SC/68D/ASI/01

Sub-committees/working group name: ASI

Spatial modeling, parameter uncertainty, and precision of density estimates from line-transect surveys: a case study with Western Arctic bowhead whales

Megan C. Ferguson, David L. Miller, Janet T. Clarke, Amelia A. Brower, Amy L. Willoughby and Audrey D. Rotrock

Spatial modeling, parameter uncertainty, and precision of density estimates from line-transect surveys: a case study with Western Arctic bowhead whales

Megan C. Ferguson[1,2*], David L. Miller[3,4], Janet T. Clarke[1,5],
Amelia A. Brower[1,5], Amy L. Willoughby[1,5], and Audrey D. Rotrock[1,5]

*Megan.Ferguson@noaa.gov
[1]Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, Seattle, WA, USA
[2]School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA
[3]Centre for Research into Ecological and Environmental Modelling, University of St. Andrews, St. Andrews, UK
[4]School of Mathematics and Statistics, University of St. Andrews, St. Andrews, UK
[5]Cooperative Institute for Climate, Ocean, and Ecosystem Studies, University of Washington, Seattle, WA, USA

ABSTRACT

Spatially-explicit models of animal density, such as density surface models (DSMs), are diverse, flexible, and powerful tools for investigating spatial patterns in animal density, examining associations between animal density and environmental covariates, and estimating abundance. Advances in spatial modeling methods and subsequent incorporation into widely accessible software allow the non-specialist to add these tools to their analytical toolbox. However, limitations in some software may prevent a thorough treatment of uncertainty. We expanded the functionality of tools for constructing DSMs from line-transect survey data to derive a population abundance estimate that honestly accounts for multiple sources of detection bias and associated uncertainty. As an illustrative case study, we used data collected during an aerial line-transect survey for Western Arctic bowhead whales (*Balaena mysticetus*) over their summering grounds in the Beaufort Sea and Amundsen Gulf during August 2019. Using spatially explicit hierarchical generalized additive models that incorporated correction factors and associated uncertainty for perception and availability bias, we estimated the abundance of the Western Arctic bowhead whale population to be 17,175 whales ($\widehat{CV}$ = 0.237; 95% confidence interval = [10,793, 27,330]). This model-based abundance estimate is similar in magnitude to the two most recent estimates for this population based on data from ice-based surveys in 2011 and 2019 . Additionally, our abundance estimate is sufficiently precise to inform management decisions for this protected species. The enhanced precision of our abundance estimate over the estimate derived using design-based analytical methods applied to the same data is due to explicit modeling of the spatial correlation in whale density. Applying the power of DSMs to the aerial line-transect survey data made this survey methodology a viable alternative to ice-based surveys, which are facing obstacles due to climate change, for updating abundance estimates for Western Arctic bowhead whales in the future. Our analytical developments can easily be applied to other line-transect datasets with similar and common challenges due to multiple survey platforms, spatial heterogeneity in animal density and environmental conditions, and habitat partitioning among groups (e.g., defined by age, sex, activity state) in the target population.

INTRODUCTION

Understanding population dynamics, assessing population status, and investigating a population's ecological role often require knowledge of the population's size and spatial distribution. Line-transect surveys are a reliable method for collecting data to address these and other questions. If certain assumptions about the survey design hold, line-transect survey data can be analyzed to derive valid abundance estimates using relatively simple "design-based" methods (Buckland et al. 2001; Hedley and Bravington 2014; Miller and Bravington 2017). In a spatially explicit modeling framework (e.g., Hedley and Buckland 2004; Johnson et al. 2010; Miller et al 2013), the number of questions that data from line-transect surveys and opportunistic surveys can address proliferates. For example, spatial models may be used to identify spatial patterns in animal density (the number of animals per unit area) at finer resolutions than design-based models; assist investigations into local or remote ecological mechanisms that shape spatial patterns in animal density; derive unbiased estimates of abundance; and provide input into protected species management, monitoring, and mitigation issues, such as spatial planning, impacts analysis, and designing effective monitoring protocols.

To estimate absolute abundance or density from line-transect survey data, a number of parameters related to the observation process must be estimated. First, the probability that an animal that is within an observer's field of view is also available to be seen may be less than 1.0. This availability bias (Marsh and Sinclair 1989) is especially relevant for cetaceans, who spend the majority of their time underwater where they cannot be detected by an aerial or vessel-based observer. Availability bias may be a function of animal behavior, observer field of view, or environmental factors such as turbidity or glare that affect the ability to see underwater animals. Second, the probability that an animal that is available to be seen is actually detected is typically less than 1.0. This issue, termed perception bias (Marsh and Sinclair 1989), is often characterized by an inverse relationship between detection probability and distance between the animal and the observer: the probability of detecting an animal located very close to the survey platform is relatively high (but often less than 1.0), whereas animals located farther away are smaller, the field of view is wider, and the animal is harder to detect (Buckland et al. 2001). Other factors that may affect perception bias include characteristics of the animal (body size, coloration, group size, behavior) or environment (glare, precipitation, clouds, water color and clarity, wind waves, presence of sea ice). None of the parameters describing availability and perception bias are known with certainty.

In a decision making context, it is crucial that scientific advice accurately and honestly depicts the uncertainty in quantities relevant to management. This is particularly true for protected or harvested species for which policy decisions hinge on both population status and uncertainty therein. In the case of line-transect data, the uncertainty stems from the variability in sightings across sample units and the parameters describing availability and perception bias. The methods available to propagate this uncertainty depend on the overall modeling framework (e.g., Hedley and Bravington 2014; Bravington and Miller 2021).

Density surface models (DSMs) can be used as a framework to both estimate abundance and propagate uncertainty. DSMs are particularly useful for estimating abundance from line-transect survey data when assumptions for design-based methods are not met (Hedley and Bravington 2014; Miller and Bravington 2017). We use "design-based" to refer to analytical methods that rely on assumptions about the survey design or sampling procedure, although other authors (e.g., Miller and Bravington 2017) refer to them as Horvitz-Thompson estimators, and Fewster and Buckland (2004) explain that estimating abundance from line-transect data fundamentally requires analytical methods that are not purely design-based. Design-based methods may produce biased results when survey coverage is not constant throughout the study area, or when sighting conditions and animal density are spatially heterogeneous. Even when design-based modeling assumptions hold, analyzing data using a DSM may improve precision or reduce bias in the abundance estimate or its estimated uncertainty (Miller and Bravington 2017). Furthermore, DSMs assume that animal density is spatially correlated and they estimate correlation parameters, allowing estimation of animal abundance (or density) and associated uncertainty at finer spatial resolutions than design-based models (Hedley and Bravington 2014). In a design-based analysis, the sample unit for estimating variance is typically the transect, and 10-20 transects located across a survey region are needed to reliably estimate precision (Fewster and Buckland 2004; Hedley and Bravington 2014). Therefore, transect number and spacing limit the spatial resolution of abundance estimates using design-based methods.

Two broad classes of DSMs for analyzing line-transect survey data are one-stage models and two-stage models. One-stage DSMs jointly estimate parameters for the observation process and the spatial density surface in a single model (e.g., Yuan et al. 2017; Johnson et al. 2010). In a two-stage DSM, parameters for the observation process are estimated first, and then bias-corrected counts of animals comprise the response variable for fitting a density surface model. We focus on the two-stage DSM because well-developed tools exist to create models that meet the needs of a variety of datasets and to evaluate model fit at each stage.

Estimating population abundance for Western Arctic bowhead whales (*Balaena mysticetus*) from data collected during a broad-scale aerial line-transect survey over the majority of the population's summer range (Clarke et al. 2020) provides an ideal case study for propagating multiple sources of uncertainty through a DSM. In this case, the additional parameters that need to be estimated and incorporated into the abundance estimate along with the DSM parameters relate to the estimates of perception and availability bias for two types of aircraft, and the effects of bowhead whale activity state (feeding, traveling, socializing or resting at the surface) on availability bias for each sighting. The Western Arctic bowhead whale population is a conservation and management priority because it is vitally important to Alaska Native subsistence and culture. The aboriginal subsistence harvest is governed by the Whaling Convention Act and co-managed by the Alaska Eskimo Whaling Commission (AEWC) and the National Marine Fisheries Service (NMFS), according to a quota from the International Whaling Commission (IWC; IWC 2018). Furthermore, this population was severely depleted due to commercial whaling that occurred between 1848 and 1914 (Bockstoce and Burns 1993). As a result, the population was listed as endangered under the US Endangered Species Conservation Act in 1970, and has been listed as endangered under the US Endangered Species Act since 1973, although prudent management has allowed the population to successfully rebound (Givens et al. 2021a; Muto et al. 2021). The population receives additional protection under the US Marine Mammal Protection Act. Lastly, bowhead whales are endemic to the Arctic, one of the most rapidly changing places on Earth (Moon et al. 2021).

Long-term monitoring and precise and unbiased abundance estimates have been integral to the management strategy for Western Arctic bowhead whales. From 1978 to 2011, the gold standard for Western Arctic bowhead whale abundance estimates was derived from data collected during spring by visual observers stationed on land-fast sea ice who recorded bowhead whales migrating past Point Barrow, Alaska, to summering grounds in the Beaufort Sea and Amundsen Gulf; during some years, concurrent passive acoustic monitoring was used to estimate availability bias (Givens et al. 2016, 2021a; Suydam et al. 2019). Up through 2011, there were 21 attempted ice-based surveys (George et al. 2013), which resulted in a series of 12 abundance estimates. Additionally, three abundance estimates (1986, 2004, and 2011) were derived from aerial imagery using photo identification (photo-ID) in a mark-recapture framework (da Silva et al. 2000; Koski et al. 2010; Givens et al. 2018).

There are limitations to the established data collection and analytical methods, especially due to rapid climate changes. Spring sea ice conditions are changing, resulting in safety and logistical challenges for the ice-based survey team and deviations in the whales' migration path. Climate forecasts predict that Arctic sea ice will continue to decline (Wang and Overland 2015). Existing methods to match photos of individual whales are time-consuming and require specialized expertise. Additionally, there is considerable uncertainty in the proportion of the population that is "unmarked", a parameter that is essential to the photo-ID mark-recapture estimator (Givens et al. 2018). Furthermore, mark-recapture abundance estimates can be subject to bias from unexplained heterogeneity (Laake et al. 2008). Therefore, during 2019, a spring ice-based visual survey (Givens et al. 2021a) and a summer aerial line-transect survey were conducted with the goal of generating independent abundance estimates for Western Arctic bowhead whales during the same year. The August 2019 aerial line-transect survey was a pilot study to determine whether this methodology provides a viable alternative to ice-based surveys for updating abundance estimates for Western Arctic bowhead whales in the future, while continuing to minimize bias and maximize precision in the estimate.

Here, we advance estimation methods from line-transect methods in two ways. First, we illustrate how to honestly account for multiple sources of bias and associated uncertainty in the line-transect observation process, and to accurately propagate that uncertainty through the density surface model. Second, we evaluate the extent to which the precision in the abundance estimate is enhanced by using spatially-explicit modeling methods compared to design-based methods. To reach these goals, we address three objectives: 1) use a spatially-explicit model to derive an unbiased and precise population abundance estimate for Western Arctic bowhead whales; 2) expand the capabilities of R packages mrds (Laake et al. 2021) and dsm (Miller et al. 2021) for analyzing line-transect data and creating density surface models, respectively; and 3) compare the precision of the model-based abundance estimate to a design-based abundance estimate (Ferguson et al. 2021) to evaluate whether the additional analytical steps were worthwhile.

METHODS

We begin with a brief summary of the field methods, followed by detailed analytical methods for the availability bias, DSM, and uncertainty estimation components of the analysis. Additional methods and results are provided in appendices. In the discussion, we examine the results from our DSM in light of existing information about Western Arctic bowhead whale ecology in the study area, address our ability to account for all known sources of bias, and compare our abundance estimate to those from the ice-based survey and photo-ID data.

Field Methods

Aerial line-transect surveys

Comprehensive descriptions of the aerial line-transect survey field methods are available in Clarke et al. (2020) and Ferguson et al. (2021). We provide a short summary here.

The Western Arctic bowhead whale abundance aerial line-transect surveys were conducted from 5 to 27 August 2019, covering most of the population's summer range, encompassing the Beaufort Sea continental shelf and a portion of Amundsen Gulf (Figure 1; total area 203,885 km^2). The surveys were conducted by the Aerial Surveys of Arctic Marine Mammals (ASAMM) project, which collected a long time series (1979-2019) of data on the distribution, density, habitat use, and behavior of marine mammals, primarily bowhead whales and other cetaceans, in the western Beaufort and eastern Chukchi seas during the open water season, July through October. The Bureau of Ocean Energy Management funded and co-managed ASAMM; the Marine Mammal Laboratory at NOAA's Alaska Fisheries Science Center conducted and co-managed surveys from 2007-2019, and contributed funding to the August 2019 surveys.

The survey design comprised systematic transects placed 19 km apart, based on a grid with a randomly selected start point (Figure 1). Transects were oriented perpendicular to the coastline, from shore to the 200-m isobath. The survey design incorporated a few transects in the western Beaufort Sea that extended to the 2000-m isobath to determine whether the bowhead whale distribution extended beyond the continental shelf break (Clarke et al. 2020).

One Turbo Commander aircraft based in Deadhorse, Alaska, USA, surveyed the western Beaufort Sea. One Turbo Commander based in Inuvik, Northwest Territories, Canada, surveyed the eastern Beaufort Sea and Amundsen Gulf. The two Turbo Commanders had identical configurations, with the exception that the aircraft based in Inuvik had a belly port with a mounted camera. One De Havilland Twin Otter aircraft based in Ulukhaktok, Northwest Territories, Canada, from 5 to 15 August and Inuvik from 16 to 27 August surveyed the eastern Beaufort Sea and Amundsen Gulf. Surveys were flown 305-460 m above ground level at a survey speed of 213 km/hr. All three aircraft had bubble windows for the left- and right-side primary observers. Bubble windows in the Twin Otter were smaller than those in the Turbo Commanders.

Each survey team comprised two primary observers and one dedicated data recorder. The data recorder used custom-built, menu-driven software to enter sighting data into a laptop computer interfaced with a global positioning system. Time and position data (latitude, longitude, altitude) were automatically recorded every 30 seconds (in time) and whenever a manual data entry was recorded. At every 5-minute time interval or whenever conditions changed, environmental and viewing conditions were recorded, including integer-valued Beaufort Sea State (wind force scale 0-6), visibility range perpendicular to the aircraft on each side of the plane (<1 km, 1-2 km, 2-3 km, 3-5 km, 5-10 km, or unlimited), sky conditions (clear, partly cloudy, overcast), integer-valued sea ice percent on each side of the plane, and impediments to visibility (glare, fog, haze, precipitation, ice on the window, low ceiling) on each side of the plane. Primary observers scanned with naked eye, using binoculars only to check potential targets or get a magnified view on a confirmed target. Declination angles from the horizon to each sighting were measured using handheld Suunto clinometers when the sighting was abeam. One "sighting" or "group" was defined as all animals of the same species within 5 body lengths of each other; therefore, a group could comprise one or more animals. Once the clinometer angle was recorded, most sightings of large cetaceans (i.e., anything larger than a beluga, *Delphinapterus leucas*) were circled to confirm species identification, obtain a final group size estimate, look for calves, and determine behavior. Sightings that could not be positively identified to species were recorded at the taxonomic level to which they could be identified (e.g., unidentified cetacean). Both initial and final group size estimates were recorded in the database; if group size could not be determined with certainty, high or low estimates were recorded. Circling did not commence in special circumstances, such as restrictions due to weather, fuel, time of day, or duty hours, or in the vicinity of subsistence hunting activities or sensitive wildlife.

Data from six survey modes (Clarke et al. 2020) were used to estimate perception bias and build the DSM: transect, circling from transect, cetacean aggregation protocols (hereafter simplified to "aggregation protocols"), circling aggregation protocols, search, and circling from search. During all six of these survey modes, observers were actively surveying and all sightings and effort data were recorded. Transect effort refers to systematic survey effort along a prescribed transect line. Search refers to non-systematic survey effort during transit or between transects. Circling from search or transect occurred when the aircraft diverted from flat and level flight to investigate a sighting or potential sightings in a localized area. Standard line-transect survey protocols (Buckland et al. 2001) were followed until bowhead whale encounter rates exceeded the observers' ability to accurately record location and clinometer angle to each sighting. In areas with extremely high densities of bowhead whales, aggregation protocols were used, wherein the survey team flew through the high-density patch in passing mode to collect accurate encounter rate data, then flew back through the patch in closing (circling aggregation protocols) mode to collect information on group size, number of calves, and behavior (Clarke et al. 2020). Only data from transect and aggregation protocols (without circling) were used to estimate encounter rate for the DSM. Data from transect, aggregation protocols (without circling), and search were used to estimate perception bias. Data from the three circling modes were used only to confirm species identification, estimate group size, and examine behavior.

Belly Port Imagery

To estimate transect detection probability for the ASAMM observers, a downward-pointing digital single lens reflex camera with a 20- or 21-mm lens mounted to the belly of a Turbo Commander aircraft during ASAMM's 2018 and 2019 field seasons collected true color (red, green, and blue [RGB]) imagery. The imagery served as an independent observer. Willoughby et al. (2021) provide detailed imagery collection and analysis methods and results. At 400 m survey altitude, a single image taken with the 21-mm lens captured a parcel of water measuring approximately 684 m perpendicular to the transect (342 m on each side of the transect) and 457 m along the transect. One image was collected every 2 to 3 seconds, resulting in each parcel of water being visible in three to four images. Metadata automatically written to each image included latitude, longitude, date, and time. Every third image collected was manually reviewed post-flight for marine mammal sightings by trained photo analysts (Willoughby et al. 2021). Any sightings detected in the imagery were manually compared to the visual survey database to determine matches based on date, time, and location (side of plane and distance from transect).

Field-of-view (FOV) Trials

To estimate the amount of time observers had to view a bowhead whale as a function of perpendicular distance to the transect, in 2018 and 2019 field-of-view (FOV) trials were flown by each aircraft type over land using a fixed structure (a Conex box for the Turbo Commander and a cabin for the Twin Otter) as a target. See Clarke et al. (2020) for additional details about the FOV field methods. These time-in-view estimates were incorporated into the availability bias correction factors explained below.

Analytical Methods

We used a combination of distance sampling, mark-recapture, and spatially explicit modeling techniques to create spatial surfaces of Western Arctic bowhead whale density across the Beaufort Sea and Amundsen Gulf. These density surfaces represented the estimated number of bowhead whales in each cell of a hexagonal lattice (10-km spatial resolution between cell midpoints) during the August 2019 survey period. To estimate total population abundance, we integrated across the density surfaces, multiplying the density estimates by geographic area. The data sources and subsets used in the analysis are shown in Figure 2, and a glossary of notations and abbreviations is provided in Appendix A. All analyses were conducted in R (R Core Team 2021), using packages mrds (Laake et al. 2021), dsm (Miller et al. 2021), mgcv (Wood 2017), sp (Pebesma and Bivand 2005; Bivand et al. 2013), maptools (Bivand and Lewin-Koh 2019), raster (Hijmans 2020), rgeos (Bivand and Rundel 2019), and rgdal (Bivand et al. 2019).

Because the resulting population abundance estimate is a function of geographic area, the geographic boundary for the analysis is important. For this analysis, the geographic boundary extended across the full longitude of the August 2019 study area (119° to 157° W; Figure 1). The coastlines of mainland Alaska and Canada, and Banks Island, defined the nearshore boundary for the analytical area. The offshore boundary reflected the expected and observed distributions of bowhead whales (Figure 1): in the western Beaufort Sea (141° to 157° W), the 2000-m isobath defined the offshore boundary; in the eastern Beaufort Sea and Amundsen Gulf (119° to 141° W), the offshore boundary was fixed at the 400-m isobath. The offshore boundary was placed 3 km outside of the closest transect to ensure that all valid bowhead whale sightings and survey effort were included in the analysis.

The basic line-transect estimator of animal density is (Buckland et al. 2001; Burt et al. 2014):

$$\widehat{D} = \frac{1}{a}\sum_{j=1}^{n_g}\frac{S_j}{\hat{p}(\hat{\theta};\boldsymbol{z}_j)} \qquad [1]$$

where

n_g = total number of groups detected;

S_j = size of group indexed by j;

a = area searched, which is equal to $2wL$;

L = total length of transects surveyed;

w = width of the strip searched on one side of the aircraft;

$\hat{p}(\hat{\theta};\boldsymbol{z}_j)$ = estimate of the overall probability that an observer detects group j, given covariates $\boldsymbol{z}_j$ that affect detectability; this term incorporates estimates of availability probability p_a, transect detection probability $p_1(0,\boldsymbol{z})$, and $p^*(\boldsymbol{z})$;

$\hat{p}_1(0,\boldsymbol{z}_j)$ = estimated probability that an observer detects a group located on the transect, given covariates $\boldsymbol{z}_j$ relating to characteristics of the sighting or environmental conditions that affect detectability;

$\widehat{p^*}(\boldsymbol{z}_j)$ = estimated probability that an ASAMM observer detects an group that is available to be seen, given covariates $\boldsymbol{z}_j$ that affect detectability, assuming transect detection probability is 1.0; $\widehat{p^*}(\boldsymbol{z}_j) = \frac{\int_0^w \hat{g}(y,\boldsymbol{z}_j)dy}{w}$;

$\hat{g}(y,\boldsymbol{z})$ = multiple covariates distance sampling detection function, which specifies the shape and scale of the observation model and assumes detection probability on the transect equals 1.0 (Appendix B);

$\hat{p}_a$ = estimated probability that a group is at the surface and within an observer's field of view; a measure of availability bias.

We define a "group" as one or more bowhead whales located within 5 body lengths of each other and recorded as a single sighting in the ASAMM database. We decomposed the problem into the following components, which are addressed sequentially below and in appendices: detection functions; availability bias correction factors, which include a state model for availability that requires information on bowhead whale surface and dive behavior, and on the aircraft field of view; DSM; and uncertainty estimation.

Detection Functions

The probability that an ASAMM observer ("observer 1") detects a group of whales located on the transect and the effects of distance (y) from the transect (and possibly other covariates $\boldsymbol{z}$) on detection probability were estimated using an observation model, $p_1(y, \boldsymbol{z})$, for each aircraft. These concepts relate to perception bias. The observation models were formulated as mark-recapture multiple covariates distance sampling detection functions (Marques and Buckland 2003; Laake and Borchers 2004; Burt et al. 2014). Complete details about the detection function methods and results are provided in Appendix B.

Availability Probability, $\widehat{p_a}$

The probability that an aerial observer will detect a cetacean during a line-transect survey is a function of the duration of time the observer has to detect the animal. Failing to account for the animal's surface and dive durations or the observer's field of view leads to availability bias in estimated density or abundance (Laake and Borchers 2004).

The state model for availability can be represented as the probability that an animal will surface within detectable range (Laake et al. 1997):

$$p_a(y) = P\{animal\ at\ y\ is\ at\ surface\} \qquad [2]$$

$$= \frac{\lambda}{\lambda + \mu} + \frac{\mu[1 - exp\{-\lambda T(y)\}]}{\lambda + \mu} \qquad [3]$$

$$= 1 - \frac{\mu exp\{-\lambda T(y)\}}{\lambda + \mu} \qquad [4]$$

where

y = perpendicular distance to the aircraft;

λ = rate parameter of the dive process;

μ = rate parameter of the surfacing process;

$T(y)$ = duration of time in which the ocean at perpendicular distance y is in the observer's view; this parameter is a function of the observer's field of view; see Appendix C for details on how this parameter was estimated.

Letting the average dive duration be $\mathbb{E}(d) = \frac{1}{\lambda}$ and the average surface duration be $\mathbb{E}(s) = \frac{1}{\mu}$, we get:

$$\hat{p}_a(y) = \frac{\mathbb{E}(s)}{\mathbb{E}(s) + \mathbb{E}(d)} + \frac{\mathbb{E}(d)\left[1 - exp\left\{-\frac{T(y)}{E(d)}\right\}\right]}{\mathbb{E}(s) + \mathbb{E}(d)} \qquad [5]$$

For $\mathbb{E}(s)$ and $\mathbb{E}(d)$, we used the corresponding mean surface and mean dive duration estimates for undisturbed bowhead whales engaged in different activity states (A) in the southern Beaufort Sea from Robertson et al. (2013). Robertson et al. (2015) provides slightly different estimates of $\mathbb{E}(s)$ and $\mathbb{E}(d)$ for equivalent activity states. We chose to use the values from Robertson et al. (2013) in our analysis because we needed associated estimates of σ_s and σ_d, the standard errors of $\mathbb{E}(s)$ and $\mathbb{E}(d)$, to estimate uncertainty in the availability probabilities.

The categories and activity states that we considered in our analysis were travel, calf, social, and feeding in deep water ("deep.feed") (Table 1). There was only a single bowhead whale sighting designated as social in the ASAMM August 2019 survey data. Socializing bowhead whales and calves tend to remain at the surface for relatively long periods. Therefore, we pooled sightings in which at least one calf was present with the sighting designated as social into a single activity state that we called "cow-calf or social" (cc.soc). To estimate availability bias for this activity state, we used $\mathbb{E}(s)$ and $\mathbb{E}(d)$ from Robertson et al.'s (2013) undisturbed calf category. In total, we computed six availability bias correction factors, one for each combination of aircraft (Turbo Commander, Twin Otter) and activity state (travel, cc.soc, deep.feed).

Density Surface Model (DSM)

Abundance was estimated using a density surface model. The basic DSM structure may be represented as (Bravington et al. 2021):

$$\mathbb{E}\left[\mathrm{W}_i | \boldsymbol{\beta}_{dsm}, \boldsymbol{\psi}, p(\hat{\theta}; \boldsymbol{z}_i)\right] = a_i p(\hat{\theta}; \boldsymbol{z}_i) exp\left(\beta_{0,dsm} + \sum_q f_q(u_{iq})\right) \qquad [6]$$

where

i = segment index;

W_i = random variable for the number of whales on segment i in the density surface model, and ω_i is the corresponding observed number;

u_{iq} = covariates that influence density (in this case, projected latitude and longitude);

a_i = segment area, computed as $2wL_i$, where w is the difference between the right-truncation distance and left-truncation distance (Appendix B);

L_i = length of transect surveyed in segment i;

$p(\hat{\theta}; \boldsymbol{z}_i)$ = overall probability that an observer detects a group in segment *i*, given covariates $\boldsymbol{z}_i$ that affect detectability; this term incorporates availability probability p_a, transect detection probability $p_1(0, \boldsymbol{z})$, and $p^*(\boldsymbol{z})$;

f_q = smooth function for the q^{th} covariate;

$\boldsymbol{\psi}$ = vector of smoothing parameters;

$\boldsymbol{\beta}_{dsm}$ = vector of coefficients in the smooth functions, including $\beta_{0,dsm}$, the intercept.

The spatial resolution of the DSM was 10 km. This analytical resolution is approximately one-half the distance between adjacent transects, which were spaced 19 km apart. The DSM was constructed using sighting and effort summaries for 10-km segments of transect and aggregation protocols effort with width w as defined above and in (Appendix B). Residual segments that were <10 km entered the model as separate analytical units (i.e., they were not merged with an adjacent 10-km segment). Predictions from the DSM were based on a hexagonal lattice with cell midpoints located 10 km apart. All geospatial data were projected into an Equidistant Conic projection (false easting: 0.0; false northing: 0.0; central meridian: -138.2°; latitude of origin: 71.4°; standard parallels: 69.5°, 72.3°; WGS84 datum; linear unit: meter [1.0]).

We evaluated DSMs formulated as single-level and hierarchical generalized additive models (Hastie and Tibshirani 1990; Wood 2017 Pedersen et al. 2019) with logarithmic link functions (Table 2). The hierarchical model is discussed further below. The basic equation for the single-level model is:

$$log\left(\mathbb{E}\left[W_i | \boldsymbol{\beta}_{dsm}, \boldsymbol{\psi}, p(\hat{\theta}; \boldsymbol{z}_i)\right]\right) = \beta_{0,dsm} + \sum_q f_q(u_{iq}) + log\left[a_i p(\hat{\theta}; \boldsymbol{z}_i)\right] \qquad [7]$$

Due to the complex coastline in the study area, which includes multiple peninsulas and estuaries (Figure 1), we evaluated candidate models that used soap film smoothers (Wood et al. 2008) and tensor products of thin plate regression splines (Wood 2017).

Because there was a high proportion of transect segments with zero bowhead whale sightings, we evaluated candidate DSMs built using negative binomial and Tweedie (Jørgensen 1987; Dunn and Smyth 2005) distributions. The negative binomial distribution is a discrete probability distribution that is commonly used to model count data that are overdispersed relative to a Poisson distribution (e.g., McCullagh and Nelder 1999). The Tweedie family comprises exponential dispersion models in which the variance is proportional to the mean (ξ) raised to a power (π) (Jørgensen 1987). If we define $\eta_i = log(\omega_i)$ and let Ω_i be the random variable associated with observation η_i, then $\Omega_i \sim Tweedie(\xi, \varphi, \pi)$, where φ is the scale or dispersion parameter, and $var(\Omega_i) = \varphi\xi^{\pi}$. The mgcv package restricts π to be between 1 and 2. The class of Tweedie distributions includes a few special cases, including the normal (π = 0), Poisson (π = 1), and gamma (π = 2) (Dunn and Smyth 2005). The Tweedie distribution offers a flexible alternative to the negative binomial distribution for modeling count data when there are a high proportion of zeros; unlike the zero-inflated negative binomial distribution, the Tweedie avoids multiple-stage modeling of zero-inflated data (e.g., Candy 2004; Miller et al. 2013).

To accommodate activity state-specific values of $\widehat{p_a}(0)$, we created and estimated parameters for a hierarchical model that allowed for factor-smooth interactions. Specifically, we used the "GS" model from Pedersen et al. (2019), which creates a global smooth, plus group-level smoothers (with the same wiggliness) corresponding to travel, cc.soc, and deep.feed. We defined an ordered factor for activity state, with travel serving as the reference level smooth because we had considerably more bowhead sightings/whales/segments classified as travel (134/157/110) compared to cc.soc (45/95/44) or deep.feed (6/23/6).

The DSMs required segment-specific estimates of detection probability, $\hat{p}\left(\hat{\theta}; \boldsymbol{z}_i\right)$. Because the detection functions for the Turbo Commander included covariates for integer-valued Beaufort Sea State and sky condition (Appendix B), effort data for these variables were summarized by segment as follows to build the DSMs. The segment-specific Sea State variable was calculated as the average value of integer-valued Beaufort Sea State for all records in the ASAMM database that were located on the segment; all records were weighted equally. The segment-specific sky condition variable was calculated by assigning each sky condition category an integer value (clear = 1; partly cloudy = 2; overcast=3), computing the average of the integer-valued sky condition variables for all records located on the segment, rounding the result, and back-transforming to the categorical sky condition variable. For example, if segment $\boldsymbol{i}$ comprised three data records with sky conditions clear, clear, and overcast, the average of their integer-valued analogs would be 1 + 1 + 3 = 1.67, which rounds to 2, so the segment would be designated "partly cloudy".

In total, five candidate DSMs were constructed and examined (Table 2). Model selection was heavily guided by expert knowledge of the system. Based on model diagnostic plots examining the relationship between the mean and variance in the residuals compared to the theoretical distribution (Ver Hoef and Boveng 2007), we selected a Tweedie distribution for the final DSM. Because of the complex coastlines, with starkly different bowhead whale habitat on opposite sides of peninsulas and capes (Figure 1), we selected soap film smoothers for the final DSM. Lastly, due to differences in bowhead whale surface and dive behavior by activity state, which ultimately affect $\widehat{p_a}(0)$, we selected the hierarchical structure for the final DSM. The default basis dimensions were used to initially parameterize the smoothing splines for all models. The mgcv function gam.check() was used to evaluate whether the basis dimensions were large enough; because the effective degrees of freedom were all much lower than the associated maximum basis complexity, there was no concern about the basis dimensions used to build the models. The full specifications for the hierarchical model used to estimate Western Arctic bowhead whale density is presented in Appendix D.

Estimation of Uncertainty and Bias

To propagate uncertainty from the detection function models for each aircraft into the DSM, we implemented Bravington et al.'s (2021) variance propagation methods using the dsm_varprop() function from the dsm package.

To estimate uncertainty in $\widehat{p_a}(0)$ for each activity state, we used the delta method approximation for multivariate data, described further below. We assumed independence between the uncertainty estimated using dsm_varprop() (for the detection function models and DSM) and the availability probabilities, allowing estimation of the overall $\widehat{CV}(\widehat{N})$ by summing the squared coefficients of variation:

$$\widehat{CV^2}(\widehat{N}) = \widehat{CV^2}(varprop) + \sum_{A} \widehat{CV^2}(\widehat{p_a}(0)) \qquad [8]$$

For consistency with Givens et al. (2016; 2021a, b), we estimated an approximate 95% confidence interval for our abundance estimate as $\left(\widehat{N}exp\{-1.96\widehat{CV}\}, \widehat{N}exp\{1.96\widehat{CV}\}\right)$.

In [8], $\widehat{CV^2}(varprop)$ was computed as the standard deviation of the abundance estimate calculated over all of the data used to build the DSM, divided by $\widehat{N}$. $\widehat{N}$ is the estimated abundance from the DSM. $\widehat{N}$ was computed by applying the predict() function to a hexagonal lattice of cells covering the study area (Figure 3), using the actual cell area as the offset, and summing predictions across all cells.

The second term in the right-hand side of [8] sums squared $\widehat{CV}$ for all relevant activity states used in the availability probabilities. For the hierarchical model, this is a sum of three terms, corresponding to activity states travel, cc.soc, and deep.feed. Because $T(0)$ for the Turbo Commander was assumed to be known, and to avoid incorporating the uncertainty in $\bar{s}$ and $\bar{d}$ into $\widehat{CV}(\widehat{N})$ multiple times, the summation term on the right-hand side of [8] was only for the Twin Otter.

The multivariate delta method approximates the sampling variance for a parameter that is composed of a function of random variables. In our case, $\widehat{p_a}(0)$ for the Twin Otter is a function of $\bar{s}$, $\bar{d}$, and $\widehat{T(0)}$. In general, the multivariate delta method can be represented as:

$$\widehat{Var}(\widehat{\Upsilon}) = \left(\frac{\partial\widehat{\Upsilon}}{\partial\widehat{\theta_\iota}}\right) V_{\widehat{\theta}} \left(\frac{\partial\widehat{\Upsilon}}{\partial\widehat{\theta_\iota}}\right)^T \qquad [9]$$

For us, $\left(\frac{\partial\widehat{\Upsilon}}{\partial\widehat{\theta_\iota}}\right)$ is a row vector with partial derivatives of $\widehat{p_a}(0)$ with respect to $\bar{s}$, $\bar{d}$, and $\widehat{T(0)}$, and $\left(\frac{\partial\widehat{\Upsilon}}{\partial\widehat{\theta_\iota}}\right)^T$ is its transpose (a column vector). The elements of $\left(\frac{\partial\widehat{\Upsilon}}{\partial\widehat{\theta_\iota}}\right)$ may be represented as follows:

$$\frac{\partial\widehat{P}_a(0)}{\partial\bar{s}} = \frac{\bar{d}exp\left(\frac{-\widehat{T(0)}}{\bar{d}}\right)}{\left(\bar{s}+\bar{d}\right)^2} \qquad [10]$$

$$\frac{\partial\widehat{P}_a(0)}{\partial\bar{d}} = \frac{-\bar{s}}{\left(\bar{s}+\bar{d}\right)^2} + \frac{1-exp\left(\frac{-\widehat{T(0)}}{\bar{d}}\right)}{\bar{s}+\bar{d}} - \frac{\widehat{T(0)}exp\left(\frac{-\widehat{T(0)}}{\bar{d}}\right)}{\bar{d}\left(\bar{s}+\bar{d}\right)} - \frac{\bar{d}\left[1-exp\left(\frac{-\widehat{T(0)}}{\bar{d}}\right)\right]}{\left(\bar{s}+\bar{d}\right)^2} \qquad [11]$$

$$\frac{\partial\widehat{P}_a(0)}{\partial\widehat{T(0)}} = \frac{exp\left(\frac{-\widehat{T(0)}}{\bar{d}}\right)}{\bar{s}+\bar{d}} \qquad [12]$$

These expressions [10, 11, 12] correct typographical errors in Robertson et al. (2015). The term $V_{\hat{\theta}}$ in [9] is a diagonal matrix with $\widehat{Var}(\bar{s})$, $\widehat{Var}(\bar{d})$, and $\widehat{Var}(\widehat{T(0)})$. We computed $\widehat{Var}(\bar{s})$ and $\widehat{Var}(\bar{d})$ using the standard error of $\bar{s}$ and $\bar{d}$: we divided the relevant standard deviation values by the associated $\sqrt{n}$ from Table 4 of Robertson et al. (2013). The term $\widehat{Var}(\widehat{T(0)})$ corresponds to the estimated variance in time-in-view at the left-truncation distance for the Twin Otter (Appendix C).

To evaluate bias, although the true abundance of the Western Arctic bowhead whale population is unknown, we compared our estimate to the ice-based survey estimates (the gold standard) and photo-ID estimates. The IWC and NMFS consider abundance estimates with a CV ≤ 0.3 to be acceptable for management advice (NMFS 2016; IWC 2003b).

RESULTS

August 2019 Bowhead Whale Abundance Aerial Line-Transect Surveys

During the August 2019 bowhead whale abundance survey, survey coverage was nearly complete (Figure 1), with the exception of portions of Amundsen Gulf that could not be surveyed due to weather and logistical issues (Clarke et al. 2020). Bowhead whale distribution and density largely matched expectations based on all available information, including Indigenous knowledge, historical whaling records, previous aerial surveys, and telemetry studies. However, there were some notable exceptions, represented by the sightings offshore of the light orange "Expected Occurrence" polygon in the Beaufort Sea in Figure 1. In the data subset used to estimate abundance, the Turbo Commander aircraft flew over twice as much survey effort (9,605 km) as the Twin Otter (4,096 km); however, the number of bowhead whales sighted from each type of aircraft were similar (102 sightings totaling 146 whales from the Turbo Commander; 83 sightings totaling 129 whales from the Twin Otter) (Table 3). The highest bowhead whale densities were observed in the eastern Beaufort Sea, where all three aggregation protocols sessions of the survey period occurred (Clarke et al. 2020; Ferguson et al. 2021). Amundsen Gulf had the lowest observed bowhead whale densities. Most bowhead whale sightings were well within the survey area boundaries (Figure 1).

A total of five sightings of single large cetaceans could not be identified to species, four in the eastern Beaufort Sea and one in the western Beaufort Sea.

Gray whales were the only other large cetacean identified to species during the August 2019 bowhead whale abundance survey period. No other species of large cetacean was expected to be encountered. The gray whales were observed during only one flight, on the Twin Otter, on 21 August. There were 8 gray whale sightings, totaling 15 whales, including 1 calf. The gray whales were observed feeding north of the Tuktoyaktuk Peninsula in 30-55 m deep water.

Availability Probability, $\widehat{p_a}$

The proportions of ASAMM bowhead whale sightings and individuals in each of the three activity states differed considerably from the proportions of samples in the corresponding activity states in Robertson et al. (2013) (Table 1). This justifies the use of activity state-specific availability bias correction factors instead of a single correction factor based on Robertson et al.'s (2013) "summer" statistics. The latter implicitly represent weighted averages of the observed surface and dive interval data for all activity states during summer that were included in the behavioral studies from

Robertson et al. (2013), where the weights corresponded to the number of summer samples for each activity state . In the ASAMM August 2019 data, 72.4% of the total sightings (134/185) and 57.1% of the total number of whales (157/275) were traveling. In contrast, in Robertson et al's (2013) data, 13.9% of the total surface interval samples (120/866) and 23.5% of the total dive interval samples (77/328) were from traveling whales. The proportions for the remaining activity states used in this analysis are provided in Table 1.

Bowhead whale availability probabilities ($\widehat{p_a}$) were higher for the Turbo Commander than the Twin Otter (Table 1). This difference was entirely due to the Turbo Commander's larger FOV and correspondingly higher time-in-view ($\widehat{T(y)}$) compared to the Twin Otter (Appendix C). Cow-calf pairs (used in our cc.soc activity state) were the most likely to be available ($\widehat{p_a}$ = 0.36 for the Turbo Commander and 0.31 for the Twin Otter) and traveling whales were the least likely to be available ($\widehat{p_a}$ = 0.17 for the Turbo Commander and 0.16 for the Twin Otter) (Table 1). The standard errors of $\widehat{p_a}$ for the Twin Otter ranged from 0.014 for traveling to 0.033 for cow-calf pairs (Table 1).

Density Surface Model

There was good concurrence between the data and the hierarchical model predictions for all activity states. Maps of predicted bowhead whale counts from the model show that it successfully identified regions of high sighting density shared by all three activity states and regions where sighting density differed (Figures 3A, B, C). The fidelity between the model predictions and data is particularly noteworthy for the deep.feed activity state, which comprised only six sightings totaling 23 whales located on six 10-km segments (Figure 3C). This result exemplifies the power of the "GS" model formulation, in which information about the general spatial distribution of whale density contained in the global smooth is shared across activity states. For the population as a whole, bowhead whale density increased from west to east, with highest densities in the eastern Beaufort Sea. Two areas of high density common to all activity states include the offshore waters between Kaktovik, Alaska, USA, and Hershel Island; and offshore waters northwest of the Tuktoyaktuk Peninsula. Additionally, traveling whales and whales feeding in deep water were concentrated in waters in and due north of Franklin Bay, east of Cape Bathurst (Figures 3A,C). The only whales sighted in the western Beaufort Sea were traveling (Figure 3A).

Abundance and Uncertainty Estimates

We estimated the abundance of the Western Arctic bowhead whale population during summer 2019 to be 17,175 whales, with $\widehat{CV}(\widehat{N}) = 0.237$ and an approximate 95% confidence interval for $\widehat{N}$ of (10,793, 27,330). The uncertainty in availability probability contributed 0.173 to $\widehat{CV}(\widehat{N})$:

$\widehat{CV}(\widehat{p_a}(0)) = \left[\widehat{CV}^2\left(\widehat{p_{a,travel}(0)}\right) + \widehat{CV}^2\left(\widehat{p_{a,cc.soc}(0)}\right) + \widehat{CV}^2\left(\widehat{p_{a,feed.deep}(0)}\right)\right]^{0.5}$ = 0.173. The remaining uncertainty in $\widehat{CV}(\widehat{N})$ was due to uncertainty in the detection functions (Appendix B) and spatial model parameter estimates.

DISCUSSION

We estimated the Western Arctic bowhead whale population in 2019 to be 17,175 whales ($\widehat{CV}(\hat{N})$ = 0.237; 95% confidence interval = [10,793, 27,330]). A spatially-explicit hierarchical generalized additive model formed the foundation of our analysis. The analytical techniques accounted for the largest known sources of bias (availability and perception bias), and seamlessly propagated uncertainty through all modeling stages to the final abundance estimate. To propagate the variance from the trial configuration mrds model, we extended the analytical capabilities of functions in the R packages mrds and dsm, which are widely used across the globe for these types of analyses.

Our two-stage density surface model comprised two detection functions in the first stage, which informed the hierarchical model in the second stage. The detection functions included a mark-recapture distance sampling model based on trial configuration with the assumption of point independence for the Turbo Commander, which incorporated data from ASAMM observers and imagery collected concurrently with the aerial surveys; and a multiple covariates distance sampling model for the Twin Otter. The variance propagation methods allowed the information about the uncertainty in detection probabilities from the detection functions to inform estimation of the density surface model paramters; analogously, the information about the spatial uncertainty in bowhead whale density was used to fine-tune the detection function parameter estimates (Bravington et al. 2021).

The abundance estimate is based on aerial line-transect surveys conducted during August 2019 across the population's primary summer range over the Beaufort Sea continental shelf and in Amundsen Gulf. Bowhead whale distribution and density in the study area during the survey period was similar to previous years based on all available information from Indigenous knowledge, historical whaling records, previous aerial surveys, and telemetry studies, although there were two notable exceptions. First, Clarke et al. (2020) found that the bowhead whale distribution in the western Beaufort sea was farther from shore during summer (July and August combined) 2019 compared to summer 2012-2018. Second, Clarke et al. (2020) also reported that the areas of highest relative density near the Tuktoyaktuk Peninsula in 2019 were farther from shore and in deeper water (51-2000 m depth) compared to 2007-2009, when Harwood et al. (2010) found greatest densities in waters 20-50 m deep.

The spatially-explicit hierarchical model structure was well suited for the Western Arctic bowhead whale case study. To minimize bias in the abundance estimate, we wanted to estimate availability probability and density surfaces separately for each activity state (travel, cc.soc, and deep.feed) to account for known differences among activity states in surface and dive behavior and suspected differences in spatial distribution. Sufficient behavioral data existed from independent studies to derive activity state-specific estimates of availability probability. However, the line-transect data for the deep.feed activity state comprised only six sightings totaling 23 whales located on six 10-km segments (Figure 3C), and that sample size was not sufficient to build a single-level DSM. To meet the analytical objectives within the constraints of the line-transect sample sizes, we constructed a hierarchical model that contained a global smooth representing the general spatial distribution of whales, plus factor-smooth interactions that faithfully represented deviations from the global pattern by whales classified into the three activity states. The agreement between the hierarchical model predictions and the sightings in the deep.feed activity state was particularly noteworthy, given the limited sample size (Figure 3C).

An estimate of absolute population abundance for Western Arctic bowhead whales from aerial line-transect survey data analyzed in a DSM framework may be affected by five fundamental sources of bias: 1) geographic extent of the population's distribution; 2) transect detection probability, $\widehat{p_1(0,\boldsymbol{z})}$; 3) availability bias, which is related to $\widehat{p_a}$; 4) species mis-identification; and 5) back-transformation bias. Below, we discuss the influence of each of these factors on the abundance estimate.

The first caveat with the present abundance estimate is that the entire summer range of Western Arctic bowhead whales was not included within the August 2019 survey area. The population's summer range stretches from Chukotka, Russia, across the Beaufort Sea to Amundsen Gulf and possibly north to Viscount Melville Sound. A small number of bowhead whales have been known to occur off Chukotka, Russia, during August (Citta et al. 2021); however, due to logistical and financial constraints, the survey area excluded waters off Chukotka. The inability to base a survey team out of Ulukhaktok, Canada, for the duration of the survey period due to lack of aviation fuel in the village resulted in limited survey coverage in Amundsen Gulf and off the west coast of Banks Island. This issue also precluded our ability to conduct a scouting flight to Viscount Melville Sound. However, the surveys that were conducted in Amundsen Gulf and all available knowledge on bowhead whale distribution in the region suggests that Amundsen Gulf is not a high-density area for bowhead whales. Similarly, all available knowledge suggests that the waters off the west coast of Banks Island and in Viscount Melville Sound do not typically have high densities of Western Arctic bowhead whales. If significant numbers of Western Arctic bowhead whales were distributed in areas outside the analysis area during August 2019, the present abundance estimate would be biased low.

No cetacean detection method is infallible and cetaceans cannot always be seen or heard. The need to estimate correction factors for perception bias and availability bias is a complication that is common to all analyses used to estimate cetacean abundance from strip- or line-transect survey data, regardless of survey platform (vessel or aircraft) and observer type (e.g., human, imagery, or acoustic). Perception and availability bias also need to be addressed in abundance estimates derived from ice-based bowhead whale surveys (e.g., Givens et al. 2016, 2021a). For the present analysis, the transect detection probability estimate, $\widehat{p_1(0,\boldsymbol{z})}$, for the Turbo Commander was applied to the data from both aircraft types, and the total sample size of bowhead whale detections in the imagery was relatively small. These issues resulted from logistical constraints and the considerable amount of time required to manually process imagery from the belly port camera. The bubble windows in the Twin Otter were smaller than in the Turbo Commander and the former had a larger left-truncation distance. It is possible that the true $\widehat{p_1(0,\boldsymbol{z})}$ for the Twin Otter could have been less than the value used in this analysis. If that were the case, the present abundance estimate would be biased low. For example, if $\widehat{p_1(0,\boldsymbol{z})}$ for the Twin Otter were 0.55 instead of the assumed value 0.65, the total abundance estimate would have been 18,406 whales instead of 17,175 whales. Using aircraft that were all identically configured would simplify analyses and would likely improve accuracy and precision of abundance estimates derived from aerial line-transect surveys because the sample sizes used to estimate transect detection probability and to construct the distance-sampling component of the detection function model would increase. Additionally, collection of additional imagery concurrent with future line-transect surveys and development of reliable algorithms to automatically detect bowhead whale sightings in imagery would undoubtedly expedite the imagery review process, ultimately resulting in more precise estimates of transect detection probability and abundance as sample size increases.

Bowhead whale surface intervals and dive intervals (key components to the availability bias estimator) are known to vary widely depending on activity state, group size and composition, and habitat (e.g., Dorsey et al. 1989; Würsig and Clark 1993; Robertson et al. 2013; Würsig and Koski 2021). We accounted for the effects of activity state and partially accounted for group composition in our corrections for availability bias by computing separate estimates of $\widehat{p_a}$ for whales that were traveling, feeding in deep water, and found with a calf or socializing. The hierarchical model generated a density surface for each activity state by modeling them as departures from an underlying smooth surface, sharing information by using the same level of wiggliness (smoothing penalty) in the smooths for all activity states. However, the bowhead whale surface interval and dive interval data we used to estimate the $\widehat{p_a}$ were for individual whales (Robertson et al. 2013), not groups of whales. The direction and magnitude of bias resulting from this application of the behavioral data depend on whether animals in groups dive synchronously or asynchronously (Hodgson et al. 2017). Additional information on bowhead whale surface and dive durations, and associated variability, would benefit any analysis that requires estimates of availability to surface (e.g., humans or imagery) or underwater (e.g., passive acoustic monitoring) "observers". Lastly, the ASAMM sighting data do not include information to estimate forward detection distance or time, and our estimates of $\widehat{p_a}$ do not account for forward detection distance. However, Borchers et al. (2013) note that failing to account for forward detection distance in estimates typically results in a lower bias for aerial surveys than shipboard surveys because animals are within viewing range for shorter periods during the former.

Due to the very small number of large whale sightings that could not be positively identified to species (n = 5 during August 2019) and the limited diversity of large whale species in the study area (Clarke et al. 2020), we did not incorporate a species-identification bias correction factor into the present analysis. It is highly likely that those five whales were bowhead whales because other large cetaceans rarely venture into the survey area during summer and autumn. However, ASAMM's sightings of gray whales off the Tuktoyaktuk Peninsula in August 2019 reinforce the idea that not all large cetaceans read the rule book. If those five "unidentified large cetacean" sightings were bowhead whales, the present abundance estimate would be biased slightly low.

Although the true abundance of the Western Arctic bowhead whale population is unknown, we can compare our estimate to the ice-based survey estimates (the gold standard) and photo-ID estimates, focusing on the recent period from 2011 to 2019 (Table 4). (See Muto et al. [2021] for a concise summary of Western Arctic bowhead whale abundance estimates and CVs during the periods prior to commercial whaling, at the end of commercial whaling, and from 1978 to 2011.) Our point estimate (17,175) is within the 95% confidence intervals for all abundance estimates except the 2011 photo-ID estimate (Givens et al. 2018). Givens et al. (2018) note that it is "reasonable to expect that the ice-based estimate is a little low and the photo-id estimate may be a little high" because bowhead whales that did not travel through the ice-based study area during the survey period were not incorporated into the former, and missed matches in the photo-id data would result in an overestimate of abundance. As noted above, the present abundance estimate is likely also biased low due to bowhead whales located outside the aerial survey study area during the August 2019 survey period.

Because this is the first time that aerial line-transect surveys were used to estimate abundance for this population, we cannot definitely estimate the recent trend in abundance. However, the magnitude of the present abundance estimate and overlapping confidence intervals compared to other recent estimates are consistent with the hypothesis that the population is relatively stable. Conducting similar aerial line-transect surveys in the future would allow estimation of trend.

The estimated CV in the abundance estimate for the present analysis ($\widehat{CV}(\hat{N})$ = 0.237; Table 4) is within the parameter space tested in the IWC's Bowhead Strike Limit Algorithm, which ranged from 0.10 to 0.34, with 0.25 for the base case (IWC 2003b). The $\widehat{CV}$ of our abundance estimate ($\widehat{CV}(\hat{N})$ = 0.237) is similar to that from the 2011 photo-ID estimate ($\widehat{CV}(\hat{N})$ = 0.217) and both estimates from the 2019 ice-based survey ($\widehat{CV}(\hat{N})$ = 0.228); however, our $\widehat{CV}(\hat{N})$ is larger than that from the 2011 ice-based survey ($\widehat{CV}(\hat{N})$ = 0.052). It is noteworthy that the $\widehat{CV}$ of our abundance estimate ($\widehat{CV}(\hat{N})$ = 0.237) is considerably lower than that from Ferguson et al.'s (2021) design-based analysis ($\widehat{CV}(\hat{N})$ = 0.540) of the same survey data, even though the latter assumed that the estimates of bowhead whale surface interval and dive interval duration (needed for the availability bias correction factor) were known constants. The spatially-explicit density surface model was able to explain small-scale variability in bowhead whale encounter rate, which was a dominant source of uncertainty in the design-based abundance estimate (Ferguson et al. 2021).

Back-transformation bias (e.g., Finney 1941; Beauchamp and Olson 1973; Smith 1993; Rothery 1988; Thorson and Kristensen 2016) is the last known source of bias that remains in our abundance estimate. Here, we explain the source of back-transformation bias, the likely effect on the abundance estimate, and potential solutions. We predicted bowhead whale density over a hexagonal lattice using a DSM relating the natural logarithm of bowhead whale counts (defined above as random variable $\boldsymbol{\Omega}$) on segments of survey effort with known length to a sum of smoothing splines. However, we were ultimately interested in an estimate of population abundance. Therefore, we applied a nonlinear function (exponentiation) to the DSM predictions of log-counts to derive predictions of whale density that we could then integrate over the study area to compute an abundance estimate. Thorson and Kristensen (2016) succinctly state the statistical issue: "Whenever a random variable is transformed by a nonlinear function, the mean and variance of the variable are also transformed." Not correcting for back-transformation bias in our estimate of Western Arctic bowhead whale population abundance likely results in a negative bias. Specifically, if the random effect distribution is symmetric in log space, then simple back-transformation results in an underestimate. Alternatively, if the random effect distribution is highly left-skewed, this could result in a positive bias. Back-transformation bias is a pervasive issue in many ecological models, yet it is often overlooked in analyses of cetacean abundance. Investigating reliable solutions to correct for back-transformation bias in DSMs and making software widely available is our next focus. Simulating from the posterior of the hierarchical model or applying the methods of Thorson and Kristensen (2016) are potential solutions.

Overall, we believe that our abundance estimate and $\widehat{CV}(\hat{N})$ are the best estimates for the Western Arctic bowhead whale population in 2019. We accounted for the dominant sources of bias in aerial line-transect survey data and propagated uncertainty from all parameter estimates to $\widehat{CV}(\hat{N})$. Givens et al. (2021a, 2021b) clearly and comprehensively detail potential sources of bias and the likely magnitude and direction of each type of bias on the resulting abundance estimates derived from the 2019 ice-based survey. Even with the correction for boat disturbance, several sources of potential bias that would result in an underestimate of true abundance remain in the 2019 ice-based survey estimate (Givens et al. 2021b): "highly unusual" ice conditions; an unusual bowhead whale migration route that was sometimes too distant from observers to detect whales; application of an availability correction factor derived from passive acoustic data collected only during previous years; failure to conduct survey effort because of closed leads in the sea ice during the early weeks of the migration when numerous whales likely passed; and an unusually short observation platform height (Givens et al. (2021a).

CONCLUSIONS

The primary contributions of this study are threefold. First, we demonstrated that abundance and uncertainty estimates for the Western Arctic bowhead whale population that meet the standards for precision and bias required for making management decisions (IWC 2003b; NMFS 2016) can be derived from aerial line-transect surveys. Second, we showed that the analytical methodology used here considerably reduced the uncertainty in the population abundance estimate compared to the design-based estimate derived using the same data (Ferguson et al. 2021). This result occurred because the DSM explicitly modeled the patterns and correlations in the bowhead whale data; in design-based models, this spatial variability typically manifests as encounter rate variance and is often a dominant contributor to the total uncertainty in the abundance estimate (e.g., Ferguson et al. 2021). Lastly, the bowhead whale case study was sufficiently complex to require enhancements to the existing R packages dsm and mrds, and this increased functionality is now freely available to the distance-sampling and density surface modeling communities.

ACKNOWLEDGMENTS

We are very grateful to Billy Adams, Craig George, and Robert Suydam at the North Slope Borough (NSB) Department of Wildlife Management for their assistance with survey planning and outreach. Geof Givens helped with both pre- and post-survey methods development. The 2019 field season was funded, in part, by the US Department of the Interior (USDOI), BOEM, Environmental Studies Program, Washington, D.C., through Interagency Agreement No. M17PG00031 with the Alaska Fisheries Science Center (AFSC), NOAA. We are eternally grateful to BOEM for providing funding to arctic marine mammal monitoring and ecological research for 41 years, and particularly to Cathy Coon and Rick Raymond for overseeing the project during the 2019 field season. NOAA-AFSC provided partial funding to enable the extension of the normal ASAMM study area to include the eastern Beaufort Sea shelf and Amundsen Gulf in 2019. Thank you NSB Mayor Harry Brower, Jr. for sharing thoughtful insights about the arctic ecosystem and communities, and for your forward-thinking prioritization of long-term monitoring of arctic marine mammals. The Alaska Eskimo Whaling Commission provided welcome enthusiasm and support to our survey teams, and that was especially heartening during the first week of the August 2019 field season when low ceilings, fog, and wind-curtailed survey efforts. The field teams during the August 2019 surveys were exceptional in every way, and included team leaders Amelia Brower, Janet Clarke, Megan Ferguson, Katie Jackson, and Amy Willoughby, and observers Corey Accardo, Lisa Barry, Laura Ganley, Suzie Hanlan, Rachel Hardee, Richard Holt, and Nick Metheny. Clearwater Air, Inc., safely, professionally, and successfully provided aircraft support to ASAMM from 2009 to 2019, and we are especially grateful to Andy Harcombe, Stan Churches, and Jake Turner for their dedication to ASAMM; the 2019 project was funded via Contract No. D15PC00102 with USDOI. The Twin Otter aircraft, pilots, and mechanical support were provided by Kenn Borek Air, Ltd., and Aklak Air via Contract No. 140D8119C0004 with USDOI. Mike Hay from XeraGIS developed and maintained the ASAMM Survey data collection and post-processing software; Mike was always available to help fix things that our team broke in the field. Ben Hou from AFSC was instrumental in helping ASAMM develop imagery protocols. Real-time flight following was provided by USDOI, Bureau of Land Management, Alaska Interagency Coordination Center, South Zone Dispatch, and Kenn Borek Dispatch. Multiple entities in Canada helped us get off the ground in foreign territory: Canada Department of Fisheries and Oceans, Fisheries Joint Management Committee Canada, and Inuvialuit Game Council. We appreciate the assistance and guidance of Julie Speegle (NOAA Public Affairs Officer), Maggie Mooney-Seus (AFSC Communications Program Manager), John Callahan (BOEM Public Affairs Officer), and Maureen Clark (U.S. Fish and Wildlife Service Public Affairs Officer) on media relations. Tim Essington and Jim Thorson assisted with analytical methods development and provided valuable feedback on earlier drafts of this manuscript. Of course, we thank the dogs of ASAMM, the most loyal supporters of all.

LITERATURE CITED

Beauchamp, J.J. and J.S. Olson. 1973. Corrections for bias in regression estimates after logarithmic transformation. *Ecology* 54(6): 1403-1407.

Bivand, R. and N. Lewin-Koh. 2019. maptools: Tools for Reading and Handling Spatial Objects. R package version 0.9-9. Available from: https://CRAN.R-project.org/package=maptools.

Bivand, R. and C. Rundel. 2019. rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-2. Available from: https//CRAN.R-project.org/package=rgeos.

Bivand, R., T. Keitt, and B. Rowlingson. 2019. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-8. Available from: https://CRAN.R-project.org/package=rgdal.

Bivand, R.S., E.J. Pebesma, and V. Gomez-Rubio. 2013. Applied Spatial Data Analysis with R, Second Edition. Springer, NY. Available from: http://www.asdar-book.org/.

Bravington, M.V., D.L. Miller, and S.L. Hedley. 2021. Variance Propagation for Density Surface Models. *Journal of Agricultural, Biological and Environmental Statistics* 26: 306-323. doi: 10.1007/s13253-021-00438-2

Bockstoce, J.R. and J.J. Burns. 1993. Commercial whaling in the North Pacific sector. pp. 563-577 *In*: J.J. Burns, J.J., Montague, and C.J. Cowles, (eds.), *The Bowhead Whale*. Special Publication No. 2, The Society for Marine Mammalogy, Lawrence, Kansas.

Borchers, D.L., W. Zucchini, M.P. Heide-Jørgensen, A. Cañadas, and R. Langrock. 2013. Using Hidden Markov Models to Deal with Availability Bias on Line Transect Surveys. *Biometrics* 69(3), 703-713. doi: 10.1111/biom.12049

Buckland, S. T., D. R. Anderson, K. P, Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations.* Oxford University Press, Oxford. 432 p.

Burt, M.L., D.L. Borchers, K.J. Jenkins, and T.A. Marques. 2014. Using mark–recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution* 5(11): 1180-1191. doi:10.1111/2041-210X.12294

Candy, S. 2004. Modelling catch and effort data using generalised linear models, the Tweedie distribution, random vessel effects and random stratum-by-year effects. *CCAMLR Science* 11: 59–80.

Citta, J.J., L. Quakenbush, and J.C. George. 2021. Distribution and behavior of Bering-Chukchi-Beaufort bowhead whales as inferred by satellite telemetry, p. 31-56. *In*: J.C. George and J.G.M. Thewissen (eds.), *The Bowhead Whale Balaena mysticetus: Biology and Human Interactions*. Academic Press, London.

Clarke, J.T., A.A. Brower, M.C. Ferguson, A.L. Willoughby, and A.D. Rotrock. 2020. Distribution and Relative Abundance of Marine Mammals in the Eastern Chukchi Sea, Eastern and Western Beaufort Sea, and Amundsen Gulf, 2019. Annual Report, OCS Study BOEM 2020-027. Marine Mammal Laboratory, Alaska Fisheries Science Center, NMFS, NOAA. 603 pp.

da Silva, C.Q., J. Zeh, D. Madigan, J. Laake, D. Rugh, L. Baraff, W. Koski, and G. Miller. 2000. Capture-recapture estimation of bowhead whale population size using photo-identification data. *Journal of Cetacean Research and Management* 2(1): 45-61.

Dorsey, E.M., W.J. Richardson, and B. Würsig. 1989. Factors affecting surfacing, respiration, and dive behaviour of bowhead whales, Balaena mysticetus, summering in the Beaufort Sea. *Canadian Journal of Zoology* 67(7): 1801-1815. doi:10.1139/z89-257.

Dunn, P.K. and G.K. Smyth. 2005. Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing* 15: 267–280. doi: 10.1007/s11222-005-4070-y

Ferguson, M.C., J.T. Clarke, A.L. Willoughby, A.A. Brower, and A.D. Rotrock. 2021. Geographically stratified abundance estimate for Bering-Chukchi-Beaufort Seas bowhead whales (*Balaena mysticetus*) from an August 2019 aerial line-transect survey in the Beaufort Sea and Amundsen Gulf. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-428. 56 pp.

Fewster, R.M. and S.T. Buckland. 2004. Assessment of distance sampling estimators. pp. 281-306 *In:* S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas (eds.), *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford.

Finney, D.J. 1941. On the distribution of a variate whose logarithm is normally distributed. *Journal of the Royal Statistical Society Supplement* 7: 155-161.

George, J.C., G.H. Givens, R. Suydam, J. Herreman, J. Mocklin, B. Tudor, R. DeLong, C. Clark, R. A. Charif, and A. Rahaman. 2013. Summary of the spring 2011 ice-based visual, acoustic, and aerial photo-identification survey of bowhead whales conducted near Point Barrow, Alaska. Paper SC/65A/BRG/11Rev presented to the IWC Scientific Committee. 25 pp.

Givens, G.H., S.L. Edmondson, J.C. George, R. Suydam, R.A. Charif, A. Rahaman, D. Hawthorne, B. Tudor, R.A. DeLong, and C.W. Clark. 2016. Horvitz-Thompson whale abundance estimation adjusting for uncertain recapture, temporal availability variation, and intermittent effort. *Environmetrics* 27:134-146. doi: dx.doi.org/10.1002/env.2379

Givens, G.G., J. Mocklin, L. Vate Brattström, B.J. Tudor, W.R. Koski, J.E. Zeh, R. Suydam, and J.C. George. 2018. Adult survival rate and 2011 abundance of Bering-Chukchi-Beaufort Seas bowhead whales form photo-identification data over three decades. Paper SC/67B/AWMP/01 Rev1 presented to the IWC Scientific Committee. 24 pp.

Givens, G.H., George, J.C., Suydam, R. and Tudor, B. 2021a. Bering-Chukchi-Beaufort Seas bowhead whale (*Balaena mysticetus*) abundance estimate from the 2019 ice-based survey. *Journal of Cetacean Research and Management* 22: 61-73.

Givens, G.H., J.C. George, R. Suydam, B. Tudor, A. Von Duyke, B. Person, and K. Scheimreif. 2021b. Correcting the 2019 survey abundance of Bering-Chukchi-Beaufort Seas bowhead whales for disturbance from powered skiffs. Paper SC/68C/ASI/01 presented to the IWC Scientific Committee [virtual meeting]. 18pp.

Harwood, L.A., J. Auld, A. Joynt, and S.E. Moore. 2010. Distribution of bowhead whales in the SE Beaufort Sea during late summer, 2007-2009. DFO Canadian Science Advisory Secretariat Research Document 2009/111. iv + 22 p.

Hastie, T.J. and R.J. Tibshirani. 1990. *Generalized Additive Models*. Monographs on Statistics and Applied Probability 43. Chapman & Hall/CRC, Boca Raton. 335 p.

Hedley, S. and M. Bravington. 2014. Comments on design-based and model-based abundance estimates for the RMP and other contexts. Paper SC/65B/RMP/11 presented to the IWC Scientific Committee. 33 pp.

Hedley, S.L., and S.T. Buckland. 2004. Spatial models for line transect sampling. *Journal of Agricultural Biological and Environmental Statistics* 9(2): 181-199.

Hijmans, R.J. 2020. raster: Geographic Data Analysis and Modeling. R package version 3.0.2. Available from: http://CRAN.R-project.org/package=raster.

Hodgson, A., D. Peel, and N. Kelly. 2017. Unmanned aerial vehicles for surveying marine fauna: Assessing detection probability. *Ecological Applications* 27(4): 1253–1267.

International Whaling Commission. 2003a. Chair's Report of the Fifty-Fourth Annual Meeting. Annex C. Report of the aboriginal subsistence whaling sub-committee. Report of the International Whaling Commission. 2002: 62–75.

International Whaling Commission. 2018. Chair's Report of the Sixty-Seventh Meeting. Annex P. Amendments to the Schedule Adopted at the 67th Meeting. Report of the International Whaling Commission.

International Whaling Commission. 2003b. Report of the Scientific Committee. *Journal of Cetacean. Research and Management* (Suppl.) 5:1–92.

Johnson, D.S., J.L. Laake, and J.M. Ver Hoef. 2010. A Model-Based Approach for Making Ecological Inference from Distance Sampling Data. *Biometrics* 66(1): 310-318. doi: 10.1111/j.1541-0420.2009.01265.x

Jørgensen, B. 1987. Exponential Dispersion Models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(2), 127–162. http://www.jstor.org/stable/2345415

Koski, W., J. Zeh, J. Mocklin, A.R. Davis, D.J. Rugh, J.C. George, and R. Suydam. 2010. Abundance of Bering-Chukchi-Beaufort bowhead whales (*Balaena mysticetus*) in 2004 estimated from photo-identification data. *Journal of Cetacean Research and Management* 11(2): 89-99.

Laake, J. L. and D. L. Borchers. 2004. Methods for incomplete detection at distance zero, p. 109-189. In: S. T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas (eds.), *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford.

Laake, J., D. Borchers, L. Thomas, D. Miller, and J. Bishop. 2021. mrds: Mark-Recapture Distance Sampling. R package version 2.2.5.9003. https://github.com/DistanceDevelopment/mrds/

Laake, J.L., J. Calambokidis, S.D. Osmek, and D.J. Rugh. 1997. Probability of detecting harbor porpoise from aerial surveys: estimating g(0). *Journal of Wildlife Management* 61(1): 1997.

Laake, J., M.J. Dawson, & J. Hone. 2008. Visibility bias in aerial survey: Mark–recapture, line-transect or both? *Wildlife Research* 35(4), 299-309. doi: 10.1071/WR07034

Marques, F. F. C., and S. T. Buckland. 2003. Incorporating covariates into standard line transect analyses. *Biometrics* 59: 924-935. doi:10.1111/j.0006-341X.2003.00107.x.

Marsh, H., and D. Sinclair. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *Journal of Wildlife Management* 53(4): 1017–1024. doi: 10.2307/3809604.

McCullagh and Nelder. 1999. *Generalized Linear Models*. Monographs on Statistics and Applied Probability 37. Chapman & Hall/CRC, Boca Raton. 511 pp.

Miller, D.L. and M.V. Bravington. 2017. When can abundance surveys be analysed with "design-based" methods?

Miller, D.L., E. Rexstad, L. Burt, M.V. Bravington, and S. Hedley. 2021. dsm: Density Surface Modelling of Distance Sampling Data. R package version 2.3.1.9000. https://github.com/DistanceDevelopment/dsm

Miller, D.L., M.L. Burt, E.A. Rexstad, and L. Thomas. 2013. Spatial models for distance sampling data: recent developments and future directions. *Methods in Ecology and Evolution* 4: 1001–1010 doi: 10.1111/2041-210X.12105

Moon, T.A., M.L. Druckenmiller, and R.L. Thoman. 2021. Executive Summary. *Arctic Report Card 2021*. doi: 10.25923/5s0f-5163/

Muto, M.M., V.T. Helker, B.J. Delean, N.C. Young, J.C. Freed, R.P. Angliss, N.A. Friday, P.L. Boveng, J.M. Breiwick, B.M. Brost, M.F. Cameron, P.J. Clapham, J.L. Crance, S.P. Dahle, M.E. Dahlheim, B.S. Fadely, M.C. Ferguson, L.W. Fritz, K.T. Goetz, R.C. Hobbs, Y.V. Ivashchenko, A.S. Kennedy, J.M. London, S.A. Mizroch, R.R. Ream, E.L. Richmond, K.E.W. Shelden, K.L. Sweeney, R.G. Towell, P.R. Wade, J.M. Waite, and A.N. Zerbini. 2021. Alaska marine mammal stock assessments, 2020. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-421, 398 pp.

NMFS. 2016. Guidelines for preparing stock assessment reports pursuant to the 1994 Amendments to the MMPA. Available at: http://www.nmfs.noaa.gov/op/pds/index.html. Date accessed: 20 February 2022.

Pebesma, E.J. and R.S. Bivand. 2005. Classes and methods for spatial data in R. R News 5 (2). Available from: http://cran.r-project.org/doc/Rnews/.

Pedersen, E.J., D.L. Miller, G.L. Simpson, and N. Ross. 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ: 7, e6876. doi:10.7717/peerj.6876

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

Robertson, F.C., W.R. Koski, J.R. Brandon, T.A. Thomas, and A.W. Trites. 2015. Correction factors account for the availability of bowhead whales exposed to seismic operations in the Beaufort Sea. *Journal of Cetacean Research and Management* 15: 35-44.

Robertson, F.C., W.R. Koski, T.A. Thomas, W.J. Richardson, B. Würsig, and A.W. Trites. 2013. Seismic operations have variable effects on dive-cycle behavior of bowhead whales in the Beaufort Sea. *Endangered Species Research* 21(2): 143-160.

Rothery, P. 1988. A cautionary note on data transformation: bias in back-transformed means. *Bird Study* 35: 219-222.

Smith, R.J. 1993. Logarithmic transformation bias in allometry. *American Journal of Physical Anthropology* 90: 215-228.

Suydam, R., J.C. George, M. Ferguson, G. Givens, and C. Coon. 2019. Update on plans for a population survey in 2019 of Bering-Chukchi-Beaufort bowhead whales. Paper SC/68A/ASW/05 presented to the IWC Scientific Committee. 4 pp.

Thorson, J.T. and K. Kristensen. 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. *Fisheries Research* 175: 66-74. doi: 10.1016/j.fishres.2015.11.016

Ver Hoef, J.M. and P.L. Boveng. 2007. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* 88(11): 2766-2772.

Wang, M., and J.E. Overland. 2015. Projected future duration of the sea-ice-free season in the Alaskan Arctic. *Progress in Oceanography*. 136, 50-59.

Willoughby, A., M. Ferguson, B. Hou, C. Accardo, A. Rotrock, A. Brower, J. Clarke, S. Hanlan, M. Foster Doremus, K. Pagan, and L. Barry. 2021. Belly port camera imagery collected to address cetacean perception bias during aerial line-transect surveys: Methods and sighting summaries. U.S. Department of Commerce, NOAA Technical Memorandum NMFS-AFSC-427. 111 pp.

Wood, S.N., M.V. Bravington, and S.L. Hedley. 2008. Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70(5): 931-955. doi:10.1111/j.1467-9868.2008.00665.x

Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC. 496 pp.

Würsig, B. and C. Clark. 1993. Behavior, p. 157-199. *In*: J. J. Burns, J.J., Montague, and C. J. Cowles, (eds.), *The Bowhead Whale*. Special Publication No. 2, The Society for Marine Mammalogy, Lawrence, Kansas.

Würsig, B. and W.R. Koski. 2021. Natural and potentially disturbed behavior of bowhead whales, p. 339-363. *In*: J.C. George and J.G.M. Thewissen (eds.), *The Bowhead Whale Balaena mysticetus: Biology and Human Interactions*. Academic Press, London.

Yuan, Y., F.E. Bachl, F. Lindgren, D.L. Borchers, J.B. Illian, S.T. Buckland, H. Rue, and T. Gerrodette. 2017. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Annals of Applied Statistics* 11(4): 2270-2297. doi: 10.1214/17-AOAS1078

Table 1. Summary of bowhead whale sightings, by activity state, in the ASAMM August 2019 data compared to those in Robertson et al. (2013). n = number of groups detected during the ASAMM surveys or number of observations in the Robertson et al. (2013) analysis. ind = number of bowhead whales. P_a = probability that a group is at the surface within an observer's field of view.

	ASAMM August 2019							**Robertson et al. 2013**							
					Twin Otter		**Turbo Commander**	**Surface Intervals**				**Dive Intervals**			
	n	**% total n**	**ind**	**% total ind**	P_a	**SE(P_a)**	P_a	**n**	**% total n**	**Mean (min)**	**SD**	**n**	**% total n**	**Mean (min)**	**SD**
Travel	134	0.72	157	0.57	0.16	0.01	0.17	120	0.14	1.51	0.92	77	0.23	11.76	8.2
Cow & Calf	44	0.24	92	0.33	0.31	0.03	0.36	164	0.19	0.93	1.09	138	0.42	3.75	4.9
Deep Feed	6	0.03	23	0.08	0.17	0.02	0.19	213	0.25	1.11	0.71	47	0.14	8.74	6.31
Social	1	0.01	3	0.01	0.26	0.02	0.30	369	0.43	1.23	0.76	66	0.20	5.44	4.47
Total	*185*	*1*	*275*	*1*				*866*	*1*			*328*	*1*		

Table 2. Summary of generalized additive models and hierarchical generalized additive models considered candidates for the density surface model.

Model Formula	**Smoothing Spline**	**Sampling Distribution**	**% Explained Deviance**	**Effective Degrees of Freedom**
te(x, y, k=15, bs="ts")	Tensor product of thin plate regression splines with shrinkage	Negative Binomial, nb()	57.8	41.9
te(x, y, k=15, bs="ts")	Tensor product of thin plate regression splines with shrinkage	Tweedie, tw()	48.1	39.3
s(x, y, k=30, bs="so", xt=list(bnd=list(bnd.list)))	Soap film	Negative Binomial, nb()	59.3	43.3
s(x, y, k=30, bs="so", xt=list(bnd=list(bnd.list)))	Soap film	Tweedie, tw()	48.8	40.6
s(x, y, k=15, bs="sf", xt=list(bnd=list(bnd.list))) + s(x, y, k=15, bs="sw", xt=list(bnd=list(bnd.list))) + ti(x, y, A.fact, k=c(15,4), bs=c("sf", "re"), d=c(2,1), xt=list(list(bnd=list(bnd.list)), NULL)) + ti(x, y, A.fact, k=c(15,4), bs=c("sw", "re"), d=c(2,1), xt=list(list(bnd=list(bnd.list)), NULL))	Hierarchical soap film: single common smoother + group-level smoothers with same wiggliness	Tweedie, tw()	51.6	67.6

Table 3. Summary statistics from August 2019 aerial line-transect survey data used to estimate Western Arctic bowhead whale abundance.

	Turbo Commander	**Twin Otter**	**Total**
number of bowhead whale sightings	102	83	185
number of bowhead whales	146	129	275
number of aggregation protocols sessions	3	0	3
number of bowhead whale sightings during aggregation protocols	20	0	20
transect and aggregation protocols effort (km)	9,605	4,096	13,701

Table 4. Summary of recent abundance estimates for the Western Arctic bowhead whale population.

Year	Data Collection Method	$\hat{N}$	$\widehat{CV}(\hat{N})$	95% Confidence Interval		Citation
				Lower Bound	**Upper Bound**	
2011	photo-ID	27,133	0.217	17,809	41,337	Givens et al. (2018)
2011	ice-based survey	16,820	0.052	15,176	18,643	Givens et al. (2016)
2019	ice-based survey - abundance estimate uncorrected for boat disturbance	12,505	0.228	7,994	19,560	Givens et al. (2021a)
2019	ice-based survey - abundance estimate corrected for boat disturbance	14,025	0.228	8,971	21,927	Givens et al. (2021b)
2019	aerial line-transect survey - design-based analysis	14,531	0.540	5,042	41,875	Ferguson et al. (2021)
2019	aerial line-transect survey - density surface model	17,175	0.237	10,793	27,330	present analysis

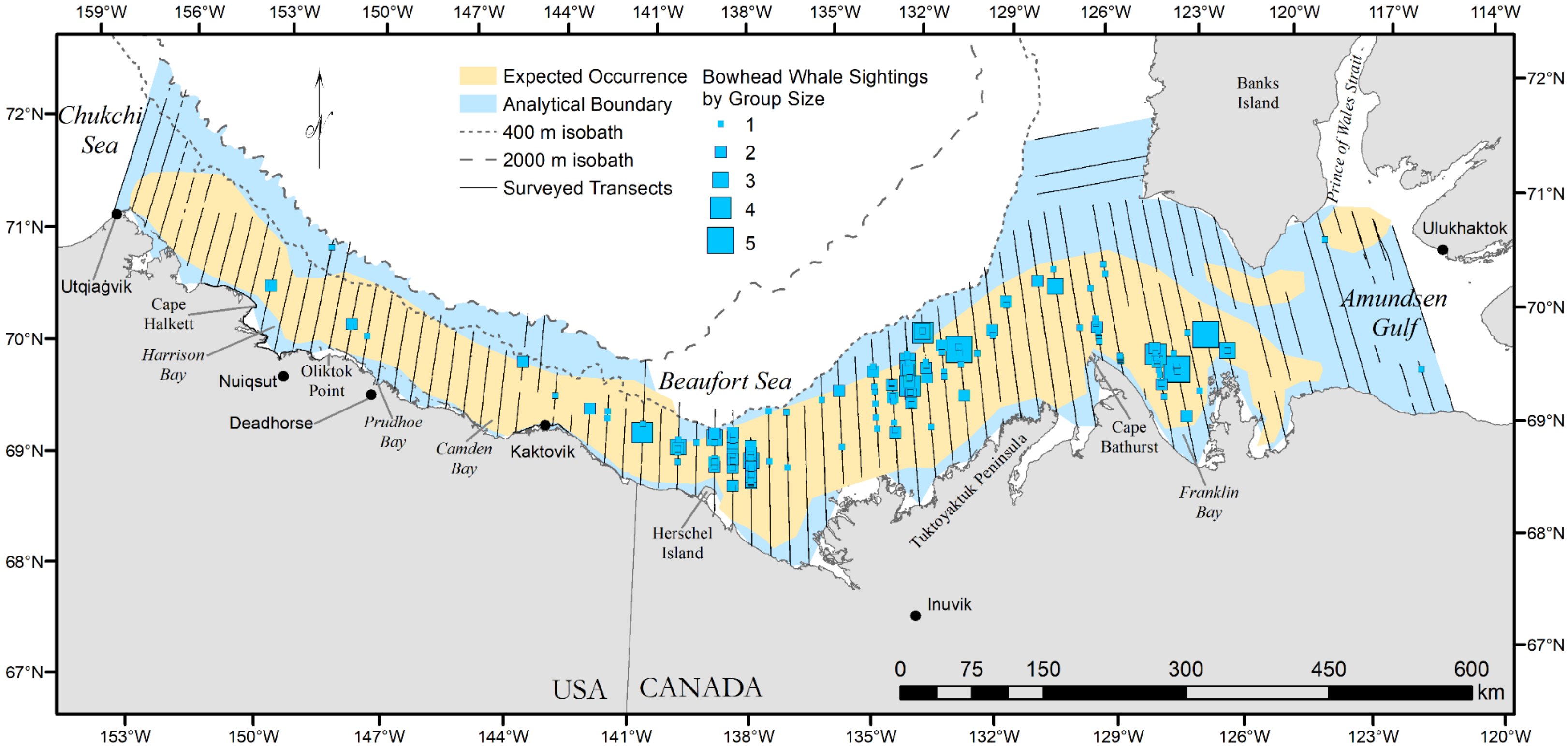

Figure 1. Study area for the Aerial Surveys of Arctic Marine Mammals (ASAMM) bowhead whale abundance survey in 2019. The expected distribution of bowhead whales in the study area during August was determined based on all available information, including Indigenous knowledge, historical whaling records, previous aerial surveys, and telemetry studies. Survey effort and bowhead whale sightings from the August 2019 survey period that were included in the abundance estimate are also shown. The primary bases of operations were Inuvik and Ulukhaktok, Northwest Territories, Canada, and Deadhorse, Alaska, USA.

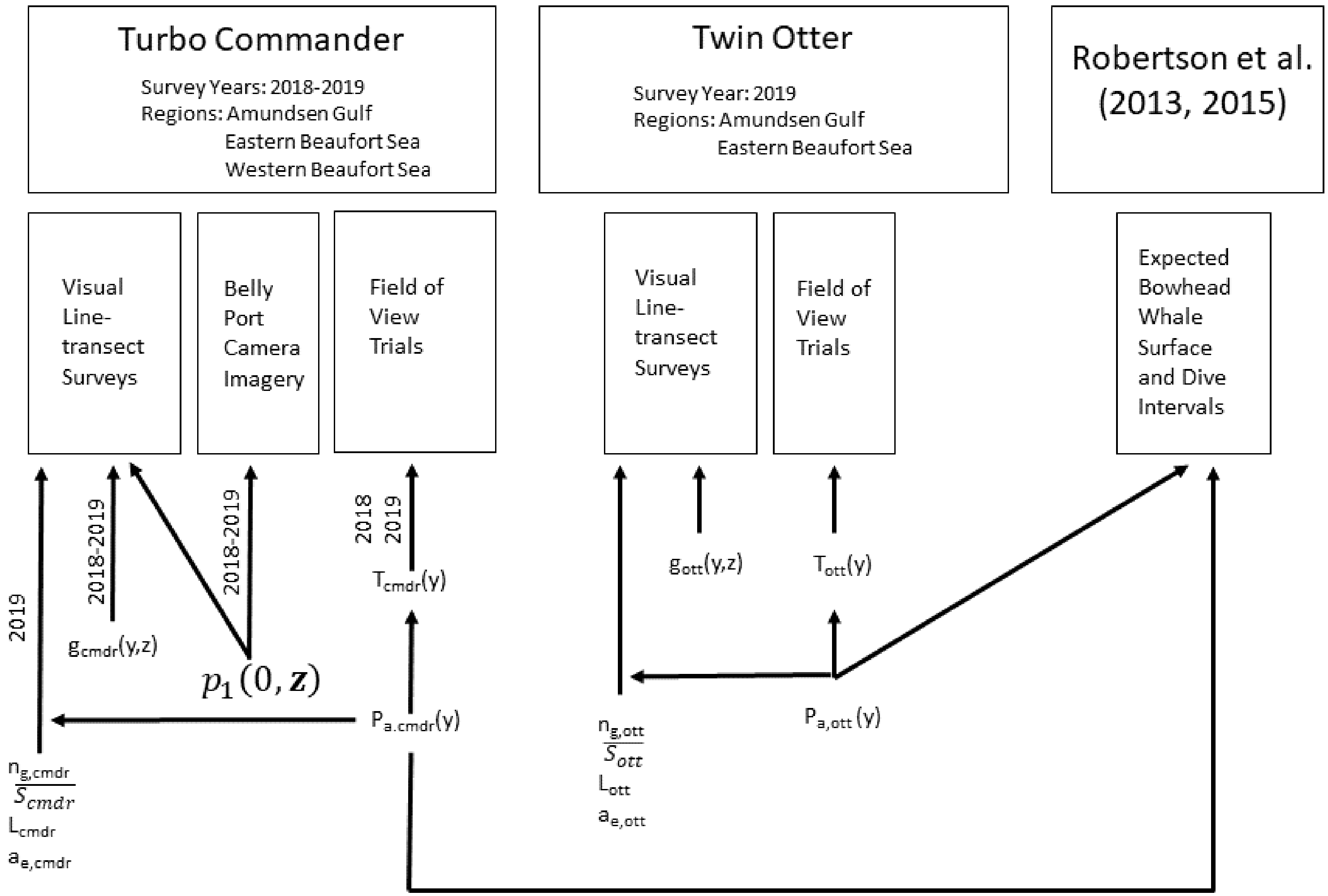

Figure 2. Data sources and subsets used to estimate parameters for the Western Arctic bowhead whale abundance estimate. cmdr = Turbo Commander. ott = Twin Otter. See text and Appendix A for definitions of other variables and parameters.

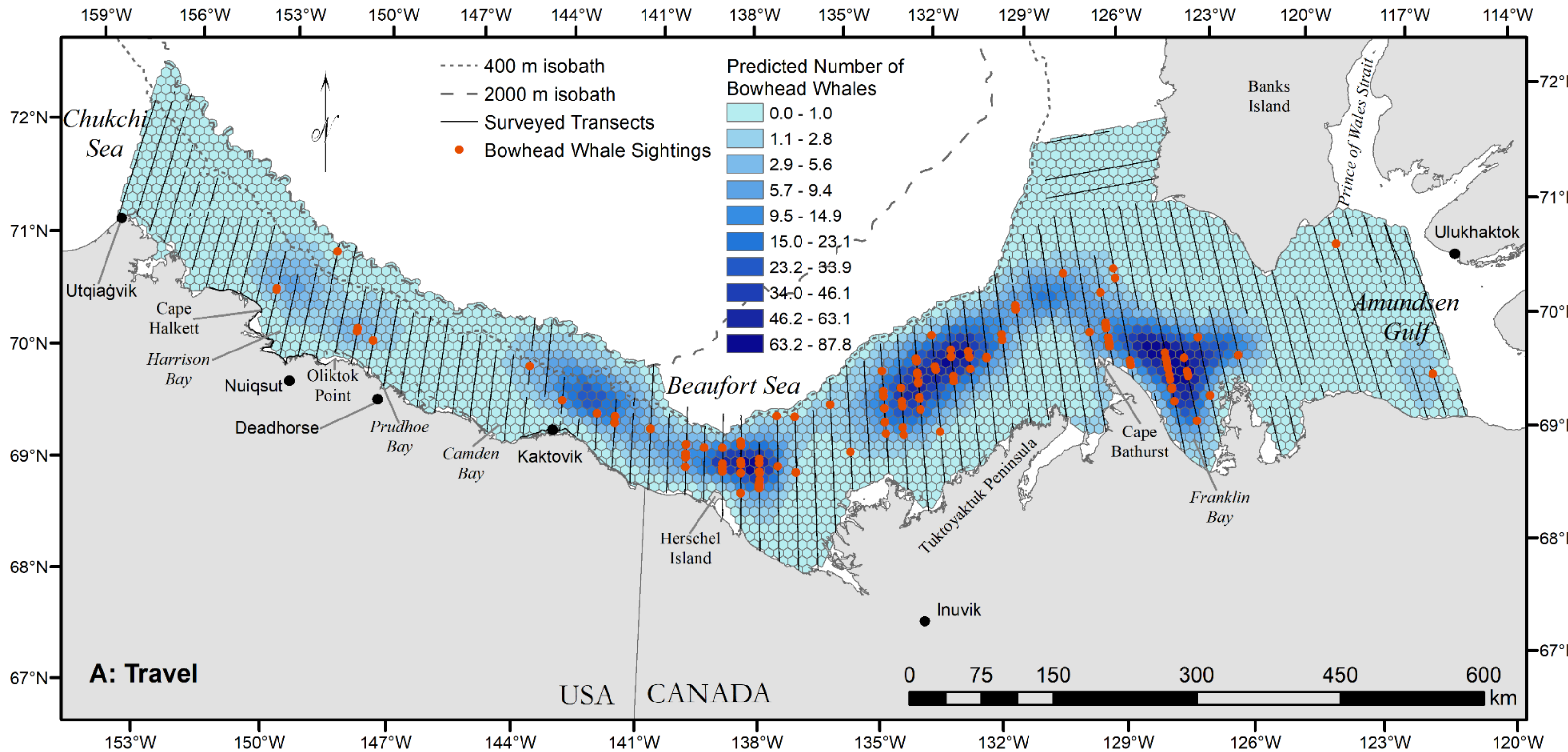

Figure 3A. Bowhead whale sightings during the August 2019 aerial line-transect survey and predicted number of bowhead whales from the density surface model, by activity state: A) travel; B) cow with calf or social; C) whales feeding in deep water.

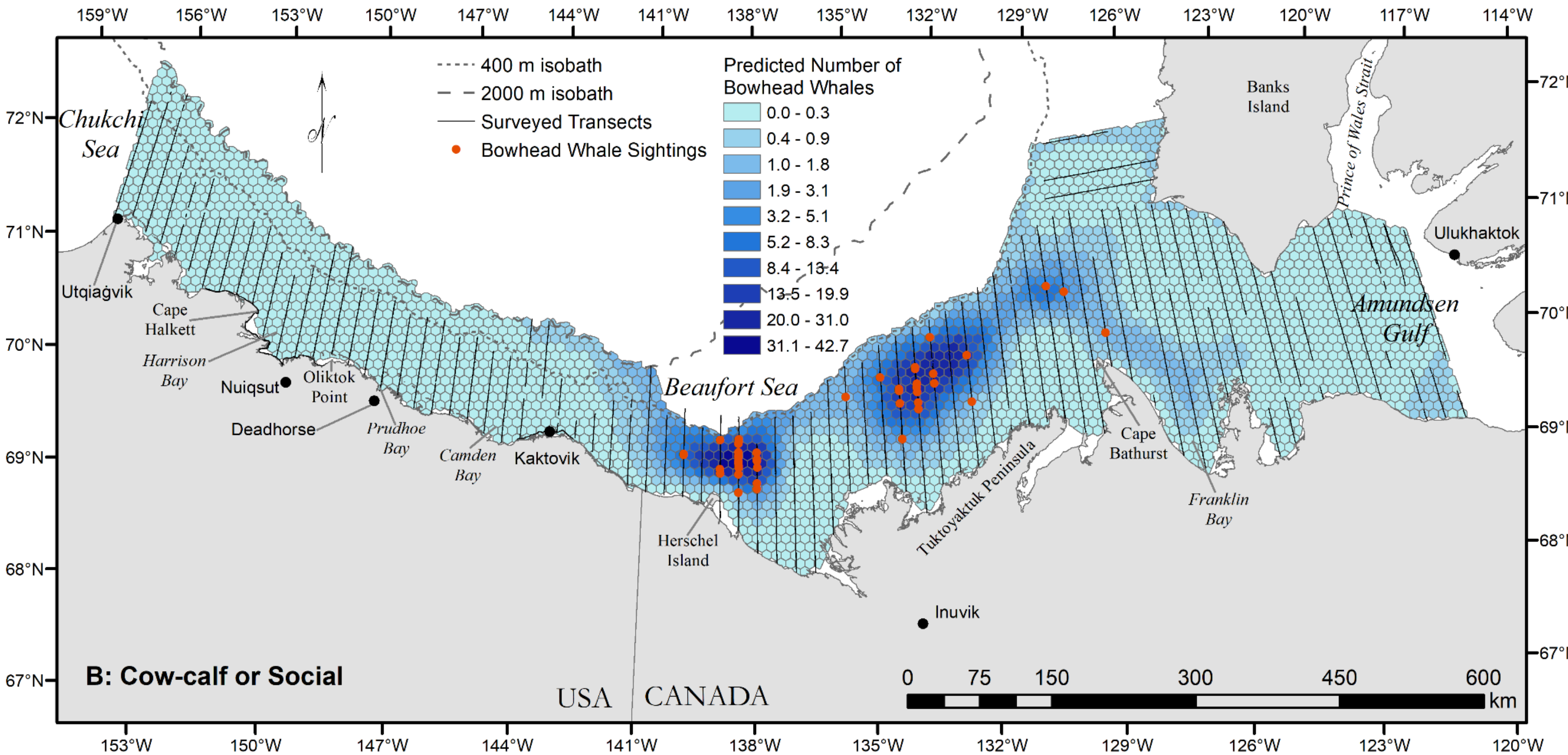

Figure 3B. Bowhead whale sightings during the August 2019 aerial line-transect survey and predicted number of bowhead whales from the density surface model, by activity state: A) travel; B) cow with calf or social; C) whales feeding in deep water.

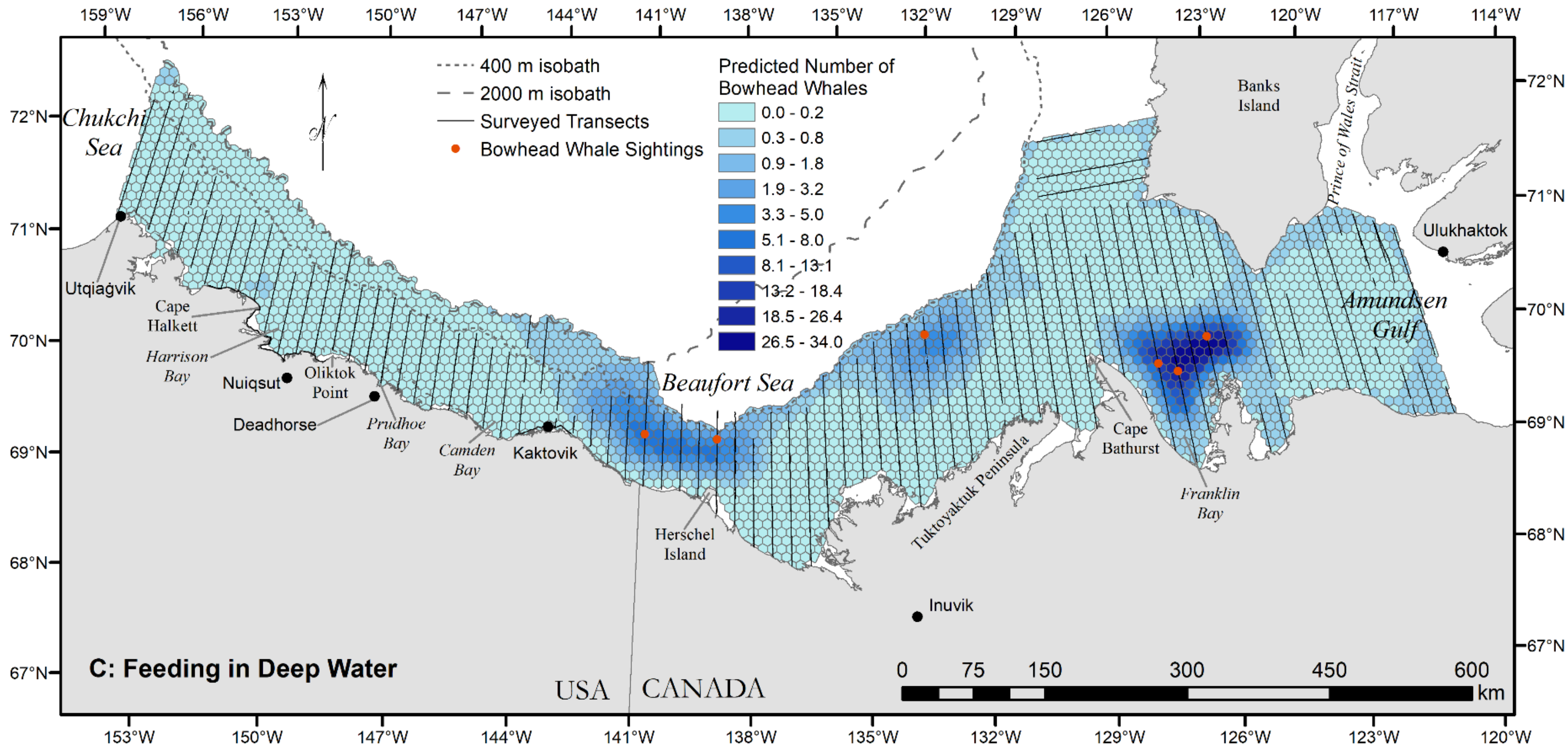

Figure 3C. Bowhead whale sightings during the August 2019 aerial line-transect survey and predicted number of bowhead whales from the density surface model, by activity state: A) travel; B) cow with calf or social; C) whales feeding in deep water.

Appendix A

Glossary of Notations and Abbreviations

a = area searched during line-transect survey

a_i = segment area, computed as $2wL_i$

A = activity state

ASAMM = Aerial Surveys of Arctic Marine Mammals project

b = a parameter in the hazard-rate detection function

BCB = Bering-Chukchi-Beaufort Seas stock of bowhead whales; also known as Western Arctic bowhead whales

$cc.soc$ = cow-calf or social activity state, used in the availability bias correction factors

cmdr = Turbo Commander

CV = coefficient of variation

$\widehat{CV}$ = estimated coefficient of variation

d = duration of a dive

$\bar{d} = \mathbb{E}(d)$= average duration of a dive

$\hat{D}$ = estimated density of whales

DSM = density surface model

f_q = smooth function for the q^{th} covariate in the density surface model

FOV = field of view

$g(y, \boldsymbol{z})$ = multiple covariates distance sampling detection function, which specifies the shape and scale of the observation model; assumes detection probability on the transect equals 1.0

h = waypoint index for the field-of-view model

i = segment index for the density surface model

j = group index in basic line-transect estimator of animal density

k = replicate index for the field-of-view model

L = total length of transects surveyed

L_i = length of transect surveyed in segment i

mcds = multiple covariates distance sampling

mrds = mark-recapture distance sampling

N = Normal (Gaussian) probability density function

$\hat{N}$ = estimated abundance of whales

n_g = number of groups detected

n_1 = number of groups detected by the ASAMM observers ("observer 1")

n_h = number of waypoints used in the field-of-view model

ott = Twin Otter

$p^*(\boldsymbol{z})$ = average probability that an ASAMM observer detects an object that is available to be seen, given covariates $\boldsymbol{z}$ that affect detectability, assuming transect detection probability is 1.0; $p^*(\boldsymbol{z}) = \frac{\int_0^w g(y,\boldsymbol{z})dy}{w}$

p_a = probability that a group is at the surface within an observer's field of view, also referred to as availability probability

$p(\hat{\theta};\boldsymbol{z})$ = overall probability that an observer detects an object, given covariates $\boldsymbol{z}$ that affect detectability; this term incorporates availability probability p_a, transect detection probability $p_1(0,\boldsymbol{z})$, and $p^*(\boldsymbol{z})$

$p_1(0,\boldsymbol{z})$ = probability that an ASAMM observer ("observer 1") detects an object located directly on the transect ($y = 0$) and available to be seen, given covariates $\boldsymbol{z}$ that affect detectability; determines the location of the intercept in the observation model

$p_1(y,\boldsymbol{z})$ = probability that an ASAMM observer ("observer 1") detects an object located at perpendicular distance y, given covariates $\boldsymbol{z}$ that affect detectability; also referred to as the observation model

$p_{1|2}(y,\boldsymbol{z})$ = probability that an ASAMM observer ("observer 1") detected an object that the photo analyst ("observer 2") also detected; derived using a mark-recapture distance sampling detection function

$\overline{pdist} = \frac{1}{n_h}\sum_h pdist_h$ = average perpendicular distance to the transect

pdist.scl = scaled perpendicular distance to the transect

q = smoothing term index for the density surface model

s = duration of a surfacing

$\bar{s} = \mathbb{E}(s)$ = average duration of a surfacing

S_j = size of group j

$^{\mathrm{T}}$ = matrix transpose operator

$T(0)$ = duration in which the ocean at perpendicular distance $y = 0$ is in the observer's view; this parameter is a function of the observer's field of view

$T(y)$ = duration in which the ocean at perpendicular distance y is in the observer's view; this parameter is a function of the observer's field of view

u_{iq} = covariate q on segment i in the density surface model

Var = variance

w = width used to build the multiple covariates distance sampling model, computed as the right-truncation distance minus the left-truncation distance

W_i = random variable for the number of whales on segment i in the density surface model

x = viewing distance (in meters) along the transect

y = perpendicular distance from the transect to the sighting

Υ = generic variable used in the definition of the multivariate delta method

$\boldsymbol{z}$ = covariates that affect detectability of a sighting

β_0 = a parameter in the mark-recapture detection function

$\boldsymbol{\beta}_{dsm}$ = vector of coefficients in the smooth functions for the density surface model

β_j = a coefficient parameter in the mark-recapture detection function

$\beta_{pdist.scl}$ = fixed effect of $pdist.scl$ on slope in the field-of-view model

β_y = the coefficient associated with perpendicular distance, y, in the mark-recapture detection function

γ = intercept in the field-of-view model

ε_k = residual error associated with the k[th] replicate in the field-of-view model

$\eta_i = log(\omega_i)$ = natural logarithm of the number of whales sighted on segment i in the density surface model

θ_0 = a parameter in the scale of the multiple covariates distance sampling detection function

θ_j = a coefficient parameter in the scale of the multiple covariates distance sampling detection function

λ = rate parameter of the dive process

μ = rate parameter of the surfacing process

ξ = mean of a Tweedie distribution

π = power parameter for a Tweedie distribution

σ_d = standard error of dive duration data

σ_{pdist} = standard deviation of perpendicular distances in the field-of-view model

σ^2_{resid} = variance of the residual error in the field-of-view model

σ_s = standard error of surface duration data

φ = dispersion or scale parameter for a Tweedie distribution

$\boldsymbol{\psi}$ = vector of smoothing parameters for the density surface model

ω_i = number of whales sighted on segment i in the density surface model

Ω_i = random variable for the natural logarithm of the number of whales on segment i in the density surface model

Appendix B

Detection Functions

Methods

The probability that an ASAMM observer ("observer 1") detects a group of whales located on the transect and the effects of distance (y) from the transect (and possibly other covariates $\boldsymbol{z}$) on detection probability were estimated using an observation model, $p_1(y, \boldsymbol{z})$, for each aircraft. Separate observation models were built for the Twin Otter and the Turbo Commander aircraft due to differences in window design and aircraft configuration that likely affected detectability. This decision was based on expert judgment rather than a formal statistical test because the latter likely would be unreliable due to the extremely unbalanced sample sizes for the two types of aircraft.

Observation models were formulated as mark-recapture multiple covariates distance sampling (mcds) detection functions (Marques and Buckland 2003, Laake and Borchers 2004). However, because belly port imagery were collected only on the Turbo Commander, the estimate of transect detection probability, $p_1(0, \boldsymbol{z})$, for the Turbo Commander was incorporated into to the observation model for the Twin Otter.

The underlying observation model was a scaled version of an mcds detection function, $g(y, \boldsymbol{z})$, (Laake and Borchers 2004):

$$p_1(y, \boldsymbol{z}) = p_1(0, \boldsymbol{z}) g(y, \boldsymbol{z}) \qquad \text{[B1]}$$

The mcds detection function assumes the probability of detecting an object on the transect equals 1.0; it specifies the functional form (shape and scale) of the observation model. The mcds model can take various forms, specified by its key function, such as the half-normal key function or hazard-rate key function. A half-normal model in which the standard deviation (scale parameter) is a linear function of covariates affecting detection probability may be represented as:

$$g(y, \boldsymbol{z}) = exp\left(\frac{-y^2}{2\left[exp\{\theta_0 + \sum_j \theta_j z_j\}\right]^2}\right) \qquad \text{[B2]}$$

An analogous hazard-rate model may be represented as:

$$g(y, \boldsymbol{z}) = 1 - exp\left[-\left(\frac{y}{exp\{\theta_0 + \sum_j \theta_j z_j\}}\right)^{-b}\right] \qquad \text{[B3]}$$

The average probability that an ASAMM observer detects an object that is available to be seen, given covariates $\boldsymbol{z}$ that affect detectability, assuming transect detection probability is 1.0, is:

$$p^*(\boldsymbol{z}) = \frac{\int_0^w g\,(y, \boldsymbol{z}) dy}{w} \qquad \text{[B4]}$$

where w is the width of the strip searched, defined in detail below.

Detection function models with half-normal and hazard-rate key functions, each with second-order cosine series adjustments, were considered. The null hazard-rate models had considerably lower AIC values and exhibited better fit than the half-normal models or models with cosine series adjustments, so forward stepwise selection of covariates, using AIC as the model selection criterion, proceeded with only the hazard-rate key function.

ASAMM line-transect data were filtered prior to building the detection functions. Only bowhead whale sightings made by primary observers during transect, aggregation protocols, and search effort (Clarke et al. 2020) with recorded declination angles were used in the detection function analyses. All analyses were limited to data collected during conditions of Beaufort Sea State 5 or less. The detection function for the Twin Otter was based on data from only 2019, the single year in which this specific type of aircraft was used to fly ASAMM surveys. The exact same configuration of Turbo Commander flew ASAMM surveys in 2018 and 2019, and belly port imagery were collected during both years. Therefore, the Turbo Commander detection functions incorporated data from all surveys in 2018 and 2019 during which imagery were concurrently collected, and from all Western Arctic bowhead whale abundance surveys, which were conducted from 5 to 27 August 2019.

Sighting data were truncated very close to and far from the transect. Data were left-truncated to account for lower sighting probabilities very close to the aircraft (Hain et al. 1999). Based on visual inspection of histograms of perpendicular sighting distances for bowhead whales, the Twin Otter data were left-truncated at 100 m (Figure B1) and the Turbo Commander data were left-truncated at 75 m (Figure B2). Based on preliminary analyses and to be consistent with ASAMM's cetacean aggregation protocols, sightings farther than 3 km from the original transect (prior to left-truncation) were omitted from the detection function analyses to minimize the effects of outliers. The strip width, w, used in the analysis was 2497 m for the Twin Otter and 2916 m for the Turbo Commander.

Belly port imagery data were also filtered prior to building the Turbo Commander detection function. Imagery sightings located on either side of the transect within the left-truncation distance for the Turbo Commander (75 m) were excluded from the mark-recapture model. Also omitted were imagery sightings collected when the ASAMM data indicated that Beaufort Sea State was > 5. For one bowhead whale detected in imagery, photo analysts could not conclusively determine whether there was a match in the ASAMM dataset; this imagery sighting was omitted from the analysis.

Covariates evaluated for inclusion in the detection function models are defined in Table B1. Detectability might depend on group size, so several group size covariates were considered. Beaufort Sea State affects an observer's ability to detect objects against the noise of whitecaps and waves, so two sea state variables were considered. Surveys were infrequently conducted when sea ice cover was greater than 10%; therefore, to provide balanced sample sizes, a categorical variable indicating only whether sea ice cover was < 10 % or ≥ 10 % was considered. Survey altitude ranged from 305-460 m above ground level, so this was included as a potential explanatory covariate. Lastly, a categorical covariate for sky condition was also considered. For instances in which there were multiple potential covariates for the same characteristic (e.g., group size), the covariate included in the univariate model with the lowest AIC value was carried through to the next round of variable selection and all of the related covariates were omitted from the rest of the analysis.

The model for $g(y, \boldsymbol{z})$ can estimate how detection probabilities vary with distance and other covariates; however, without the mark-recapture component $p_1(0, \boldsymbol{z})$, the intercept of the mcds detection function cannot be estimated (Laake and Borchers 2004). Hence, $p_1(0, \boldsymbol{z})$ determines the location of the intercept in the observation model and represents the probability that an ASAMM observer detects an object located on the transect, at $y = 0$ (the left-truncation point).

To derive $p_1(0, \boldsymbol{z})$, a mark-recapture detection function was used to find the probability that an ASAMM observer detected an object that the photo analyst ("observer 2") detected, $p_{1|2}(y, \boldsymbol{z})$. The model for $p_{1|2}(y, \boldsymbol{z})$ was based on trial configuration of observers, with the assumption of point independence (Laake and Borchers 2004). Trial configuration is appropriate here because imagery were used to estimate transect detection probability for the ASAMM observers; there was no need to derive a detection function for the photo analysts. Point independence requires that detections of objects located on the transect are independent between the ASAMM observers and photo analysts, but not necessarily elsewhere. Due to the assumption of point independence, $p_1(0, \boldsymbol{z}) = p_{1|2}(y, \boldsymbol{z})$.

Mark-recapture estimators are inherently plagued by bias due to unmodeled heterogeneity in detection probability, and distance is one of the largest sources of detection probability heterogeneity in distance-sampling data (Laake and Borchers 2004). The model for $p_{1|2}(y, \boldsymbol{z})$ allows detection probability to depend on perpendicular distance from the aircraft and other covariates. The additional covariates considered for inclusion in the mark-recapture model related to Beaufort Sea State and survey altitude (Table B1). The logistic model was used for the mark-recapture detection function:

$$p_{1|2}(y, \boldsymbol{z}) = \frac{exp\{\beta_0 + \beta_y y + \sum_j \beta_j z_j\}}{1 + exp\{\beta_0 + \beta_y y + \sum_j \beta_j z_j\}} \qquad [B5]$$

Transect detection probability for the ASAMM observers, averaged over all $\boldsymbol{z}$, was estimated by:

$$\hat{p}_1(0) = \frac{\sum_i^{n_1} \hat{p}_1(0, \boldsymbol{z}_i)/\hat{\mathbb{E}}(\hat{p}_1(\boldsymbol{z}_i))}{\sum_i^{n_1} 1/\mathbb{E}(\hat{p}_1(\boldsymbol{z}_i))} \qquad [B6]$$

where n_1 = number of groups detected by the ASAMM observers.

For each aircraft type, a single best observation model was selected based on AIC. For the Twin Otter, this involved selecting only an mcds model and incorporating the results from the mark-recapture detection function for the Turbo Commander. For the Turbo Commander, the mcds and mark-recapture models were selected by minimizing their respective AIC values. If a model with fewer covariates was within 2 AIC units of the model with the lowest AIC, the simpler model was chosen as the final model.

To evaluate model fit, for the mcds models we examined histograms of perpendicular sighting distances overlaid with model fit (Figures B3, B4) and conducted Cramer von Mises goodness-of-fit tests (Twin Otter test statistic = 0.0552879, p = 0.84; Turbo Commander test statistic = 0.0477863, p = 0.89). For the Turbo Commander's mrds model, we conducted a chi-square goodness-of-fit test (chi-square = 2.7436e-27, p = 1 with 4 degrees of freedom).

Results

For the Twin Otter, the mcds detection function was based on 85 bowhead whale sightings. The final model was the null hazard-rate model (Table B2), relying on perpendicular distance alone to estimate detection probabilities. For the Twin Otter, $\widehat{CV}\left(\widehat{p^*}(\boldsymbol{z})\right)$ = 0.122.

For the Turbo Commander, the mcds detection function was based on 297 bowhead whale sightings and the final model incorporated sky conditions and perpendicular distance (Table B2). Detection probabilities were highest under clear skies and lowest under overcast skies. The mark-recapture detection function for the Turbo Commander was based on a total of 305 unique observations, 297 of which were detected by the ASAMM observers, 53 were detected in the imagery, and 45 were detected by both. The final mark-recapture model incorporated distance and integer-valued Beaufort Sea State (Table B2). In this model, detection probability near the transect increased with increasing Sea States, which could indicate that observers focused their scans closer to the aircraft when surface waters were rough. The overall transect detection probability, $\boldsymbol{p_1(0, z)}$, was estimated to be 0.65. For the Turbo Commander, $\widehat{CV}\left(\widehat{p^*}(\boldsymbol{z})\right)$ = 0.076 and $\widehat{CV}\left(\hat{p}(0, \boldsymbol{z})\right)$ = 0.005.

Literature Cited

Hain, J.H.W., S.L. Ellis, R.D. Kenney, and C.K. Slay. 1999. Sightability of right whales in coastal waters of the southeastern United States with implications for the aerial monitoring program, p. 191-208. *In*: G.W. Garner, S.C. Amstrup, J.L. Laake, B.F.J. Manly, L.L. McDonald, and D.G. Robertson (eds.), *Marine Mammal Survey and Assessment Methods.* Proceedings of the Symposium on Surveys, Status, and Trends of Marine Mammal Populations, Seattle, WA, USA, 25-27 February 1998.

Laake, J. L. and D. L. Borchers. 2004. Methods for incomplete detection at distance zero, p. 109-189. *In*: S. T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers, and L. Thomas (eds.), *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford.

Marques, F. F. C., and S. T. Buckland. 2003. Incorporating covariates into standard line transect analyses. *Biometrics* 59: 924-935. doi:10.1111/j.0006-341X.2003.00107.x.

Table B1. Definitions of covariates considered for inclusion in the multiple covariates distance sampling detection function models for the Twin Otter (Ott) and Turbo Commander (Cmdr) aircraft. *Also considered in the mark-recapture distance sampling detection function model for the Turbo Commander. Perpendicular distance to sighting was included in all models.

Covariate Name	Definition	Categories	Aircraft
size	observed group size of the sighting		Ott, Cmdr
loggs	log10(size)		Ott, Cmdr
catsize	categorical group size	{1, >1}	Ott, Cmdr
catsizeGT2	categorical group size	{1, 2, >2}	Ott, Cmdr
catsize10	categorical group size	{1, 2, >2 & ≤10, >10}	Cmdr
iBeauf*	Beaufort sea state, as an integer-valued numeric variable ranging from 1 to 5		Ott, Cmdr
f5Beauf*	Beaufort sea state, as a categorical variable	{0 to 2, 3 to 5}	Ott, Cmdr
best.Alt*	aircraft altitude from the GPS, if available; otherwise, barometric altitude; scaled by 1/1000		Ott, Cmdr
SkyCon	sky condition	clear, partly cloudy, overcast	Ott, Cmdr
catIcePct	percent sea ice cover	{<10%, ≥10%}	Cmdr
Observer	Observer Initials	{LB, RH}	Ott

Table B2. Detection function parameter estimates for the Twin Otter and Turbo Commander aircraft. MCDS = multiple covariates distance sampling. MRDS = mark-recapture distance sampling.

Twin Otter MCDS Model

		Estimate	SE
Scale Coefficients			
	Intercept	-0.424	0.174
Shape Coefficients			
	Intercept	0.953	0.182

Turbo Commander MCDS Model

		Estimate	SE
Scale Coefficients			
	Intercept	0.436	0.197
	SkyCon overcast	-0.742	0.252
	SkyCon partly cloudy	-0.640	0.226
Shape Coefficients			
	Intercept	0.722	0.134

Turbo Commander MRDS Model

	Estimate	SE
Intercept	-2.190	1.502
distance	12.367	6.245
iBeauf	1.138	0.508

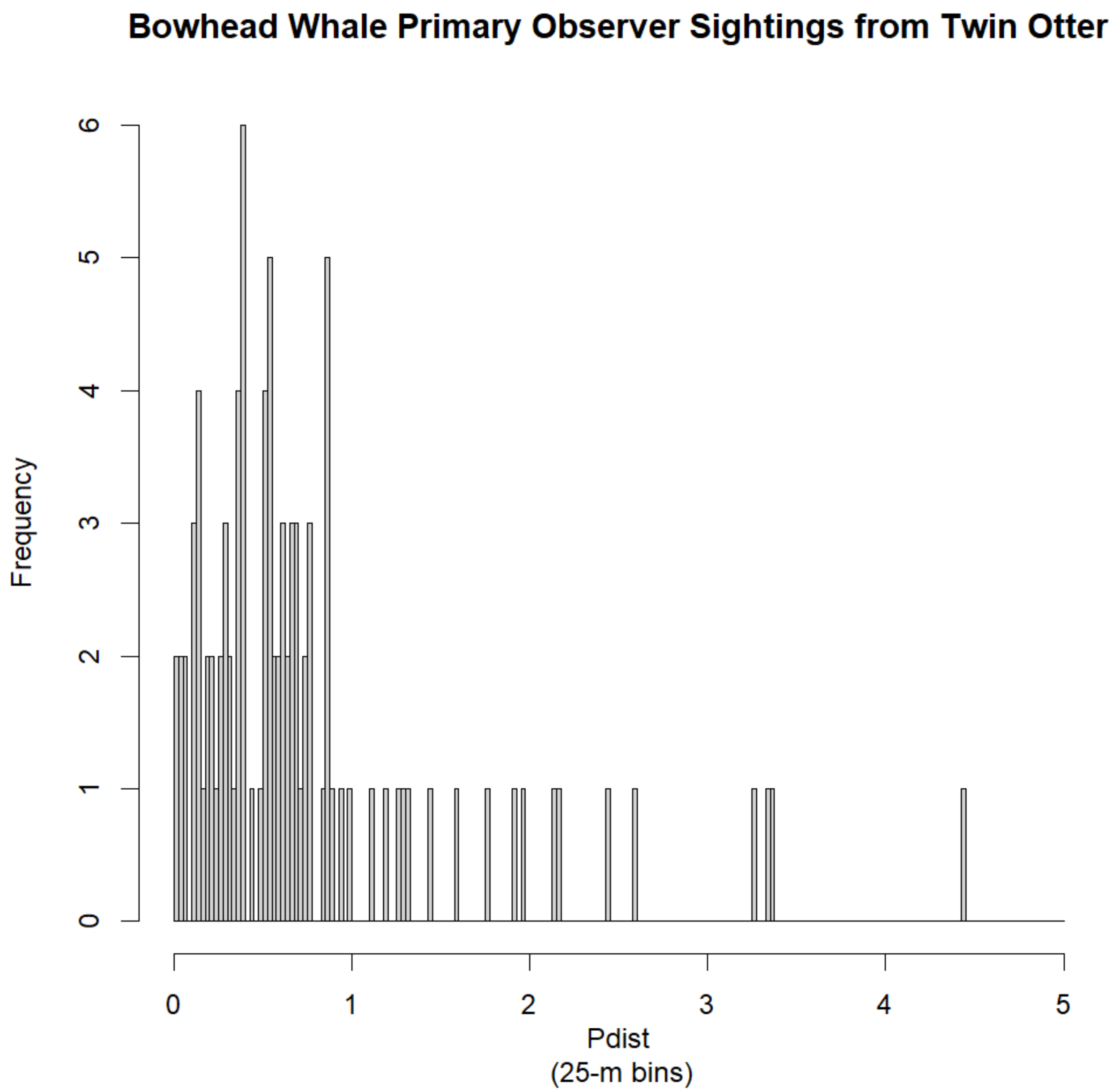

Figure B1. Histogram of perpendicular distance to bowhead whale sightings by primary observers on the Twin Otter aircraft during the August 2019 line-transect surveys.

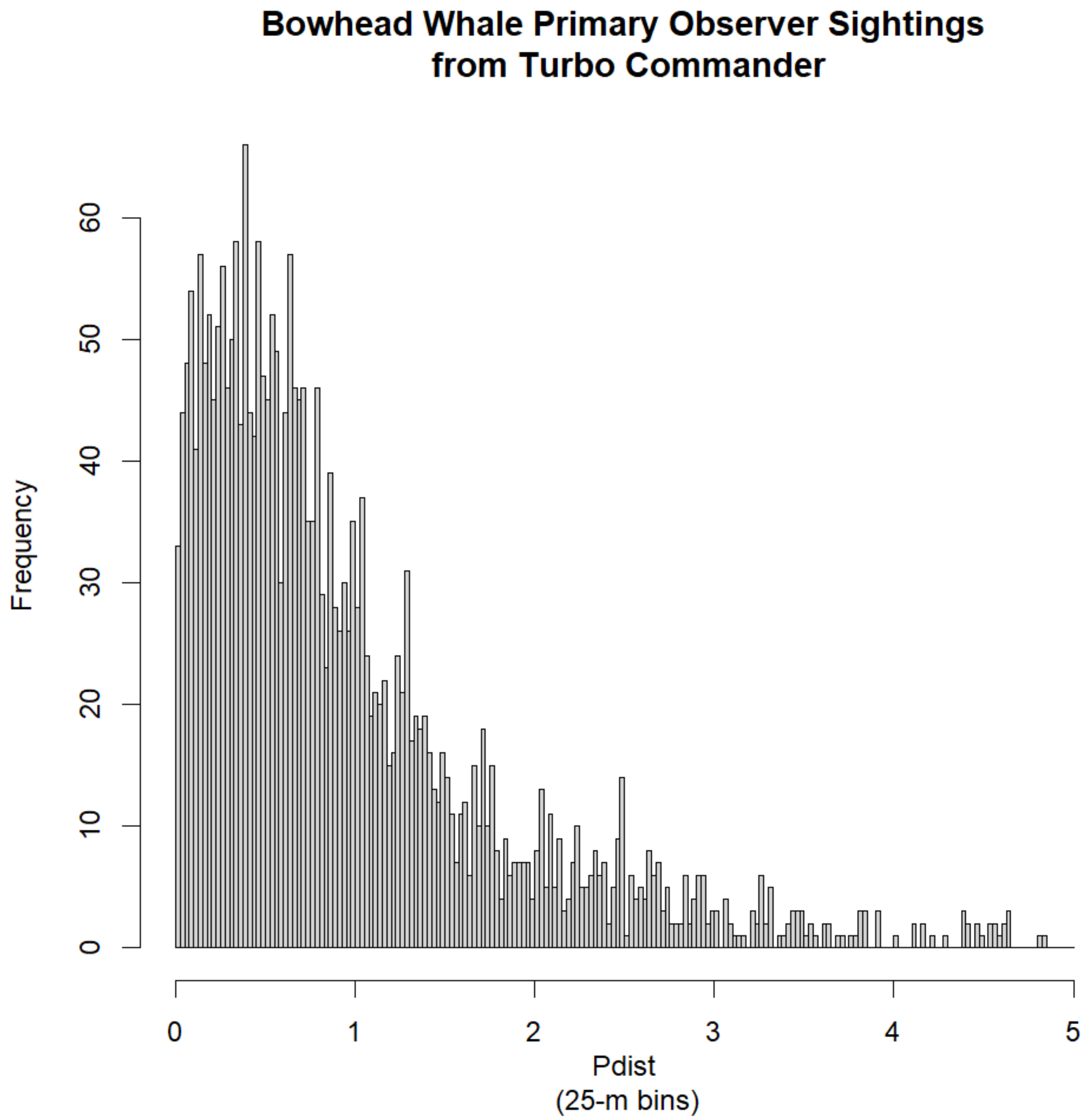

Figure B2. Histogram of perpendicular distance to bowhead whale sightings by primary observers on the Turbo Commander aircraft during the August 2019 line-transect surveys and other surveys conducted in 2018 and 2019 during which belly port imagery were collected.

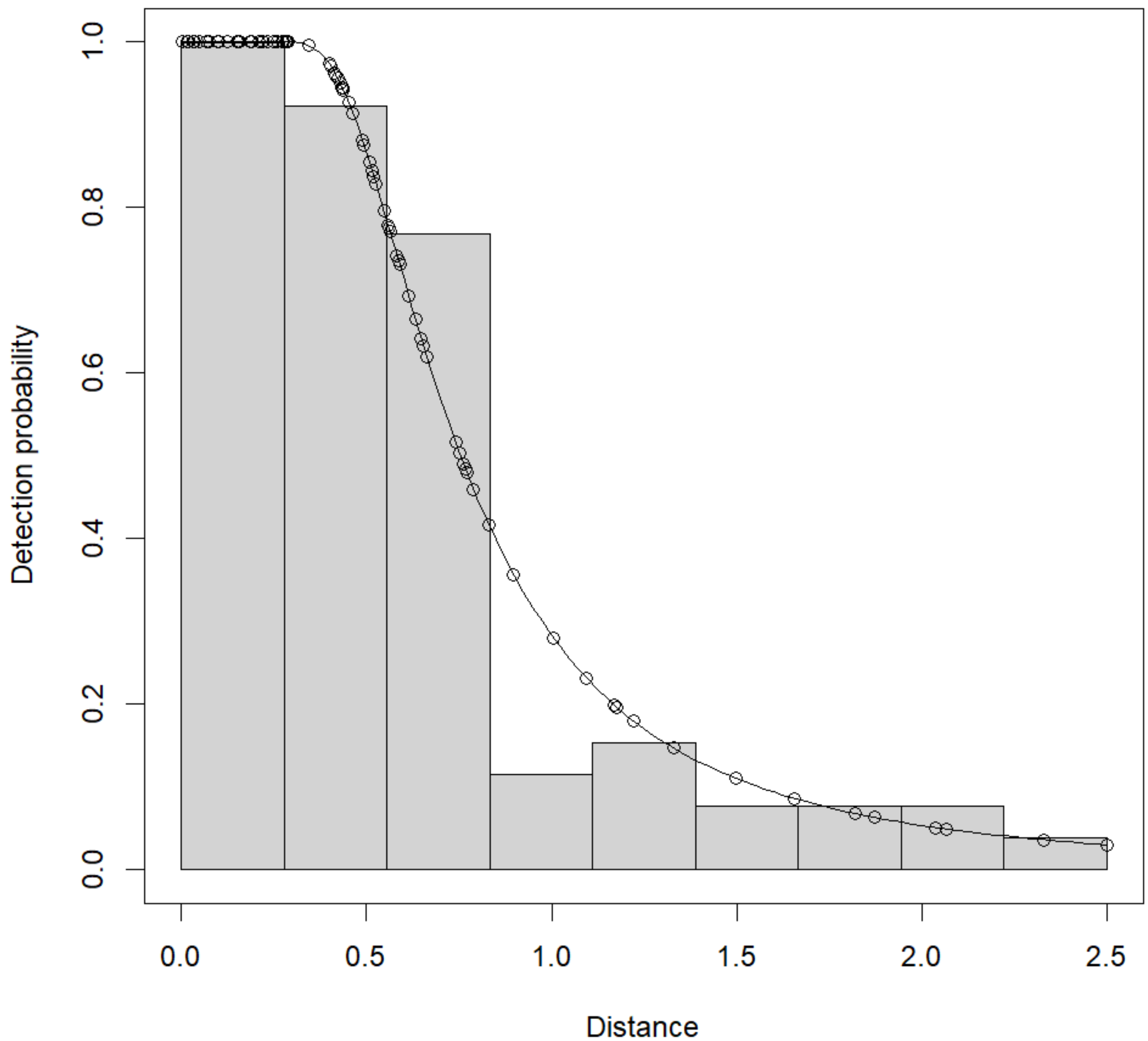

Figure B3. Histogram of perpendicular sighting distances to bowhead whale sightings from the Twin Otter, overlaid with the best detection model fit.

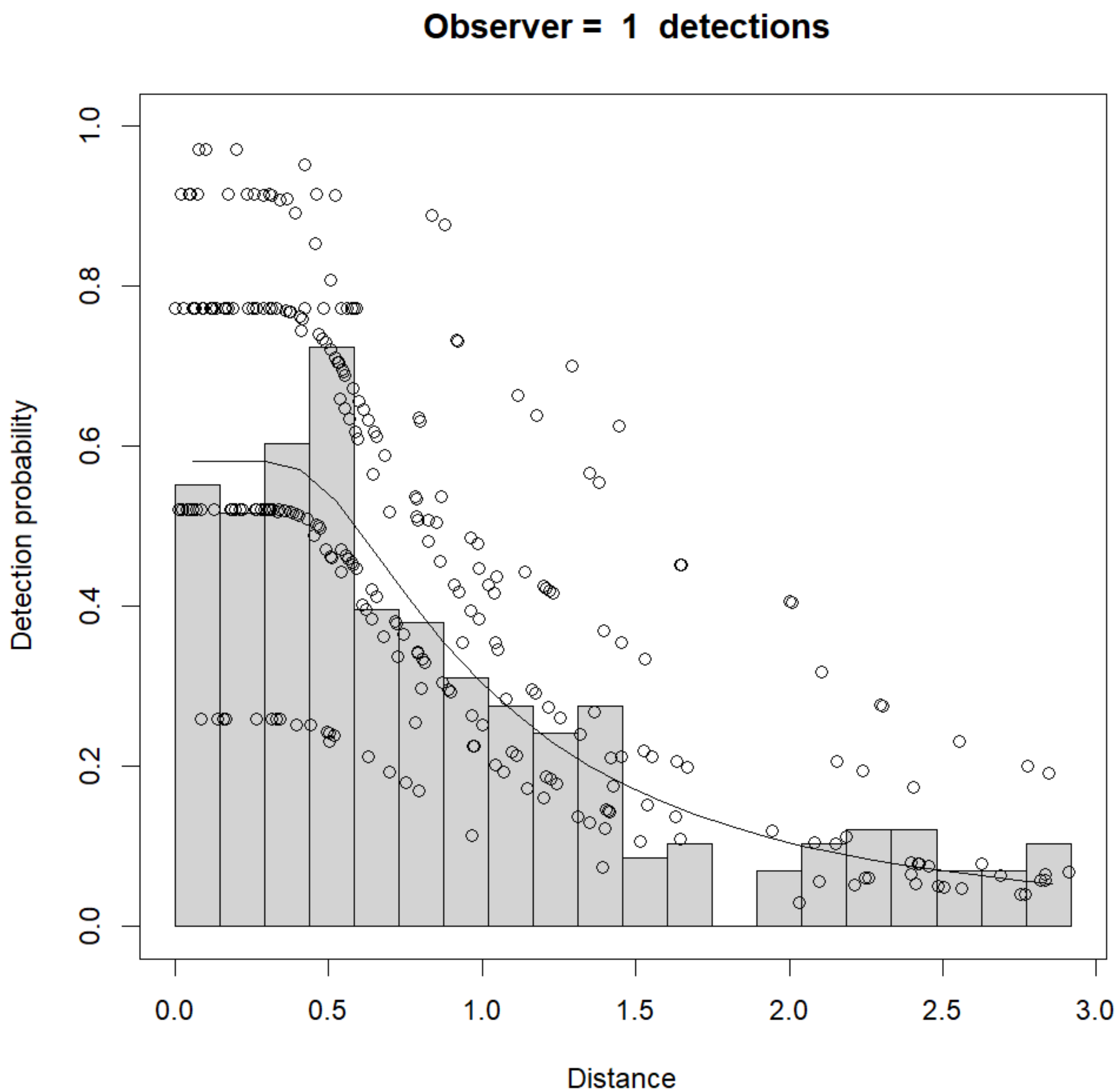

Figure B4. Histogram of perpendicular sighting distances to bowhead whale sightings made by marine mammal observers on the Turbo Commander, overlaid with the best detection model fit.

Appendix C

Field of View

Methods

Time-in-view estimates were incorporated into availability bias correction factors for the Western Arctic bowhead whale abundance estimate. To estimate the amount of time observers had to view a bowhead whale as a function of perpendicular distance to the transect ($T(y)$), in 2018 and 2019 the survey aircraft flew field-of-view (FOV) trials over land using fixed structures (a Conex box for the Turbo Commander and a cabin for the Twin Otter) as targets. See Clarke et al. (2020) for details about the FOV field methods.

Time-in-view at perpendicular distance (*pdist*) y, $T(y)$, increases linearly with viewing distance along the transect (*x*) as a function of aircraft speed (Robertson et al. 2015; Figure C1). The FOV model was defined using viewing distance rather than time as the response variable so that the results would be applicable at any aircraft speed. The forward time-in-view is the relevant parameter for deriving an availability bias correction factor for ASAMM data because sightings initially detected in the aft field of view are considered to have been "missed" by the ASAMM primary observers and were excluded from the abundance estimate analysis.

Because the field of view from the windows in the Turbo Commander was unobstructed ahead of the plane at the left-truncation distance (Ferguson et al. 2021), $T(0)$ was assumed to be a function of the distance at which a bowhead whale can be detected. Therefore, $T(0)$ for the Commander was computed by dividing the right-truncation distance (2.92 km) used to build the multiple covariates distance sampling detection function model (Appendix B) by the survey speed (213 km/h). The resulting estimate of time-in-view for the Turbo Commander was 49.3 sec.

For the Twin Otter, due to the relatively short viewing distance near the aircraft and the considerable variability in the FOV data, viewing distance was estimated from a linear model. Only two primary observers flew in the Twin Otter and sample sizes from the FOV trials were limited due to logistical constraints; therefore, data from both observers on the Twin Otter were pooled in the FOV model for this aircraft. Furthermore, the left and right bubble windows in the Twin Otter were identical in size and placement, so data from both sides of the aircraft were pooled.

The FOV model for the Twin Otter was based on scaled perpendicular distance to the transect, *pdist.scl*:

$$pdist.scl_h = \frac{(pdist_h - \overline{pdist})}{\sigma_{pdist}}, \qquad \text{[C1]}$$

where

h = waypoint index;

$\overline{pdist} = \frac{1}{n_h} \sum_h pdist_h$;

n_h = number of waypoints; and

σ_{pdist} = standard deviation among *pdist*.

The linear model for the FOV of the Twin Otter was defined as:

$$x_k = \gamma + \beta_{pdist.scl} * pdist.scl_k + \varepsilon_k \,, \qquad \text{[C2]}$$

where

x = viewing distance (in meters) along the transect;

k = replicate index;

γ = intercept;

$\beta_{pdist.scl}$ = fixed effect of *pdist.scl* on slope; and

$\varepsilon_k \sim N(0, \sigma^2_{resid})$.

Diagnostic tests run on the final Twin Otter FOV model exhibited no concerns about the data meeting the required assumptions of normally and independently distributed data for a linear model. The Shapiro-Wilk normality test was not statistically significant (W = 0.95888; p = 0.6416). There were no points with large leverages. A test of the Studentized residuals failed to identify outliers (Bonferroni-corrected p = 0.38563). Lastly, the half-normal plot of Cook's Distance failed to identify outliers.

The estimate of $\widehat{T(0)}$ for the Twin Otter used in the Western Arctic bowhead whale abundance analysis corresponds to the median from 10,000 parametric bootstrap samples from the Twin Otter FOV linear model (Ferguson 2020). The corresponding standard deviation was computed from the same set of bootstrap samples.

Results

The Twin Otter FOV data and the resulting model of forward viewing distance suggested that the target remained in view longer from the farthest (2000 m perpendicular distance) transect compared to the closest (500 m perpendicular distance) transect. The estimated intercept of the FOV model was 2180.5 (SE = 169.9), and the estimated slope was 131.7 (SE = 175.5). The corresponding values of the estimated intercept and slope for unscaled perpendicular distance, *pdist*, were 1975.1 and 0.17, respectively. The model estimated that a target located at the left-truncation distance (100 m) was visible to an observer on the Twin Otter for approximately 33.6 seconds (SD = 5.1 sec).

Literature Cited

Clarke, J.T., A.A. Brower, M.C. Ferguson, A.L. Willoughby, and A.D. Rotrock. 2020. Distribution and Relative Abundance of Marine Mammals in the Eastern Chukchi Sea, Eastern and Western Beaufort Sea, and Amundsen Gulf, 2019. Annual Report, OCS Study BOEM 2020-027. Marine Mammal Laboratory, Alaska Fisheries Science Center, NMFS, NOAA. 603 pp.

Ferguson, M.C. 2020. Bering-Chukchi-Beaufort Seas bowhead whale (*Balaena mysticetus*) abundance estimate from the 2019 aerial line-transect survey. Paper SC/68B/ASI/09 presented to the International Whaling Commission Scientific Committee, May 2020.

Ferguson, M.C., J. T. Clarke, A. L. Willoughby, A. A. Brower, and A. D. Rotrock. 2021. Geographically stratified abundance estimate for Bering-Chukchi-Beaufort Seas bowhead whales (*Balaena mysticetus*) from an August 2019 aerial line-transect survey in the Beaufort Sea and Amundsen Gulf. U.S. Dep. Commer., NOAA Technical Memorandum NMFS-AFSC-428. 54 p.

Robertson, F.C., W.R. Koski, J.R. Brandon, T.A. Thomas, and A.W. Trites. 2015. Correction factors account for the availability of bowhead whales exposed to seismic operations in the Beaufort Sea. *Journal of Cetacean Research and Management* 15: 35-44.

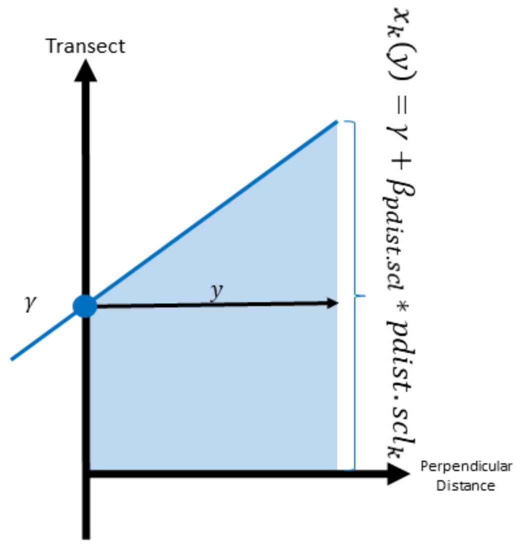

Figure C1. Schematic representation of the simple linear model for estimating parameters defining the forward (fwd) field of view for a primary observer on the right side of the aircraft. x is the viewing distance. γ is the intercept. $pdist.scl$ is scaled perpendicular distance .$\beta_{pdist.scl}$ is the slope. y is the perpendicular distance to the transect. *k* indexes the field-of-view trial number.

Appendix D

Hierarchical Generalized Additive Model Specification

In the language of the dsm() and gam() functions from the dsm (Miller et al. 2021) and mgcv packages (Wood 2017) in R, the full hierarchical generalized additive model used to estimate Western Arctic bowhead whale density can be represented as:

```
modl <- dsm(formula = count ~
        s(x.coord, y.coord, k=15, bs="sf",xt=list(bnd=list(bnd.list))) +
        s(x.coord, y.coord, k=15, bs="sw", xt=list(bnd=list(bnd.list))) +
        ti(x.coord, y.coord, A.fact, k=c(15,4), bs=c("sf", "re"), d=c(2,1),
            xt=list(list(bnd=list(bnd.list)), NULL)) +
        ti(x.coord, y.coord, A.fact, k=c(15,4), bs=c("sw", "re"), d=c(2,1),
            xt=list(list(bnd=list(bnd.list)), NULL)) +
        offset(log(ap)),
        knots = soap.knots,
        family = tw(),
        method = "REML",
        ddf.obj = list(cmdr.ddf, ott.ddf),
        observation.data = obs.dat,
        segment.data = seg.dat)
```

See the dsm and mgcv package helpfiles for additional information about each of the arguments.

Literature Cited

D.L. Miller, E. Rexstad, L. Burt, M.V. Bravington and S. Hedley. 2021. dsm: Density Surface Modelling of Distance Sampling Data. R package version 2.3.1.9000. https://github.com/DistanceDevelopment/dsm

Wood, S.N. 2017. *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC. 496 pp.