abundance produced estimates that largely conformed to our a priori expectation. Although care must be taken to tailor models to match the study population and survey data available, we argue that hierarchical spatiotemporal statistical models represent a powerful way forward for estimating abundance and explaining variation in the distribution of dynamical populations.

APPENDIX 4. - Conn, P. B., D. S. Johnson, and P. L. Boveng. 2015. On extrapolating past the range of observed data when making statistical predictions in ecology. PLoS ONE 10(10): e0141416. doi:10.1371/journal.pone.0141416

In this paper we developed statistical diagnostics for spatial regression applications in ecology to help gauge the reliability of model predictions, using an example of the ribbon seal data from the BOSS surveys.

PLOS ONE

RESEARCH ARTICLE

On Extrapolating Past the Range of Observed Data When Making Statistical Predictions in Ecology

Paul B. Conn*, Devin S. Johnson, Peter L. Boveng

National Marine Mammal Laboratory, NOAA, National Marine Fisheries Service, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115 United States of America

* paul.conn@noaa.gov

OPEN ACCESS

Citation: Conn PB, Johnson DS, Boveng PL (2015) On Extrapolating Past the Range of Observed Data When Making Statistical Predictions in Ecology. PLoS ONE 10(10): e0141416. doi:10.1371/journal.pone.0141416

Editor: Thierry Boulinier, CEFE, FRANCE

Received: April 17, 2015

Accepted: October 8, 2015

Published: October 23, 2015

Data Availability Statement: All data and R code used in our analysis have been compiled into an R package, SpatPred. Current releases are available at https://github.com/pconn/SpatPred/releases, and a static version of the package has been publicly archived via figshare. The relevant citation is Conn, P. B. (2015): SpatPred. figshare. http://dx.doi.org/10.6084/m9.figshare.1381915 Retrieved 22:14, Apr 16, 2015 (GMT).

Funding: Funding for Bering Sea aerial surveys was provided by the U.S. National Oceanic and Atmospheric Administration (NOAA; www.noaa.gov) and by the U.S. Bureau of Ocean Energy Management (BOEM; www.boem.gov) via

Abstract

Ecologists are increasingly using statistical models to predict animal abundance and occurrence in unsampled locations. The reliability of such predictions depends on a number of factors, including sample size, how far prediction locations are from the observed data, and similarity of predictive covariates in locations where data are gathered to locations where predictions are desired. In this paper, we propose extending Cook's notion of an independent variable hull (IVH), developed originally for application with linear regression models, to generalized regression models as a way to help assess the potential reliability of predictions in unsampled areas. Predictions occurring inside the generalized independent variable hull (gIVH) can be regarded as interpolations, while predictions occurring outside the gIVH can be regarded as extrapolations worthy of additional investigation or skepticism. We conduct a simulation study to demonstrate the usefulness of this metric for limiting the scope of spatial inference when conducting model-based abundance estimation from survey counts. In this case, limiting inference to the gIVH substantially reduces bias, especially when survey designs are spatially imbalanced. We also demonstrate the utility of the gIVH in diagnosing problematic extrapolations when estimating the relative abundance of ribbon seals in the Bering Sea as a function of predictive covariates. We suggest that ecologists routinely use diagnostics such as the gIVH to help gauge the reliability of predictions from statistical models (such as generalized linear, generalized additive, and spatio-temporal regression models).

Introduction

In ecology and conservation, a common goal is to make predictions about an unsampled random variable given a limited sample from the target population. For instance, given a model ($\mathcal{M}$), estimated parameters ($\hat{\boldsymbol{\theta}}$), and a covariate vector $\mathbf{x}_i$, we often desire to predict a new observation y_i at i (where i can be a design point or a spatial location). For instance, we might use a generalized linear model (GLM; [1]) or one of its extensions to predict species density or occurrence in a new location. Such predictions can be of direct use to conservation and

Interagency Agreement M12PG00017. NOAA provided salary for all three authors and funded surveys. BOEM had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

management, for instance, in estimating population abundance or distribution, and for projecting shifts in species range as a function of climate change. Spatially explicit estimates of abundance are also useful for testing theory related to biogeography or biodiversity (e.g., neutral theory; [2]), and for accurately estimating the strength of density dependence [3].

Early in their training, ecologists and statisticians are warned against extrapolating statistical relationships past the range of observed data. This caution is easily interpreted in the context of single-variable linear regression analysis; one should be cautious in using the fitted relationship to make predictions at some new response y_i whenever $x_i < \min(\mathbf{x})$ or $x_i < \max(\mathbf{x})$ (where x_i is an independent variable measured at point i). But what about more complicated situations where there are multiple explanatory variables, or when one uses a spatial regression model to account for the residual spatial autocorrelation that is inevitably present in patchy ecological data [4]? How reliable are spatially- or temporally-explicit predictions in sophisticated models for animal abundance and occurrence?

Statisticians have long struggled with the conditions under which fitted regression models are capable of making robust predictions at new combinations of explanatory variables. The issue is sometimes considered more of a philosophical problem than a statistical one, and has even been likened to soothsaying [5]. In our view, the reliability of predictions from statistical models is likely a function of several factors, including (i) the intensity of sampling, (ii) spatial or temporal proximity of the prediction location to locations where there are data, (iii) variability of the ecological process, and (iv) the similarity of explanatory covariates in prediction locations when compared to the ensemble of covariates for observed data locations.

In this paper, we investigate one possibility for defining extrapolation in the GLM and its extensions, including generalized additive models (GAMs; [6, 7]) and spatio-temporal regression models (STRMs). In particular, we exploit some of the same ideas used in multiple linear regression regarding leverage and outliers [8] to operationally define "extrapolation" as making predictions that occur outside of a generalized independent variable hull (gIVH) of observed data points. Application of the gIVH and related criterion (e.g., prediction variance) can provide intuition regarding the reliability of predictions in unobserved locations, and can aid in model construction and survey design. We illustrate use of the gIVH on simulated count data, and on several species distribution model (SDM) formulations for ribbon seals (*Histriophoca fasciata*) in the eastern Bering Sea. In particular, we examine the performance of the gIVH in identifying problematic extrapolations when modeling survey counts using GAMs, GLMs, and STRMs.

Materials and Methods

All data collected and research activities described in this manuscript were performed under National Marine Fisheries Service research permit number 15126.

Generalizing the independent variable hull

Extrapolation is often distinguished from interpolation. In a prediction context, we might define (admittedly quite imprecisely) that extrapolation consists of making predictions that are "outside the range of observed data" while interpolation consists of making predictions "inside the range of observed data." But what exactly do we mean by "outside the range of observed data"? Predictions outside the range of observed covariates? Predictions for locations that are so far (in either geographical or covariate space) from places where data are gathered that we are skeptical that the estimated statistical relationship still holds? To help guide our choice of an operational definition, we turn to early work on outlier detection in simple linear regression analysis.

In the context of outlier detection, Cook [8] defined an independent variable hull (IVH) as the smallest convex set containing all design points of a full-rank linear regression model.

Linear regression models are often written in matrix form; that is,

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \epsilon,$$

where $\mathbf{Y}$ are observed responses, $\mathbf{X}$ is a so-called design matrix that includes explanatory variables [9], and ϵ represent normally distributed residuals (here and throughout the paper, bold symbols will be used to denote vectors and matrices). Under this formulation, the IVH is defined relative to the hat matrix, $\mathbf{V}_{\mathrm{LR}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (where the subscript "LR" denotes linear regression). Letting ν denote the maximum diagonal element of $\mathbf{V}_{LR}$ (i.e., $\nu = \max(\mathrm{diag}(\mathbf{V}_{LR}))$), one can examine whether a new design point, $\mathbf{x}_0$ is within the IVH. In particular, $\mathbf{x}_0$ is within the IVH whenever

$$\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0 \leq \nu. \tag{1}$$

Cook [8] used this concept to identify influential observations and possible outliers, arguing that design points near the edge of the IVH are deserving of special attention. Similarly, points outside the IVH should be interpreted with caution.

We simulated two sets of design data to help illustrate application of the IVH (Fig 1). In simple linear regression with one predictor variable, predictions on a hypothetical response variable obtained at covariate values slightly outside the range of observed data are also outside the IVH. However, fitting a quadratic model exhibits more nuance; if there is a large gap between design points, intermediate covariate values may also be outside of the IVH and thus more likely to result in problematic predictions. Fitting a model with two covariates and both linear and quadratic effects, the shape of the IVH is somewhat more irregular, and even includes a hole in the middle of the surface when interactions are modeled (Fig 1). These simple examples highlight the sometimes counterintuitive nature of predictive inference, a problem that can only become worse as models with more dimensions are contemplated (including those with temporal or spatial structure). Fortunately, the ideas behind the IVH provide a potential way forward.

Cook's [8] formulation for the IVH is particular to linear regression analysis, which assumes independent and identically distributed (*iid*) Gaussian error. Thus, it is not directly applicable to generalized models, such as those including alternative response distributions (e.g., Poisson, binomial) or spatial random effects. Further, the hat matrix is not necessarily well defined for more complicated models with prior distributions on parameters, as with hierarchical models. However, since the hat matrix is proportional to prediction variance, Cook [8] notes that design points with maximum prediction variance will be located on the boundary of the IVH. We therefore define a generalized independent variable hull (gIVH) as the set of all predicted locations $\mathcal{S}_0$ for which

$$\mathrm{var}(\lambda_i) \leq \max[\mathrm{var}(\boldsymbol{\lambda}_{\mathcal{S}})], \tag{2}$$

where $i \in \mathcal{S}_0$, λ_i corresponds to the mean prediction at i, $\mathcal{S}$ denotes the set of locations where data are observed, and $\boldsymbol{\lambda}_{\mathcal{S}}$ denotes predictions at $\mathcal{S}$.

Generalizations of the linear model are often written in the form

$$Y_i \sim f_Y(g^{-1}(\mu_i)), \tag{3}$$

where f_Y denotes a probability density or mass function (e.g., Bernoulli, Poisson), g gives a link function, and μ_i is a linear predictor. For many such generalizations, it is possible to specify the μ_i as

$$\boldsymbol{\mu} = \mathbf{X}_{aug}\boldsymbol{\beta}_{aug}, \tag{4}$$

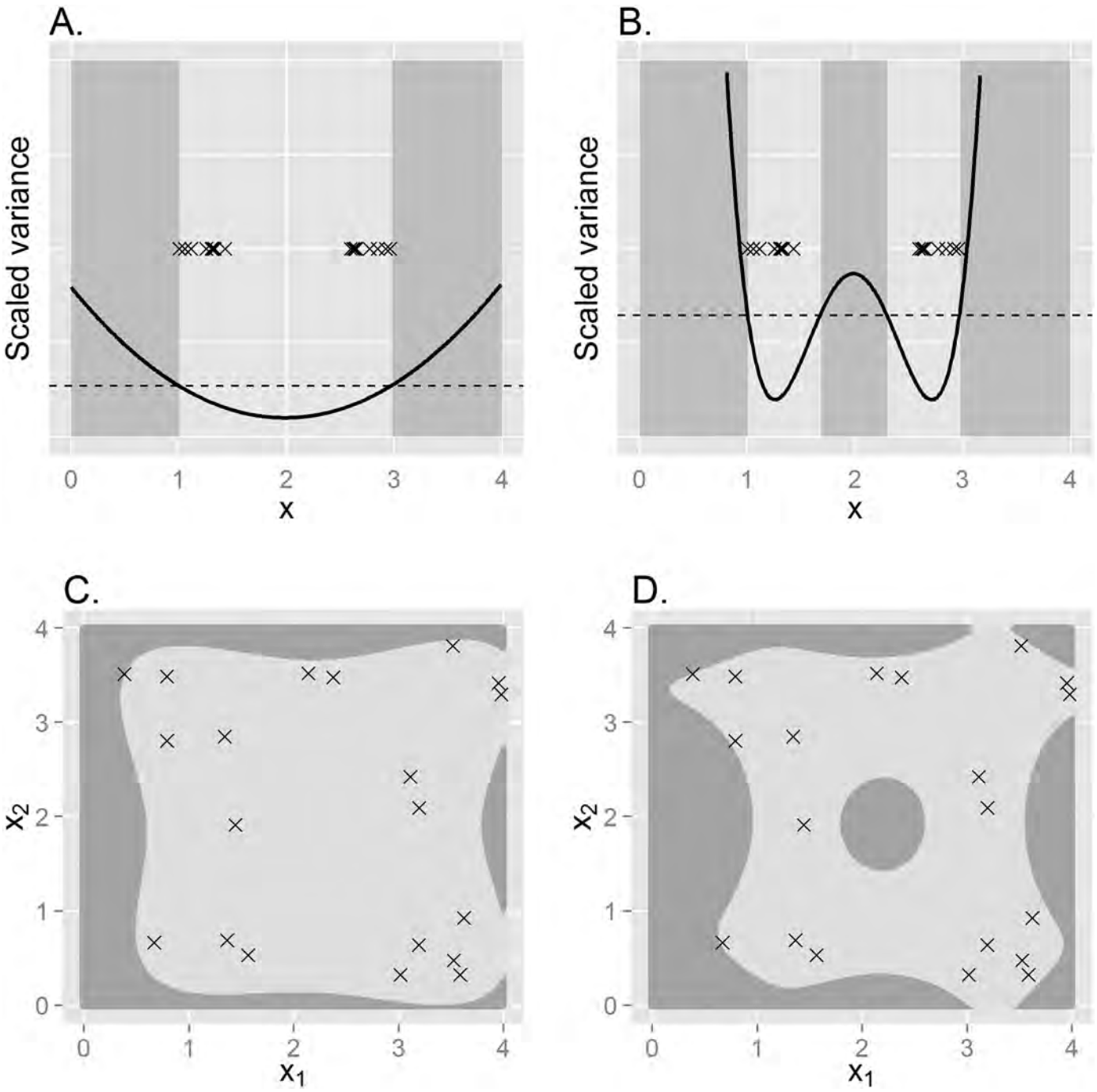

Fig 1. Example IVHs constructed from simulated data. In (A) and (B), linear regression is used to relate a response variable to a single covariate, x, obtained at locations denoted with an "x". Using x as a simple linear effect (A), only predictions less than the minimum observed value of x or greater than the maximum value of x are outside the IVH (shaded area), as scaled prediction variance in these areas (solid line) is greater than the maximum scaled prediction variance for observed data (dashed line). Using both linear and quadratic effects (B), some intermediate points are also outside the IVH. When both linear and quadratic effects of two covariates (x_1 and x_2) are modeled, the IVH is more nuanced and depends on whether interactions are omitted (C) or included (D).

doi:10.1371/journal.pone.0141416.g001

where $\mathbf{X}_{aug}$ denotes an augmented design matrix, and $\boldsymbol{\beta}_{aug}$ denote an augmented vector of parameters. For instance, in a spatial model, $\boldsymbol{\beta}_{aug}$ might include both fixed effect parameters and spatial random effects in a reduced dimension subspace (see S1 Text for examples of how numerous types of models can be written in this form).

When models are specified as in Eq 4, we can write prediction variance generically as

$$\text{var}(\hat{\boldsymbol{\mu}}) = \mathbf{X}_{aug}\text{var}(\hat{\boldsymbol{\beta}}_{aug})\mathbf{X}'_{aug}, \tag{5}$$

where it is understood that the exact form of $\mathbf{X}_{aug}$ and $\text{var}(\hat{\boldsymbol{\beta}}_{aug})$ depends on the model chosen (i.e., GLM, GAM, or STRM; S1 Text). Alternative model structures for $\boldsymbol{\mu}$ can also be accommodated; for instance, in Bayesian models $\text{var}(\boldsymbol{\mu})$ can be set equal to posterior predictive variance, $\text{var}(\tilde{\boldsymbol{\mu}}|\mathbf{Y}, \boldsymbol{\theta})$ (where $\boldsymbol{\theta}$ represent hyperparameters).

The expression for prediction variance in Eq 5 is on the linear predictor scale. If a non-identity link function is used, an additional step is needed to convert prediction variance to the response scale (i.e., to calculate $\text{var}(\boldsymbol{\lambda})$ as needed to define the gIVH in Eq 2). One approach for calculating variance on the response scale is simply to use the delta method [10, 11]. In particular, we can write the variance of the expected responses as

$$\begin{aligned} \text{var}(\hat{\boldsymbol{\lambda}}) &= \text{var}(g(\hat{\boldsymbol{\mu}})) \\ &\approx \boldsymbol{\Delta}\text{var}(\hat{\boldsymbol{\mu}})\boldsymbol{\Delta}', \end{aligned} \tag{6}$$

where $\boldsymbol{\Delta}$ is a matrix of partial derivatives of the function $g(\boldsymbol{\mu})$ with respect to its parameters, evaluated at the estimators, $\hat{\mu}$. Specifically, the rth row and cth column of $\boldsymbol{\Delta}$ is given by $\Delta_{rc} = \partial g(\mu_r)/\partial \mu_c|_{\boldsymbol{\mu}=\hat{\boldsymbol{\mu}}}$. Under common univariate link functions (e.g., log, logit, probit), $\boldsymbol{\Delta}$ has a diagonal form, while for multivariate links (e.g., multinomial logit) $\boldsymbol{\Delta}$ will be dense.

Alternatively, one can use a simulation based method for determining variance of the predictive mean response vector. In Bayesian analysis of hierarchical models, this is easily accomplished via posterior predictive inference [12]. In a similar spirit, it is also possible to use parametric bootstrapping instead of the delta method to approximate prediction variance on the response scale for frequentist models [13–15].

We propose to use the gIVH in much the same manner as Cook [16]. In particular, we use the gIVH to determine whether spatial predictions are interpolations (predictive design points lying inside the gIVH) or extrapolations (predictive design points lying outside the gIVH). For most of the following treatment, we shall assume that data have already been collected (see Discussion for comments on the potential use of the gIVH in survey planning). For further details on how the gIVH was calculated for specific models in this paper, see S1 Text.

Computing

We developed a package `SpatPred` in the R statistical programming environment [17] to simulate data and conduct all analyses. The seal dataset is included as part of this package, and is available at https://github.com/pconn/SpatPred/releases. The R package has also been archived via figshare [18].

Simulation study

We conducted a simulation study to investigate whether the gIVH (and accompanying prediction variance) was useful in diagnosing prediction biases when analyzing animal count data. For each of 100 simulations, we generated animal abundance over a 30×30 grid assuming that animal density was homogeneous in each grid cell. Animal abundance was generated as a function of three hypothetical spatially autocorrelated habitat covariates (S2 Text). For each

simulated landscape, we conducted virtual surveys of $n = 45$ survey units using two different designs: (1) a spatially balanced sample [19], and (2) a convenience sample where the probability of sampling was greater for cells closer to a "base of operations" located in the middle of the survey grid. The former approach preserves randomness while seeking a degree of regularity when distributing sampling locations across the landscape, while the latter may be easier to implement logistically.

We configured virtual sampling quadrats such that they encompassed 10% of the area of each selected grid cell. For ease of presentation and analysis, we assumed detection probability was 1.0 in each quadrat. Once animal counts were simulated, three different estimation models were fitted to the data: a GLM, a GAM, and an STRM (S1 Text). The fixed effects components of the GLM and STRM were configured to have both linear and quadratic covariate effects and first-order interactions, while the GAM expressed log-density as a function of smooth terms for each covariate (S2 Text). Each model was provided with two of the three covariates used to generate the data.

For each simulated data set and model structure, we calculated the posterior predictive variance and resulting gIVH as in Eq 2. We then calculated posterior predictions of animal abundance within and outside of each gIVH in order to gauge bias as a function of this restriction. Specifically, the performance of the gIVH may help decide its utility in limiting the scope of inference once data have been collected and analyzed, and perhaps point out areas worthy of additional sampling. A fuller, technical description of the simulation study design is provided in S2 Text; a visual depiction of a single simulation replicate is displayed in Fig 2.

Ribbon seal SDM

As part of an international effort, researchers with the U.S. National Marine Fisheries Service conducted aerial surveys over the eastern Bering Sea in 2012 and 2013. Agency scientists used infrared video to detect seals that were on ice, and collected simultaneous digital photographs to provide information on species identity. For this study, we use spatially referenced count data from photographed ribbon seals, *Phoca fasciata* on a subset of 10 flights flown over the Bering Sea from April 20–27, 2012. We limited flights to a one week period because sea ice melts rapidly in the Bering Sea in the spring, and modeling counts over a longer duration would likely require addressing how sea ice and seal abundance changes over both time and space [20]. However, limiting analysis to a one week period makes the assumption of static sea ice and seal densities tenable [21].

Our objective with this dataset will be to model seal counts on transects through 25km by 25km grid cells as a function of habitat covariates and possible spatial autocorrelation. Estimates of apparent abundance can then be obtained by summing predictions across grid cells. Fig 3 show explanatory covariates gathered to help predict ribbon seal abundance. These data are described in fuller detail by [21], who extend the modeling framework of STRMs to account for incomplete detection and species misidentification errors. Since our focus in this paper is on illustrating spatial modeling concepts, we devote our efforts to the comparably easier problem of estimating apparent abundance (i.e., uncorrected for vagaries of the detection process).

Inspection of ribbon seal data (Fig 4) immediately reveals a potential issue with spatial prediction: abundance of ribbon seals appears to be maximized in the southern and/or southeast quadrant of the surveyed area. Predicting abundance in areas farther south and west may thus prove problematic, as the values of several explanatory covariates (Fig 3) are also maximized in these regions.

We start by fitting hierarchical GLMs and STRMs to the ribbon seal data. To accommodate incomplete coverage of grid cells and account for non-target habitat, we adapted Eqs 3 and 4 as

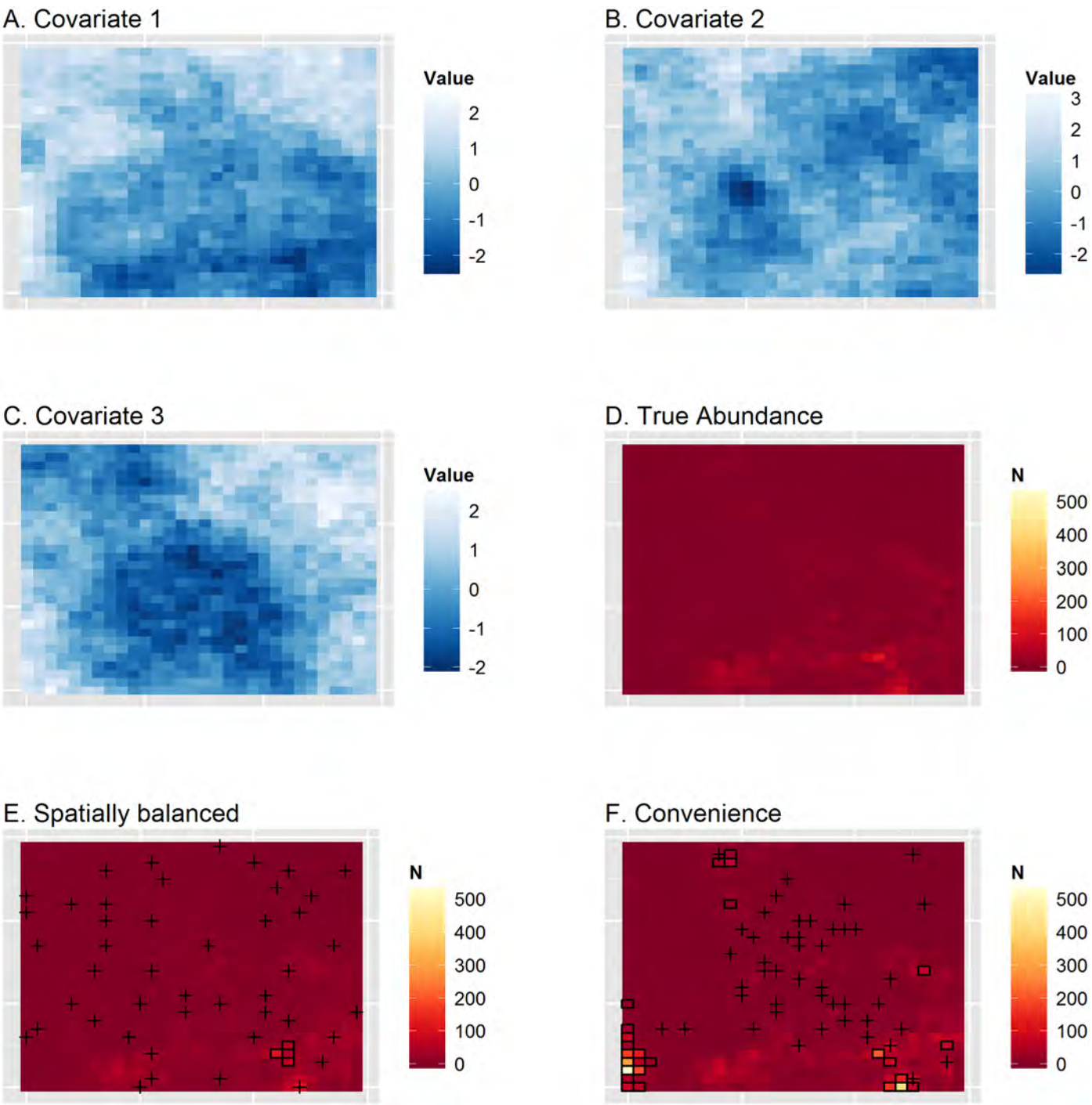

Fig 2. Depiction of a single simulation replicate where problematic extrapolation occurs. Panels (A-C) give simulated covariate values, panel D gives true animal abundance, (E) gives estimated abundance from a GLM run on count data from a spatially balanced survey design, and (F) gives abundance from a GLM applied to count data from a convenience survey. In (E-F), predictions outside the gIVH are represented by black boxes, and sampling locations are represented with an x. For the convenience sample, there was considerable positive bias, particularly in cells outside of the gIVH. In this case, the median posterior abundance prediction for the entire survey area is 57% greater than true abundance when inference is made to the whole study area. When inference is restricted to cells within the gIVH, median posterior abundance is 16% greater than true abundance.

doi:10.1371/journal.pone.0141416.g002

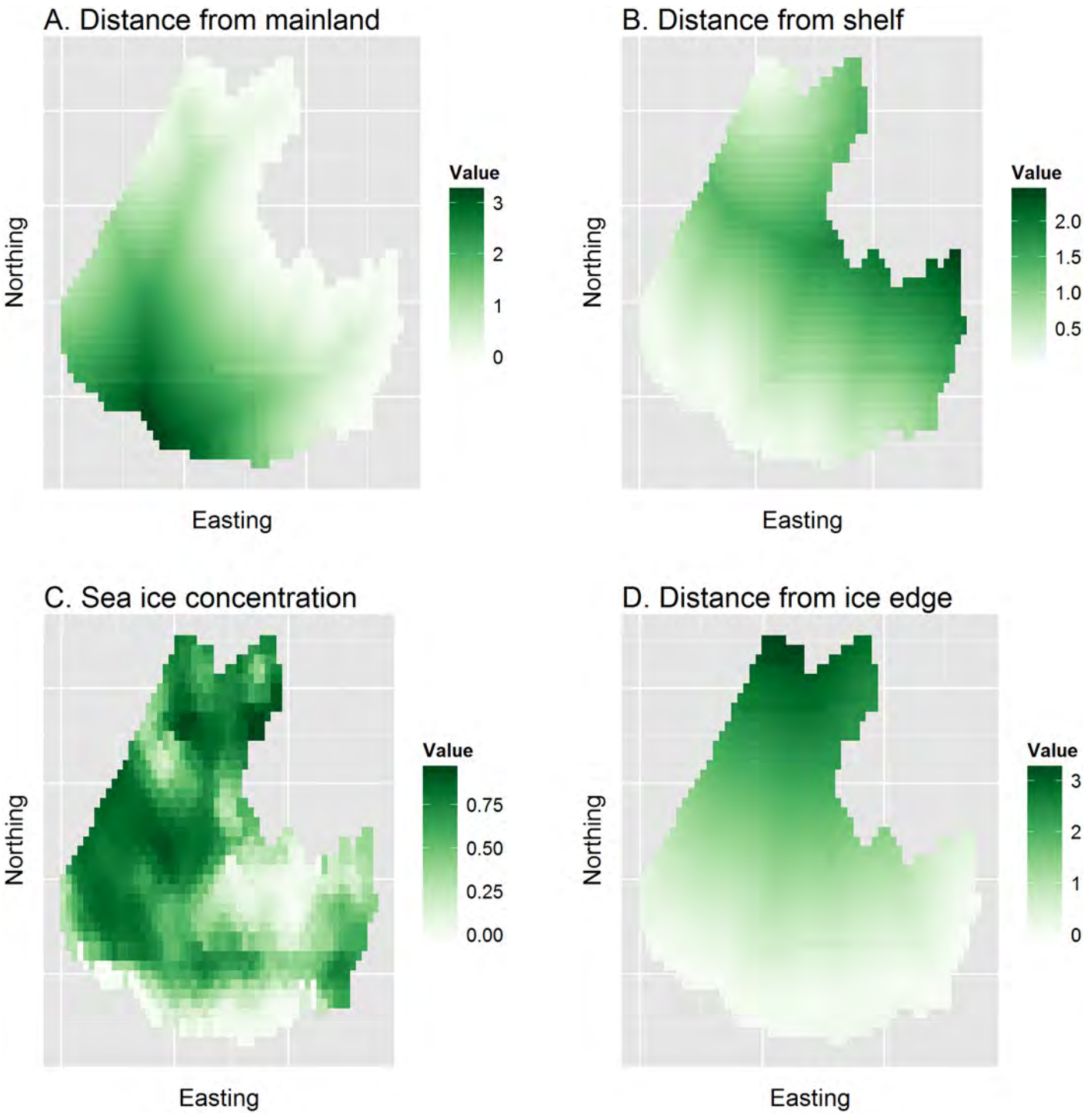

Fig 3. Assembled covariates used to help explain and predict ribbon seal relative abundance in the eastern Bering Sea. Covariates include distance from mainland (`dist_mainland`), distance from 1000m depth contour (`dist_shelf`), average remotely sensed sea ice concentration while surveys were being conducted (`ice_conc`), and distance from the southern sea ice edge (`dist_edge`). All covariates except ice concentration were standardized to have a mean of 1.0 prior to plotting and analysis.

doi:10.1371/journal.pone.0141416.g003

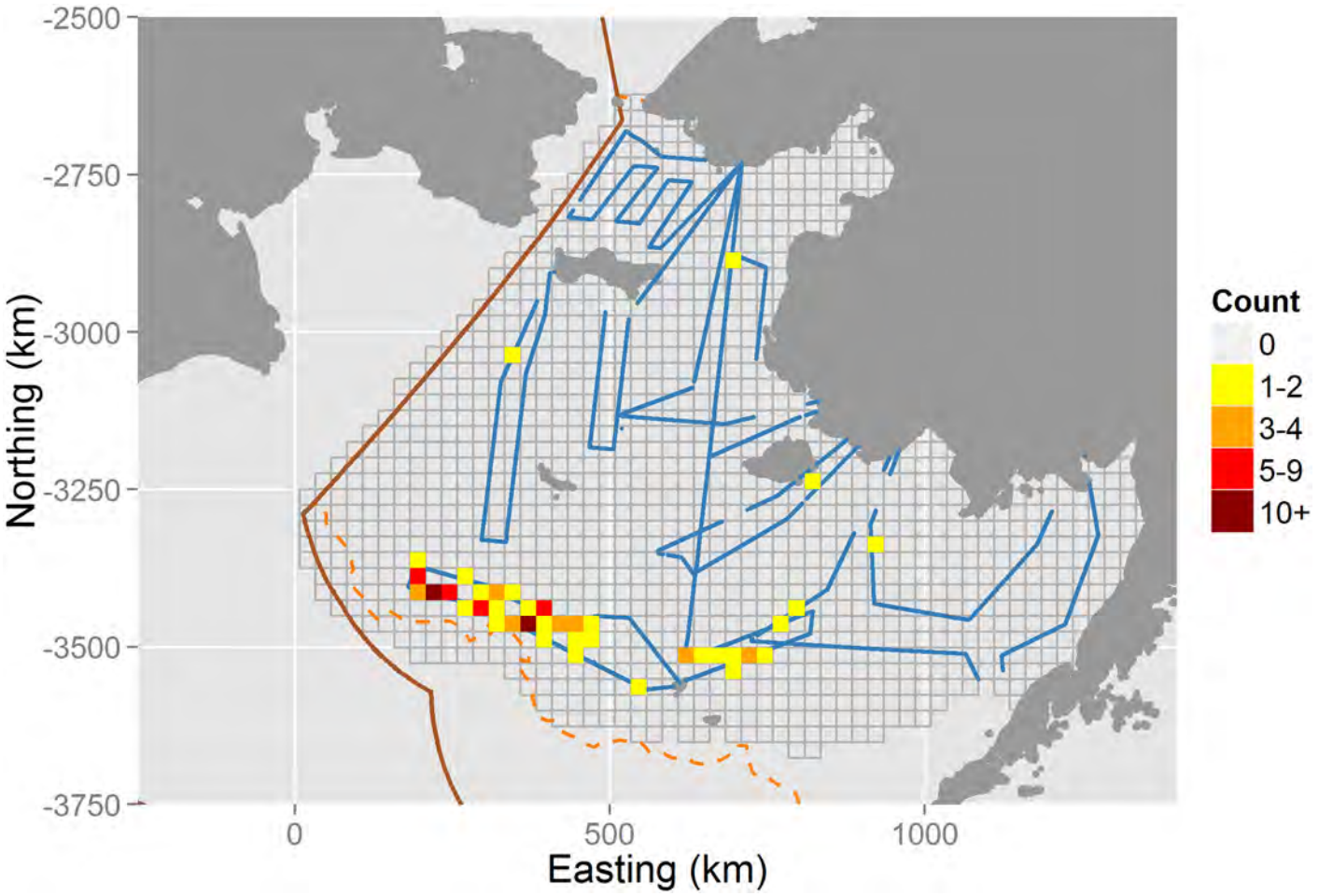

Fig 4. Aerial survey tracks over the Bering Sea, April 22–29, 2012. Survey tracks are shown in blue, and are overlayed on a tesselated study area consisting of 25km by 25km grid cells (gray lines). Dark gray indicates land, while the orange dashed line indicates a 1000m depth contour, and the solid brown line shows the U.S Exclusive Economic Zone (EEZ) boundary. Colored pixels indicate ribbon seal counts along aerial transects. The average effective area surveyed in each grid cell was approximately 2.6km^2 (0.4%). Note that surveys were designed to target multiple seal species, several of which had high densities further north (results not shown).

doi:10.1371/journal.pone.0141416.g004

follows. First, let Y_i denote the ribbon seal count (Y_i) obtained in sampled grid cell i. Suppose that counts arise according to a log-Gaussian Cox process, such that

$$\begin{aligned} Y_i &\sim \text{Poisson}(\lambda_i) \text{ and} \\ \log(\lambda_i) &= \log(P_i) + \log(A_i) + \theta_i + \epsilon_i, \end{aligned} \tag{7}$$

where P_i gives the proportion of area surveyed in grid cell i, A_i gives the proportion of cell i that is seal habitat, θ_i is a linear predictor, and ϵ_i is normally distributed *iid* error. By formulating θ_i differently, we can arrive at representations characteristic of GLMs and STRMs (see S1 Text).

The fixed effects component of the GLM and STRM included linear effects of all explanatory covariates (Fig 3), as well as a quadratic effect for sea ice concentration. For the STRM, we imposed a restricted spatial regression (RSR) formulation for spatially autocorrelated random effects, where dimension reduction was accomplished by only selecting eigenvectors of the spectral decomposition associated with eigenvalues that were greater than 0.5 (see S1 Text for additional information on model structure). Adopting a Bayesian perspective, we

estimated parameters for these models using MCMC (see S1 Text for algorithm details and information on prior distributions) with 60,000 iterations where the first 10,000 iterations were discarded as a burn-in. We generated posterior predictions of ribbon seal abundance across the landscape as

$$N_i \sim \text{Poisson}(A_i\lambda_i), \tag{8}$$

and calculated the gIVH as in Eq 2, with delta method modifications as specified in Eq 6.

We also fitted a frequentist GAM to seal data using the `mgcv` R package [7]. We included smooth terms for all explanatory covariates; however, owing to relative data sparsity, we provided `mgcv` with the smallest basis size allowable (k = 3) for the default thin plate spline smoother. We used a quasipoisson error structure in `mgcv` for this analysis, which was the most similar option available to the log-Gaussian Cox formulation chosen for the GLM and STRM models. For more information on the procedure used to generate parameter estimates and abundance predictions on the response scale, see S1 Text.

Initial spatial predictions using two of the three models (GLM, STRM) produced extremely high, unbelievable predictions along the southern boundary of the study area (Fig 5). Predictions in this region were also largely out of the gIVH, indicating the potential utility for the gIVH in revealing problematic extrapolations. We considered several possible alternatives for trying to obtain more robust abundance estimates before settling on a preferred alternative. First, one could refine the study area to eliminate predictions outside of the gIVH (as in the simulation study). However, this is not ideal in that one does not get an abundance estimate for the whole study area, and it may be difficult to compare abundance from one year to the next using this approach. Second, one could try different predictive covariate models (e.g., by altering the combination or polynomial degree of covariates included in the model). Finally, one could build in a priori knowledge of habitat preferences into the model structure. We adopted the latter solution, incorporating presumed absences (i.e., zero counts where sampling was not conducted) in locations where it would have been (nearly) impossible to detect seals. Specifically, we inserted presumed absences in cells where ice concentrations were $<0.1\%$. This solution seemed the most logical, as many of the large, anomalous predictions were over open water along the southern edge of the study area, where we would have obtained zero counts had they been surveyed. This approach effectively requires that sea ice concentration be included as a predictive covariate to help model absences in cells without ice.

Results

Simulation study

Posterior predictions from simulations indicated that the distribution for proportional error in total abundance was right skewed when statistical inference was made with regard to the entire survey area (Fig 6). Although median bias was close to zero, this right skew translated into positive mean bias, and was exacerbated when convenience sampling was employed. The magnitude of mean absolute bias was either the same or reduced (often substantially so) when inference was constrained to the gIVH. Positive proportional bias was the rule, and was of concerning magnitude (≈ 0.3) for GLMs and STRMs when convenience sampling was employed and inference was not restricted to the gIVH. By contrast, proportional bias was close to zero when inference was restricted to the gIVH, although there appeared to be a small negative bias (Fig 6). Interestingly, bias for frequentist GAMs was of smaller magnitude than the Bayesian GLM or STRM models for the particular model structures used here.

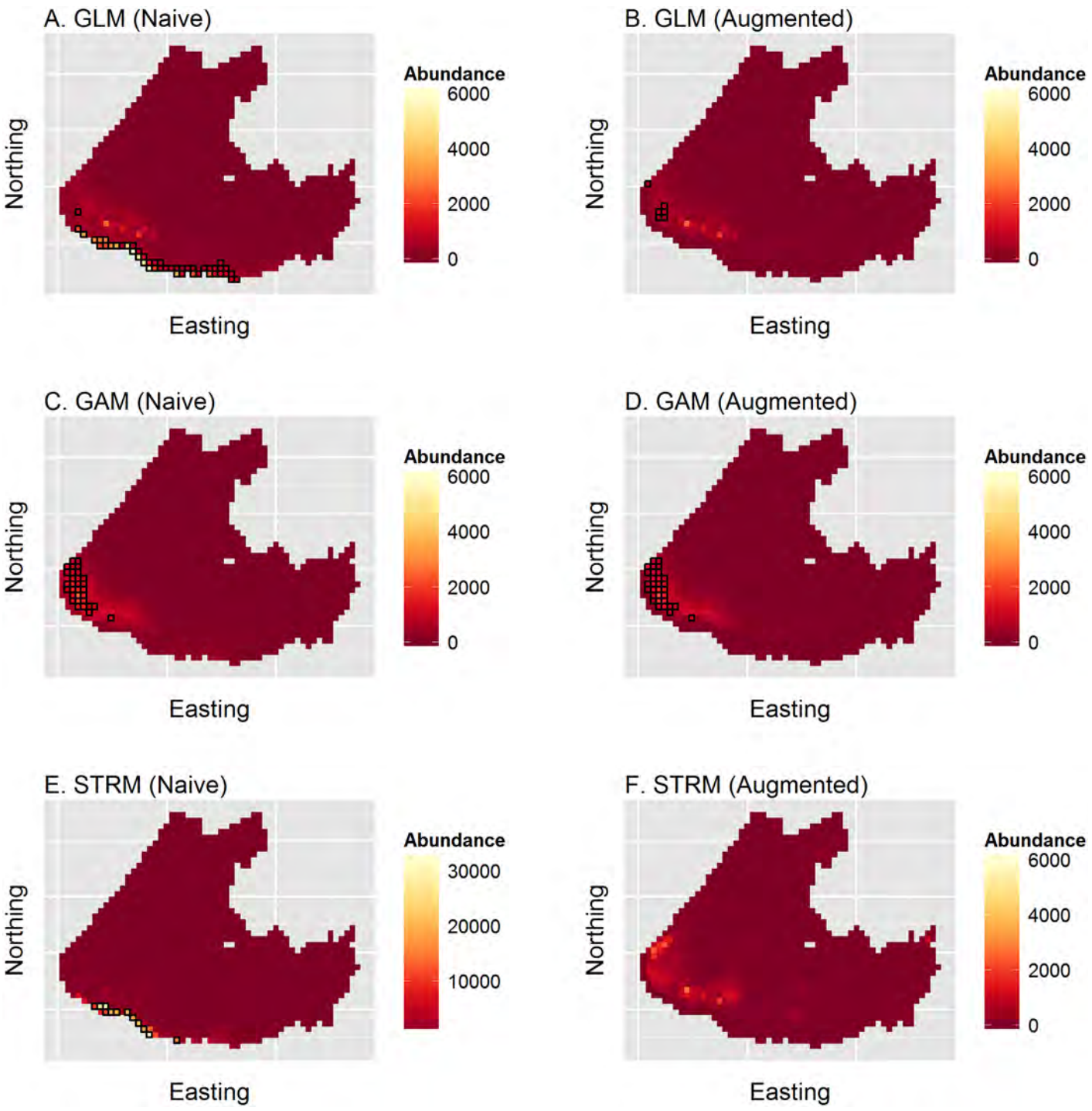

Fig 5. Predictions of ribbon seal apparent abundance across the eastern Bering sea from models fit to survey data. Predictions were obtained using the posterior predictive mean for GLM and STRM models, and for the GAM using the `predict.gam` function in the R `mgcv` package [7]. Each row gives result for different model types (GLM, GAM, or STRM, respectively); left column plots give results for naive runs without presumed absences, while plots in the right column give predictions for runs where presumed absence data (i.e., 0 counts in cells with <0.1% ice) were included. Cells highlighted in black indicate those where predictions were outside the generalized independent variable hull (gIVH).

doi:10.1371/journal.pone.0141416.g005

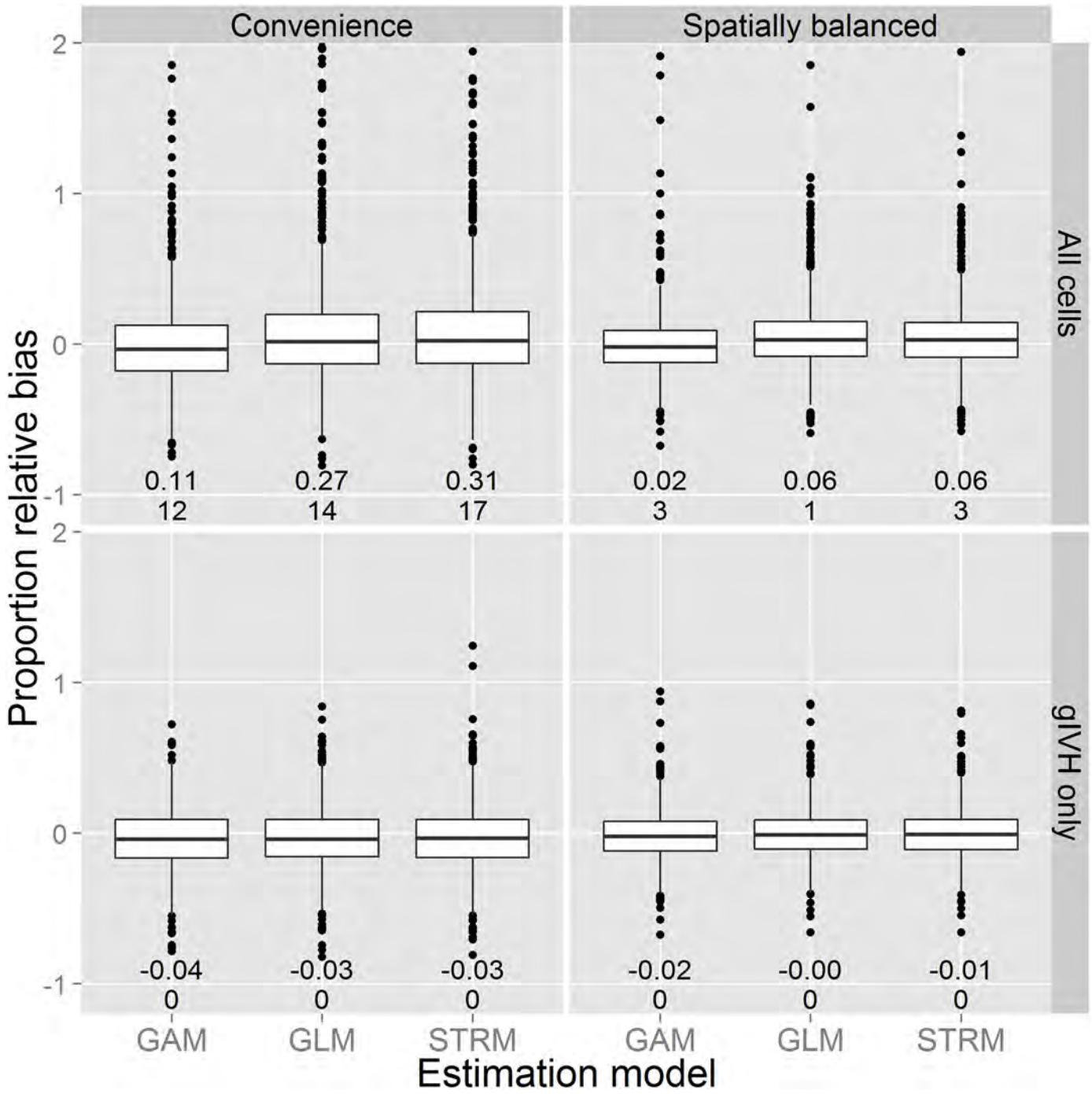

Fig 6. Boxplots summarizing proportional error in abundance from the simulation experiment. Each boxplot summarizes the distribution of proportional error in the posterior predictive median of abundance as a function of estimation model (x-axis), survey design (columns) and whether or not inference was restricted to the gIVH (rows). The lower and upper limits of each box correspond to first and third quartiles, while whiskers extend to the lowest and highest observed bias within 1.5 interquartile range units from the box. Outliers outside of this range are denoted with points. Horizontal lines within boxes denote median bias. The two numbers located below each boxplot indicate mean bias (upper number) and the number of additional outliers for which proportional bias was greater than 2.0 (lower number).

doi:10.1371/journal.pone.0141416.g006

Ribbon Seal SDM

Fitting our three ribbon seal SDMs to the augmented dataset with presumed absences, most predictions occurred within the gIVH (Fig 5). Posterior summaries of abundance across the entire study area were of similar magnitude, with 5%, 50%, and 95% posterior prediction quantiles as follows: GLM (48,686, 64,836, 93,927); STRM(41,039, 63,717, 194,095). The GAM produced an estimate of 92,277 (90% bootstrap CI: 63,090, 129,367). The largest differences among the three models was in the southwest corner of the study area in the area where predictions often occurred outside the gIVH. Restricting comparison of abundance to those cells that occur within the gIVH in all three models (i.e., only cells not highlighted in the right column of Fig 5), posterior prediction quantiles for the GLM were (41,750, 52,863, 69,557) and for the STRM were (39,446, 56,520, 135,427); estimated GAM abundance from `mgcv` was 59,104 (90% bootstrap CI:52,629, 73,076). There thus appears to be substantial between-model variation in predicted abundance when summed over the entire study area, but much better agreement (albeit with a heavier right tail for the STRM) when restricting inference to locations where predictions occur within the gIVH.

We note that these estimates are for example illustration only, as they are uncorrected for imperfect detection (e.g., incomplete detection of thermal cameras, animals that were unavailable for sampling because they were in the water, species misidentification; [21]). Our approach here was to examine extrapolation and prediction error using relatively simple models, with the understanding that such effects are also likely to occur in complex models with more realistic observation processes. Standard diagnostics (e.g., q-q plots in `mgcv`) also suggested some lack of fit associated with the quasipoisson error distribution; future work should investigate alternate error structures such as the Tweedie distribution [15]. Although not reported here, additional model fitting suggested sensitivity to model complexity and choice of basis, both of which are worthy of additional investigation.

Discussion

We have demonstrated the capacity of certain classes of statistical models to produce biased predictions of animal abundance when extrapolating past the range of observed data. In simulations, commonly used models exhibited substantial mean positive bias when predictions were required for the entire study area, particularly when convenience sampling was employed. Median bias in the simulation study was close to zero, but the bias distribution was right skewed, indicating the possibility of considerably biased overestimates in a substantial proportion of simulation replicates. By contrast, restricting inference to locations within the gIVH led to small negative bias. Although this negative bias is undesirable, it may be preferable from a conservation and management standpoint. For instance, making management decisions (e.g., harvest, restoration efforts) based on estimates that have a small negative bias are much less likely to lead to catastrophic population collapse than are decisions based on overestimates.

In the ribbon seal example, naive extrapolation of fitted statistical relationships produced high positive bias along the southern boundary of the study area for the GAM and STRM models. However, the gIVH appeared useful in diagnosing places where extrapolations from the fitted statistical model were problematic. For ribbon seal relative abundance, it was useful for confirming that the naive models needed to be reformulated. Reformulated models (with presumed absence data) still yielded estimates of total abundance with considerable between-model variation in the southwest corner of the study area. However, when inference was restricted to locations within the gIVH for all three fitted models, abundance estimates were quite comparable.

When estimating species distributions, researchers often stress the need for prediction locations to be similar to the locations used for model development [22]. One way to accomplish

this is through a prediction envelope, whereby a specific criterion is used to limit predictions of animal density or occurrence to the range of conditions and covariates encountered during surveys [23]. Using the gIVH for this purpose will likely be more conservative than envelope specifications based on other criterion (e.g., in contrast to minimum and maximum observed covariate values as in [23]), but is more in line with linear modeling theory. A comparison of envelope specification methods is beyond the scope of this paper, but we suspect there are cases where seemingly intuitive envelope strategies result in problematic extrapolations, particularly when the form of prediction models is of high dimension or includes multiple interaction terms.

In SDMs and model-based abundance estimation, the goal for analysts is often to build predictive maps of species abundance or occurrence using a limited number of sample locations. In such applications, the ultimate aim of analysts should be to build models that have low bias and high precision. However, traditional approaches to quantifying bias (e.g., goodness-of-fit statistics) only work with observed data points. When inference is extended to unsampled locations, the gIVH appears to be a useful diagnostic for whether bias for predictions in unsampled locations can be expected. In some cases, biological knowledge and intuition may be sufficient to diagnose anomalous predictions. However, such determinations are likely to be quite subjective, and may prove insufficient when there are a large number of regression coefficients and interaction terms. For instance, even relatively simple regression models may exhibit non-intuitive patterns (e.g., Fig 1). Further, relying on expert opinion alone in successive rounds of model formulation and fitting may lead to investigators choosing models based on how much they like the results, which is clearly not ideal scientific practice.

Our intent is to raise awareness of potential problems with extrapolation bias in statistical models, and to provide an additional tool (the gIVH) to help diagnose its presence. Other methods for selecting models to enhance predictive performance, such as cross validation [24], are also useful for this purpose, but may not entirely eliminate the problem (particularly for sparse datasets). One approach that might be useful in practice is to combine the cross validation and gIVH paradigms—for instance, using cross validation to narrow down the field to a suite of models with good predictive performance at test locations, and then calculating the gIVH to examine the potential for anomalously high predictions in unsampled locations.

The analyses in this paper focused on abundance estimation, which is necessarily non-negative. As such, counts are usually analyzed with a log link function, and there is a much greater potential for positive bias than negative. Since prediction variance tends to increase as a function of the magnitude of the prediction, the gIVH will only tend to be able to diagnose predictions that are anomalously large. However, one could also apply the gIVH when predicting species occurrence from presence/absence data. In this case, common link functions (e.g., probit or logit) are symmetric, and potential for positive and negative bias in predictive maps seem equally likely. Future research should be directed to examine conditions under which the gIVH is a useful diagnostic in such applications.

One area that gIVH ideas may also prove useful is in formulating survey designs. The topic of optimal (or near-optimal) spatial design has received considerable attention in the statistical literature, often in the context of designing environmental monitoring programs [25]. Optimal designs can be sensitive to the structure of the estimation model that is used, so that tailoring a survey design to a particular model can be somewhat dangerous if there is uncertainty about the ultimate "best" structure for the model used to relate animal abundance and occurrence to available covariates. Nevertheless, one could still think about augmenting a given sampling design with a number of locations which are known or thought to have high prediction variance as a function of available covariates (e.g., the southwest corner of the study area in the ribbon seal example). This could potentially be done as an exercise before any data (or perhaps

data from a pilot study) have been collected. We are excited about this prospect, and it is a subject of current research.

Supporting Information

S1 Text. Model formulation, Gibbs sampling algorithms, and gIVH calculations for certain classes and extensions of the generalized linear model.
(PDF)

S2 Text. Full details of simulation study examining predictive extrapolation.
(PDF)

Acknowledgments

We thank J. Thorson and J. Ver Hoef for helpful discussion on an earlier version of this manuscript. David L. Miller also provided an insightful review that led us to incorporate frequentist GAMs into the suite of estimation models considered. The views and conclusions in this article do not necessarily represent findings or policy of the U.S. National Oceanic and Atmospheric Administration. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author Contributions

Conceived and designed the experiments: PBC DSJ PLB. Performed the experiments: PBC. Analyzed the data: PBC. Contributed reagents/materials/analysis tools: PBC DSJ PLB. Wrote the paper: PBC DSJ PLB.

References

1. McCullagh P, Nelder JA. Generalized Linear Models. New York: Chapman and Hall; 1989.
2. Hubbell SP. The Unified Neutral Theory of Biodiversity and Biogeography. Princeton, New Jersey, USA: Princeton University Press; 2001.
3. Thorson JT, Skaug H, Kristensen K, Shelton AO, Ward EJ, Harms J, et al. The importance of spatial models for estimating the strength of density dependcence. Ecology. In press;p.
4. Lichstein JW, Simons TR, Shriner SA, Franzreb KE. Spatial autocorrelation and autoregressive models in ecology. Ecological Monographs. 2002; 72:445–463. doi: 10.1890/0012-9615(2002)072%5B0445:SAAAMI%5D2.0.CO;2
5. Ehrenberg A, Bound J. Predicatibility and prediction. Journal of the Royal Statistical Society A. 1993; 156:167–206. doi: 10.2307/2982727
6. Hastie TJ, Tibshirani RJ. Generalized Additive Models. Boca Raton, Florida: Chapman & Hall/CRC; 1999.
7. Wood SN. Generalized additive models. Boca Raton, Florida: Chapman & Hall/CRC; 2006.
8. Cook RD. Influential observations in linear regression. Journal of the American Statistical Association. 1979; 74:169–174. doi: 10.1080/01621459.1979.10481634
9. Draper NR, Smith H. Applied Regression Analysis. New York: John Wiley & Sons; 1966.
10. Dorfman R. A note on the delta-method for finding variance formulae. The Biometric Bulletin. 1938; 1:129–137.
11. Ver Hoef JM. Who invented the delta method? The American Statistician. 2012; 66(2):124–127. doi: 10.1080/00031305.2012.687494
12. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis, 2nd Edition. Boca Raton: Chapman and Hall; 2004.
13. Marra G, Miller DL, Zanin L. Modelling the spatiotemporal distribution of the incidence of resident foreign population. Statistica Neerlandica. 2012; 66(2):133–160. doi: 10.1111/j.1467-9574.2011.00500.x
14. Augustin NH, Trenkel VM, Wood SN, Lorance P. Space-time modelling of blue ling for fisheries stock management. Environmetrics. 2013; 24(2):109–119. doi: 10.1002/env.2196

15. Miller DL, Burt ML, Rexstad EA, Thomas L. Spatial models for distance sampling data: recent developments and future directions. Methods in Ecology and Evolution. 2013; 4:1001–1010. doi: 10.1111/2041-210X.12105

16. Connor EF, McCoy ED. The Statistics and Biology of the Species-Area Relationship. The American Naturalist. 1979; 113(6):pp. 791–833. Available from: http://www.jstor.org/stable/2460305.

17. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2012. ISBN 3-900051-07-0. Available from: http://www.R-project.org.

18. Conn PB. SpatPred; 2015. Http://dx.doi.org/10.6084/m9.figshare.1381915.

19. Stevens DL Jr, Olsen AR. Spatially balanced sampling of natural resources. Journal of the American Statistical Association. 2004; 99:262–278. doi: 10.1198/016214504000000250

20. Conn PB, Johnson DS, Ver Hoef JM, Hooten MB, London JM, Boveng PL. Using spatio-temporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. Ecological Monographs. In Press;.

21. Conn PB, Ver Hoef JM, McClintock BT, Moreland EE, London JM, Cameron MF, et al. Estimating multispecies abundance using automated detection systems: ice-associated seals in the eastern Bering Sea. Methods in Ecology and Evolution. 2014; 5:1280–1293. doi: 10.1111/2041-210X.12127

22. Elith J, Leathwick JR. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics. 2009; 40:677–697. doi: 10.1146/annurev.ecolsys.110308.120159

23. Mannocci L, Monestiez P, Spitz J, Ridoux V. Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. Journal of Biogeography. 2015; 42:1267–1280. doi: 10.1111/jbi.12530

24. Picard RR, Cook RD. Cross-validation of regression models. Journal of the American Statistical Association. 1984; 79:575–583. doi: 10.1080/01621459.1984.10478083

25. Müller WG. Collecting Spatial Data; Optimum Design of Experiments for Random Fields, Third Edition. Berlin: Springer; 2007.

APPENDIX 5. - Conn, P. B., J. T. Thorson, and D. S. Johnson. *In review*. Confronting preferential sampling in wildlife surveys: diagnosis and model-based triage. Methods in Ecology and Evolution

In this paper we used data from bearded seal density and distribution in the BOSS surveys to demonstrate important considerations in the design and analysis of model-based surveys when it is not always possible to adhere to pre-planned survey routes.

Methods in Ecology & Evolution 2016 doi: 10.1111/...

ARTICLE

Confronting preferential sampling in wildlife surveys: diagnosis and model-based triage †

Paul B. Conn[1*], **James T. Thorson**[2], **Devin S. Johnson**[1]

[1] *National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle, WA 98115 USA;* [2] *Fisheries Resource Assessment and Monitoring Division (FRAM), Northwest Fisheries Science Center, National Marine Fisheries Service (NMFS), NOAA, 2725 Montlake Boulevard E, Seattle, WA 98112, USA*

Summary

1. Wildlife surveys are often used to estimate the density, abundance, or distribution of animal populations. Recently, model-based approaches to analyzing survey data have become popular because one can more readily accommodate departures from pre-planned survey routes and construct more detailed maps than one can with design-based procedures.
2. Species distribution models fitted to wildlife survey data often make the implicit assumption that locations chosen for sampling and animal abundance at those locations are conditionally independent given modeled covariates. However, this assumption is likely violated in many cases when survey effort is non-randomized, leading to preferential sampling.
3. We develop a hierarchical statistical modeling framework for detecting and alleviating the biasing effects of preferential sampling in species distribution models fitted to count data. The approach works by jointly modeling wildlife state variables and the locations selected for sampling, and specifying a dependent correlation structure between the two models.
4. Using simulation, we show that moderate levels of preferential sampling can lead to large (e.g. 40%) bias in estimates of animal density, and that our modeling approach can considerably reduce this bias.
5. We apply our approach to aerial survey counts of bearded seals (*Erignathus barbatus*) in the eastern Bering Sea. Models that included a preferential sampling effect led to lower estimates of abundance than models without, but the effect size of the preferential sampling parameter decreased in models that included explanatory environmental covariates.
6. When wildlife surveys are conducted without a well-defined sampling frame, ecologists should recognize the potentially biasing effects of preferential sampling. Joint models, such as those described in this paper, can be used to test and correct for such biases. Predictive covariates are also useful for bias reduction, but ultimately the best way to avoid preferential sampling bias is to incorporate design-based principles such as randomization and/or systematic sampling into survey design.

Word count: 6038

Key-words: count data, preferential sampling, spatial autocorrelation, species distribution model

Prepared using besauth.cls *[Version: 2011/03/08 v1.00]*

Introduction

Surveys of unmarked animal populations are often used to estimate abundance and occurrence of animal populations and to predict species distributions, enterprises central to conservation, ecology, and management. For studies of abundance, researchers historically relied on design-based statistical inference (e.g. Cochran 1977), which requires adoption of a pre-defined sampling frame (e.g. using systematic random sampling, stratified random sampling, or some variant thereof). Designing animal surveys is relatively straightforward in such applications, and unbiased point and variance estimators are available. Recently, however, there has been a surge in research describing model-based procedures for estimating abundance, density, and occupancy from surveys of unmarked animals, including N-mixture and Dail-Madsen models for repeated point counts (Royle 2004; Dail & Madsen 2011), occupancy models for presence-absence surveys (MacKenzie *et al.* 2002; Johnson *et al.* 2013), and various model-based formulations for distance-sampling data (Hedley & Buckland 2004; Johnson *et al.* 2010; Miller *et al.* 2013). In such applications, it is common to use habitat or environmental covariates together with spatial effects (e.g. via trend surfaces or spatial random effects) to predict density or distributions across the landscape. We shall refer to the amalgam of model-based approaches for making spatially explicit inference about animal populations as "species distribution models" (SDMs; *sensu* Elith & Leathwick 2009), even though this term is more often used to refer to animal occurrence than it is to density or abundance.

One of the main advantages of using SDMs is that one is no longer beholden to predetermined sampling frames, and can potentially use data gathered from non-randomized designs or platforms of opportunity to make inferences about animal populations (Johnson *et al.* 2010). However, in a recent paper, Diggle *et al.* (2010) emphasized that spatially explicit statistical models can easily provide biased estimates when sampling disproportionately targets locations where the response of interest is higher (or lower) than expected given a particular set of explanatory covariates. In the context of SDMs, this might occur if sampling disproportionately occurs in locations where animals are known to be present or of high abundance. For example, if volunteer inventory participants have access to multiple sites with similar covariate values, bias might arise if they consistently choose sites where species are thought or known to be present. Bias might also arise if surveying effort is higher near bases of operations, and if animal abundance is higher (or lower) near bases of operations than elsewhere in the landscape.

In this article, we explore potential for bias in SDMs resulting from preferential sampling (hereafter, PS), and describe several model-based approaches for detecting and correcting for such biases. We start by describing a common currency for notation and basic model structures considered in this paper. Second, we review PS bias in a mathematical light, and describe prior approaches to coping with its effects. Third, we introduce a novel generalization of previously proposed PS models, allowing the investigator to jointly model animal encounter data and the locations chosen for sampling, including possible dependence structure between these two types of observations. Fourth, we conduct a simulation study to examine the performance of traditional SDMs and our newly developed PS model when data are gathered preferentially. Finally, we demonstrate our modeling approach by analyzing aerial survey counts of bearded seals (*Erignathus barbatus*) in the Bering Sea.

Materials and methods

NOTATION AND BASIC MODEL STRUCTURES

We focus here exclusively on discrete space (areal) models for animal encounter data as these seem to be the dominant form used in design and analysis of animal population surveys, although we note that PS is likely to affect analyses similarly regardless of the choice of spatial domain. We suppose that the investigator intending to fit a SDM to animal encounter data breaks their study area up into S survey units (label these $U_1, U_2, \ldots, U_S$), of which n are selected for sampling (call the set of sampled locations $\mathcal{S}$). Each survey unit i is assigned a vector of covariates, $\mathbf{x}_i$, and an indicator R_i that takes on the value 1.0 if location i is sampled (i.e. if $U_i \in \mathcal{S}$), and is 0 otherwise. To formulate a "traditional" SDM, one could then write animal abundance or occurrence as a stochastic realization of a probability mass function $f()$:

$$Z_i \quad \sim \quad f(g^{-1}(\mu_i)). \qquad \text{eqn 1}$$

*Correspondence author. E-mail: paul.conn@noaa.gov

In this example, Z_i denotes the state variable of interest (e.g. occupancy or abundance), $g()$ is a link function (e.g. probit or logit for occupancy, log for count data), and μ_i is a link-scale intensity value. In applications described in this paper, we write the intensity as

$$\mu_i \quad = \quad \beta_0 + \mathbf{x}_i\boldsymbol{\beta} + \delta_i, \qquad \text{eqn 2}$$

where β_0 is an intercept parameter, $\mathbf{x}_i$ is a row vector of m predictive covariates associated with site i, $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_m\}$ is a column vector of m regression parameters, and $\boldsymbol{\delta} = \{\delta_1, \delta_2, \ldots, \delta_S\}$ are spatially autocorrelated random effects. For occupancy, $f()$ would typically be Bernoulli, while the Poisson or negative binomial are typically choices for analysis of count data; common forms for δ_i include geostatistical specifications (Cressie 1993; Diggle *et al.* 1998), Gaussian Markov random fields (e.g. conditionally autoregressive models; Rue & Held 2005), or low rank alternatives such as predictive process (Banerjee *et al.* 2008; Latimer *et al.* 2009) or restricted spatial regression models (Reich *et al.* 2006; Hughes & Haran 2013).

The model for Z_i describes variation in the process of interest and is often described as the "process" model. However, it is usually impossible to observe the system perfectly even in locations where sampling occurs, so it is customary to include an observation model describing incomplete detection. For occupancy studies, the response variable $Y_i = 1$ if the species of interest is detected and is 0 otherwise, and is modeled with a Bernoulli distribution (Royle & Dorazio 2008):

$$Y_i \quad \sim \quad \text{Bernoulli}(Z_i p_i). \qquad \text{eqn 3}$$

Here, the detection probability p_i is possibly a function of survey and observer specific covariates. Replicate surveys of the same sampling unit provide the necessary information to estimate p_i. For count surveys, a possible model is

$$Y_i \quad \sim \quad \text{Poisson}(Z_i A_i p_i), \qquad \text{eqn 4}$$

where the Y_i now represents the count of animals obtained while surveying unit i, A_i denotes the proportion of sample unit i that is surveyed, and p_i gives detection probability. Additional information will often be needed to estimate p_i in this context, such as data from double observers, distance observations, or double sampling (see e.g. Buckland *et al.* 2001; Royle *et al.* 2004; Borchers *et al.* 2006; Conn *et al.* 2014).

For the remainder of this treatment, we use bold symbols to denote vector-valued quantities or matrices. We also use standard bracket notation to denote probability mass and density functions. For instance $[\mathbf{Z}]$ denotes the marginal probability mass function for $\mathbf{Z}$, and $[\mathbf{Z}|\mathbf{Y}]$ represents the conditional distribution of $\mathbf{Z}$ given $\mathbf{Y}$. We use $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ to denote log-scale abundance and the logit of the probability of sampling, so that $Z_i \sim f(\mu_i)$, and $R_i \sim f(\nu_i)$. We use the notation Z_i when describing the state process in general terms, but often switch to the conventional notation N_i when animal abundance is the explicit focus of interest.

PREFERENTIAL SAMPLING: A PRIMER

One of the appealing aspects of model-based estimation is that there is no requirement that surveys rely on a pre-planned survey design selected probabilistically from an underlying sampling frame. For instance, investigators can reallocate sampling effort if weather or logistics preclude surveying in a desired location. This can be a crucial advantage in surveys covering large areas with frequent inclement weather. It also opens the door for using platforms of opportunity, presence only, and citizen science data for estimation.

However, the manner in which effort is ultimately allocated can potentially have profound influence on SDM estimator performance. With respect to nonrandom sampling, two possible problems seem particularly likely in discrete spatial domains: coarse scale preferential sampling (CSPS), and fine scale preferential sampling (FSPS) (Fig. 1). FSPS arises when the observations taken at a particular sampling unit are non-random with respect to the density of animals within that sampling unit. For instance, when allocating line transect survey effort, it may be tempting to place the transect in a manner that targets habitat or landscape features that maximize the number of animals that will be encountered. However, this strategy will clearly lead to positive bias when estimating density or abundance.

By contrast, CSPS (hereafter, PS), the primary focus of this article, arises when the locations being sampled and the process of interest (e.g. density, occupancy) are conditionally dependent given modeled covariates (Diggle *et al.* 2010). For instance, PS can occur when the investigator uses a priori knowledge or observations of the state variable obtained during sampling to allocate survey effort in places where abundance or occurrence is known to be high. Diggle *et al.* (2010) showed that this type of PS can lead to bias when this extra information is not included in models for the state variable of interest. Specifically, PS arises when we consider the set of sampled locations as stochastic and when $[\mathbf{R}, \mathbf{Z}|\mathbf{x}] \neq [\mathbf{R}|\mathbf{x}][\mathbf{Z}|\mathbf{x}]$, where $\mathbf{R}$ is an indicator vector whose elements R_i are 1.0 if sampling unit i is sampled and are zero otherwise. We use this definition of PS throughout the rest of the manuscript, noting that it is somewhat different than has sometimes been used in the SDM literature. For instance, Merckx *et al.* (2011) use the term "preferential sampling" to refer to the process of visiting some sites more often than others, while Manceur & Kühn (2014) define it as occurring when the locations selected for sampling are a function of an environmental covariate. Neither of these latter conditions are problematic outside of the specialized field of presence-only modelling.

Diggle *et al.* (2010) demonstrated PS with an environmental monitoring problem, whereby pollutant monitoring stations were more highly clustered around urban areas with high concentrations of pollutants than in rural areas with comparably low levels of pollutants. Fitting simple geostatistical models without fixed effects led to positively biased estimates of landscape-level pollutant concentrations. Presumably (and as noted by discussants of the article) including a fixed effect associated with a relevant covariate (e.g. a development index) would likely reduce or eliminate bias. However, the primary point of Diggle *et al.* (2010) is well taken: inclusion of spatially autocorrelated random effects in a statistical model is insufficient to remove the potentially biasing effects of PS.

As in the pollution example, having good explanatory covariates may also reduce bias when fitting SDMs to animal encounter data under PS. However, in many ecological applications, predictive covariates explain only a small portion of variation present in the data. If the locations selected for sampling are a function of some unmodelled factor related to abundance (intentionally or unintentionally) , bias may still occur. Despite the clear potential for bias in SDMs, there are few examples where PS (*sensu* Diggle *et al.* 2010) is discussed with regard to SDMs. One exception is Chakraborty *et al.* (2010), who acknowledged the likely presence of PS when fitting SDMs to data obtained using nonrandomized designs. However, they did not attempt to account for PS in their models.

In design-based sampling, unequal sampling intensity is often accommodated via stratification or unequal probability sampling, as with Horvitz-Thompson-like estimators where the probability of inclusion varies by sampling unit (Cochran 1977). However, in the case of PS, this inclusion probability also depends on the value of the response associated with the sampling unit. Evidently, any approach to account for PS should also account for the dependence between the state variable of interest and the locations chosen for sampling.

Several authors have attempted model-based corrections for PS in the statistical literature. For Gaussian models in a continuous spatial domain, Diggle *et al.* (2010) and Pati *et al.* (2011) jointly modeled the locations that are chosen for sampling and the underlying random field of interest. In particular, they expressed sampled locations as an inhomogeneous Poisson point process where the underlying log-scale intensity depended linearly on spatially-referenced random field values. For instance, writing observations of the spatial random field at a location i as

$$Z_i \quad = \quad \mu_i + \epsilon_i, \qquad \text{eqn 5}$$

the spatially continuous relative intensity (ψ_i) of sampling locations at i could be written as

$$\psi_i \quad \propto \quad \exp(\xi_i + b\mu_i). \qquad \text{eqn 6}$$

Here, the parameter b describes the level of PS; $b = 0$ implies no PS, $b > 0$ implies a greater level of sampling in locations where the spatial process (e.g. animal density) is high, and $b < 0$ implies greater sampling where the spatial process is low. Importantly, when explanatory covariates are used in models for μ_i and ξ_i, Pati *et al.* (2011) show that ". . . accounting for informative sampling is only necessary when there is an association between the spatial surface of interest and the sampling density that cannot be explained by

the shared spatial covariates." Pati *et al.* (2011) also consider a simpler, plug-in based estimator, where the log of a nonparameteric estimate of sampling density (specifically, a two dimensional kernel density estimate) is used as an additional fixed effect in Eq. 5, finding that this approach helped reduce bias associated with PS, but did not perform as well as the full joint model.

A GENERALIZED PREFERENTIAL SAMPLING MODEL

The models considered by Diggle *et al.* (2010) and Pati *et al.* (2011) are a useful first step in addressing and modeling PS. However, they are somewhat limited since they are specific to continuous spatial domains, continuous data (as opposed to presence/absence or count data), and Gaussian error distributions. Also, they require the linear predictor of the PS model to be written as a simple linear function of the the spatial process model for density. In real world applications, we can envision cases where sampling is strongly preferential in certain areas of the landscape, and not in others. For instance, sampling may be more strongly preferential close to bases of operations, (e.g. landing strips in the case of aerial surveys), but less so in areas that are harder to get to.

Given these limitations, our present task is to generalize PS models to the types of data more typical of SDMs, and to allow the degree of PS to vary across the landscape. Like Diggle *et al.* (2010) and Pati *et al.* (2011), we impose a joint model for the process of interest (animal abundance or occurrence) and the locations chosen for sampling. For the abundance process model, we start with eq. 1 as a general formulation for non-Gaussian data, writing the link-scale expectation as in eq. 2. Next, recalling that R_i is a binary indicator taking on the value 1.0 if survey unit i is selected for sampling, and is 0.0 otherwise, we model R_i using a Bernoulli distribution:

$$R_i \quad \sim \quad \text{Bernoulli}(h^{-1}(\nu_i)), \qquad \text{eqn 7}$$

where $h()$ denotes a link function appropriate for binary data (e.g. logit, probit). We then write the intensity for this model as

$$\boldsymbol{\nu} \quad = \quad \beta_0^* + \mathbf{x}^*\boldsymbol{\beta}^* + \boldsymbol{\eta} + \mathbf{B}\boldsymbol{\delta}. \qquad \text{eqn 8}$$

In a similar fashion to the model for the state process, the sampling intensity model has an intercept (β_0^*), explanatory covariates ($\mathbf{x}_i^*$), fixed effect regression parameters ($\boldsymbol{\beta}^*$) and spatially autocorrelated random effects ($\boldsymbol{\eta}$ and $\boldsymbol{\delta}$). The predictive covariates $\mathbf{x}_i$ from Eq. 2 and $\mathbf{x}_i^*$ from Eq. 8 may or may not be the same. Note also that the spatially autocorrelated random effects $\boldsymbol{\delta}$ are included in both Eqs. 2 and 8, allowing for dependency in the two models, with the matrix $\mathbf{B}$ describing the strength and type of dependence between the sampling process and underlying density. The spatially autocorrelated random effects $\boldsymbol{\eta}$ are assumed independent of the $\boldsymbol{\delta}$. In practice, we find we often need to fix $\beta_0^* = 0.0$ when random effects in Eq. 8 are estimated to permit parameter identification.

The formulation in Eq. 8 is similar to the one previously proposed for hierarchical multivariate models with spatial dependence (cf. Royle & Berliner 1999). There are multiple ways of structuring $\mathbf{B}$ depending on the complexity of spatial dependence desired for the PS process (Royle & Berliner 1999). For instance, setting $\mathbf{B} = \mathbf{0}_{S\times S}$ corresponds to an absence of spatial dependence (and thus no PS). Setting $\mathbf{B} = b\mathbf{I}$, where b is an estimated parameter and $\mathbf{I}$ is an $(S \times S)$ identity matrix corresponds to the linear PS model suggested by Diggle *et al.* (2010) and Pati *et al.* (2011). Alternatively, we could allow the degree of PS to vary across the landscape. For instance, one can contemplate a trend surface model for PS by specifying a diagonal matrix for $\mathbf{B}$, with entries given by $b_0 + b_1\text{lat}_i + b_2\text{long}_i$, where b_0, b_1, and b_2 are estimated parameters and lat_i and long_i give latitude and longitude, respectively (Royle & Berliner 1999). Theoretically, one could include more highly parameterized structures for spatial dependence, such as higher order trend surface or spline formulation (Royle & Berliner 1999), but the ability to robustly estimate the parameters of such a model is likely dependent on having a rich, spatially balanced dataset, which is often not the case in ecological applications.

A comparison of the performance of models with different sets of constraints on $\mathbf{B}$ can serve as a test of PS. In particular, if one can demonstrate that models with $\mathbf{B} = \mathbf{0}$ perform similarly or better than models with $\mathbf{B} \neq \mathbf{0}$, then PS is likely not worth modeling and inference can proceed using standard SDMs (i.e. not modeling sampling intensity).

SIMULATION STUDY

To illustrate PS and demonstrate that our proposed model has reasonable performance, we conducted a small simulation experiment. For each of 500 simulations, we generated abundance of a hypothetical species over a 25×25 grid as

$$[N_i|\mu_i] \quad = \quad \text{Poisson}(\exp(\mu_i)),$$

where i indexes survey unit i, and μ_i is determined according to Eq. 2. Abundance was generated as a function of a single spatially autocorrelated landscape covariate, as well as residual spatial autocorrelation (δ_i) and overdispersion (fig. 2). Specific details of data generation procedures are provided in Appendix S1.

For each simulated landscape we generated three virtual count surveys using eqs. 7 and 8. Each survey had $\boldsymbol{\beta}^* = \eta_i = 0$ (that is, no covariate or spatially autocorrelated random effects), but differed in how the matrix $\mathbf{B}$ was parameterized. In the first, we set $\mathbf{B} = 0$, so that surveyed locations were selected independently of the abundance generating process. For the second and third, we set $\mathbf{B}$ to be a diagonal matrix with entries $b = 1$ and $b = 5$, respectively, so that the probability of sampling a given survey unit (grid cell) was explicitly dependent on the latent abundance in that unit. We refer to these scenarios as moderate and pathological PS, respectively (see fig. 3). Simulations were configured so that $n = 50$ of the 625 survey units were sampled; each survey was set to cover half of the target cell.

We fitted two different models to each count dataset, both of which were provided the habitat covariate used (in part) to generate the data for which a log-linear coefficient β was estimated. In the first model, the elements of $\mathbf{B}$ in eq. 8 were all set to zero. In this case, the abundance and sampling process submodels were independent, as is the case canonical SDMs (at lest when fitted to presence-absence or count data). In the second model, we included an explicit connection between the distribution of animal abundance and the sampling process by setting $\mathbf{B} = b\mathbf{I}$, where b is an estimated parameter, and $\mathbf{I}$ is an identity matrix.

We used maximum likelihood to conduct statistical inference. In particular, we used Template Model Builder (TMB; Kristensen *et al.* 2016), interfaced with the R programming environment, to conduct maximization. The TMB software uses a Laplace approximation to integrate out random effects ($\boldsymbol{\eta}$ and $\boldsymbol{\delta}$), and a bias correction algorithm (Tierney *et al.* 1989; Thorson & Kristensen In Press) to obtain abundance estimates and standard errors that properly account for nonlinear transformations of random effects. This approach resulted in a facile implementation and speedy computing times, allowing us to conduct simulation and model testing with greater efficiency than would have been possible with Bayesian simulation. Further detail on statistical methods are provided in Appendix S1; requisite R and TMB code will be published to a publicly accessible repository upon acceptance, and is also available at `https://github.com/NMML/pref_sampling/`.

BEARDED SEAL COUNT SURVEYS

We applied our modeling technique to counts of bearded seals obtained on aerial transects flown over the eastern Bering Sea from 10-16 April, 2012 (Fig. 5). These counts were gathered as part of a larger survey designed to estimate abundance of four species of ice-associated seals; the survey is described in greater detail elsewhere (Conn *et al.* 2014, 2015). The survey area considered here consists of 25 by 25 km grid cells bordered to the north by the Bering strait, to the west by the international date line, to the south by maximal April ice extent, and to the east by the Alaska, USA mainland. Here, we limit counts to those gathered within a one week period so that relative abundance will remain relatively constant throughout the study area. Our primary focus in this application is to diagnose PS (rather than to estimate absolute abundance). As such, we do not attempt to correct for nuisance processes such as incomplete detection or species misclassification, which requires models of increased sophistication (Conn *et al.* 2014).

Our choice to model bearded seal counts, as opposed to one of the other seal species, is based on the observation that bearded seal densities tend to be highest in the northern portion of the study area. This is also the location of one of the primary airports used to prosecute surveys (Nome, Alaska, USA). Higher survey coverage in areas of high bearded seal density could potentially lead to positive bias in apparent abundance owing to PS.

To test for such an effect, we modeled bearded seal counts using the formulation

$$\begin{aligned}[Y_i|Z_i] &= \text{Poisson}(P_i Z_i), \text{ where} \\ Z_i &= A_i \exp(\mu_i),\end{aligned}$$

where P_i defines the proportion of grid cell i that is sampled, A_i gives the proportion of grid cell i that is composed of salt water habitat, and μ_i is defined in Eq. 2. We modeled the grid cells that were chosen for sampling using Eqs. 7-8.

We fitted a total of six models ($M_{cov=0,b=0}$, $M_{cov=0,b=1}$, $M_{cov=1,b=0}$, $M_{cov=1,b=1}$) to bearded seal count data using the same estimation framework as in the simulation study. Models varied by (i) whether or not habitat and landscape variables were used as predictors of bearded seal density ($cov = 1$ and $cov = 0$, respectively), and (ii) the form of PS ($b = 0$ indicates no PS; $b = 1$ indicates $\mathbf{B} = b\mathbf{I}$, where b is an estimated parameter). We also attempted to fit models where the PS $\mathbf{B}$ matrix varied over the landscape using a trend surface specification, but parameter identification was suspect in these models and are not reported here (see Appendix S2 for more information). When habitat and landscape variables were included, we used three log-linear predictors: linear and quadratic functions of sea ice concentration, and distance from the southern ice edge. Remotely sensed sea ice data were obtained at a 25×25 km resolution from the National Snow and Ice Data Center, Boulder, CO, USA, as described by Conn *et al.* (2014). Models for μ_i and ν_i both utilized spatially autocorrelated random effects with a Matérn covariance function between grid cell centroids (see Appendix S1 for further details). When covariates were included, they were included in both models (i.e. for μ_i and ν_i).

Results

SIMULATION STUDY

Estimates of cumulative animal abundance across simulated landscapes were median unbiased for both estimation methods when the sites selected for sampling were independent of animal density, though when b was estimated, abundance estimates were more right skewed and had higher variance (fig. 4). Under moderate PS ($b = 1$), estimation of the PS parameter b led to a median bias of 5%, while the canonical SDM model ignoring preferential sampling had a median bias of 40%. Under pathological PS ($b = 5$), both estimation methods were extremely biased, but was even more severe for the naive model ignoring PS (fig. 4).

BEARDED SEAL ANALYSIS

Marginal AIC strongly favored models with covariate effects, but for such models the presence of PS was equivocal (Table 1). Further intuition can be gained by examining estimates of the PS parameter, b. For the PS model without predictive covariates, $\hat{b} = 0.27$ (SE 0.11), and for the model with predictive covariates, $\hat{b} = 0.19$ (SE 0.13). Thus, it appeared that including predictive covariates decreased the PS effect size, as suggested by Pati *et al.* (2011). Estimates of abundance were substantially higher for models without a PS effect, with the non-PS model having a 49% higher estimate when covariates were not modeled, and a 19% higher estimate when covariate effects were included.

Note that unlike the other models, $M_{cov=1,b=0}$ predicted anomalously high bearded seal abundance in the extreme southern portion of the study area where sea ice was absent (where there was no habitat for seals). Thus, while we present original likelihood and AIC values to permit direct comparison with other models, we refitted $M_{cov=1,b=0}$ to produce an estimate of apparent abundance without this feature. Specifically, we refitted the model with 20 pseudo-absences in this portion of the study area to better inform abundance-covariate relationships.

The fact that models with and without a PS effect garner approximately equal weight suggests a need to account for PS when producing abundance estimates from this data set. A model averaged estimate calculated using AIC machinery (Burnham & Anderson 2002) is 54854 (SE 9351), which is 7.5% less than the estimate assuming no PS. Notably, the standard error of the model averaged estimate was 79% higher than the model assuming no PS.

Discussion

In this study, we showed that coarse-grained preferential sampling (Fig. 1) can have a profound impact on the quality of estimates (e.g. animal abundance) when sampling is non-randomized. In simulations, estimators were increasingly positively biased as PS

increased. When PS was present, we were able to substantially reduce bias by conducting estimation under a framework where the state variable of interest and the sites chosen for sampling were jointly modeled under a dependent covariance structure. In absence of PS, simulations suggest that this structure results in lower precision than a model without a PS effect; thus the need to account for PS reduces the quality of inference.

Bias attributed to PS may seem counterintuitive, especially given the maxim in survey sampling to allocate more effort to strata for which animal density is high. For instance, in large scale line transect surveys under stratified sampling, the optimal amount of effort that should be allocated to stratum s is $A_s D_s^{0.5}$, where A_s is the area of s and D_s is the anticipated density (Buckland *et al.* 2001; eqn 7.7). Thus, there are theoretical reasons to sample more in high density areas than in low density areas. The obvious solution in this instance is to account for variation in sampling intensity with explanatory covariates or post hoc stratification. However, it is not always clear how to perform post hoc stratification when effort is allocated in a subjective manner.

When applied to bearded seal count data, approximately equal support was given to models with and without a PS effect. The PS effect size was estimated to be positive and to produce considerably lower abundance estimates than models without a PS effect. Differences between apparent abundance estimates decreased when covariates were added to model structure, supporting previous theoretical results (Diggle *et al.* 2010; Pati *et al.* 2011) that covariates serve to decrease the conditional dependence between site selection and the state variable of interest. However, in our data set, adding covariates did not eliminate evidence of PS. Accounting for PS in a model averaging framework led to a moderate decrease in our apparent abundance estimate for bearded seals in this region, and markedly decreased precision. As in our simulations, the need to account for PS thus appeared to have a real cost in terms of variance inflation.

We attempted to fit models to bearded seal data where the degree of PS changed over the landscape, in a similar manner to multivariate spatial models (Royle & Berliner 1999). However, such models led to difficulties with parameter identification in our bearded seal application (see Appendix S2). Evidently, such models may require greater spatial balance or richer data sets. At this time, we suggest limiting initial consideration to models with a single, estimated b parameter as a composite adjustment to abundance or occupancy. Although more complex models are clearly identifiable in some situations (Royle & Berliner 1999), further research on the viability of such models as a function of data quality appears warranted. It would also be worthwhile to investigate whether parameter identification varies as a function of the support of the state variable being modeled (e.g. binary vs. count data).

The models we have developed here are specific to spatial models with discrete support, as when data are gathered at a plot level, or aggregated prior to analysis. However, it should be possible to extend our approach to continuous space. One approach would be to model sampling locations as realizations from a spatial point process in a manner similar to Warton & Shepherd (2010). Another possible extension would be to consider models for the sampling process where sampling occurs without replacement for a fixed sample size. For instance, the Bernoulli sampling model makes the implicit assumption that sample size is random. If, instead, a fixed number of locations are sampled, the Bernoulli model is somewhat misspecified. Our simulations suggest some robustness to this misspecification, as the Bernoulli model performed reasonably well when sampling was without replacement for a fixed sample size (Fig. 4). Still, a more precise treatment would need to rely on an extended hypergeometric distribution with variable inclusion probabilities when formulating the sampling model; this extension is nontrivial.

Our conception of PS is related, but not equivalent to "sample selection bias" (e.g. Phillips *et al.* 2009) in presence-only models. In such models, absence of a species at a given site is never directly observed. To draw inference about space use, it is thus necessary to produce a background sample representing the range of locations and habitats that could have been sampled. Sample selection bias then results if the characteristics of sites selected for sampling (e.g. by a volunteer or museum collector) differ systematically from the assumed background sample. In our case, we use PS to refer to the case where absences are available, but where the probability of sampling is dependent on some unknown factor that is also related to abundance or presence of the target species.

Conclusion

Model-based approaches to estimation of abundance or occurrence have become popular in recent years. We (the authors) have noticed a tendency for analysts to assume that inclusion of spatial covariates or random effects into predictive models will make the underlying sampling design ignorable. We have shown in this paper that this is not the case, although our results do suggest

that including predictive covariates can indeed decrease bias from preferential sampling. We have also shown that it is possible to further diagnose and adjust for preferential sampling by jointly modeling dependence between the data collection mechanism and the process of interest (e.g. abundance or occupancy). However, such models can be considerably less precise and have greater instability than models without a preferential sampling parameter. Where possible, we suggest that survey planners incorporate design-based elements (e.g. random or systematic sampling) into their survey designs to reduce the need for model-based triage.

Acknowledgements

Funding for aerial surveys was provided by the U.S. National Oceanic and Atmospheric Administration and by the U.S. Bureau of Ocean Energy Management. Views expressed are those of the authors and do not necessarily represent findings or policy of any government agency. Use of trade or brand names does not indicate endorsement by the U.S. government.

Data accessibility

R scripts and data necessary to recreate analyses have been collated into an R package, which is currently available at https://github.com/NMML/pref_sampling. We plan to publish the package to an online archive/repository upon acceptance.

References

Banerjee, S., Gelfand, A.E., Finley, A.O. & Sang, H. (2008) Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society B*, **70**, 825–848.

Borchers, D.L., Laake, J.L., Southwell, C. & Paxton, C.G.M. (2006) Accomodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics*, **62**, 372–378.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001) *Introduction to Distance Sampling: Estimating the abundance of biological populations*. Oxford University Press, Oxford, U.K.

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

Chakraborty, A., Gelfand, A.E., Wilson, A.M., Latimer, A.M. & Silander Jr, J.A. (2010) Modeling large scale species abundance with latent spatial processes. *The Annals of Applied Statistics*, 1403–1429.

Cochran, W. (1977) *Sampling Techniques, 3rd Edition*. Wiley, New York.

Conn, P.B., Johnson, D.S., Ver Hoef, J.M., Hooten, M.B., London, J.M. & Boveng, P.L. (2015) Using spatio-temporal statistical models to estimate animal abundance and infer ecological dynamics from survey counts. *Ecological Monographs*, **85**, 235–252.

Conn, P.B., Ver Hoef, J.M., McClintock, B.T., Moreland, E.E., London, J.M., Cameron, M.F., Dahle, S.P. & Boveng, P.L. (2014) Estimating multi-species abundance using automated detection systems: ice-associated seals in the eastern Bering Sea. *Methods in Ecology and Evolution*, **5**, 1280–1293.

Cressie, N.A.C. (1993) *Statistics for spatial data, revised edition.* Wiley, New York.

Dail, D. & Madsen, L. (2011) Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, **67**(2), 577–587.

Diggle, P.J., Tawn, J.A. & Moyeed, R.A. (1998) Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **47**(3), 299–350.

Diggle, P.J., Menezes, R. & Su, T.l. (2010) Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **59**(2), 191–232, doi:10.1111/j.1467-9876.2009.00701.x, URL http://dx.doi.org/10.1111/j.1467-9876.2009.00701.x.

Elith, J. & Leathwick, J.R. (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677–697.

Hedley, S.L. & Buckland, S.T. (2004) Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, **9**, 181–199.

Hughes, J. & Haran, M. (2013) Dimension reduction and alleviation of confounding for spatial generalized mixed models. *Journal of the Royal Statistical Society B*, **75**, 139–159.

Johnson, D.S., Conn, P.B., Hooten, M., Ray, J. & Pond, B. (2013) A probit approach for spatio-temporal modeling of ecological occupancy data. *Ecology*, **94**, 801–808.

Johnson, D.S., Laake, J.L. & Ver Hoef, J.M. (2010) A model-based approach for making ecological inference from distance sampling data. *Biometrics*, **66**, 310–318.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H. & Bell, B.M. (2016) TMB: Automatic differentiation and Laplace approximation. *Journal of Statistical Software*, **70**, 1–21.

Latimer, A.M., Banerjee, S., Sang, H., Moshner, E.S. & Silander Jr., J.A. (2009) Hierarichical models facilitate spatial analysis of large data sets: a case study on invasive plant species in the northern United States. *Ecology Letters*, **12**, 144–154.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, **83**, 2248–2255.

Manceur, A.M. & Kühn, I. (2014) Inferring model-based probability of occurrence from preferentially sampled data with uncertain absences using expert knowledge. *Methods in Ecology and Evolution*, **5**(8), 739–750.

Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M. & Vanaverbeke, J. (2011) Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. *Ecological Modelling*, **222**(3), 588 – 597, doi:http://dx.doi.org/10.1016/j.ecolmodel.2010.11.016, URL http://www.sciencedirect.com/science/article/pii/S0304380010006216.

Miller, D.L., Burt, M.L., Rexstad, E.A. & Thomas, L. (2013) Spatial models for distance sampling data: recent developments and future directions. *Methods in Ecology and Evolution*, **4**, 1001–1010.

Pati, D., Reich, B.J. & Dunson, D.B. (2011) Bayesian geostatistical modelling with informative sampling locations. *Biometrika*, **98**(1), 35–48.

Phillips, S.J., Dudík, M., Elith, J., Graham, C.H., Lehmann, A., Leathwick, J. & Ferrier, S. (2009) Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications*, **19**(1), 181–197.

Reich, B.J., Hodges, J.S. & Zadnik, V. (2006) Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics*, **62**, 1197–1206.

Royle, J.A. (2004) N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, **60**, 108–115.

Royle, J.A., Dawson, D.K. & Bates, S. (2004) Modeling abundance effects in distance sampling. *Ecology*, **85**, 1591–1597.

Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology*. Academic Press, London, U.K.

Royle, J.A. & Berliner, L.M. (1999) A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, **4**, 29–56.

Rue, H. & Held, L. (2005) *Gaussian Markov Random Fields*. Chapman & Hall/CR, Boca Raton, Florida, USA.

Thorson, J.T. & Kristensen, K. (In Press) Implementing a generic method for bias correction in statistical models using random effects, with spatial and popualtion dynamics examples. *Fisheries Research*.

Tierney, L., Kass, R.E. & Kadane, J.B. (1989) Fully exponential Laplace approximations to expectations and vvariance of non positive functions. *Journal of the American Statistical Association*, **84**, 710–716.

Warton, D.I. & Shepherd, L.C. (2010) Poisson point process models solve the "pseudo-absence" problem for presence-only data in ecology. *Annals of Applied Statistics*, **4**, 1383–1402.

Table 1. A summary of model selection results and estimated abundance for the four models fitted to bearded seal counts. The models include formulations with or without predictive covariates ($cov = 1$ or 0, respectively) , and with or without the preferential sampling parameter b estimated ($b = 1$ or 0, respectively) . All models included spatially autocorrelated random effects on log-scale abundance intensity. Shown are the log integrated likelihood, the number of fixed effect parameters, ΔAIC, AIC model weights, and estimated apparent abundance over the landscape ($\hat{N}$) together with a Hessian-based standard error estimate.

Model	Log likelihood	Params	ΔAIC	Wgt	$\hat{N}$(SE)
$M_{cov=0,b=0}$	-2667.1	3	21.7	0.00	68556 (7408)
$M_{cov=0,b=1}$	-2665.3	4	20.1	0.00	45857 (5114)
$M_{cov=1,b=0}$	-2650.3	9	0.0	0.53	59312† (5231)
$M_{cov=1,b=1}$	-2649.4	10	0.3	0.47	49826 (10369)

† Refitted model; see *Results*.

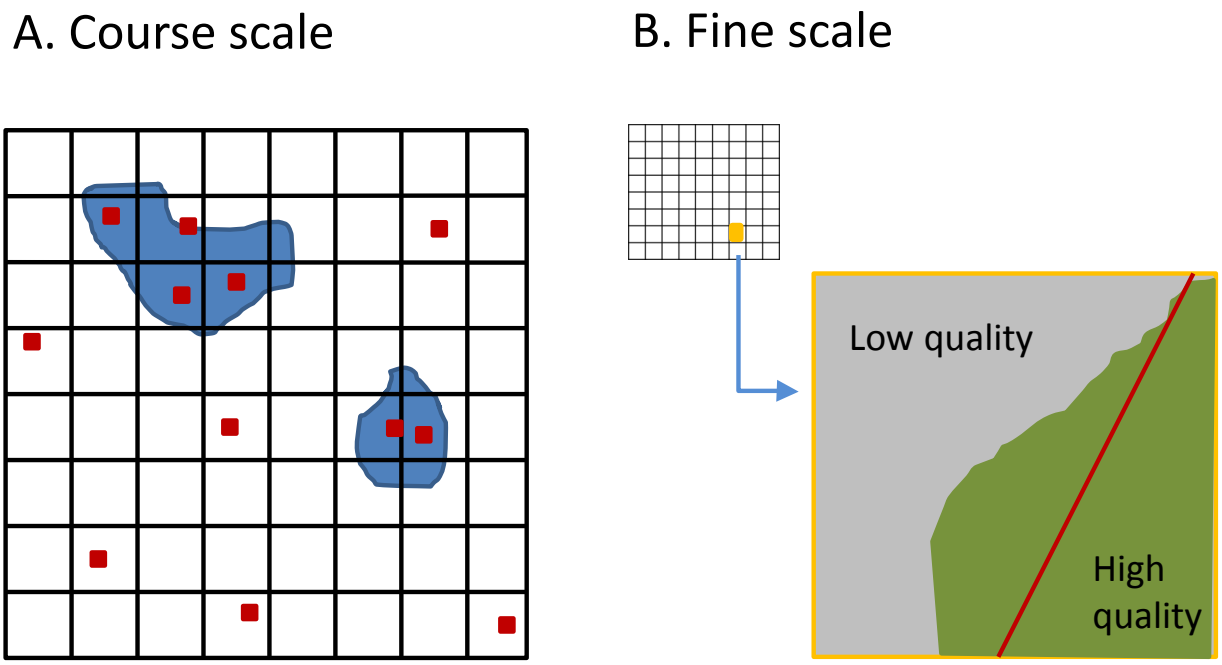

Fig. 1. A depiction of two types of preferential sampling. In (A), an investigator preferentially places point transects (red squares) within regions of high known animal density (blue polygons). This can cause bias in abundance or occupancy estimators unless this a priori knowledge about density is explicitly modeled. In (B), a fine scale version of preferential sampling occurs when a line transect (red line) is intentionally placed across a region of high quality habitat. If a landscape is discretized into homogeneous survey units for analysis (as in a grid), it is essential that the habitat surveyed within each survey unit be randomly determined when estimating abundance. If not, bias (usually positive) can be expected.

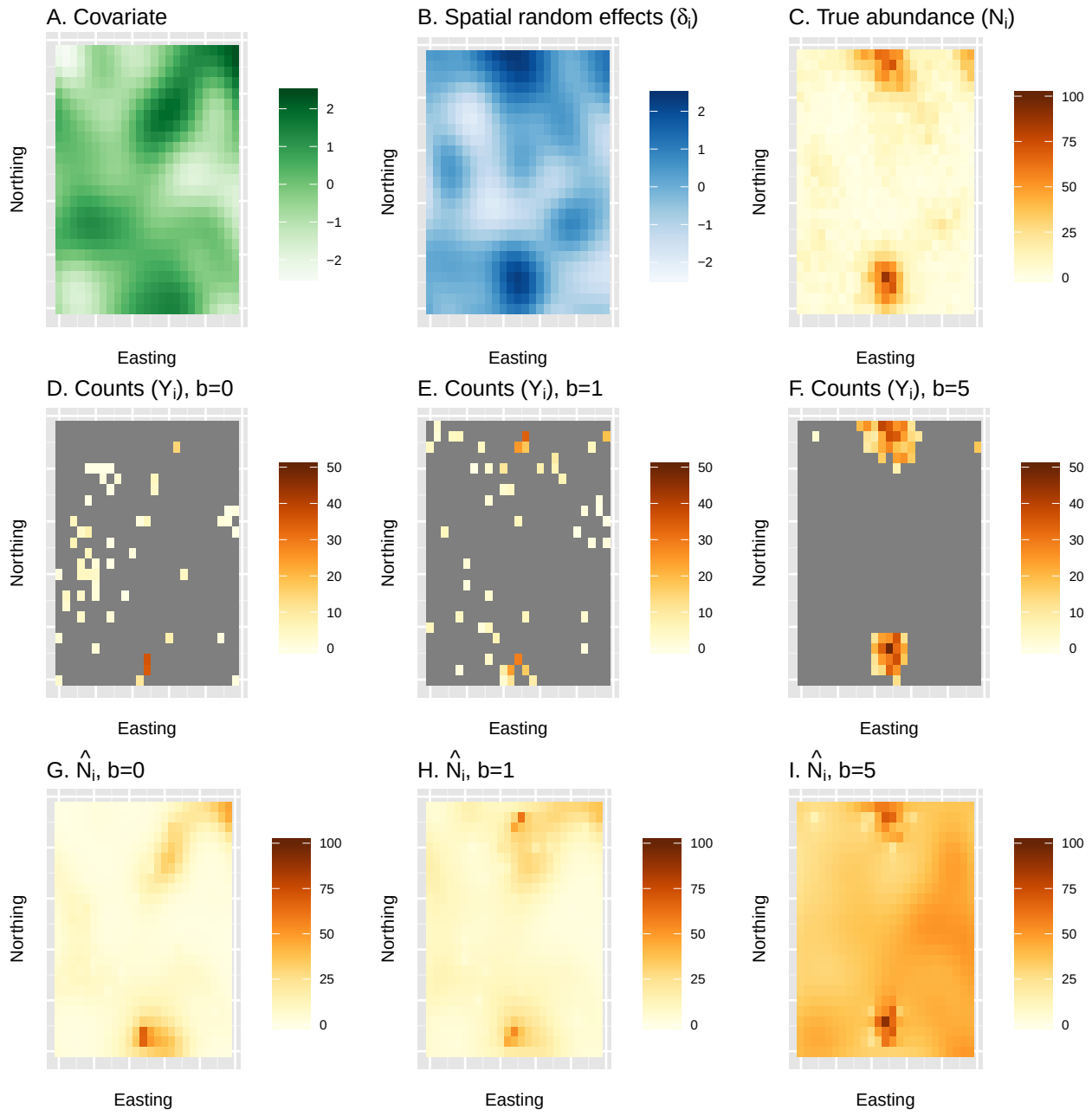

Fig. 2. An example of a single simulation replicate examining estimates of abundance from a naive species distribution model under preferential sampling. First, true abundance (C) is generated as a function of a spatially autocorrelated covariate (A) and a spatially autocorrelated random effect (B). Second, counts are generated for three different types of surveys, including a simple random sample ($b = 0$; D) and surveys with moderate ($b = 1$; E) or pathological ($b = 5$; F) levels of preferential sampling. Finally, spatially explicit estimates of abundance are generated using a traditional SDM (with b set to 0.0) to each of the count datasets (G-I). In this particular simulation replicate, cumulative abundance was underestimated by 18% when $b = 0$, overestimated by 17% when $b = 1$, and overestimated by 293% when $b = 5$. For a summary of bias over 500 simulation replicates, see fig. 4.

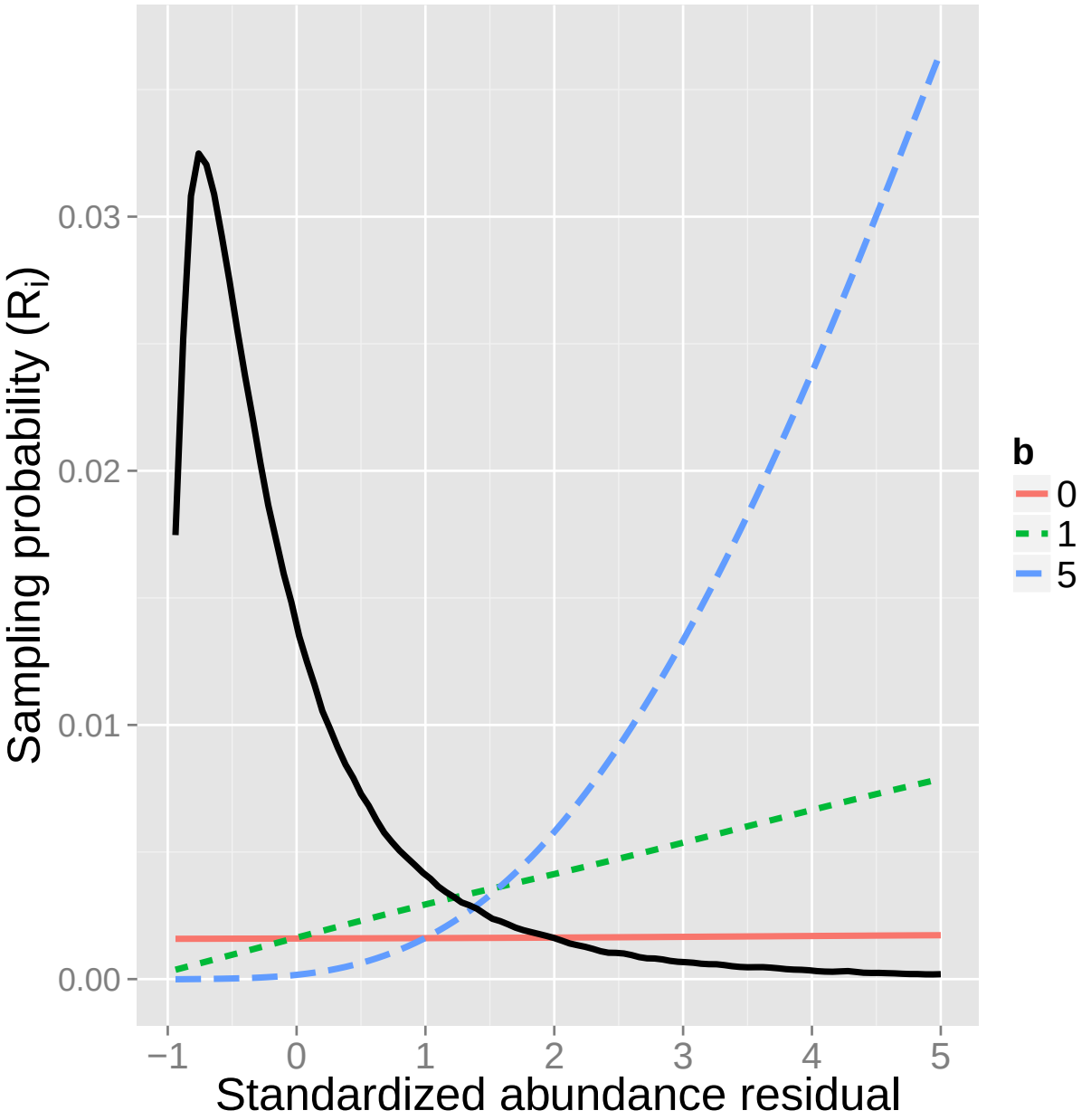

Fig. 3. Expected relationship between the probability of a survey unit being selected for sampling and its abundance residual in the simulation study. The base case $b = 0$ represents simple random sampling, while $b = 1$ and $b = 5$ represent moderate and pathological levels of preferential sampling, respectively. Also shown are is the realized distribution (smoothed histogram) of abundance residuals among survey units in the simulation study, scaled to fit in the plot margins (solid black line).

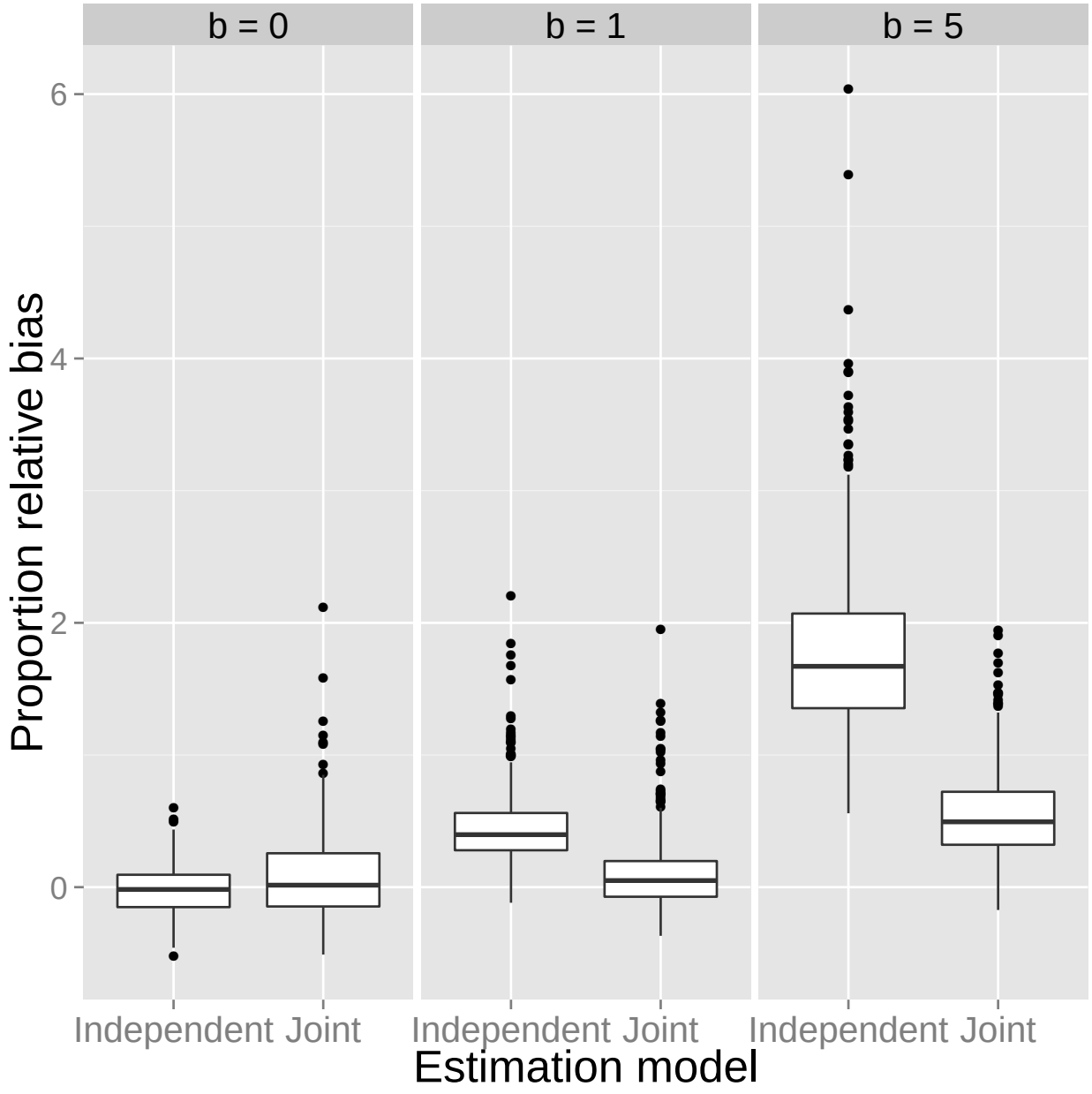

Fig. 4. Relative proportional error in abundance from the simulation experiment as computed with respect the posterior mode with a bias correction. Each boxplot summarizes the distribution of relative proportional error as a function of the type of sampling, including simple random sampling ($b = 0$), moderate preferential sampling ($b = 1$), and pathological preferential sampling ($b = 5$). Results vary by the type of estimation model; in the "independent" model, b is set to 0.0; in the "joint" model, b is estimated. Lower and upper limits of each box correspond to first and third quartiles, while whiskers extend to the lowest and highest observed bias within 1.5 interquartile range units from the box. Points denote outliers outside of this range. Horizontal lines within boxes denote median bias.

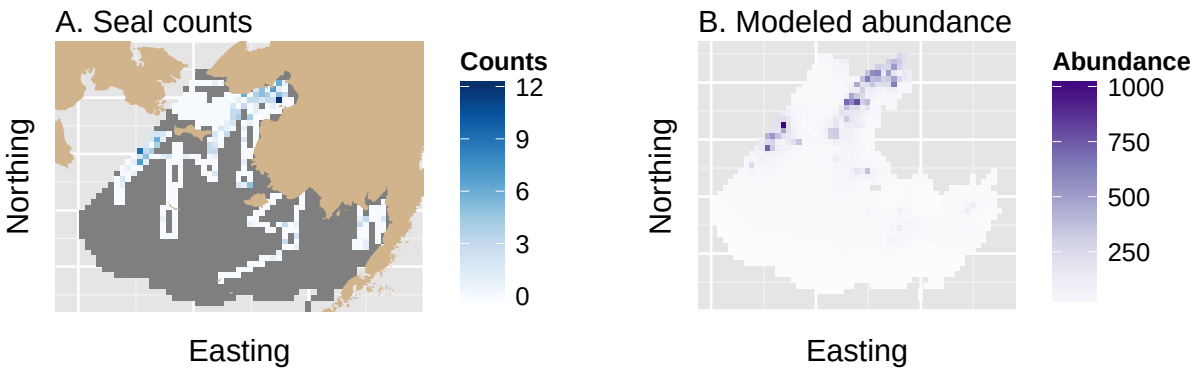

Fig. 5. Aerial survey counts and estimated apparent abundance of bearded seals in the eastern Bering Sea, April 10-16, 2012. Counts and estimates are shown relative to a survey grid that extends south from the Bering Strait and borders the Alaska, USA mainland to the east. In (A), tan shading denotes land, unsurveyed grid cells appear in dark gray, and counts appear in a white-blue spectrum. Apparent bearded seal abundance estimates (B) are presented from the model with the lowest integrated AIC score, which included covariate effects but no preferential sampling effect. Apparent abundance estimates are uncorrected for imperfect detection or species misclassification.