SC/69B/SM/01

Sub-committees/working group name: SM

Spatially-explicit models of density improve estimates of Eastern Bering Sea beluga (Delphinapterus leucas) abundance and distribution from line-transect surveys

M.C. Ferguson, P.B. Conn, And J.T. Thorson

INTERNATIONAL
WHALING COMMISSION

Spatially-explicit models of density improve estimates of Eastern Bering Sea beluga (*Delphinapterus leucas*) abundance and distribution from line-transect surveys

Megan C. Ferguson[1, 2], **Paul B. Conn**[2], **and James T. Thorson**[3]

[1]**Biodiversity Research Institute, Portland, Maine, USA**
[2]**Marine Mammal Laboratory, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington, USA**
[3]**Resource Ecology and Fisheries Management, Alaska Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, Washington, USA**

Corresponding author:
Megan C. Ferguson[1]

Email address: megan.ferguson@briwildlife.org

ABSTRACT

We investigate spatially explicit models and ensemble modeling techniques for estimating animal abundance from line-transect survey data. Spatially explicit models are expected to be statistically more efficient, resulting in more precise abundance estimates, than design-based abundance estimators that rely heavily on assumptions about survey design. Ensemble models allow model selection uncertainty to propagate to the abundance estimator. We develop density surface models using stochastic partial differential equations and basis-penalty smoothers for a case study, belugas (*Delphinapterus leucas*) from the Eastern Bering Sea (EBS) stock. EBS belugas are upper trophic level predators in a rapidly changing ecosystem and are a vital nutritional and cultural resource for Alaska Natives. Effective management of this stock requires regular monitoring to derive accurate and unbiased estimates of abundance. Since 1992, aerial line-transect surveys have been the primary means of surveying and estimating abundance of EBS belugas in the region. We compared EBS beluga abundance estimates for 2017 and 2022 that were derived using post-stratified, design-based abundance estimators with analogous estimates derived using spatially explicit and ensemble modeling methods. Although the design-based estimators were less precise than individual spatially-explicit models (with one exception), precision (CV) was essentially equivalent between the design-based and ensemble model-averaged abundance estimators. The design-based models estimated that there were 12,269 belugas in 2017 (CV = 0.12) and 20,635 belugas in 2022 (CV = 0.31; the study area was larger in 2022). The ensemble spatial models estimated that there were 11,597 belugas in 2017 (CV = 0.12) and 17,197 belugas in 2022 (CV = 0.33). Among the individual spatially-explicit models, abundance estimates ranged from 11,242 to 11,962 (CV = 0.11 to 0.12) in 2017 and 12,593 to 21,508 (CV = 0.18 to 0.29) in 2022. Because spatial models identify spatial patterns in beluga density (the number of belugas per unit area) at finer resolutions than design-based models, we argue that ensembles of spatially-explicit density models provide a reasonable path forward for estimating EBS beluga abundance and distribution in a way that is useful to management and conservation efforts.

INTRODUCTION

Effective management of wild populations requires an accurate estimate of uncertainty, including both of its components, bias and precision. This is because natural resource management inevitably involves choosing among alternative actions that may have different effects on a population in the future. Our ability to predict the future depends on how well we know the ecosystem today and the magnitude and direction of cascading effects that may result from a particular management action. Transparent communication about scientific uncertainty is particularly important when managing populations that are

hunted for Native subsistence, due to the animals' nutritional, cultural, and spiritual value to Indigenous peoples.

For decades, line-transect data were analyzed using a two-stage process invoking model-based inference to estimate detection probability, followed by design-based inference to extrapolate an estimate of the number of animals on the surveyed transects to an estimate of the number of animals throughout the study area. Design-based inference has a rich history in sampling (Cochran, 1977), and is appealing in its simplicity. In particular, random or systematic placement of transects ensures that simple extrapolations of densities from sampled to unsampled areas (using, e.g., simple random sampling or stratified random sampling estimators) will be unbiased, assuming that the specified design was correctly followed during field-sampling.

We explore the use of post-stratified design-based abundance estimators and density surface models (DSMs) to estimate abundance. Since the early 2000s, model-based approaches to inference from line-transect survey data (Hedley and Buckland, 2004; Johnson et al., 2010; Miller et al., 2013; Yuan et al., 2017) have become more popular for a number of reasons. For instance, modeling animal density as a function of spatial or environmental covariates may increase precision and reduce bias in the overall abundance estimate for the survey area. This applies particularly to cases in which animal density is spatially heterogeneous and achieved survey coverage is non-uniform, for example due to incomplete survey effort or spatially heterogeneous detection probability (Hedley and Bravington, 2014). Additionally, DSMs can be used to create high-resolution maps of animal density, which are useful for marine spatial planning, estimating potential impacts from anthropogenic activities, and investigating ecological relationships. We are particularly interested in the sensitivity of abundance estimates to DSM model structure, how this variance propagates through ensemble models, and how model-based abundance estimates compare with conventional post-stratified design-based estimators. We identify similarities and differences among different analytical approaches both theoretically and with a case study, the Eastern Bering Sea (EBS) stock of belugas (*Delphinapterus leucas*), which is hunted for subsistence by Alaska Natives.

EBS belugas are vital to Indigenous communities near Norton Sound and the Yukon River Delta in northwestern Alaska (Figure 1). The northern Bering Sea ecosystem is experiencing rapid ecological changes (Siddon, 2023) and increased human activities. EBS belugas are one of four beluga stocks that have been co-managed since 1988 by the Alaska Beluga Whale Committee and NOAA Fisheries (Adams et al., 1993; Frost et al., 2021). The ABWC includes hunters, resource managers, and scientists. The goals of the ABWC are to maintain healthy beluga populations in Alaskan waters, provide adequate subsistence harvest of beluga whales, and protect hunting privileges for Alaskan subsistence hunters (Frost et al., 2021). Since its founding, the ABWC has believed that education, maintaining accurate harvest data, and conducting surveys to estimate stock abundance on a regular schedule are critical to the health of northwestern Alaska's beluga stocks and the communities who depend on them.

The distribution and movement patterns of EBS belugas are primarily known from Indigenous and other local knowledge (Huntington and Communities of Buckland, Elim, Koyuk, Point Lay, and Shaktoolik, 1999; Oceana and Kawerak Inc., 2014; Lowry et al., 2017), aerial surveys (Lowry et al., 2017; Ferguson et al., 2023), telemetry studies (Citta et al., 2017), and genetics (e.g., O'Corry-Crowe et al., 2018; O'Corry-Crowe et al., 2021). EBS belugas predictably occur in the Norton Sound/Yukon Delta region during the period from shortly after sea ice breakup (usually mid-May) until freeze-up (usually November) (Lowry et al., 2017; Citta et al., 2017). Belugas from this stock are hunted by more than 20 villages during spring, summer, and autumn (Lowry et al., 2019). EBS beluga distribution from spring through autumn reflects high densities of prey, particularly fishes (Lowry et al., 2017), and also can be affected by sea ice conditions and human disturbance (Huntington and Communities of Buckland, Elim, Koyuk, Point Lay, and Shaktoolik, 1999; Oceana and Kawerak Inc., 2014).

To obtain an estimate of EBS beluga abundance that could be used to evaluate the sustainability of beluga subsistence harvests, the ABWC conducted aerial surveys in Norton Sound and along the Yukon River Delta each year from 1992 to 1995 and 1999 to 2000 (Lowry et al., 2017). In 1992, aerial surveys were conducted in May, June, and September to determine the best month for conducting future surveys. Based on those results, aerial surveys for all remaining years were conducted in June, when belugas tend to concentrate near Pastol Bay and the Yukon River Delta (Figure 1). Lowry et al. (2017) estimated EBS beluga abundance to be 6,994 belugas (95% confidence interval 3,162-15,472) based on the aerial surveys conducted in June 2000. This estimate included a correction factor of 2.0 to account for availability bias

(Marsh and Sinclair, 1989), which arises when belugas in the area searched during the surveys are out of the observers' field of view or are underwater when the aircraft flies over them.

During June 2017 and 2022, ABWC and NOAA Fisheries collaborated to conduct aerial line-transect survey in the Norton Sound/Yukon Delta region to collect data to derive updated abundance estimates for the EBS beluga stock. Ferguson et al. (2023) presented an estimate of EBS beluga abundance of 12,269 belugas (CV = 0.12) based on the 2017 surveys. Their abundance estimate incorporated correction factors for availability bias and transect detection probability, and was derived using design-based methods with post-stratification (Ferguson et al., 2023). Compared to 2017, the 2022 surveys included less survey effort, and beluga sightings were more patchily distributed. We were interested in whether DSMs could provide a reasonable alternative to design-based abundance estimators for these two most recent survey years. There are a considerable number of ways to formulate DSMs. Therefore, we also examined how different types of DSMs performed on the same dataset. We expected to see differences among density surfaces and total population abundance estimates across DSMs due to differing assumptions about spatial covariance.

This paper compares different analytical methods used to estimate population abundance from line-transect survey data for the purpose of effectively managing a population. The remainder of this paper is structured as follows: i) introduction to basic estimators of animal density and abundance from line-transect survey data; ii) definition of the marginal likelihood that forms the core of the DSMs; iii) definition of the random effects that form the basis of the different DSMs that we compared; iv) description of methods for predicting abundance from individual DSMs; v) explanation of model validation and evaluation methods; vi) description of methods for calculating uncertainty in abundance estimates for each individual DSM; vii) definition of the ensemble modelling approach that was used to account for model selection uncertainty; and, finally, viii) application to the EBS beluga case study, focusing on how results from the individual DSMs, ensemble DSMs, and design-based estimators of abundance compare. We assume that the reader is familiar with basic terminology and definitions associated with distance sampling (Buckland et al., 2001).

MATERIALS & METHODS

Fundamentally, our density surface model uses assumptions about the spatial relationships among animals in a particular geographic area to extrapolate from what is known about the number and distribution of animals sighted on transects during a survey to an estimate of the total number of animals that were truly present in the geographic area during the survey period. We estimate abundance independently for a specified point or period in time and do not explicitly model changes in abundance over time. A density surface represents the estimated density (number of individuals per unit area) of animals in each cell of a grid. To estimate total abundance during a given survey period, we integrate across the density surface, which involves multiplying each cell's estimate of animal density by its geographic area and then summing cellwise abundances across all cells in the study area.

The analytical methods that we present below may be used for a wide range of datasets and taxa. To understand the flexibility in the methods and critical elements that were included to accommodate the *Eastern Bering Sea beluga case study*, we note that three independent datasets were used in the case study: 1) aerial line-transect marine mammal observer (i.e., "aerial observer") data from the eastern Bering Sea in 2017 and 2022 (Supplement 2, hereafter "S2") were used to estimate a multiple covariates distance-sampling (MCDS) detection function (S4) and to construct the DSMs; 2) aerial imagery collected in the eastern Chukchi and western Beaufort seas during July through October in 2018 and 2019 (S3) were used to estimate the probability of detecting a beluga group on the transect line (Buckland et al., 2015; Laake and Borchers, 2004; S4); and 3) VHF telemetry data from Bristol Bay, Alaska, in June 1983, and Cunningham Inlet, Somerset Island, Canada, in July 1988 (Frost et al., 1985; Frost and Lowry, 1995) were used to estimate availability probability (S4).

Unless otherwise stated, the following text uses unbolded symbols to denote scalars, lower case bolded symbols to denote vectors, and upper case bolded symbols to denote matrices. See S1 for a *Glossary of notation and abbreviations*.

Design-based estimator

Although our primary focus is on developing density surface models for EBS belugas, previous abundance estimates for this stock were generated using a design-based estimator. A basic Horvitz-Thompson-like

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

$$\hat{D} = \frac{1}{a} \sum_{j=1}^{n_g} \frac{S_j}{\hat{p}(\mathbf{z}_j; \hat{\theta})} \tag{1}$$

where

- n_g : number of groups detected;
- S_j : size of group indexed by j;
- a : area searched during line-transect survey, where $a = 2Lw$, L is the total length of transects surveyed, and w is the width of the strip searched on one side of the aircraft;
- $\hat{p}(\mathbf{z}_j; \hat{\theta})$: model-based estimate of the overall probability that an observer detects group j, given covariates $\mathbf{z}_j$ that affect detectability. This term accounts for all sources of perception and availability bias (Marsh and Sinclair, 1989; S4);
- $\hat{\theta}$: parameter estimates required to estimate detection probabilities.

To derive an estimate of the total number of animals in the study area ($\hat{N}$), we multiply the total study area size, A, by the density estimate from Eq. 1:

$$\hat{N} = A\hat{D} \tag{2}$$

Under this formulation, inference proceeds by first fitting detection function models to observed distances and other covariates to produce estimates of detection parameters (i.e., $\hat{\theta}$), before applying Eq. 1 in a second step. The abundance estimator in Eq. 2 is unbiased if certain assumptions about the survey design hold (Buckland et al., 2001; Hedley and Bravington, 2014). Hence, this $\hat{N}$ is referred to as a design-based estimator.

In addition to allowing calculation of the design-based estimator, we also use these estimates of detection probability when fitting DSMs. In the following, we shorten notation such that $p_j = \hat{p}(\mathbf{z}_j; \hat{\theta})$ for sightings and $p_i = \hat{p}(\mathbf{z}_i; \hat{\theta})$ for segments. For more information about detection probability calculations, see S4.

DSMs: Marginal likelihood

As with most DSM implementations, we construct a spatial model for counts of individuals, which in our case were summarized over 10-km transect segments (see *Eastern Bering Sea beluga case study* for further information). For all DSMs, we write a generic marginal likelihood of a parameter vector, ξ, given observed counts of individual animals, $\mathbf{c}$, and other known variables, $\mathbf{x}$, as

$$\mathscr{L}(\xi; \mathbf{c}, \mathbf{x}) = \int_{\eta} [\mathbf{c}|\xi, \eta, \mathbf{x}][\eta|\mathbf{x}, \xi] d\eta. \tag{3}$$

Here, $[\mathbf{c}|\xi, \eta, \mathbf{x}]$ is the conditional probability density function of observed counts, given parameters, random effects (η), and known covariates. The counts represent the number of animals detected on rectangular transect segments (with width $2w$, as in the design-based estimator). The component $[\eta|\mathbf{x}, \xi]$ represents the distribution of random effects. We use the integral to indicate that the random effects will be integrated out of the joint likelihood - in our case using the Laplace approximation available in TMB software (Kristensen et al., 2015). As is usual in likelihood-based inferential statistics, the likelihood is viewed as a function of the unknown parameters, ξ.

In order to derive $[\mathbf{c}|\xi, \eta, \mathbf{x}]$, we must first specify a suitable probability mass or density function. Although it is customary to specify probability mass functions for count data, initial exploration of Poisson and negative binomial distributions indicated considerable lack-of-fit when applied to our EBS beluga data set. Specifically, model diagnostic plots examining the relationship between the mean and variance in the residuals compared to the theoretical distribution (Ver Hoef and Boveng, 2007), and quantile residuals computed using a probability integral transform (PIT; Dunn and Smyth, 1996) and visualized using the R package DHARMa (Hartig, 2022), showed that Tweedie distributions (Jørgensen, 1987; Dunn and Smyth, 2005; Kendal, 2004) provided a better fit to the data. Therefore, we adopted a parameterization based on the Tweedie distribution.

The Tweedie distribution provides increased flexibility compared to the Poisson and negative binomial distributions, allowing a diversity of shapes and accommodating zero-inflation. It is a specific case of an exponential dispersion model, with mean μ, and variance $V(\mu) = \phi\mu^{\rho}$ (Dunn and Smyth, 2005). We specifically set the range of ρ to be on $1 < \rho < 2$, a parameterization variously known as "compound Poisson," "compound gamma," or "Poisson-gamma" (Dunn and Smyth, 2005; Kendal, 2004). This distribution has support on the non-negative real line, although authors often use this distribution for non-negative integers (e.g., counts; Kendal, 2002; Miller et al., 2013; Sigourney et al., 2020), which is our approach in this paper. Kendal (2002) and Kendal (2004) discuss the relationship between the Tweedie distribution and Taylor's power law in ecology, which explains clustered spatial distributions as manifestations of power function relationships between the variance and mean number of organisms in an area (Taylor, 1961).

For $1 < \rho < 2$, the Tweedie distribution does not have a closed form, but can be evaluated numerically (e.g., using the 'dtweedie' function in the TMB library). Therefore, we symbolically write

$$[\mathbf{c}|\xi,\eta,\mathbf{x}] \equiv \prod_i \text{Tweedie}(c_i;\mu_i,\phi,\rho). \tag{4}$$

That is, the joint likelihood of observed counts on transect segments indexed by i is a product of conditionally independent univariate Tweedie density functions, with mean μ_i and constant dispersion and power parameters, ϕ and ρ. The mean, μ_i, is a function of fixed and random effects, and "known" detection probability and survey coverage offsets, such that

$$\mu = \exp(\beta_0 + \delta + \log(\mathbf{a}) + \log(\mathbf{p})). \tag{5}$$

Here, β_0 represents an intercept parameter (no other fixed effects were included in our models), δ is a vector of 'realized' random effects for transect counts, $\mathbf{a}$ is a vector of the area surveyed for each transect segment ($a_i = 2L_i w$), and $\mathbf{p}$ is a vector of the overall detection probability (including both availability and perception bias corrections) for each segment (p_i, see S4 Eq. 6).

Note that this parameterization requires that any covariates used to estimate detectability relate only to the transect segment; observation-specific covariates (e.g., color, group size) cannot be used in this parameterization. See Miller et al. (2013) for an alternative parameterization that allows observation-specific covariates by specifying the response variable to be an estimate of bias-corrected abundance.

The actual dimension of random effects often differs from the number of transect segments. Specifically, we model $\delta = \mathbf{A}\eta$, where the matrix $\mathbf{A}$ has dimension (n_i, n_η), with n_η denoting the true number of random effects, and n_i the number of transect segments. Next, we elaborate on the random effects specifications.

Random effects specifications

We have yet to describe $[\eta|\mathbf{x},\xi]$ in Eq. 3. This component defines the specification of spatially autocorrelated random effects. For a given survey period (year) the data likelihood (Eq. 4) is the same for all the models that we considered, so the random effects specification is the only difference among the DSMs we developed.

For all models, random effects were assumed to be drawn from a multivariate normal distribuion with mean zero, and a spatially patterned covariance matrix, Σ:

$$\eta \sim \text{Multivariate normal}(\mathbf{0},\Sigma).$$

Spatial autocorrelation is imparted by constraints on the $(n_\eta \times n_\eta)$ Σ matrix. In practice, we chose to work with a precision matrix $\mathbf{Q} = \Sigma^{-1}$, which was often sparse, enabling greater computational efficiency.

We employed two related, but conceptually different types of models to specify $\mathbf{Q}$: stochastic partial differential equations (SPDEs) to approximate Matérn geostatistical models (Lindgren et al., 2011), and basis-penalty smoothers. The latter are commonly used in generalized additive models (e.g., Wood, 2006), where smooth terms are often viewed as penalized fixed effects. However, it is also possible to conceptualize smooths as mean-zero random effects, with an associated precision matrix (Miller et al.,

2020), which is the approach we used in this paper. Further details on how $\mathbf{Q}$ is specified for individual models is provided below.

We conducted all analyses in the R programming environment (R Core Team 2023), using the TMB package (Kristensen et al. 2015) to formulate marginal log-likelihoods and generate DSM parameter estimates, the mgcv package (Wood 2015) to set up spatial spline bases, and R-INLA (Rue et al. 2009) to create Delaunay triangulation meshes for SPDE models. All data and code used in this paper have been uploaded to github at URL, and will be publicly archived upon manuscript acceptance.

SPDE models

SPDE Matérn model We used an SPDE approximation to model the precision matrix associated with the Matérn covariance function (Lindgren et al., 2011). This required establishing a set of n_η vertices at locations $\mathbf{s} \subset \mathbb{R}^2$ (often termed "knots"). The knot locations $\mathbf{s}$ are each associated with one of the n_η random effects. We used the function "inla.mesh.2d" from R-INLA to specify knot locations, creating a triangular mesh (i.e., spatial basis) that allows animal density to be predicted at any location on the mesh.

A particular mesh is defined by a number of characteristics, including the spatial domain (i.e., geographic boundary), location of knots, and maximum and minimum distances between knots. Additionally, boundary conditions must be imposed to create SPDE solutions on bounded domains (Lindgren and Rue, 2015). R-INLA uses Neumann conditions, which results in variance inflation by a factor of two along straight boundaries and a factor of four near right-angled corners. At a distance equal to the geostatistical range (i.e., the distance at which the correlation between two points approaches zero), the boundary effect is negligible. To eliminate boundary effects in the area of interest, Lindgren and Rue (2015) recommend extending the outer boundary of the spatial domain by a distance at least equal to the range. Knot density can be reduced in the outer buffer area to minimize the additional computational burden of the knots located far from the data.

Different meshes could result in different estimates of animal density. There are no strict rules on how to create a mesh for a particular dataset. Therefore, we used preliminary analyses with a variety of meshes to guide our decisions on which mesh was best suited to each year's data. For example, a certain mesh might be a poor fit to the data because the numerical optimization of the geostatistical range parameter can fail if spatial autocorrelation occurs at a much finer scale than the minimum distance between knots. In general, we followed the suggestions of Belmont (2022) to create the meshes. Initial knot locations were placed at the transect segment midpoints. The maximum distance between knots in the buffer area was twice that in the aerial survey boundaries. The minimum distance between knots equaled $\frac{1}{5}$ the maximum distance between knots. The extension radius used to set the overall boundary of the spatial domain (hence, the width of the buffer area) was approximately 35% of the diameter of the aerial survey study area.

Interpolations of random effects between knot locations and data locations are made with a bilinear interpolation matrix ($\mathbf{A}$), where the data location is taken to be the centroid of each 10-km transect segment. We used the R-INLA function "inla.spde.make.A" to create interpolation matrices. Interpolation matrices are completely determined by the underlying mesh and the data locations, and are nonzero for only three elements of each row (corresponding to the three triangular vertices that surround a given point).

To define a precision matrix for the Matérn covariance model at knot locations, we used the function "inla.spde2.matern" in R-INLA (Rue et al., 2009; Lindgren et al., 2011), which generates three structure matrices, M_0, M_1, and M_2. The precision matrix is then specified using these three matrices, together with two unknown parameters, τ and κ:

$$\mathbf{Q} = \tau^2(\kappa^4\mathbf{M}_0 + 2\kappa^2\mathbf{M}_1 + \mathbf{M}_2) \tag{6}$$

Here, τ can be be interpreted as a precision parameter and κ as an inverse geostatistical range parameter. Eq. 6 results from applying finite-element methods to approximate a stochastic partial differential equation representing diffusion. Using notation from Lindgren et al. (2011), we see that a diffusive SPDE for second-order adjacency results in a precision with the form:

$$\mathbf{Q} = \mathbf{K}\mathbf{C}^{-1}\mathbf{K} \tag{7}$$

where $\mathbf{K} = \kappa^2(\mathbf{C}+\mathbf{G})$ and where $\mathbf{C}$ and $\mathbf{G}$ are both sparse matrices. Replacing $\mathbf{C}$ with a diagonal matrix $\tilde{\mathbf{C}}$ (so that $\tilde{\mathbf{C}}^{-1}$ remains sparse), plugging in $\mathbf{K}$, and simplifying, we then obtain Eq. 6 where $\mathbf{M_0} = \tilde{\mathbf{C}}$, $\mathbf{M_1} = \mathbf{G}$, and $\mathbf{M_2} = \mathbf{G}\tilde{\mathbf{C}}^{-1}\mathbf{G}$. However, we retain M-notation in Eq. 6 to maintain consistency with terminology that is common when using the R-INLA package.

SPDE Matérn model with barriers The SPDE model defined above approximates a stationary, isotropic Matérn covariance function. Conditional on τ and κ, the only variable affecting spatial autocorrelation is the distance between knots. However, in areas with complex coastlines (such as islands, bays, peninsulas, and points), it is plausible that spatial connectivity for the distribution and density of marine animals would be interrupted by land barriers, making points that are close together in the contiguous ocean more "alike" than two points separated at the same distance on the opposite sides of a land barrier. Therefore, we implemented an alternative SPDE Matérn model that accounts for land-based barriers. Specifically, we followed the approach outlined by Bakka et al. (2019), where locations that occur on land are assigned a small, fixed effective range value, and the range parameter for locations at sea is estimated during model fitting. To implement the $\mathbf{Q}$ matrix for this model in TMB, we used the code template at `https://github.com/skaug/tmb-case-studies/tree/master/spdeBarrier`. This essentially specifies a high value for decorrelation rate κ for knots over land, to ensure that correlations between locations in water are calculated from the set of paths over water.

Basis-penalty smoother models

We considered several types of basis-penalty smoothers as alternatives for specifying spatial random effects. In each case, we used mgcv to construct spline bases and appropriate penalization matrices, then passed these into TMB when formulating our marginal log-likelihood. We implemented three types of bivariate smoothing splines: isotropic, thin plate regression splines (tprs) with shrinkage (Wood, 2003); anisotropic tensor product splines (Wood, 2017) comprising tprs with shrinkage; and isotropic soap film smoothing splines (Wood et al., 2008). The first two types of splines treat spatial correlation as depending on distance only (isotropic tprs), or distance and direction (tensor product splines). The soap film smoother allows spatial correlation to be interrupted when there are barriers, such as land, between suitable habitat. We reasoned this would be a desirable property given the complex coastline in our study area, which included multiple peninsulas and estuaries (Figure 1).

Bivariate and isotropic thin plate regression spline For the bivariate and isotropic thin plate regression spline, we used the gam() function in the mgcv R package to construct a bivariate "ts" spline basis of easting and northing for observed data ($\mathbf{A}$, typically referred to as a design matrix in this context), an interpolation matrix for predictions ($\mathbf{A}^{pred}$), and a penalization matrix, $\mathbf{S}$. We then set $\mathbf{Q} = \lambda\mathbf{S}$ in our TMB optimization, where the smoothing parameter λ was treated as an estimated parameter. This procedure follows the example by H. Skaug at `https://github.com/skaug/tmb-case-studies/tree/master/pSplines`.

Tensor product smoother The tensor product smoother produces an anisotropic spline basis, allowing different correlations on the dimensions corresponding to easting and northing in our analysis. For the tensor product smoother, we again used a gam() function in the mgcv R package to construct a "ts" spline basis for observed data ($\mathbf{A}$) and an interpolation matrix for predictions ($\mathbf{A}^{pred}$). In this case, mgcv produces two penalization matrices, $\mathbf{S_1}$ and $\mathbf{S_2}$ (one for easting and one for northing). Following code from D. Miller (`https://github.com/dill/mgcvminusminus`), we set $\mathbf{Q} = \lambda_1\mathbf{S}_1 + \lambda_2\mathbf{S}_2$, where λ_1 and λ_2 are treated as estimated parameters.

Soap film smoother The soap film smoother (Wood et al., 2008; Miller and Wood, 2014) is another approach to constructing a smooth surface over space where correlation does not persist over boundaries (e.g., penninsulas). To produce $\mathbf{A}$, $\mathbf{A}^{pred}$, and penalization matrices, we again used mgcv. In particular, we supplied the gam() function with a data frame delineating study area boundaries. Like the tensor product smoother, the soap film smoother option in mgcv produces two penalization matrices, $\mathbf{S}_1$ and $\mathbf{S}_2$, this time associated with boundaries and internal space, respectively. However, we constructed the precision matrix in the same manner (i.e., $\mathbf{Q} = \lambda_1\mathbf{S}_1 + \lambda_2\mathbf{S}_2$, where λ_1 and λ_2 are treated as estimated parameters).

Prediction

For each model and year of analysis, we used Monte Carlo integration to generate abundance predictions of the number of belugas in each hexagonal grid cell h in our study area (see *Eastern Bering Sea beluga*

case study for how these were defined). Specifically, we calculated

$$\hat{N}_h = \exp(\hat{\beta}_0 + \hat{\delta}_h + \log(a_h)). \tag{8}$$

Note that a_h gives the area of ocean in hexagon h (i.e., omitting land). There is no offset for detection probability because we are interested in all belugas, not just those that are detectable and detected. The vector of "realized" random effects $\hat{\delta}$ are calculated as $\hat{\delta} = \mathbf{A}^{pred}\hat{\eta}$ where $\hat{\eta}$ is the value of η that maximizes the joint likelihood conditional upon the MLE $\hat{\xi}$ for fixed effects (termed the empirical Bayes estimator for η). This predictor for $\hat{N}_h$ is called the "plug-in estimator" because it plugs in the empirical Bayes estimator as if it were fixed. For SPDE models, the (n_h, n_η) interpolation matrix, $\mathbf{A}^{pred}$ was constructed using the "inla.spde.make.A" function in R-INLA, using the centroid of each hexagon as the prediction location. For basis-penalty smoothers, we obtained $\mathbf{A}^{pred}$ using mgcv's "predict()" function with "type=lpmatrix," again using the centroids of each hexagon as prediction locations. An estimate of total abundance is then calculated as $\hat{N} = \sum_h \hat{N}_h$.

Model evaluation and final candidate model selection

To evaluate DSMs and select the final candidate DSMs for the ensemble model, we advocate using several criteria, including: examining PIT residuals via the DHARMa package (Hartig, 2022); extrapolation metrics (defined below); and visual examination of maps showing the DSM predictions overlaid with the sightings used to build the models. We provide a detailed example in the *Eastern Bering Sea beluga case study*.

To identify models whose predictions might be unreliable due to extrapolation bias, we considered two types of ad hoc metrics. First, for each combination of model and cell (i.e., hexagon, h), we computed the following ratio:

$$\frac{\lambda_{m,h}}{\lambda_{m,max}} \tag{9}$$

Where $\lambda_{m,h}$ is the predicted abundance from model m for unsampled location h, and $\lambda_{m,max}$ is the maximum predicted abundance across all sampled cells. Second, for each model we counted the number of unsampled cells (i.e., cells that did not have line-transect survey effort) with predicted abundance exceeding the maximum predicted abundance in sampled cells. These procedures are motivated by a generalized version of Cook's independent variable hull (Cook, 1979; Conn et al., 2015).

Uncertainty estimation

Correcting for detransformation bias

Random effects η are treated as random variables (and marginalized across) during maximum likelihood estimation, but are then treated as if they were fixed at the modes of their distributions (conditional on the MLEs for the fixed effects) by the plug-in estimator. However, η will generally have substantial variance and skewness, and this will cause the plug-in estimator to be a poor estimator for the expectation of N when integrating across the distribution for random effects.

To better estimate the expectation for N, we employed the epsilon bias-correction procedure described by Thorson and Kristensen (2016) and implemented in TMB to obtain estimates and standard errors. This epsilon method corrects for both the nonlinearity of the transformation (i.e., exponentiation in Eq. 8) and the variance and skewness of random effects. Thorson and Kristensen (2024, Chap. 6) shows a closed-form calculation for the epsilon method in a simplified scenario involving a single (scalar-valued) random effect, and confirms that it provides very close to the known expectation when transforming skewed random variables with a number of different nonlinear functions.

Variance estimation

We relied on the law of total variance to construct an unconditional variance estimator for each DSM that includes uncertainty from the MCDS contribution to detection probability, $\mathbf{p}_g$ (S5). Specifically, we calculated

$$\hat{\mathrm{Var}}(\hat{N}) = \mathbb{E}(\hat{\mathrm{Var}}(\hat{N}|\tilde{\mathbf{p}}_g)) + \hat{\mathrm{Var}}(\mathbb{E}(\hat{N}|\tilde{\mathbf{p}}_g)). \tag{10}$$

The first part of Eq. 10, $\mathbb{E}(\hat{\text{Var}}(\hat{N}|\tilde{\mathbf{p}}_g))$, is the expected variance of the abundance estimator given a particular realization of detection probability, $\tilde{\mathbf{p}}_g$. We approximated this component with $\hat{\text{Var}}(\hat{N}|\hat{\mathbf{p}}_g)$, which is the variance of the abundance estimator conditioned on the MLEs for detection probability from the MCDS analysis. This variance estimate is produced by the TMB software, using the algorithm detailed below (see *Conditional variance of abundance estimator*).

The second component of Eq. 10, $\hat{\text{Var}}(\mathbb{E}(\hat{N}|\tilde{\mathbf{p}}_g))$, in effect gives the variance of the mean, representing how estimates of abundance vary depending on the values of $\mathbf{p}_g$ that are sampled. To approximate $\hat{\text{Var}}(\mathbb{E}(\hat{N}|\tilde{\mathbf{p}}_g))$, we used the following bootstrap procedure (see S5 for pseudocode):

1. For $k \in 1, 2, \ldots, K$, sample $\mathbf{p_g}^{(k)} \sim f(\mathbf{p}_g)$, where $f(\mathbf{p}_g)$ is the joint predictive distribution of detection probabilities from the MCDS detection function analysis. In practice, each sample $\mathbf{p}_g^{(k)}$ was obtained by assuming that the parameters of the detection function had a multivariate normal distribution on the logit scale.

2. For each k, fit a TMB DSM to the beluga data, treating $\mathbf{p}_g = \mathbf{p}_g^{(k)}$ as a fixed value, and record the abundance estimate, $\hat{N}^{(k)}$.

3. Approximate $\hat{\text{Var}}(\mathbb{E}(\hat{N}|\tilde{\mathbf{p}}_g))$ as $K^{-1}\sum_k(\hat{N}^{(k)} - \bar{N})^2$, where $\bar{N}$ is the mean abundance estimate from all K bootstrap iterations.

Following application of this procedure, to generate estimates of total uncertainty in the abundance estimate from each individual DSM (i.e., $CV_{tot}(\hat{N}_m)$ in Eq. 5 of S5), the delta method (Dorfman 1938) could be used to incorporate the uncertainty due to independent estimates of transect detection probability or availability probability.

Conditional variance of abundance estimator We compute an estimator for the variance of Eq. 8 that accounts for uncertainty in both fixed and random effects. We call this a conditional estimator because we are specifically conditioning on a fixed vector of detection probabilities. Although different estimators are available, TMB software uses the estimator from Kass and Steffey (1989). This involves calculating the joint precision $\mathbf{Q}_{joint}$ for fixed and random effects:

$$\mathbf{Q}_{joint} = \begin{pmatrix} \mathbf{H}_1 & -\mathbf{H_1}\nabla \\ -\nabla^{\mathbf{t}}\mathbf{H_1} & \nabla^{\mathbf{t}}\mathbf{H_1}\nabla + \mathbf{H_2} \end{pmatrix} \tag{11}$$

where $\mathbf{H}_2$ is the matrix of second derivatives for $\log\mathscr{L}(\xi|\mathbf{c},\mathbf{x})$ (the "outer Hessian matrix"), $\mathbf{H}_1$ is the matrix of second derivatives for $\log([\mathbf{c}|\xi,\eta,\mathbf{x}][\eta|\mathbf{x},\xi])$ conditional upon the MLE for fixed effects ξ (the "inner Hessian matrix"), and ∇ is the matrix of gradients of predicted random effects with respect to fixed effects (the "outer Jacobian matrix").

We then compute the variance for derived quantity $\hat{N}$ from this joint precision. We specifically calculate the gradient $\mathbf{J}$ of $\hat{N}$ with respect to the vector of fixed and random effects. We then compute

$$\hat{\text{Var}}(\hat{N}|\hat{\mathbf{p}}_g) = \mathbf{J}\mathbf{Q}_{joint}^{-1}\mathbf{J}^t \tag{12}$$

Ensemble model

Fitting multiple DSMs to sightings raises the question of which model, or collection of models, should be used to generate a final abundance estimate and density surface. The question is particularly important when different models produce markedly different estimates of abundance. We chose to base ultimate inference on an ensemble (Araújo and New, 2007), whereby estimates from different models are averaged to produce a final estimate. Specifically, we compute

$$\hat{N}_{ens} = \sum_m w_m \hat{N}_m \tag{13}$$

where $\hat{N}_m$ is the MLE of abundance from TMB for each model m. The model weight is w_m, where $\sum_m w_m = 1.0$. The advantage of averaging models is that there is often a reduction in prediction error (Burnham and Anderson, 2002; Dormann et al., 2018).

There are different approaches for setting the model weights (Dormann et al., 2018). For instance, a common approach is to use Akaike's information criterion (AIC) associated with fitted models to calculate weights (Burnham and Anderson, 2002). However, calculation of AIC weights relies on the complexity of a model, often computed as the effective degrees of freedom from a generalization of the hat-matrix, and this is difficult to compute in a hierarchical model using maximum-likelihood methods. Instead, we used equal model weights, which have been shown to perform well in prediction of species distributions (Dormann et al., 2018). This procedure has the added advantage that a single model with an extremely high or low abundance estimate will not dominate inference.

The variance of model-averaged predictions was calculated using the standard unconditional variance estimator (i.e., Burnham and Anderson, 2002):

$$\hat{Var}(\hat{N}_{ens}) = \left[\sum_m w_m \sqrt{Var(\hat{N}_m) + (\hat{N}_m - \hat{N}_{ens})^2}\right]^2 \tag{14}$$

Because $Var(\hat{N}_m)$ did not include uncertainty from the estimate of transect detection probability in the beluga case study (i.e., $\hat{p}_{MR}(0, \mathbf{z}_j; \hat{\theta}_{MR})$ in S4), we apply the delta method to add this component of uncertainty to $\hat{Var}(\hat{N}_{ens})$, resulting in $\hat{Var}_{tot}(\hat{N}_{ens})$ (see Eq. 4 of S5).

Eastern Bering Sea beluga case study

The data collection methods and sighting and effort summaries are presented in S2 for the aerial line-transect surveys and S3 for the aerial imagery. The analytical methods used to estimate detection probabilities are presented in S4. See Frost et al. (1985) and Frost and Lowry (1995) for details about the VHF telemetry data and analyses. There were no estimates of uncertainty for availability probability (Ferguson et al., 2023); therefore, this parameter was included as a known constant in the offset for the DSM (Eq. 5).

To derive spatially-explicit estimates of EBS beluga abundance, we constructed density surface models separately for each year, 2017 and 2022. DSMs were constructed using aerial line-transect sighting and effort summaries for 10-km segments of transect effort. This segment length is approximately the distance between adjacent transects (9.3 km). The segments were created by sequentially slicing transect effort conducted in Beaufort Sea State ≤ 4, beginning with the start of each transect. End segments < 10 km were added to adjacent segments so that all segments used in the analysis were ≥ 10 km. Predictions from the DSM were based on a hexagonal grid 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: -164.0^o; latitude of origin: 63.5^o; standard parallels: 62.5^o, 64.5^o; WGS84 datum; linear unit: kilometer), and this projection was used when calculating cell areas and distances for the spatial correlation functions or splines.

The DSMs required segment-specific estimates of detection probability, p_i (Eq. 5). The best-fitting MCDS detection function for EBS belugas included covariates for Beaufort Sea State (integer-valued) and turbidity (binary) (S4). To build the DSMs, effort data for these variables were summarized by segment. The segment-specific Beaufort Sea State variable was calculated as the average value of integer-valued Beaufort Sea State for all records that were located on the segment; all records were weighted equally. The segment-specific turbidity variable was calculated by assigning the binary turbidity variable an integer value (no = 0; yes = 1), computing the average of the integer-valued turbidity values for all records located on the segment, and rounding the result. For example, if segment i comprised three data records with turbidity "yes", "yes", and "no", the average of their integer-valued analogs would be 1 + 1 + 0 = 0.67, which rounds to 1, so the segment would be designated as turbidity = "yes".

For 2017 and 2022, 13 and 12 DSMs, respectively, were constructed and examined. Overviews of key aspects of each model are provided in Tables S6.1 and S6.2. We followed the guidance provided in the helpfile for the mgcv function "gam.check" to determine whether the basis dimensions for the basis-penalty smooths were sufficient. In particular, we used "gam.check" to compare the estimated degrees of freedom (EDF) to the maximum possible EDF, and examined the value of the p-value associated with the reported k-index. For all basis-penalty smooths, the EDF was much smaller than the maximum possible EDF. The p-values (although approximate) were considerably larger than 0.1 for all models except the isotropic bivariate thin plate regression spline, whose simulated p-value was 0.085. Based on these diagnostics, the basis dimensions for the basis-penalty DSMs were adequate.

Identical DSMs were fit in mgcv and TMB, with the exception of the barrier SPDE models, for which mgcv functions defining this type of basis were not available, so they were constructed only in TMB. We constructed identical models in both software platforms for two reasons: 1) to apply the methods presented in Miller et al. (2020) to the EBS beluga data and confirm that nearly identical results could be derived using mgcv and TMB; and 2) to evaluate whether the existing "ds_varprop" function in the dsm package (Bravington et al., 2021; Miller et al., 2022) would be an alternative to Eq. 10 and the methods presented in S5 for propagating uncertainty from the MCDS detection function model into the overall estimate of uncertainty in the abundance estimate. However, as of dsm version 2.3.3, the "dsm_varprop" function needs to be modified to propagate errors through SPDE models (pers. comm. M. Ferguson with D. Miller 7 September 2023).

For comparison with the 2017 EBS beluga DSM results, we examined the post-stratified design-based abundance estimator of Ferguson et al. (2023). Their variance estimator had three components: 1) variation from uncertainty in estimating the MCDS parameters; 2) variation from uncertainty in estimating transect detection probability; and 3) variation in abundance due to random sample selection. The first and third components were estimated by the "dht" function from the R package mrds (Laake et al., 2023), using the delta method to compute the MCDS variance and the default encounter rate estimator $\hat{N}/L$ for the random sample variability. Ferguson et al. (2023) used the delta method to incorporate uncertainty from the estimate of transect detection probability into the overall estimate of uncertainty in the abundance estimator.

The detection function model used in the design-based estimator for 2017 was based on only a single year of data, whereas the detection function model used in the model-based estimators for 2017 and 2022 was based on data from both years pooled (S4). However, the CV of the former detection function model was 0.043 and the CV of the latter was 0.037, only trivially smaller; therefore, we do not believe that this difference in detection function models affected our overall comparison of the precision in the different abundance estimators.

To derive an analogous design-based estimator of abundance for 2022, we used the methods of Ferguson et al. (2023) with an MCDS detection function model based on the pooled data from 2017 and 2022.

RESULTS

Here, we focus on results of the EBS beluga case study. None of the 13 candidate models for 2017 exhibited signs of extrapolation bias based our extrapolation diagnostics (Table S6.1). For 2022, four of the 12 candidate models had at least one cell with extrapolation ratios >1.0 (Eq. 9). Of those four models, the total number of cells per model with outliers ranged from one to three, out of a total of 554 cells in the prediction grid (Table S6.2). Based on these metrics, we did not find evidence for concern about extrapolation bias.

The number of models per year that were selected for inclusion in the ensemble model average was not chosen a priori. Rather, we examined PIT residuals via the R package DHARMa (Hartig, 2022; S6), extrapolation metrics, and visual inspection of maps of $\hat{N}$ predictions and sightings and effort to narrow the field to four candidate models per year (Table 1; S6). For both years, the SPDE Matérn models with maximum edge length less than 60 km exhibited a number problems in the DHARMa residual analyses, including significant dispersion tests, quantile deviations, and significant combined adjusted quantile tests. The SPDE Matérn model with barriers was eliminated from the ensemble model for 2017 because the residual analyses from the DHARMa package showed quantile deviations and the combined adjusted quantile tests were significant. The tensor product smoother was eliminated from the ensemble model for 2022 because it predicted that there were 51,645 belugas in the study area; this represented an extreme outlier and was consdiered to be biologically implausible.

The candidate DSMs included in the ensemble for 2017 were the SPDE Matérn, soap film smoother, tensor product smoother, and bivariate isotropic thin plate regression spline (Figure 2). The candidate DSMs included in the ensemble for 2022 were the SPDE Matérn with and without barriers, soap film smoother, and bivariate isotropic thin plate regression spline (Figure 3). The number of random effects used to fit each candidate model in the ensembles for 2017 and 2022 are shown in Table 2. The percent deviance explained is also shown in Table 2, and it was computed as $100 * (1 - \frac{R1}{R0})$, where $R1$ is the sum of squared deviance residuals for model m and $R0$ is the sum of squared deviance residuals for the null (intercept-only) model. The candidate DSMs for 2017 explained between 45.7% (bivariate and isotropic

thin plate regression spline) and 56.7% (tensor product smoother) of the deviance. Among the 2022 candidate DSMs, the percent deviance explained ranged from 62.1% (bivariate and isotropic thin plate regression spline) to 83.4% (SPDE Matérn).

The maximum predicted abundance per cell was lower in 2017 than 2022. In 2022, the year with less transect effort and fewer beluga sightings, the SPDE models tended to result in relatively confined clusters of high abundance, whereas the predictions from the basis-penalty smoothers were relatively more diffuse (Figure 3). To investigate these differences among the 2022 DSMs further, we conducted pairwise comparisons of predicted beluga abundance for all pairs of DSMs selected for the ensemble model. Specifically, for each cell h, we computed scaled differences in predicted abundance between models m_1 and m_2 as $(\hat{N}_{m_1,h} - \hat{N}_{m_1,h})/max(abs(\hat{N}_{m_1,h} - \hat{N}_{m_1,h}))$. The resulting values are constrained to [-1,1]. Two points are worth highlighting from these comparisons (Figure 4). First, predicted abundance was largely consistent between models throughout the overwhelming majority of the study area. The largest discrepancies in model predictions were in areas of high beluga sighting density (Figure S2.2): basis-penalty smoothers tended to estimate higher abundance in a small cluster of cells near Scammon Bay (where some of the largest beluga groups were detected in 2022) and their high-abundance hotspots were more dispersed than the predictions from the SPDE models around Stuart Island.

The area-integrated estimates of EBS beluga abundance (with and without detransformation bias correction via the epsilon method) are shown in Table 1 for the following: each of the candidate DSMs included in the 2017 and 2022 ensemble models; the ensemble models; and the conventional (post-stratified design-based) estimator. For the candidate DSMs in 2017, epsilon bias correction resulted in abundance estimates that were 7% to 13% larger than the plug-in estimators. For the candidate DSMs in 2022, the corresponding increase was higher, ranging between 18-33%. In 2017, the epsilon-corrected area-integrated abundance estimates ranged from 11,242 to 11,962 (CV = 0.11 to 0.12). In 2022, the analogous range was wider, from 12,593 to 21,508 (CV = 0.18 to 0.29). Nevertheless, within a survey year, the 95% lognormal confidence intervals for abundance overlapped across all candidate DSMs. The ensemble spatial models estimated that there were 11,597 belugas in 2017 (CV = 0.12) and 17,197 belugas in 2022 (CV = 0.33). For comparison, the conventional, design-based models estimated that there were 12,269 belugas in 2017 (CV=0.12; Ferguson et al., 2023) and 20,635 belugas in 2022 (CV = 0.31).

The abundance estimates for the full 2022 study area that were derived from each of the candidate DSMs, the ensemble model, and the conventional estimator were all higher than the corresponding estimates for 2017. To investigate how much of this larger abundance in 2022 was due to the larger study area, we used each individual candidate DSM from 2022 to compute area-integrated abundances for the area corresponding to the geographic strata from the 2017 analysis. The results of this investigation were split between modeling paradigms: the SPDE DSMs estimated that there were fewer belugas in that subarea during 2022 compared to 2017, whereas the basis-penalty DSMs estimated that there were more belugas within that subarea in 2022 (Table 1).

The Tweedie parameter estimates from each of the candidate DSMs in 2022 provide some insight into the question of whether there were spatially-restricted clusters of high abundance in a few locations or high abundances over a broader spatial area. Specifically, estimates of the dispersion and power parameters (ϕ and ρ, respectively) were higher for the basis-penalty DSMs compared to the SPDE DSMs (Table 3). This suggests that the spatial random effects in the SPDE DSMs might have been able to better match the patchiness in the data. This is consistent with the variability in percent deviance explained among the 2022 models: the SPDE DSMs explained a higher percentage of the deviance than the basis-penalty DSMs (Table 2).

DISCUSSION

In this paper, we present detailed methods for constructing and evaluating hierarchical spatially-explicit density models to estimate abundance from line-transect survey data using two leading model frameworks, SPDE approximations to geostatistical models and basis-penalty smoothers. Critical issues that we addressed include: 1) accounting for the precision and bias of all components of the hierarchical model in the final estimate of uncertainty in abundance via a parametric bootstrap; 2) applying the epsilon bias correction factor (Thorson and Kristensen, 2016) to account for detransformation bias in the DSM; 3) implementing a thorough model evaluation and selection process that incorporated examination of PIT residuals, maps of model predictions, and extrapolation diagnostic metrics; and 4) using ensemble model averaging techniques to derive the ultimate estimates of abundance and uncertainty to account for model

selection uncertainty, which is especially important in situations for which different structurally sound models seem to produce widely ranging results.

We demonstrated our methods using Eastern Bering Sea belugas as a case study. This was a particularly informative case study because aerial line-transect surveys were conducted using identical protocols during two years in which survey effort and beluga distribution differed dramatically, 2017 and 2022. The goals of the EBS beluga analysis were to: 1) assess whether DSMs represent an improvement over conventional, post-stratified design-based estimators; 2) produce updated estimates of abundance for this stock; and 3) produce detailed maps of beluga density in the survey area during the 2017 and 2022 survey periods. For the case study, we constructed identical DSMs using two alternative methods to construct basis functions representing spatially correlated variation in population density, i.e., the SPDE approximation to the Matèrn correlation function and bivariate splines.

For EBS belugas, DSMs produced similar estimates to the conventional design-based estimator in 2017, and averaged lower but had a range that included the design-based estimator in 2022. Precision of individual DSM models was higher than for the design-based estimator, but precision of the ensemble of DSM models was equivalent to the design-based estimator. Because density surface modeling paradigm also enables estimation of higher resolution maps of species density (Figures 2,3), we view DSMs as an improvement in statistical methodology for analyzing EBS beluga data to maximize utility in management and conservation decisions.

Our ensemble estimate of abundance in 2022 ($\hat{N}_{ens}$=17,089) was larger than for 2017 ($\hat{N}_{ens}$=11,553). There are several possible reasons for this difference, including random error, actual population decrease or increase between years, emigration or immigration from the system, and increased survey area between 2017 and 2022. ABWC advocated for the larger survey area in 2022. They noted that Indigenous knowledge has confirmed that the southern extent of the EBS beluga stock's distribution during early summer extends farther south than the historical survey boundaries. We recommend the 2022 ensemble abundance estimate as the most pertinent for management at present, recognizing that there are still some unaddressed issues (e.g., no sampling of belugas in rivers; Castellote et al., 2023) that likely make it a slight underestimate.

Although we recommend the 2022 estimate for EBS beluga management, we note that the estimated precision is considerably less than for our 2017 estimate. Additionally, the CV of the 2022 ensemble abundance estimate (CV = 0.33) does not meet the precision threshold recommended by regulatory entities (CV=0.3; International Whaling Commission, 2003; National Marine Fisheries Service, 2023). Decreased precision in 2022 is likely a function of both decreased survey effort and increased beluga clustering, particularly in the southern end of the survey area that had not be covered by aerial line-transect surveys for belugas in the past. The three extremely large groups of belugas (67, 87, and 120 belugas) detected nearshore, North of Scammon Bay, in 2022 (Figure S2.2) increased overdispersion in the data (Table 3), and thus added to uncertainty about whether there are large groups in unsurveyed locations (i.e., between transects within the existing survey area boundaries). Increased survey effort, redistribution of survey effort, or some combination thereof will be needed in future surveys to reduce uncertainty in the abundance estimate.

Our study offers a number of lessons for researchers seeking to implement DSMs, whether with belugas or other species. First, it was apparent from our analysis of the 2022 data that models with different spatial basis function formulations have the potential to produce quite different abundance estimates. This is likely due to the way in which estimated abundance is interpolated (and extrapolated) into unsampled areas. The tendency for this to occur may be affected by sampling intensity (lower in 2022) and the level of overdispersion (several large group sizes in 2022). Therefore, we strongly caution against employing just one form of spatial model; instead, we recommend that investigators routinely fit models with different spatial basis functions, and consider ensemble modeling (Araújo and New, 2007) if models produce different predictions.

Our approach in this paper was to employ ensemble models with equal weighting. Alternatives, such as using an information criterion (Burnham and Anderson, 2002) to weight models, are certainly possible. However, computing the effective degrees of freedom as a measurement of model complexity can be difficult in spatial models with random effects. To our knowledge, the performance of marginal AIC (i.e., ignoring spatial random effects when counting parameters) for model weighting has not been rigorously evaluated. In our case, using marginal AIC to weight models would have placed virtually all model weight on a single model for 2022. To be conservative, we thus adopted an equal weighting strategy, which has

been shown to be reasonable in practice (Dormann et al., 2018). Alternative strategies for DSM ensemble weighting would make for useful future research.

Although we recommend spatial DSMs for EBS belugas, this is not a disavowal of general principles of survey design. Such principles (e.g., randomization and replication; Buckland et al., 2001) help to ensure that model-based estimators will be unbiased and should be regarded as good practice in transect surveys, no matter the method used to analyze the data (Hedley and Bravington, 2014). Design-based concepts in survey design (e.g. systematic random samples) are still important for the quality of inference in DSMs.

CONCLUSIONS

Density surface models (DSMs) are commonly fitted to counts obtained during line-transect surveys of marine mammal populations as an alternative to conventional design-based estimators. For EBS belugas, we found DSMs to be preferable, given the extra information one gains through maps of spatial distributions. However, when fitting DSMs, researchers need to be cognizent that different spatial basis functions can result in different estimates, particularly when animals are patchily distributed. In such cases, use of ensemble predictions are likely warranted. Further, investigators should take care to properly account for uncertainty by propagating uncertainty in detection probability into resultant estimates, and to account for possible detransformation bias.

ACKNOWLEDGMENTS

We thank the hunters and scientists of the Alaska Beluga Whale Committee who provided invaluable knowledge and support for the Eastern Bering Sea beluga aerial line-transect surveys in 2017 and 2022. The aerial survey observers included A. Brower (2017 and 2022), K. Shelden (2017 and 2022), C. Sims (2017), and M. Ferguson (2022). Aircraft support in both years was provided by Clearwater Air, Inc.; we appreciate the dedication and skills of the Clearwater pilots A. Harcombe, J. Turner, C. Wilson, S. Corbin, K. Vacendak, and D. McDonald. H.Skaug and D. Miller provided detailed, functional, and educational examples of DSMs in TMB and R that formed the basis for all of the models evaluated in this paper. We thank A. Warlick and J. Ver Hoef for comments on a previous version of this manuscript that enhanced its quality. Finally, M. Ferguson would like to thank R. Ferguson for dedicating large portions of his weekends to analyses and writing; the trails await!

REFERENCES

Adams, M., Frost, K. J., and Harwood, L. A. (1993). Alaska and Inuvialuit Beluga Whale Committee (AIBWC)—an initiative in "at home management". *Arctic*, pages 134–137.

Araújo, M. B. and New, M. (2007). Ensemble forecasting of species distributions. *Trends in ecology & evolution*, 22(1):42–47.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., and Rue, H. (2019). Non-stationary Gaussian models with physical barriers. *Spatial statistics*, 29:268–288.

Belmont, J. (2022). Building the mesh. https://rpubs.com/jafet089/886687.

Bravington, M. V., Miller, D. L., and Hedley, S. L. (2021). Variance propagation for density surface models. *Journal of Agricultural, Biological and Environmental Statistics*, 26:306–323.

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

Buckland, S. T., Rexstad, E. A., Marques, T. A., and Oedekoven, C. S. (2015). *Distance sampling: methods and applications*, volume 431. Springer.

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

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

Castellote, M., Garner, C., Okitkun, M., and Kameroff, B. (2023). *Acoustic monitoring of beluga whales in the Yukon Delta: 2023 pilot study*.

Citta, J. J., Richard, P., Lowry, L. F., O'Corry-Crowe, G., Marcoux, M., Suydam, R., Quakenbush, L. T., Hobbs, R. C., Litovka, D. I., Frost, K. J., et al. (2017). Satellite telemetry reveals population specific winter ranges of beluga whales in the Bering Sea. *Marine Mammal Science*, 33(1):236–250.

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

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

Cook, R. D. (1979). Influential observations in linear regression. *J. Amer. Statist. Assoc.*, 74:169–174.

Dormann, C. F., Calabrese, J. M., Guillera-Arroita, G., Matechou, E., Bahn, V., Bartoń, K., Beale, C. M., Ciuti, S., Elith, J., Gerstner, K., et al. (2018). Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. *Ecological Monographs*, 88(4):485–504.

Dunn, P. K. and Smyth, G. K. (1996). Randomized quantile residuals. *J Comput Graph Stat*, 5(3):236–244.

Dunn, P. K. and Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15:267–280.

Ferguson, M., Brower, A., Willoughby, A., and Sims, C. (2023). Distribution and estimated abundance of eastern Bering Sea belugas from aerial line-transect surveys in 2017. *NOAA technical memorandum NMFS-AFSC*, 471.

Frost, K. and Lowry, L. (1995). Radio tag based correction factors for use in beluga whale population estimates. Working paper for Alaska Beluga Whale Committee Scientific Workshop, Anchorage, AK, 5-7 April 1995.

Frost, K. J., Gray, T., Goodwin Sr, W., Schaeffer, R., and Suydam, R. (2021). Alaska Beluga Whale Committee—a unique model of co-management. *Polar Research*, 40.

Frost, K. J., Lowry, L. F., and Nelson, R. R. (1985). Radiotagging studies of belukha whales (Delphinapterus leucas) in Bristol Bay, Alaska. *Marine Mammal Science*, 1(3):191–202.

Hartig, F. (2022). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*. R package version 0.4.6.

Hedley, S. and Bravington, M. (2014). Comments on design-based and model-based abundance estimates for the rmp and other contexts. *Paper SC/65b/RMP11 presented to the 65b IWC Scientific Committee*, 9.

Hedley, S. and Buckland, S. (2004). Spatial models for line transect sampling. *J. Agric. Biol. Environ. Stat.*, 9:181–199.

Huntington, H. P. and Communities of Buckland, Elim, Koyuk, Point Lay, and Shaktoolik (1999). Traditional knowledge of the ecology of beluga whales (Delphinapterus lencas) in the eastern Chukchi and northern Bering Seas, Alaska. *Arctic*, 52:49–61.

International Whaling Commission (2003). Report of the scientific committee. *J Cetacean Res Manage (Suppl)*, 5:1–92.

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

Jørgensen, B. (1987). Exponential dispersion models. *J. R. Stat. Soc. B*, 49:127–162.

Kass, R. E. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, 84(407):717–726.

Kendal, W. S. (2002). Spatial aggregation of the colorado potato beetle described by an exponential dispersion model. *Ecological Modelling*, 151(2):261–269.

Kendal, W. S. (2004). Taylor's ecological power law as a consequence of scale invariant exponential dispersion models. *Ecological Complexity*, 1(3):193–209.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. (2015). TMB: Automatic differentiation and Laplace approximation. *J Stat Softw*, 70:doi: 10.18637/jss.v070.i05.

Laake, J. and Borchers, D. (2004). Methods for incomplete detection at distance zero. In Buckland, S., Anderson, D., Burnham, K., Laake, J., Borchers, D., and Thomas, L., editors, *Advanced Distance Sampling*, pages 108–189. Oxford University Press, Oxford, U.K.

Laake, J., Borchers, D., Thomas, L., Miller, D., Bishop, J., and McArthur, J. (2023). *mrds: Mark-Recapture Distance Sampling*. R package version 2.3.0.

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of statistical software*, 63(19).

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. B*, 73(4):423–498.

Lowry, L. F., Citta, J. J., O'corry-Crowe, G., Quakenbush, L. T., Frost, K. J., Suydam, R., Hobbs, R. C., and Gray, T. (2019). Distribution, abundance, harvest, and status of western Alaska beluga whale, Delphinapterus leucas, stocks. *Marine Fisheries Review*, 81:54–71.

Lowry, L. F., Kingsley, M. C., Hauser, D. D., Clarke, J., and Suydam, R. (2017). Aerial survey estimates of abundance of the eastern Chukchi Sea stock of beluga whales (Delphinapterus leucas) in 2012. *Arctic*, pages 273–286.

Marsh, H. and Sinclair, D. (1989). Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *J Wildlife Manage*, 53:1017–1024.

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

Miller, D. L., Glennie, R., and Seaton, A. E. (2020). Understanding the stochastic partial differential equation approach to smoothing. *Journal of Agricultural, Biological and Environmental Statistics*, 25(1):1–16.

Miller, D. L., Rexstad, E., Burt, L., Bravington, M. V., Hedley, S., Ferguson, M., and Kelly., N. (2022). *dsm: Density Surface Modelling of Distance Sampling Data*. R package version 2.3.3.

Miller, D. L. and Wood, S. N. (2014). Finite area smoothing with generalized distance splines. *Environmental and Ecological Statistics*, 21:715–731.

National Marine Fisheries Service (2023). Guidelines for Preparing Stock Assessment Reports Pursuant to the 1994 Amendments to the MMPA. Available at `https://www.fisheries.noaa.gov/national/marine-mammal-protection/guidelines-assessing-marine-mammal-stocks`. Accessed 7 February 2024.

Oceana and Kawerak Inc. (2014). Bering Strait marine life and subsistence use data synthesis. Available at `https://oceana.org/publications/reports/the-bering-strait-marine-life-and-subsistence-data-synthesis`.

O'Corry-Crowe, G., Ferrer, T., Citta, J. J., Suydam, R., Quakenbush, L., Burns, J. J., Monroy, J., Whiting, A., Seaman, G., Goodwin Sr, W., et al. (2021). Genetic history and stock identity of beluga whales in Kotzebue Sound. *Polar Research*, 40:`https://doi.org/10.33265/polar.v40.7623`.

O'Corry-Crowe, G., Suydam, R., Quakenbush, L., Potgieter, B., Harwood, L., Litovka, D., Ferrer, T., Citta, J., Burkanov, V., Frost, K., et al. (2018). Migratory culture, population structure and stock identity in north Pacific beluga whales (Delphinapterus leucas). *PLoS One*, 13(3):e0194201.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent Gaussian models

by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):319–392.

Siddon, E. (2023). Ecosystem assessment. In Siddon, E., editor, *Ecosystem Status Report 2023: Eastern Bering Sea, Stock Assessment and Fishery Evaluation Report*. North Pacific Fishery Management Council, 1007 West 3rd Ave., Suite 400, Anchorage, Alaska 99501.

Sigourney, D. B., Chavez-Rosales, S., Conn, P. B., Garrison, L., Josephson, E., and Palka, D. (2020). Developing and assessing a density surface model in a Bayesian hierarchical framework with a focus on uncertainty: insights from simulations and an application to fin whales (Balaenoptera physalus). *PeerJ*, 8:e8226.

Taylor, L. (1961). Aggregation, variance and the mean. *Nature*, 189:732–735.

Thorson, J. and Kristensen, K. (2024). *Spatio-Temporal Models for Ecologists*. Chapman and Hall/CRC.

Thorson, J. T. and Kristensen, K. (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.

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

Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(1):95–114.

Wood, S. N. (2006). *Generalized additive models*. Chapman & Hall/CRC, Boca Raton, Florida.

Wood, S. N. (2017). P-splines with derivative based penalties and tensor product smoothing of unevenly distributed data. *Statistics and Computing*, 27:985–989.

Wood, S. N., Bravington, M. V., and Hedley, S. L. (2008). Soap film smoothing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(5):931–955.

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

		$\hat{N}$ in 2017 Strata		$\hat{N}$ in 2022 Strata	
Year	Model	Uncorrected	Corrected	Uncorrected	Corrected
2017	SPDE	10140	11242 (0.11)		
2017	soap	10583	11729 (0.12)		
2017	te	10539	11962 (0.12)		
2017	s	10664	11453 (0.11)		
2017	Ensemble		11597 (0.12)		
2017	Conventional	12269 (0.12)			
2022	SPDE	7856	9425	10687	12593 (0.18)
2022	SPDE with barriers	8422	10985	10900	14525 (0.23)
2022	soap	10513	14133	15272	20162 (0.26)
2022	s	11794	14837	17079	21508 (0.29)
2022	Ensemble				17197 (0.33)
2022	Conventional			20635 (0.31)	

Table 1. Estimated abundance ($\hat{N}$) of Eastern Bering Sea belugas from models fitted to 2017 and 2022 aerial line-transect survey data. For spatial models, we present both uncorrected estimates and those that employed epsilon bias correction ("Corrected"). As precision estimates were numerically intensive to calculate, and only of primary interest for epsilon bias-corrected models, we provide estimated CVs (parentheses) for epsilon bias-corrected estimates and for the full area surveyed each year. We also provide point estimates of abundance based on the 2022 survey data that were restricted to the area within the 2017 strata to allow comparison between years for the same region. SPDE = SPDE Matérn model. SPDE with barriers = SPDE Matérn model with barriers. soap = Soap film smoother. te = Tensor product smoother. s = Bivariate and isotropic thin plate regression spline.

	2017		2022	
Model	# RE	Pct. Dev. Expl.	# RE	Pct. Dev. Expl.
SPDE Matérn	198	53.6	307	83.4
SPDE Matérn with barriers	NA	NA	312	80.3
Soap film smoother	79	51.1	146	72.6
Tensor product smoother	224	56.7	NA	NA
Bivariate isotropic thin plate regression spline	29	45.7	29	62.1

Table 2. Number of random effects (# RE) used in the density surface models fitted to 2017 and 2022 Eastern Bering Sea beluga aerial line-transect survey data and included in the ensemble model for each year. For a given model type, the number of random effects can differ between years due to differences in the sample sizes available for fitting the models and in the study area extent. The percent deviance explained (Pct. Dev. Expl.) for each candidate model in the ensemble is also shown. Models showing "NA" in the table were not included in the ensemble model for that year.

Year	Model	ϕ	ρ
2017	SPDE	5.80	1.42
2017	Soap film smoother	5.95	1.43
2017	Tensor product smoother	5.56	1.42
2017	Bivariate isotropic thin plate regression spline	6.34	1.44
2022	SPDE	5.26	1.40
2022	SPDE with barriers	6.24	1.46
2022	Soap film smoother	7.42	1.49
2022	Bivariate isotropic thin plate regression spline	9.27	1.53

Table 3. Estimates of Tweedie dispersion (ϕ) and power (ρ) parameters for density surface models fitted to Eastern Bering Sea beluga aerial line-transect survey data, where $Var(Y) = \phi\mu^{\rho}$.

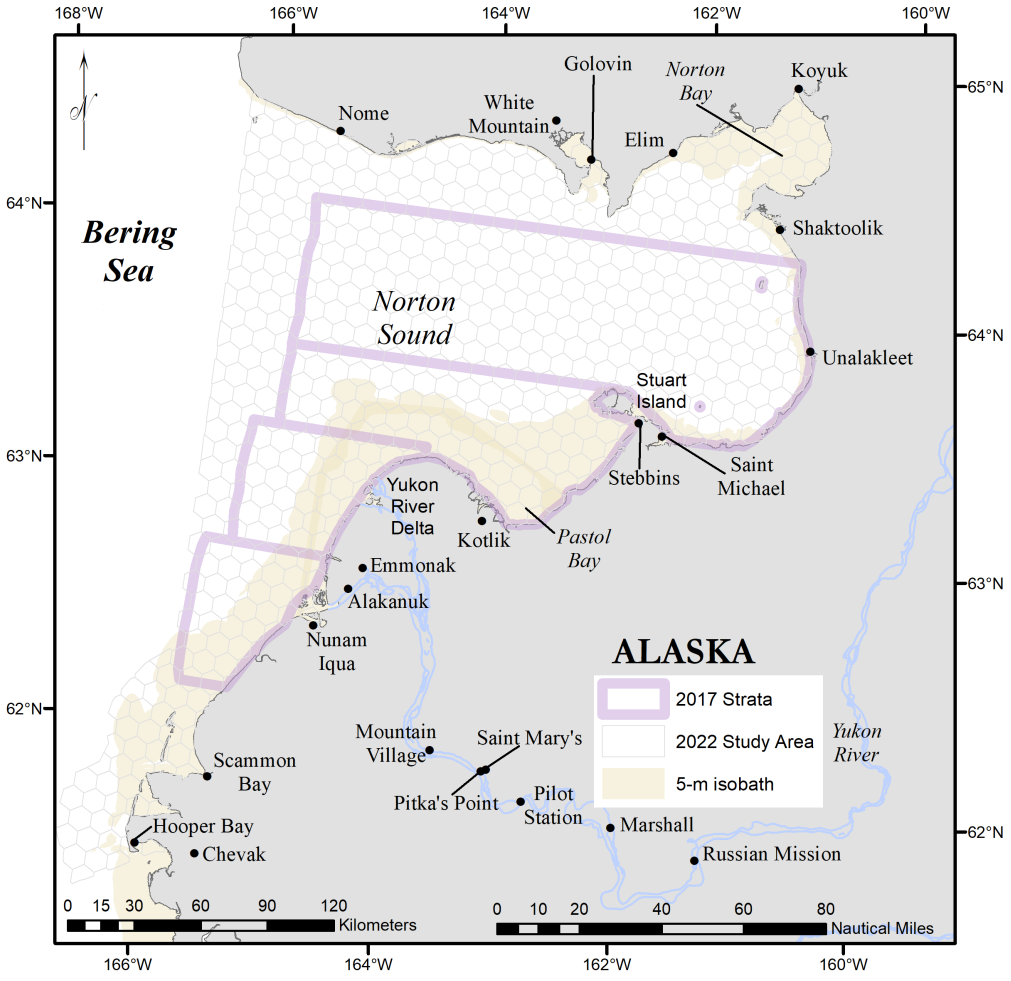


Figure 1. Study area for the Eastern Bering Sea beluga case study.

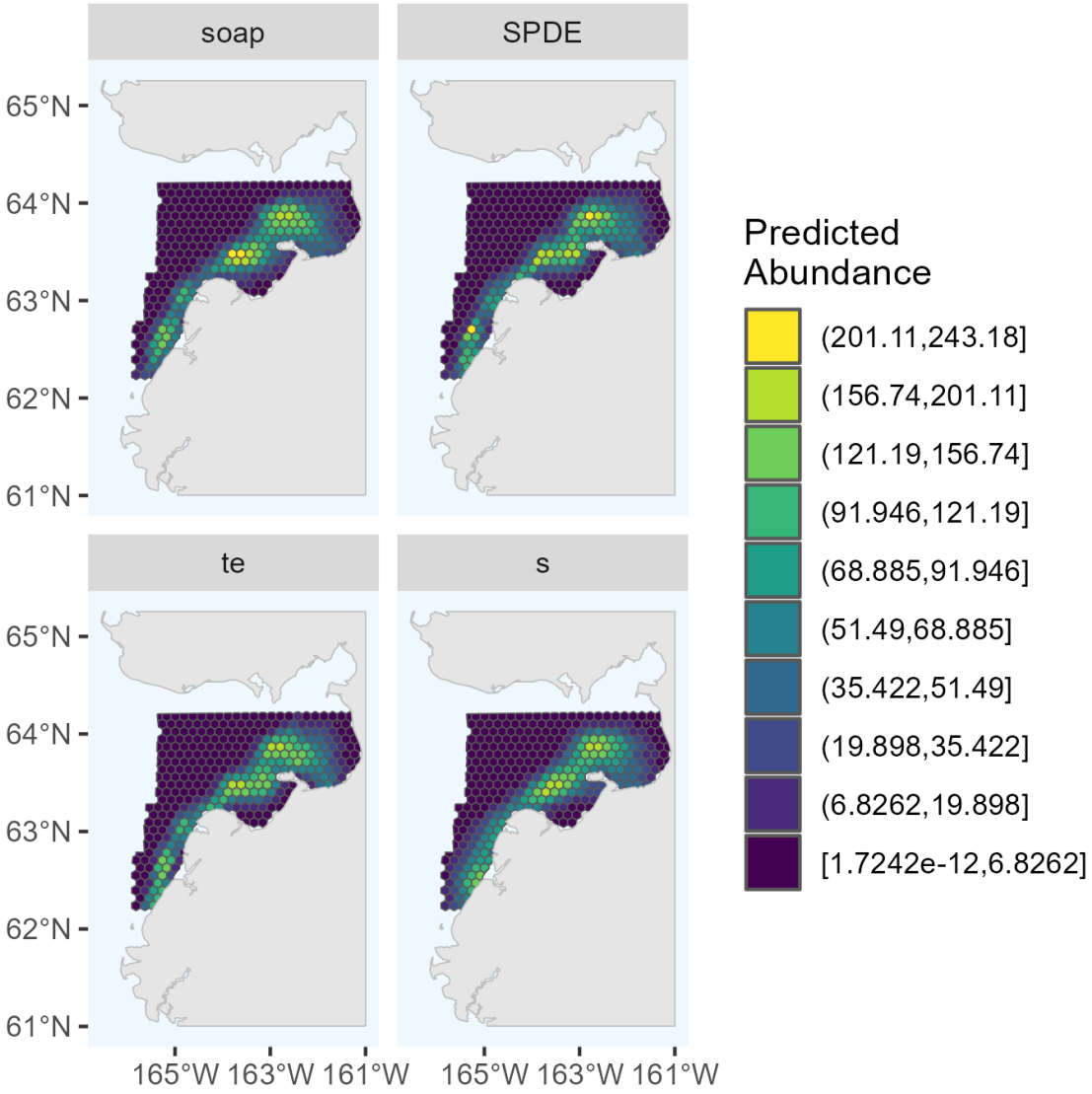


Figure 2. Predicted abundance of Eastern Bering Sea belugas in 2017 based on the four candidate density surface models selected for the ensemble model. soap: soap film smoother. SPDE: SPDE Matérn model. te: tensor product smoother. s: bivariate and isotropic thin plate regression spline.

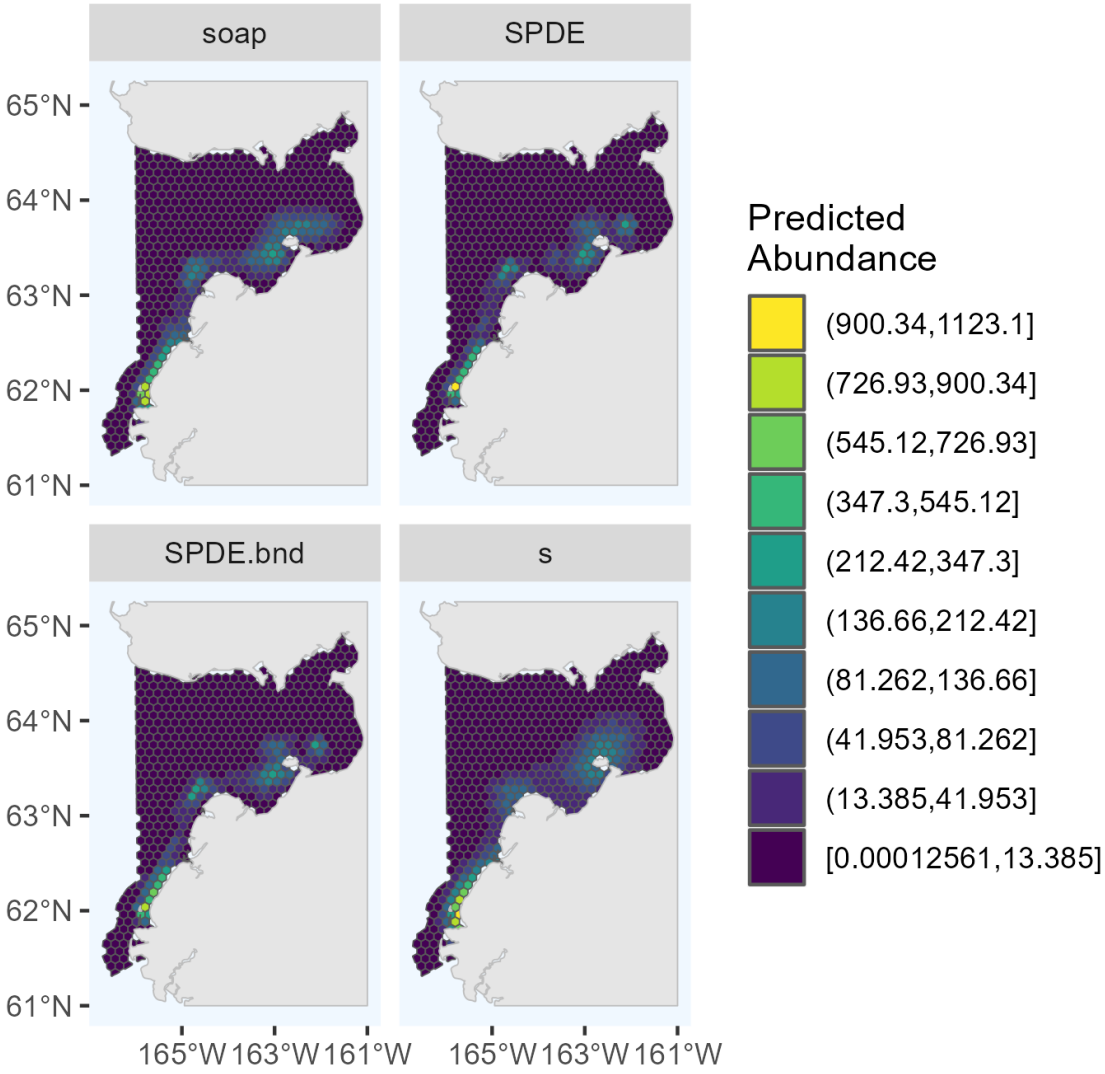


Figure 3. Predicted abundance of Eastern Bering Sea belugas in 2022 based on the four candidate density surface models selected for the ensemble model. soap: soap film smoother. SPDE: SPDE Matérn model. SPDE.bnd: SPDE Matérn model with barriers. s: bivariate and isotropic thin plate regression spline.

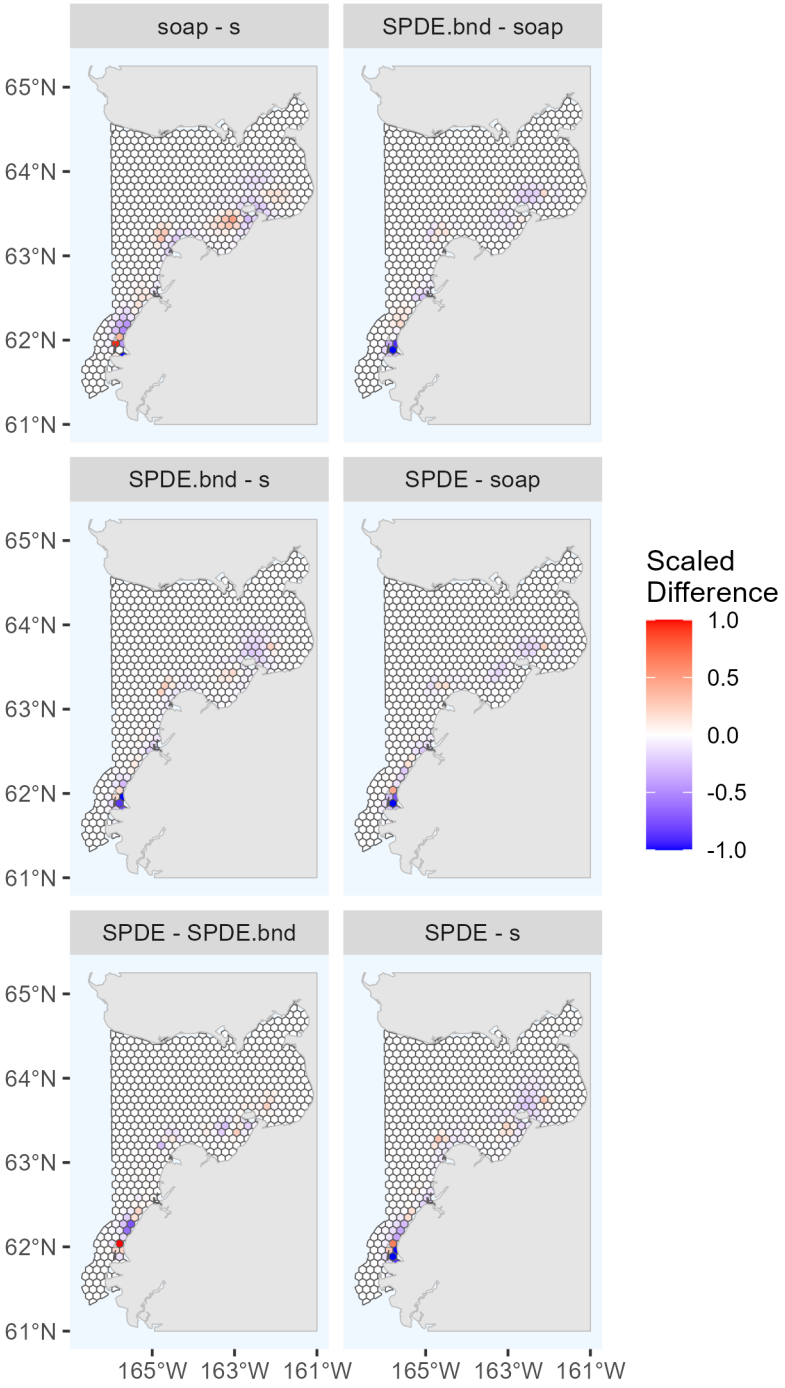


Figure 4. Pairwise comparisons of density surface model (DSM) predictions for Eastern Bering Sea belugas in 2022. Each map shows the scaled differences between two DSMs (m_1, m_2) in predicted beluga abundance by cell (h): $(\hat{N}_{m_1,h} - \hat{N}_{m_1,h})/max(abs(\hat{N}_{m_1,h} - \hat{N}_{m_1,h}))$. Comparisons between all pairs of DSMs selected for the ensemble model are shown. soap: soap film smoother. SPDE: SPDE Matérn model. SPDE.bnd: SPDE Matérn model with barriers. s: bivariate and isotropic thin plate regression spline.

Supplement 1
Glossary of Notation and Abbreviations

a : area searched during line-transect survey, where $a = 2Lw$, L is the total length of transects surveyed, and w is the width of the strip searched on one side of the aircraft

$\mathbf{a}$: vector of the area surveyed for each transect segment ($a_i = 2L_i w$)]

A : total study area size

$\mathbf{A}$: an (n_i, n_η) interpolation matrix used to convert raw random effects into transect-specific values. For SPDE models, this is constructed with a Delauney triangulation. For basis-penalty smooth models, $\mathbf{A}$ is simply the design matrix associated with spatial smooth parameters.

$\mathbf{A}^{pred}$: an (n_h, n_η) interpolation matrix used to convert raw random effects into grid-cell specific predictions. For SPDE models, this is constructed with a Delauney triangulation. For basis-penalty smooth models, $\mathbf{A}$ is a design matrix associated with locations of grid-cell centroids (obtained using the "predict" function in `mgcv`).

$\mathbf{c}$: vector of observed counts of individual animals

c_i : observed number of individual animals on transect segment i

$[\mathbf{c}|\xi, \boldsymbol{\eta}, \mathbf{x}\ ]$: conditional probability density function of observed counts, given parameters, random effects, and known covariates

CV : coefficient of variation

$\hat{D}$: estimate of density of animals (number of animals per unit area)

DSM : density surface model

EBS belugas : Eastern Bering Sea belugas

$g(y_j, \mathbf{z}_j; \hat{\boldsymbol{\theta}}_g)$: probability of detecting an animal at distance y_j, given that it is available to be seen and is associated with covariates $\mathbf{z}_j$, assuming perfect detection on the transect

h : grid cell index

i : segment index

j : group index

k : bootstrap replicate index

L : transect length

n_g : number of groups detected

n_i : number of transect segments in DSM

n_h : number of grid cells in DSM

n_η : number of random effects in DSM

$\hat{N}$: estimate of the total number of animals in the study area

$\hat{p}(\mathbf{z}_j; \hat{\boldsymbol{\theta}})$: model-based estimate of the overall probability that an observer detects a group of whales, given covariates $\mathbf{z}_j$ that affect detectability. This term accounts for all sources of perception and availability bias (Marsh and Sinclair 1989; S4).

$\hat{p}_g(\mathbf{z}_j; \hat{\boldsymbol{\theta}}_g)$: average probability that an observer detects an object that is available to be seen in the area searched, given covariates $\mathbf{z}_j$ that affect detectability, assuming transect detection probability is 1.0

$\mathbf{p}_g$: vector of $\hat{p}_g(\mathbf{z}_j; \hat{\boldsymbol{\theta}}_g)$ for a collection of sightings indexed by j

p_i : shorthand for $\hat{p}(\mathbf{z}_i; \hat{\boldsymbol{\theta}})$ for transects. Note that making the change from group-level detection probability (subscript j) to transect-level detect probability (subscript i) requires that we omit group-specific covariates, such as group size.

$\mathbf{p}_i$: vector of the overall detection probability (including both availability and perception bias corrections) for each segment, p_i.

p_j : shorthand for $\hat{p}(\mathbf{z}_j; \hat{\boldsymbol{\theta}})$ for groups

$p^*(y_j, \mathbf{z}_j; \hat{\boldsymbol{\theta}})$: probability of detecting an animal at distance y_j, given that it is associated with covariates $\mathbf{z}_j$

$\hat{p}_A$: estimate of availability probability, defined as the probability that a group is at the surface within an observer's field of view

$\hat{p}_{MR}(0, \mathbf{z}_j; \hat{\boldsymbol{\theta}}_{MR}$: estimate of transect detection probability, defined as the probability of detecting an animal on the transect (or left-truncation point, if applicable)

$\mathbf{s}$: vector of knot locations for SPDE models

S_j : size of group indexed by j

SPDE : stochasic partial differential equation

$\mathbf{Q}$: precision (inverse covariance) matrix for random effects

Var : variance

w : distance (width) searched on one side of the transect

$\mathbf{x}$: vector of known covariates used in the DSM

y_j : perpendicular distance from the transect line to the sighting of group j

z_i : covariates that affect detectability on segment i

z_j : covariates that affect detectability of group j

β_0 : DSM intercept parameter

$\boldsymbol{\delta}$: vector of 'realized' random effects for transect counts

$\boldsymbol{\eta}$: vector of random effects for the DSM

$[\boldsymbol{\eta}|\mathbf{x}, \boldsymbol{\xi}$]: probability density function of random effects for the DSM

$\hat{\theta}$: parameter estimates required to estimate detection probabilities

μ : mean of a Tweedie probability density function

μ_i : expected number of whales encountered on transect segment i

$\boldsymbol{\xi}$: vector of unknown parameters for the DSM

ρ : power parameter for the Tweedie probability density function

ϕ : dispersion parameter for the Tweedie probability density function

τ : Matérn precision parameter

κ : Matérn inverse range parameter

λ : penalization parameter(s) for basis-penalty smooths

$\lambda_{m,h}$: predicted abundance from model m for unsampled location h

$\lambda_{m,max}$: maximum predicted abundance across all sampled cells

Supplement 2

Aerial line-transect surveys for Eastern Bering Sea Belugas

in 2017 and 2022

Survey methods

The Eastern Bering Sea (EBS) beluga study area encompasses Norton Sound and the Yukon River Delta (Figure S2.1). Norton Sound is a shallow bay (average depth 13 m) located along western Alaska, south of the Seward Peninsula, spanning approximately 160 km from Cape Nome to the Yukon River Delta (Figure S2.1). The Sound is seasonally covered with sea ice. During June, sea ice is usually absent, which was the case in June 2017 and 2022. Outflow from the Yukon River creates a nearshore zone of turbid water, extending approximately 40 km offshore, bounded by a sharp oceanographic front, beyond which the waters are more clear and it is possible to see below the surface of the water from an aerial platform.

Aerial line-transect surveys were flown in Norton Sound and along the Yukon River Delta, from 16 to 29 June 2017 (Figure S2.1) and from 24 to 30 June 2022 (Figure S2.2). Following Lowry et al. (2017), systematic transects were placed 9.3 km apart, based on a grid with a randomly selected start point. Transect length varied from approximately 10 to 250 km. Transects were oriented east-west, along lines of latitude, from shore to 166º W. The northernmost transect in both years was located in Norton Bay. The southernmost transect in 2017 was located at 62.3º N, approximately 50 km north of Scammon Bay (Figure S2.1). In 2022, the southern boundary of the study area extended farther south than any previous aerial line-transect survey of EBS belugas had flown (Figures S2.2; Lowry et al. 2017). This southward extension was incorporated into the survey design to determine the southern extent of the range of EBS belugas. The total study area was 41,417 km^2 in 2017 and 47,381 km^2 in 2022 (Table S2.1).

During the surveys conducted in 2017 and 2022, the Turbo Commander aircraft provided and flown by Clearwater Air, Inc., was based in Nome, Alaska. Unalakleet, Alaska, was an alternate airport that was used for refueling when conducting surveys of the central and southern transects. The Turbo Commander is a twin-turbine, high-wing aircraft. The plane had bubble windows for the left- and right-side primary observers, allowing unobstructed views from directly beneath the plane out to the horizon. Surveys were conducted at 320 m altitude at 213 km/h.

The survey team comprised two primary observers and one dedicated data recorder. The data recorder input sighting data related to detectability and species density into a laptop computer, connected to a GPS, running specialized, menu-driven software (Clarke et al. 2020; MML unpublished report). Time and position data (latitude, longitude, altitude) were automatically

recorded in 30-sec intervals or whenever a manual data entry was recorded. Environmental and viewing conditions, including integer-valued Beaufort Sea State, turbidity (binary, yes or no), 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 (the average from both sides of the plane), and impediments to visibility (glare, fog, haze, precipitation, ice on the window, low ceiling) on each side of the plane were recorded in 5-min intervals or whenever conditions changed.

Primary observers scanned with the 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 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. Belugas in the study area during June are typically distributed in small groups comprising only a few animals. If group size could not be determined with confidence, high and low estimates could also be recorded.

Beluga calves were identified primarily based on size: calves were noticeably smaller than the other animals. In addition to being smaller, coloration (typically grayish or brownish pigmentation) and close proximity to an adult helped observers identify beluga calves. However, it is not always possible for aerial observers to distinguish beluga calves of the year from juveniles; therefore, animals recorded as beluga calves likely include belugas up to a few years old.

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 or small unidentified pinniped).

Aerial observers watched for any abrupt and unexpected changes in marine mammals' initially observed behavior, presumably due to the aircraft. Observed responses and the number of animals that responded were recorded in the database.

Weather permitting, survey effort along each transect was uninterrupted; the aircraft diverted from the transect to circle sightings only in exceptional situations (e.g., to photograph carcasses or investigate sightings of cetaceans that were not belugas, and to confirm species identification). Four survey modes were used for data collection: deadhead, transect, circling from transect, and search. No sighting data were collected during transit or when weather was not conducive to surveying (i.e., during "deadhead" effort). During the remaining three survey modes, observers were actively surveying and all sightings and environmental data were recorded. Transect effort refers to systematic survey effort along a prescribed transect line. Search refers to non-systematic survey effort between transects. Circling from transect occurred when the aircraft diverted from flat and level flight to circle a localized area to investigate a sighting or potential sightings.

Sighting and effort summaries

There was more line-transect survey effort and beluga sightings in 2017 compared to 2022. Ferguson et al. (2023) present detailed results for the 2017 survey; therefore, we provide only an overview here. Between 16 and 29 June 2017, a total of 16 survey flights (62 flight hours) were conducted over 12 days. Each transect in the study area was surveyed at least once, and most transects were surveyed twice (Figure S2.1). The total number of living belugas detected in 2017 was 1,897 (Table S2.1), including 95 calves (as defined above); an additional 2 beluga carcasses were detected. Beluga group sizes during the 2017 surveys ranged from 1 to 39 whales. The geographic stratum with the largest average group size (3.5 belugas per group) was located south of the Yukon Delta (Figure S2.1).

Between 24 and 30 June 2022, a total of 8 survey flights (30 flight hours) were conducted over 4 days. All transects from Pastol Bay to the southern end of the study area (Hooper Bay) were surveyed (Figure S2.2). Most of the transects north of Pastol Bay where relatively high densities of belugas had been detected in previous years (Lowry et al. 2017; Ferguson et al. 2023) were surveyed (Figure S2.2). Due to poor weather, transects in the northern portion of the study area, between Shaktoolik and Stuart Island, were not completed. Figure S2.3 shows all transect effort completed during Beaufort Sea State ≤ 4, color-coded by Beaufort Sea State conditions at the time the survey was conducted. A total of 821 living belugas were detected in 2022 (Table S2.1), including 5 calves; an additional 1 beluga carcass was detected. Beluga group sizes during the 2022 surveys ranged from 1 to 120 whales. The three largest groups were sighted near the barrier islands north of Scammon Bay, an area that had not been surveyed during any previous beluga surveys conducted by the Alaska Beluga Whale Committee or NOAA Fisheries (Lowry et al. 2017; Ferguson et al. 2023; Figure S2.2). The aerial survey observers estimated that these large groups comprised 67, 87, and 120 belugas.

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. 603 p.

Ferguson, M.C., A.A. Brower, A.L., Willoughby, and C.K. Sims. 2023. Distribution and estimated abundance of eastern Bering Sea belugas from aerial line-transect surveys in 2017. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-471, 50 p.

Lowry, L.F., A. Zerbini, K.J. Frost, D.P. DeMaster, and R.C. Hobbs. 2017. Development of an abundance estimate for the eastern Bering Sea stock of beluga whales (*Delphinapterus leucas*). J. Cetacean Res. Manage. 16: 39-47.

Table S2.1 Summary statistics from the line-transect aerial surveys for belugas conducted in the Norton Sound/Yukon Delta region between 1992 and 2022 by the Alaska Beluga Whale Committee and NOAA Fisheries.

Survey Dates	Transect Effort (km)	Belugas Counted	Encounter Rate (belugas/km)	Study Area (km^2)
17-21 June 1992	7,278	1,625	0.223	6,145
14-18 June 1993	5,539	374	0.068	10,975
11-16 June 1994	5,746	370	0.064	13,965
5-8 June 1995	4,450	750	0.169	19,983
20-22 June 1995	1,776	456	0.257	3,352
15-17 June 1999	3,366	589	0.175	15,794
17-20 June 2000	4,226	428	0.101	38,104
16-29 June 2017	8,587	1,897	0.221	41,416
24-30 June 2022	3,557	821	0.231	47,381

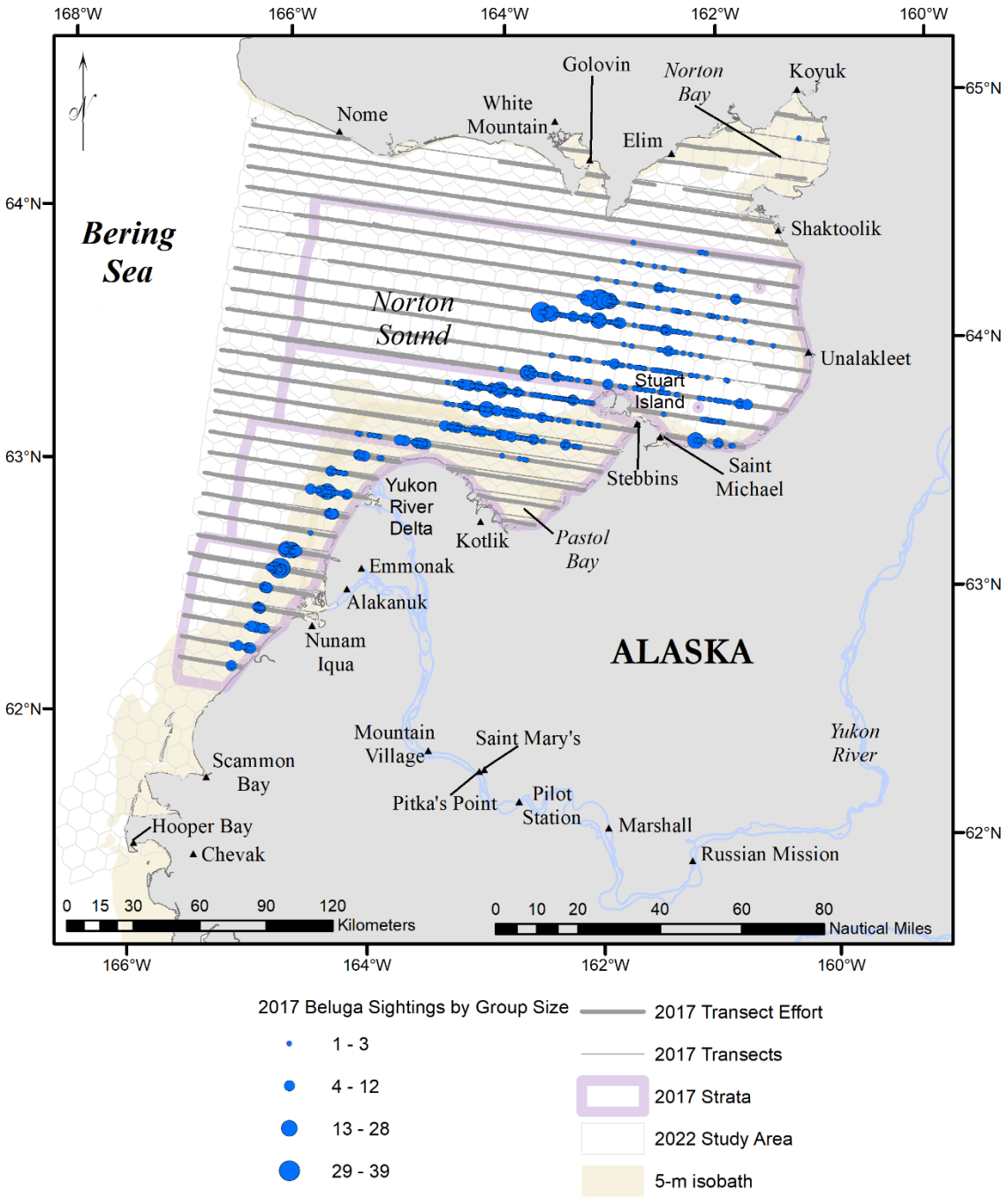


Figure S2.1. 2017 Eastern Bering Sea beluga aerial line-transect survey study area, survey design, and field results. All live beluga sightings and transects flown during Beaufort Sea State ≤ 4 are shown. Waters shallower than 5 m are shaded. The outlines of the geographic strata defined in Lowry et al. (2017) and used in the present analysis are shown. The extent of the 2022 Eastern Bering Sea study area is shown using hexagonal cells.

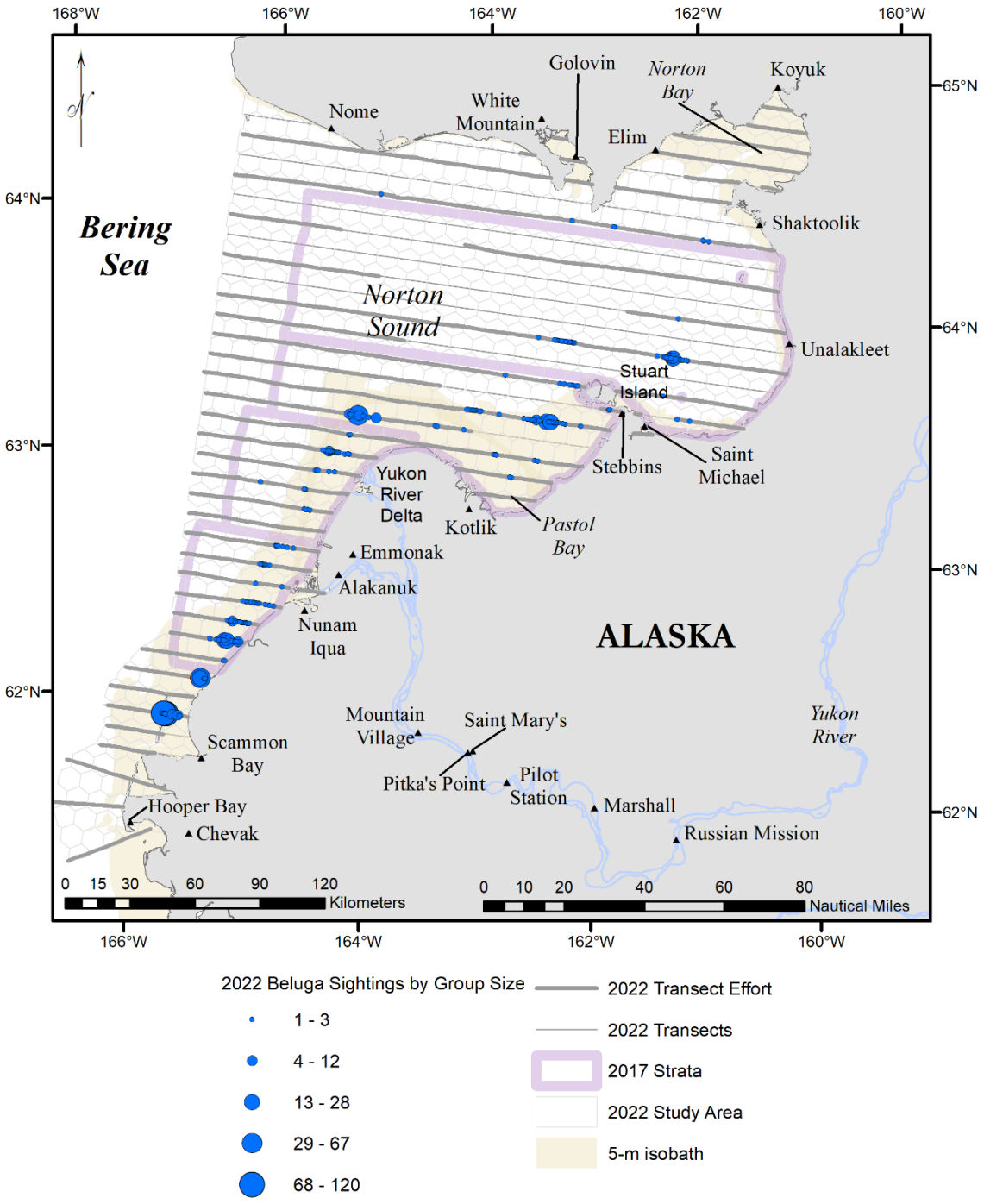


Figure S2.2. 2022 Eastern Bering Sea beluga aerial line-transect survey study area, survey design, and field results. All live beluga sightings and transects flown during Beaufort Sea State ≤ 4 are shown. Waters shallower than 5 m are shaded. The outlines of the geographic strata defined in Lowry et al. (2017) are shown.

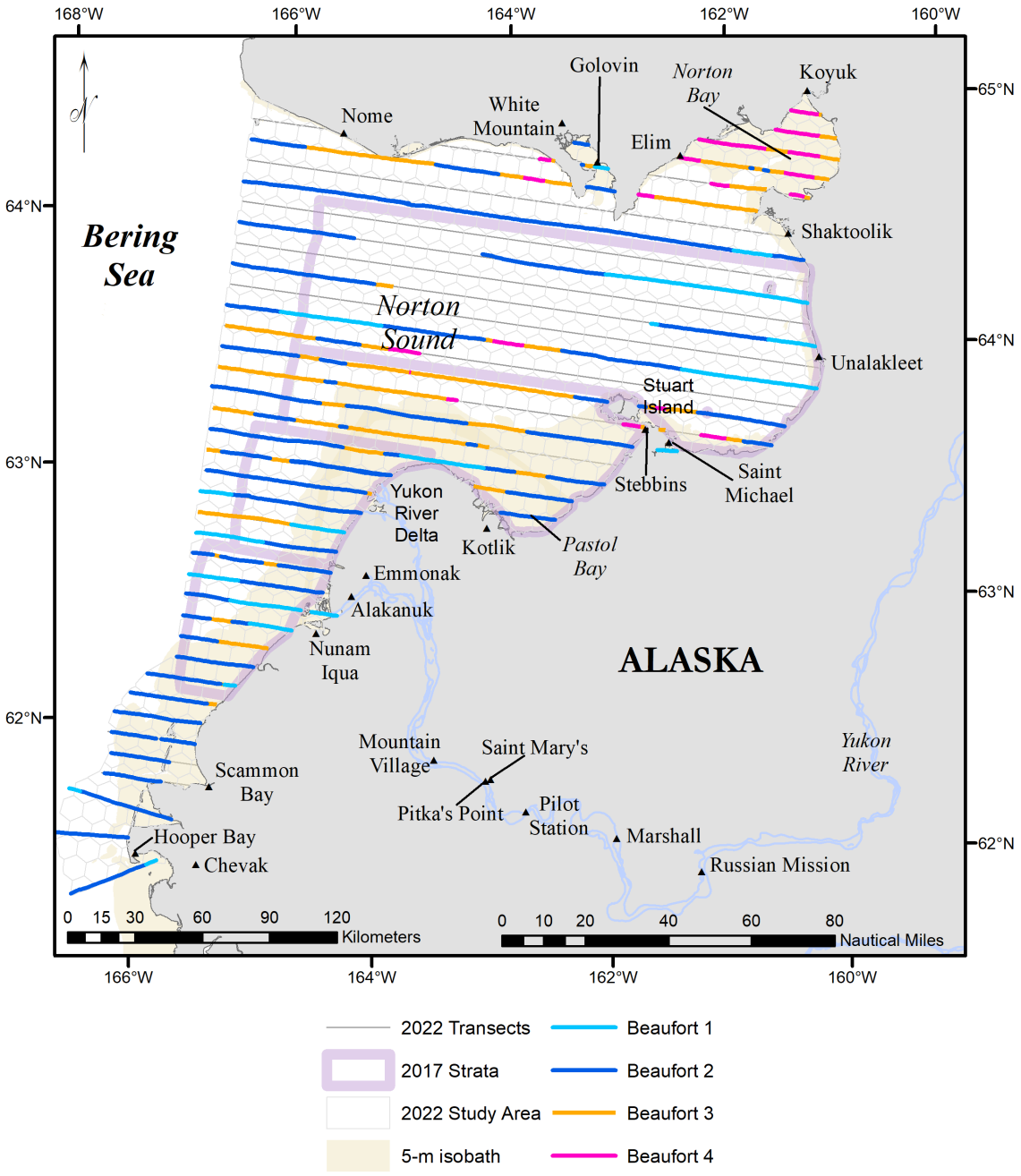


Figure S2.3. 2022 Eastern Bering Sea beluga aerial line-transect survey design. All transects flown during Beaufort Sea State ≤ 4 are shown, color-coded by Beaufort Sea State conditions at the time the survey was conducted.

Supplement 3
Aerial Imagery Collection and Processing Methods

A brief overview of the aerial imagery collection and processing methods is provided here. See Clarke et al. (2019, 2020) and Willoughby et al. (2021) for comprehensive details. The methods used to estimate transect detection probability from these data are detailed in Ferguson et al. (2023).

During the 2017 and 2022 Eastern Bering Sea (EBS) beluga surveys, data were not collected to estimate the detection probability for beluga groups on the transect line (defined as $\hat{p}_{MR}(0, \boldsymbol{z_j}; \widehat{\boldsymbol{\theta}}_{MR})$ below and in S4). Therefore, we relied on the best information available to us, which was collected during the Aerial Surveys of Marine Mammals (ASAMM) line-transect surveys conducted in the eastern Chukchi and western Beaufort seas from July through October in 2018 and 2019 (Clarke et al. 2019, 2020). We believe the estimates of transect detection probability from ASAMM provide reasonable approximations to the actual value for the 2017 and 2022 EBS beluga aerial surveys based on similarities in survey protocols and beluga group size distributions, which affect detectability. ASAMM surveys targeted belugas and larger cetaceans; marine mammal observer and imagery data were collected concurrently. ASAMM line-transect survey protocols were comparable to those used during the 2017 and 2022 EBS beluga aerial surveys and are detailed in Clarke et al. (2019, 2020). Additionally, the same aircraft (including bubble windows) and marine mammal observer configuration used during the 2017 and 2022 EBS beluga aerial surveys were also used to conduct the ASAMM flights that collected the data we describe below. ASAMM surveys were flown at the same target speed (213 km/h speed) and a similar target altitude (400 m) as the 2017 and 2022 EBS beluga surveys (320 m). Beluga group size distributions were comparable in the 2017 and 2022 EBS beluga data and the relevant ASAMM survey data. For the 2017 EBS beluga survey, 54.7% (338/618) of the sightings were of single belugas, 23.6% (146/618) comprised two belugas, 21.4% (132/618) had 3-10 belugas, and < 1% (2/618) had more than 10 belugas. For the 2022 EBS beluga survey, 75% (180/241) of the sightings were of single belugas, 13% (31/241) comprised two belugas, 8% (19/241) had 3-10 belugas, and 5% (11/241) had more than 10 belugas. In the 2018-2019 ASAMM survey data, 70.9% (720/1015) of the sightings were of single belugas, 17.1% (174/1015) comprised two belugas, 10.6% (108/1015) had 3-10 belugas, and 1.3% (13/1015) had more than 10 belugas.

To estimate transect detection probability for marine mammal observers during ASAMM line-transect surveys, a downward-pointing digital single lens reflex camera with a 20- or 21-mm lens mounted to the belly of the aircraft collected true color (red, green, and blue [RGB]) imagery (Clarke et al. 2019, 2020; Willoughby et al. 2021). 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. The imagery served as an "independent observer" for a mark-recapture analysis of the ASAMM aerial observer data.

Willoughby et al. (2021) provide detailed imagery collection and analysis methods and results; here, we present a brief overview. 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. All sightings detected in the imagery were manually compared to the aerial observer database to determine matches based on date, time, and location (side of plane and distance from transect). The results of the matching analysis could be one of three categories: matched, not matched, and "inconclusive results" (abbreviated "IR"). Inconclusive results meant that the photo analyst could not determine for certain whether an imagery sighting was also detected by the aerial observers.

See Ferguson et al. (2023) for comprehensive details on estimating transect detection probability from these data.

Literature Cited

Clarke, J.T., A.A. Brower, M.C. Ferguson, and A.L. Willoughby. 2019. Distribution and relative abundance of marine mammals in the Eastern Chukchi and Western Beaufort Seas, 2018. Annual Report, OCS Study BOEM 2019-021. 451 p.

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. 603 p.

Ferguson, M.C., A.A. Brower, A.L., Willoughby, and C.K. Sims. 2023. Distribution and estimated abundance of eastern Bering Sea belugas from aerial line-transect surveys in 2017. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-471, 50 p.

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. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-427, 111 p.

Supplement 4
Eastern Bering Sea Beluga Detection Probabilities

Basic density estimator

A basic Horvitz-Thompson-like line-transect estimator of animal density is (Buckland et al. 2001; Burt et al. 2014; Marques and Buckland 2003):

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

where

n_g = total number of groups detected;

S_j = size of group indexed by j;

a = area searched, equal to $2wL$, where L is the total length of transects surveyed and w is the width of the strip searched on one side of the aircraft;

$\hat{p}\left(\boldsymbol{z}_j;\widehat{\boldsymbol{\theta}}\right)$ = estimate of the overall probability that an aerial observer detects group j located in the area searched. This probability depends on estimated parameters ($\widehat{\boldsymbol{\theta}}$) and other possible covariates ($\boldsymbol{z}_j$) that affect detectability. This term accounts for all sources of perception and availability bias (Marsh and Sinclair 1989). Note that $\hat{p}\left(\boldsymbol{z}_j;\widehat{\boldsymbol{\theta}}\right)$ does not depend on distance from the transect line, since we integrate the detection function over the truncation width of the transect as suggested by Marques et al. (2003) to reduce overall variability in the estimator (see Eq. 5 below).

Eastern Bering Sea beluga observation model

The underlying observation model for the Eastern Bering Sea (EBS) beluga aerial survey was a scaled version of a multiple covariates distance sampling (MCDS) detection function, $g\left(y_j\,,\boldsymbol{z}_j;\widehat{\boldsymbol{\theta}}_g\right)$, (Marques and Buckland 2003, Laake and Borchers 2004):

$$\hat{p}^*\left(y_j,\boldsymbol{z}_j;\widehat{\boldsymbol{\theta}}\right) = \hat{p}_A\hat{p}_{MR}\left(0,\boldsymbol{z}_j;\widehat{\boldsymbol{\theta}}_{MR}\right)g\left(y_j\,,\boldsymbol{z}_j;\widehat{\boldsymbol{\theta}}_g\right). \quad [2]$$

where

$g(y_j, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_g)$ = probability of detecting an animal at distance y_j, given that it is available to be seen and is associated with covariates $\boldsymbol{z}_j$, assuming perfect detection on the transect;

$\hat{p}^*(y_j, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}})$ = probability of detecting an animal at distance y_j, given that it is associated with covariates $\boldsymbol{z}_j$. This probability allows for intermittent availability and imperfect detection on the transect.

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 scaling factors in the observation model include an estimated availability probability, $\hat{p}_A$, and a mark-recapture component, $\hat{p}_{MR}(0, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_{MR})$.

Availability probability is the probability that a beluga is at the water's surface and within the observers' field of view sometime during the period that a plane passes overhead. The mark-recapture component is the probability that a group is detected on the transect line, given that it is at the surface and within the observers' field of view. Together, $\hat{p}_A$ and $\hat{p}_{MR}(0, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_{MR})$ determine the location of the intercept in the observation model.

Values for $\hat{p}_A$ and $\hat{p}_{MR}(0, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_{MR})$ were taken from Ferguson et al. (2023) and are described in more detail below. First, we describe how the MCDS detection function was estimated using data from EBS beluga surveys.

Multiple covariates distance sampling detection function for the EBS beluga aerial surveys

The MCDS detection function was constructed using data only from the 2017 and 2022 EBS beluga aerial surveys, which were filtered prior to fitting the model. Only beluga sightings made by primary observers during transect effort conducted in Beaufort Sea State 0-4 that had recorded declination angles were used to construct the detection function.

Additionally, sighting data were truncated 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). The histogram of perpendicular distances to beluga sightings indicated fewer than expected sightings within 75 m of the transect; therefore, the data were left-truncated at 75 m (Figure S4.1). The farthest 5% of sightings were omitted from the detection function analysis to minimize the effects of outliers. This right-truncation distance was 5.2 km. The width of the strip searched on one side of the aircraft equals the right-truncation distance minus the left-truncation distance, w = 4.45 km.

A MCDS model can take various forms, specified by its key function, such as the half-normal key function or hazard-rate key function. For the EBS beluga aerial surveys, MCDS detection function models with half-normal and hazard-rate key functions were considered. A half-normal model in which the standard deviation (scale parameter) is a linear function of covariates affecting detection probability may be represented as:

$$g(y_j, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_g) = \exp\left(\frac{-y_j^2}{2\left[\exp\{\theta_0 + \sum_l \theta_j z_{jl}\}\right]^2}\right) \quad [3]$$

An analogous hazard-rate model may be represented as:

$$g(y_j, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_g) = 1 - \exp\left[-\left(\frac{y_j}{\exp\{\theta_0 + \sum_l \theta_j z_{jl}\}}\right)^{-b}\right] \quad [4]$$

We used the R package mrds (Laake et al. 2021) to fit MCDS detection functions to observed perpendicular distances of beluga groups, including effects of additional covariates that potentially affect detectability. The covariates that we considered were four different group size variables, turbidity, and Beaufort Sea State (Table S4.1). mrds uses maximum likelihood to fit models. AIC was used to compare the two functional forms for distance data (half-normal vs. hazard-rate), conduct model selection, and ultimately derive maximum likelihood estimates and variances of parameters, $\widehat{\boldsymbol{\theta}}_g$**.** The null hazard-rate models had considerably lower AIC values and exhibited better fit (based on visual inspection of the detection function curve overlaid on the histogram of perpendicular sighting distances) than the half-normal models, so covariate selection proceeded with only the hazard-rate key function. Although four different group size covariates were considered in the initial univariate model, only the covariate with the lowest AIC value among the univariate group size models was retained for further consideration in the model fitting and selection process. The best-fitting MCDS detection function model for EBS belugas included covariates for Beaufort Sea State and turbidity.

The average probability that an aerial observer detects an object that is available to be seen in the area searched, given covariates $\boldsymbol{z}_j$ that affect detectability, assuming transect detection probability is 1.0, is (Marques and Buckland 2003):

$$\hat{p}_g(\boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_g) = \frac{\int_0^w g(y, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_g)\, dy}{w} \quad [5]$$

The effective strip half-width (ESW) equals $\hat{p}_g(\boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_g)w$ and it is noteworthy because as many groups are detected within the ESW as are missed beyond the ESW. Due to the left-truncation in the EBS beluga analysis, the ESW stretches from the left-truncation distance to the right-truncation distance.

Mark-recapture detection probability

The MCDS detection function does not account for animals that are missed on the transect line (or left truncation point, in our case). In absence of a dedicated double-observer study during the EBS beluga surveys, we relied on estimates of $\widehat{\boldsymbol{\theta}}_{MR}$ from previous surveys of belugas in the Chukchi and Beaufort seas in 2018 and 2019 (Aerial Surveys of Arctic Marine Mammals [ASAMM]; Ferguson et al. 2023) to determine $\hat{p}_{MR}\left(0, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_{MR}\right)$. Briefly, during ASAMM aerial line-transect surveys, images of whales were collected from a camera system concurrently with visual line-transect survey data. Based on a mark-recapture distance sampling (MRDS) analysis, the proportion of whales detected in the imagery that were also detected by human observers was estimated to be 75.3% ($\hat{p}_{MR}\left(0, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_{MR}\right) = 0.753$). Comprehensive details on the MRDS analysis are provided in Ferguson et al. (2023).

Availability probability

Availability probability, p_A, is the probability that a group is at the surface within an observer's field of view (Marsh and Sinclair 1989). Animals that spend a low proportion of their time at the surface where observers can detect them (i.e., low availability) will be detected infrequently relative to their true density (i.e., raw counts will have considerable bias relative to true density). The inverse of availability probability is the availability bias correction factor. Availability probability is a function of the animals' respiratory patterns and the duration of time in which the ocean at perpendicular distance y is in the observer's view (i.e., viewing time). We used the estimate of availability probability from Ferguson et al. (2023); therefore, we present only a brief summary of the analytical methods here.

We assumed that the effect of distance on detectability was captured by the MCDS detection function model. Therefore, we used the estimate of availability probability on the transect, $\hat{p}_A(0)$, effectively scaling the transect detection probability (Ferguson et al. 2023). Because the field of view from the windows in the EBS beluga survey aircraft was unobstructed ahead of the plane at the left-truncation distance (Ferguson et al. 2021), the length of time for which a sighting was in view on the transect was assumed to be a function of the distance at which a beluga can be detected. The resulting estimate of viewing time on the transect was 15.9 sec.

The best available information on beluga respiration patterns was from behavioral observations made on three adult female belugas tagged with VHF radio tags: one beluga tagged in Bristol Bay, Alaska, in June 1983; and two belugas tagged in Cunningham Inlet, Somerset Island, Canada, in July 1988 (Frost et al. 1985; Frost and Lowry 1995). Using data from these belugas, we determined that, on average, 50% of belugas would be at the surface sometime during the 15.9 sec it took the survey aircraft to pass overhead, resulting in $\hat{p}_A = 0.5$, equating to an availability bias correction factor of 2.0. There were no estimates of uncertainty for availability probability (Ferguson et al. 2023).

Now, we can decompose the detection probability term $\hat{p}(\boldsymbol{z}_j; \widehat{\boldsymbol{\theta}})$ from Eq. 1 into its component parts:

$$\hat{p}(\boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}) = \hat{p}_A \, \hat{p}_{MR}(0, \boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_{MR}) \hat{p}_g(\boldsymbol{z}_j; \widehat{\boldsymbol{\theta}}_g). \quad [6]$$

Literature cited

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.

Ferguson, M.C., A.A. Brower, A.L., Willoughby, and C.K. Sims. 2023. Distribution and estimated abundance of eastern Bering Sea belugas from aerial line-transect surveys in 2017. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-AFSC-471, 50 p.

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 Tech. Memo. NMFS-AFSC-428, 54 p.

Frost, K. J., and L. F. Lowry. 1995. Radiotag based correction factors for use in beluga whale population estimation. Working paper for Alaska Beluga Whale Committee Scientific Workshop, Anchorage, AK, 5-7 April 1995.

Frost, K. J., L. F. Lowry, and R. R. Nelson. 1985. Radiotagging studies of beluga whales (*Delphinapterus leucas*) in Bristol Bay, Alaska. Mar. Mammal Sci. 1: 191-202.

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 A. A. Balkema, Rotterdam.

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. Available from: https://github.com/DistanceDevelopment/mrds/.

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.

Table S4.1. Definitions of covariates considered for inclusion in the multiple covariates distance sampling detection function models for the Eastern Bering Sea beluga line-transect aerial surveys in 2017 and 2022.

Covariate Name	Definition	Categories
size	Observed group size of the sighting	
loggs	$\log_{10}(size)$	
catsize	Categorical group size	{1, >1}
catsize3	Categorical group size	{1, 2, >2}
iBeauf	Integer-valued Beaufort Sea State	
Turb	Turbidity	yes, no

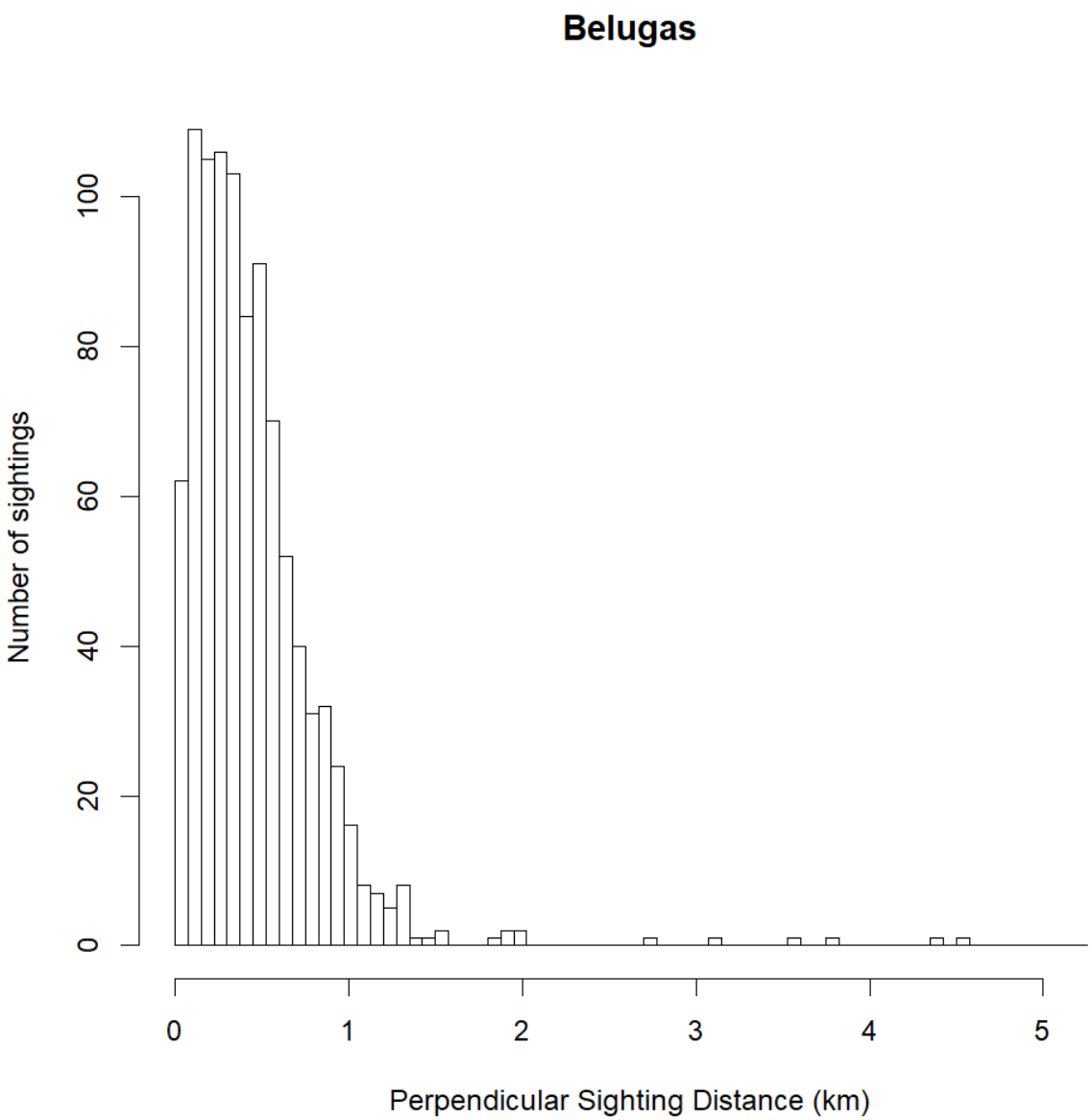


Figure S4.1. Perpendicular distances (km) to beluga sightings during the 2017 and 2022 Eastern Bering Sea beluga aerial line-transect surveys. Bin widths are 75 m.

Supplement 5

Bootstrap Pseudo-code for Estimating Uncertainty in Model-based Estimates of Beluga Abundance

Goal: Incorporate uncertainty from the multiple covariates distance sampling (MCDS) detection function (ddf) model, the spatially-explicit density surface model (DSM), and the mark-recapture distance sampling detection function estimate of transect detection probability into the estimate of uncertainty for the model-based estimated abundance ($\widehat{N}$) of belugas from aerial line-transect surveys.

1. Fit MCDS model to EBS beluga sighting data from 2017 and 2022. Obtain: (1) estimated regression parameters, $\widehat{\boldsymbol{\beta}}$ and their associated covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\beta}$; and (2) predictions of detection probability $\tilde{p}_{g,i}$ for each transect segment *i*. In this case, we limit the MCDS coefficients to segment variables, not observation variables. The value $\tilde{p}_{g,i}$ is the average detection probability, computed as the integral of the detection function over the 1/2-width, divided by the 1/2-width (S4 Eq. 5). In S4 Eq. 5, we refer to $\tilde{\boldsymbol{p}}_g$ using the longhand notation $\hat{p}_g\left(\boldsymbol{z_j}; \widehat{\boldsymbol{\theta}}_g\right)$. This is the value returned from using the mrds::predict( ) function on a ddf object.
2. Fit each of the *m=1,…,M* DSMs to beluga sighting and effort data in TMB, using the $\tilde{\boldsymbol{p}}_g$ from step 1 when computing offsets for each count model. The estimated beluga density surfaces from the TMB models are used to compute model-specific estimates of abundance, $\widehat{N_m}$. We also obtain a conditional variance, $Var(\widehat{N_m}|\tilde{\boldsymbol{p}}_g)$, from the TMB DSMs. These estimates of $\widehat{N_m}$ and $Var(\widehat{N_m}|\tilde{\boldsymbol{p}}_g)$ incorporate a correction for detransformation bias via the epsilon algorithm (Thorson and Kristensen 2016). We are using this approach rather than simulating from the DSMs posterior distributions of random effects because the latter led to problems with extremely high outliers (i.e., implausibly high abundance estimates).
3. To account for increased variance due to uncertainty in $\tilde{\boldsymbol{p}}_g$, we rely on the law of total variance, which in our context states

$$Var\left(\widehat{N_m}\right) = E(Var\left(\widehat{\boldsymbol{N_m}}|\tilde{\boldsymbol{p}}_{\boldsymbol{g}}\right)) + Var(E(\widehat{N_m}|\tilde{\boldsymbol{p}}_g)). \qquad \text{Eqn 1}$$

For $E(Var\left(\widehat{N_m}|\tilde{\boldsymbol{p}}_g\right))$, we simply substitute our conditional variance estimate from the original TMB DSM, $Var(\widehat{N_m}|\tilde{\boldsymbol{p}}_g)$, obtained in step 2.

Heuristically, the additional piece, $Var(E(\widehat{N_m}|\tilde{\boldsymbol{p}}_g))$, describes how abundance estimates might vary with different values of $\tilde{\boldsymbol{p}}_g$. To estimate this variance, we apply the following bootstrap procedure for each candidate DSM:

   a. Extract the covariance matrix, $\widehat{\boldsymbol{\Sigma}}_{\beta}$ for the estimated detection function model (ddf.obj) parameters:

      ddf.sigma <- solve(ddf.obj$hessian)

   b. Extract the detection function model parameters, $\widehat{\boldsymbol{\beta}}$:

      ddf.beta <- ddf.obj$par

c. Implement bootstrap algorithm:

for(k in 1:500){ #repeat the following steps 500 times

- Create bootstrap sample bs.beta for iteration i from the detection function model parameters, assuming a multivariate normal distribution with covariance ddf.sigma:

  bs.beta <- mgcv::rmvn(1, ddf.beta, ddf.sigma)

- Generate new bootstrap detection probabilities $\tilde{p}_{g,(k)}$:
  bs.ddf <- ddf.obj
  bs.ddf$par <- bs.beta
  bs.p <- predict(bs.ddf, gam.data)$fitted

- Re-fit and optimize the TMB DSM for model *m* using the bootstrap detection probabilities bs.p in the calculation of the offset. Compute the bootstrap estimate of abundance for iteration *i*, $\widehat{N}_{m,(k)}$.

} #end bootstrap algorithm

d. For bootstrapped DSMs that exhibited numerical convergence, compute $Var\left(\widehat{N}_{m,(k)}\right)$.

e. Use Eqn 1 to compute $Var\left(\widehat{N}_m\right)$ from $E(Var\left(\widehat{N_m}|\tilde{\boldsymbol{p}}_g\right))$ generated in step 2 and $Var\left(\widehat{N}_{m,(k)}\right) = Var(E(\widehat{N_m}|\tilde{\boldsymbol{p}}_g))$ from step 3d. This value $Var\left(\widehat{N}_m\right)$ is an unconditional estimate of variance that includes uncertainty attributable to MCDS parameters and to the distribution of observed counts.

4. The estimate of abundance for the ensemble model, $\widehat{N}_{ens}$, is calculated as the arithmetic average of the $\widehat{N}_m$ from the candidate DSMs:

$$\widehat{N}_{ens} = \frac{1}{M}\sum_{m=1}^{M}\widehat{N}_m \qquad \text{Eqn 2}$$

5. Calculate the variance of model-averaged predictions using the standard unconditional variance estimator (i.e., Burnham and Anderson 2004, Eq. 4.9):

$$\widehat{Var}\left(\widehat{N}_{ens}\right) = \left[\sum_{m=1}^{M} w_m \sqrt{Var\left(\widehat{N}_m\right) + \left(\widehat{N}_m - \widehat{N}_{ens}\right)^2}\right]^2 \qquad \text{Eqn 3}$$

6. Because our estimate of transect detection probability, $p_{MR}\left(0, \boldsymbol{z_j}; \widehat{\boldsymbol{\theta}}_{MR}\right)$, was derived from aerial line-transect survey data and imagery collected in a different study area during a different year, we assumed that it was independent of the MCDS detection function and the DSMs. Therefore, we used the delta method to incorporate uncertainty from transect detection probability into the estimate of total uncertainty for the ensemble model estimate of abundance:

$$CV_{tot}\left(\widehat{N}_{ens}\right) = \sqrt{\frac{\widehat{Var}(\widehat{N}_{ens})}{\widehat{N}_{ens}^{\,2}} + \left\{CV\left[p_{MR}\left(0, \boldsymbol{z_j}; \widehat{\boldsymbol{\theta}}_{MR}\right)\right]\right\}^2} \qquad \text{Eqn 4}$$

Similarly, the delta function was used to estimate the total CV in estimated abundance from any single DSM:

$$CV_{tot}\left(\widehat{N}_m\right) = \sqrt{\frac{\widehat{Var}(\widehat{N}_m)}{\widehat{N}_m^{\,2}} + \left\{CV\left[p_{MR}\left(0, \boldsymbol{z_j}; \widehat{\boldsymbol{\theta}}_{MR}\right)\right]\right\}^2} \qquad \text{Eqn 5}$$

7. There were no estimates of uncertainty for availability probability, $\hat{p}_A$ (Ferguson et al. 2013); therefore, this parameter did not contribute to the estimated uncertainty in abundance.

Supplement 6
DSM Overview and Evaluation for the Eastern Bering Sea Beluga Case Study

Table S6.1. Key aspects of all 2017 Eastern Bering Sea beluga density surface models constructed and evaluated. Blue highlighting indicates models included in the ensemble.

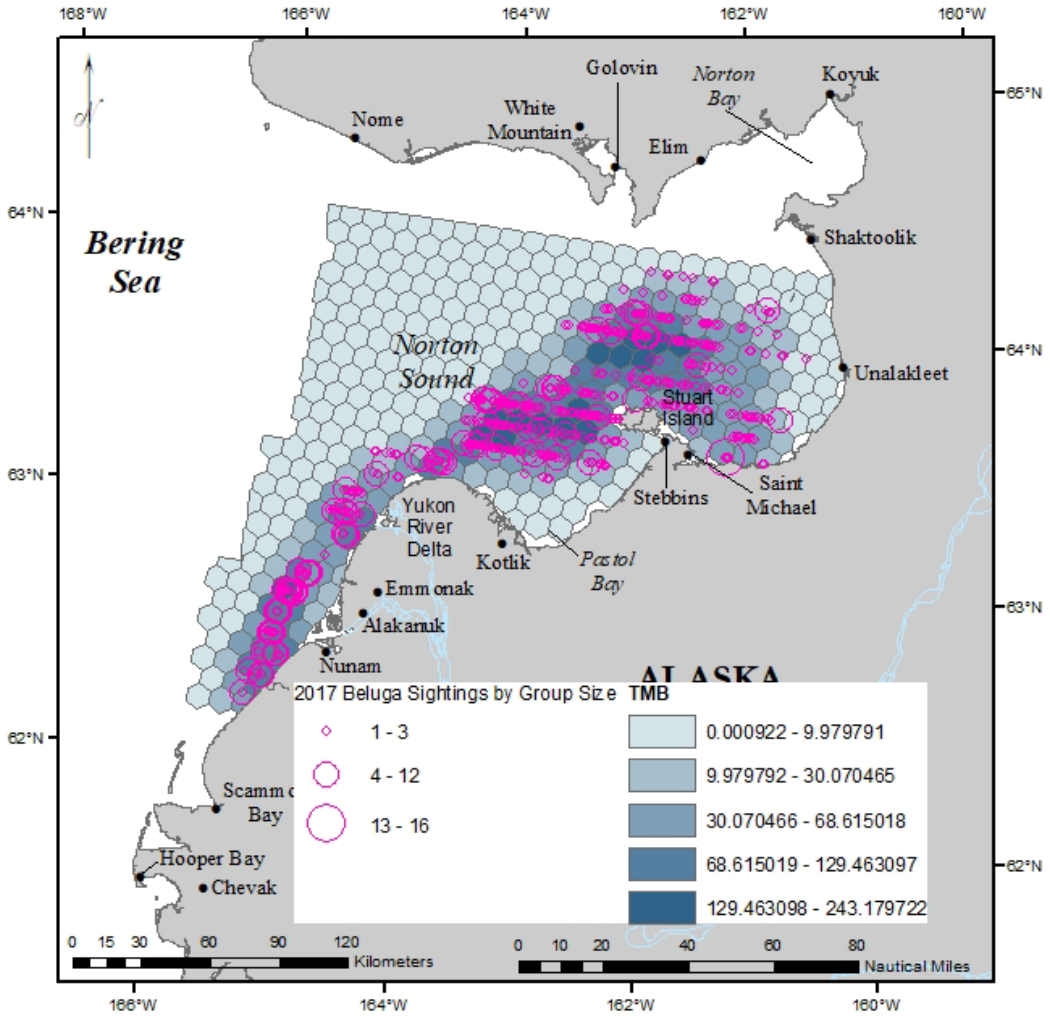
Year	Spatial Formulation	Number of Observations	Number of Transect Segments	Number of Prediction Cells	Number of Nodes	max.edge Inside (Outside)	Cutoff	Inner Boundary	Outer Boundary	Smoothing Spline Variables	number unsampled cells with extreme predicted abundance
2017	SPDE Matérn	598	604	342	595	25 (50)	0.2*max.edge	non-convex hull around all segment midpoints, plus a buffer of 25 km	offset from the inner boundary by a distance equal to 1/3 the latitudinal range of the study area	NA	0
2017	SPDE Matérn	598	604	342	564	25 (50)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2017	SPDE Matérn	598	604	342	528	25 (50)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 20% of the inner boundary diameter	NA	0
2017	SPDE Matérn	598	604	342	289	50 (100)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2017	SPDE Matérn	598	604	342	160	75 (150)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2017	SPDE Matérn	598	604	342	199	60 (120)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2017	SPDE Matérn with barriers	598	604	342	675	25 (50)	0.2*max.edge	coastline and study area boundaries	NA	NA	0
2017	SPDE Matérn with barriers	598	604	342	295	50 (100)	0.2*max.edge	coastline and study area boundaries	NA	NA	0
2017	SPDE Matérn with barriers	598	604	342	177	75 (150)	0.2*max.edge	coastline and study area boundaries	NA	NA	0
2017	SPDE Matérn with barriers	598	604	342	213	60 (120)	0.2*max.edge	coastline and study area boundaries	NA	NA	0
2017	Soap film smoother	598	604	342	NA	NA	NA	NA	NA	s(easting, northing, bs="so", k=30)	0
2017	Tensor product smoother	598	604	342	NA	NA	NA	NA	NA	te(easting, northing, bs="ts", k=15)	0
2017	Bivariate isotropic tprs	598	604	342	NA	NA	NA	NA	NA	s(easting, northing, bs="ts")	0

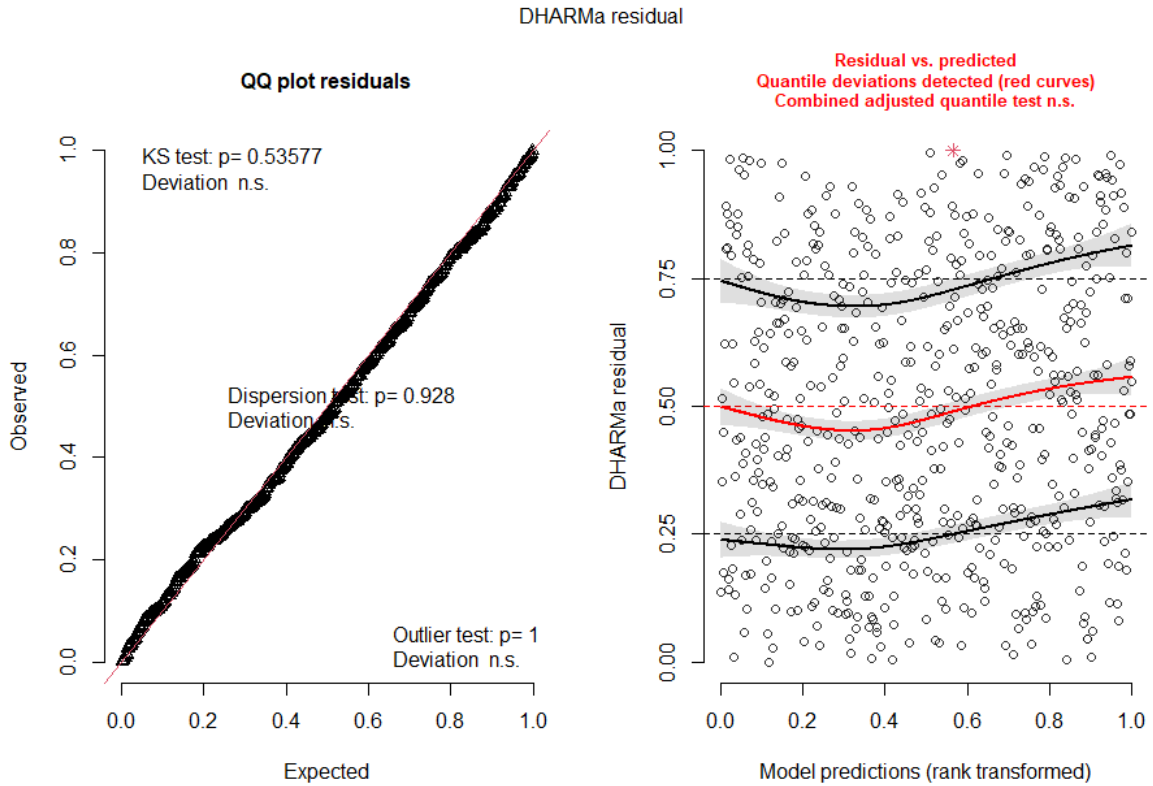
Table S6.2. Key aspects of all 2022 Eastern Bering Sea beluga density surface models constructed and evaluated. Blue highlighting indicates models included in the ensemble.

Year	Spatial Formulation	Number of Observations	Number of Transect Segments	Number of Prediction Cells	Number of Nodes	max.edge Inside (Outside)	Cutoff	Inner Boundary	Outer Boundary	Smoothing Spline Variables	number unsampled cells with extreme predicted abundance
2022	SPDE Matérn	241	317	554	831	25 (50)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2022	SPDE Matérn	241	317	554	257	75 (150)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2022	SPDE Matérn	241	317	554	351	50 (100)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2022	SPDE Matérn	241	317	554	308	60 (120)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2022	SPDE Matérn	241	317	554	326	55 (110)	0.2*max.edge	non-convex hull around all segment midpoints	offset from the inner boundary by a distance equal to 35% of the inner boundary diameter	NA	0
2022	SPDE Matérn with barriers	241	317	554	350	55 (110)	0.2*max.edge	coastline and study area boundaries	NA	NA	0
2022	SPDE Matérn with barriers	241	317	554	312	60 (120)	0.2*max.edge	coastline and study area boundaries	NA	NA	0
2022	SPDE Matérn with barriers	241	317	554	251	75 (150)	0.2*max.edge	coastline and study area boundaries	NA	NA	1
2022	SPDE Matérn with barriers	241	317	554	642	40 (80)	0.2*max.edge	coastline and study area boundaries	NA	NA	0
2022	Soap film smoother	241	317	554	NA	NA	NA	NA	NA	s(easting, northing, bs="so",k=60)	2
2022	Tensor product smoother	241	317	554	NA	NA	NA	NA	NA	te(easting, northing, bs="ts", k=15)	3
2022	Bivariate isotropic tprs	241	317	554	NA	NA	NA	NA	NA	s(easting, northing, bs="ts")	2

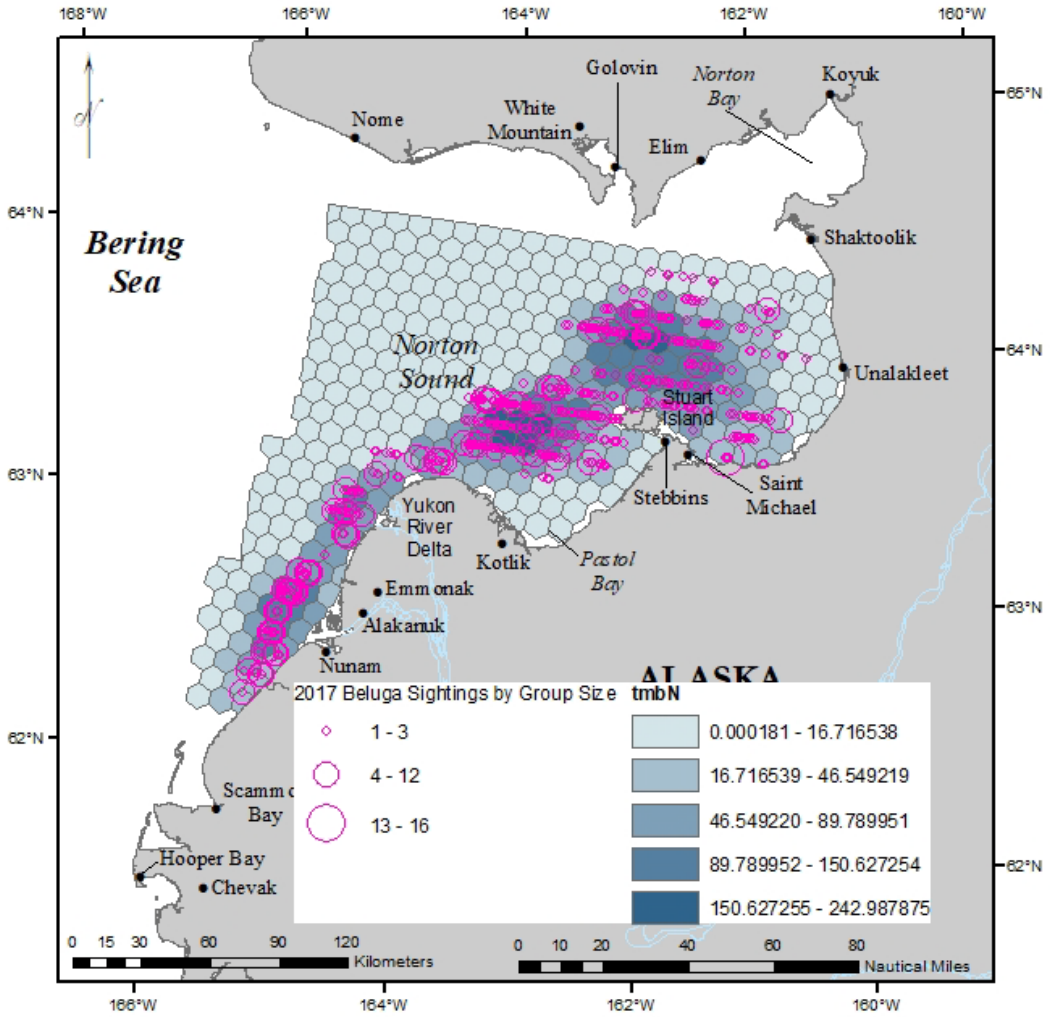
2017 candidate models in ensemble

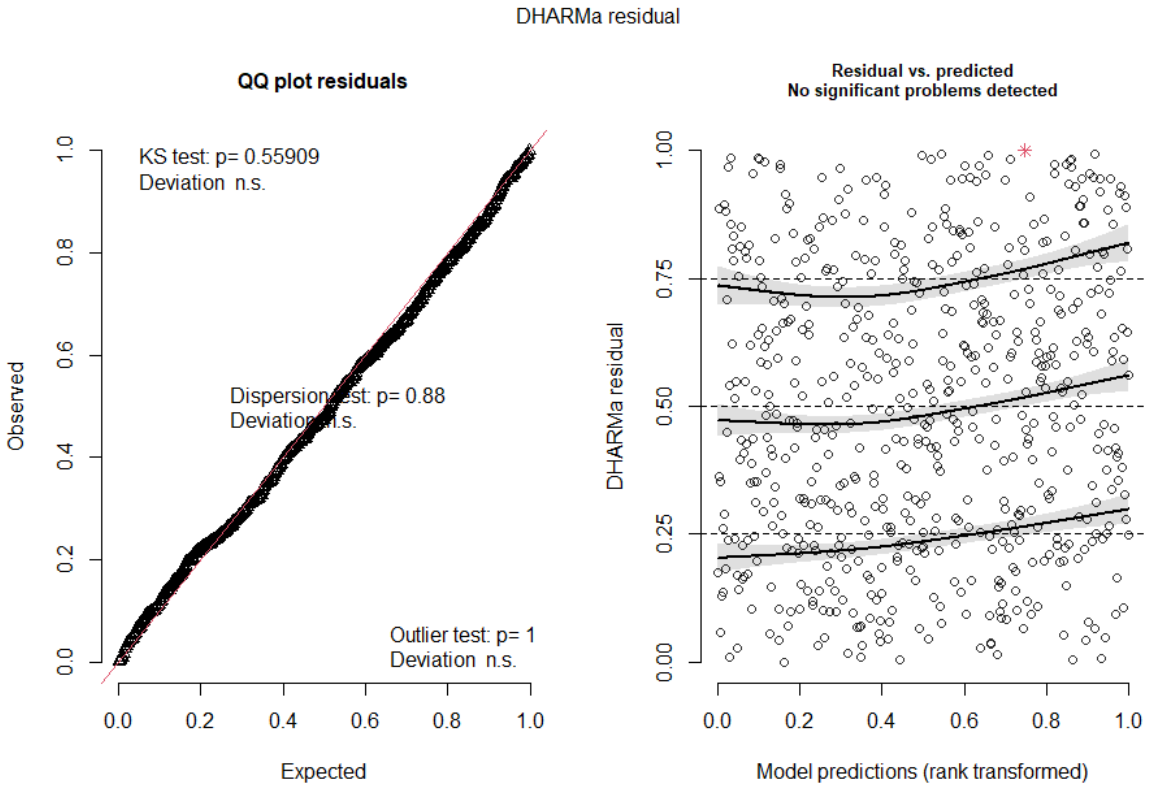
2017 TMB SPDE mesh6 (max.edge = 60 km)



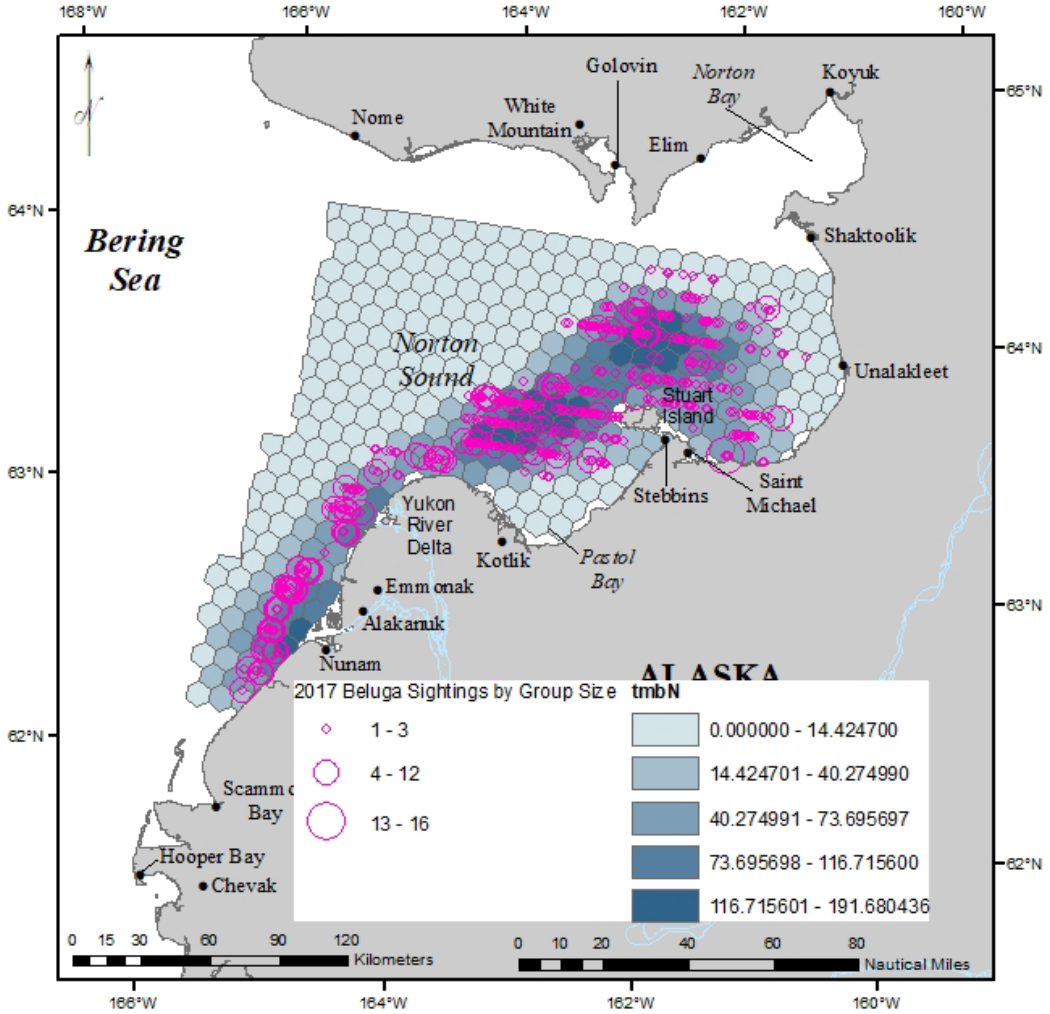


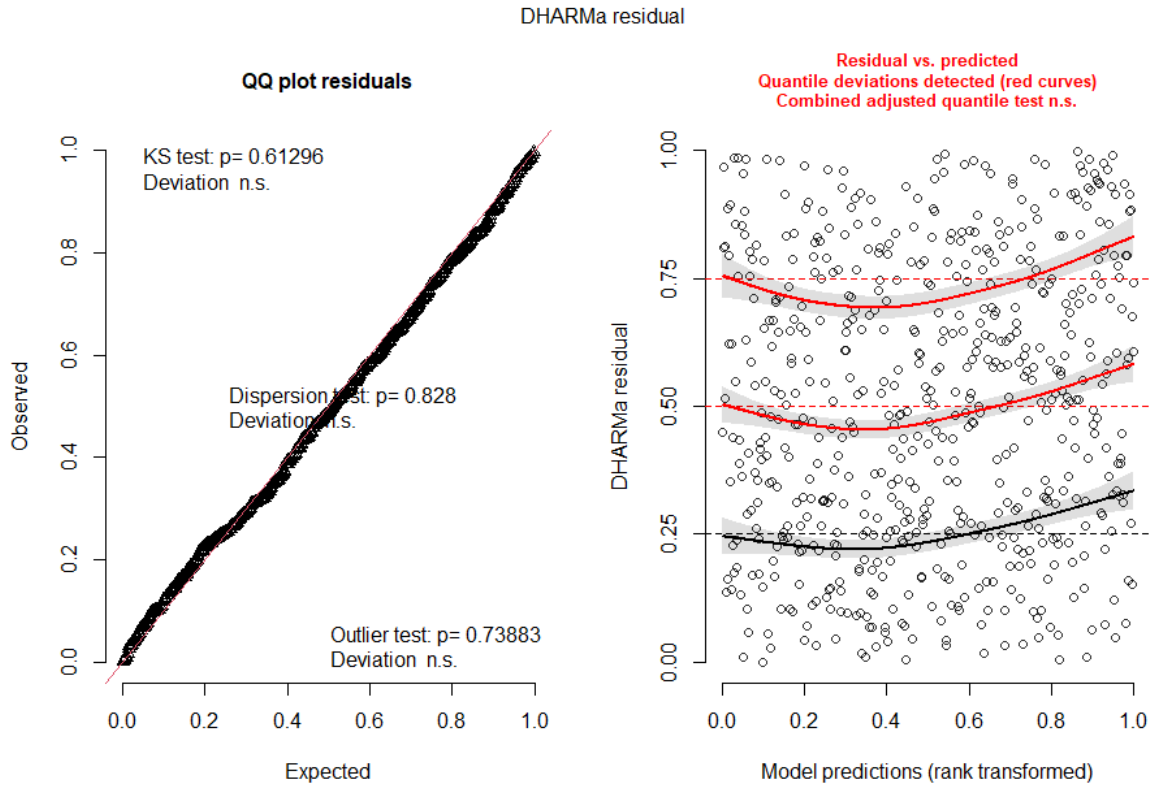
2017 TMB soap(easting,northing)



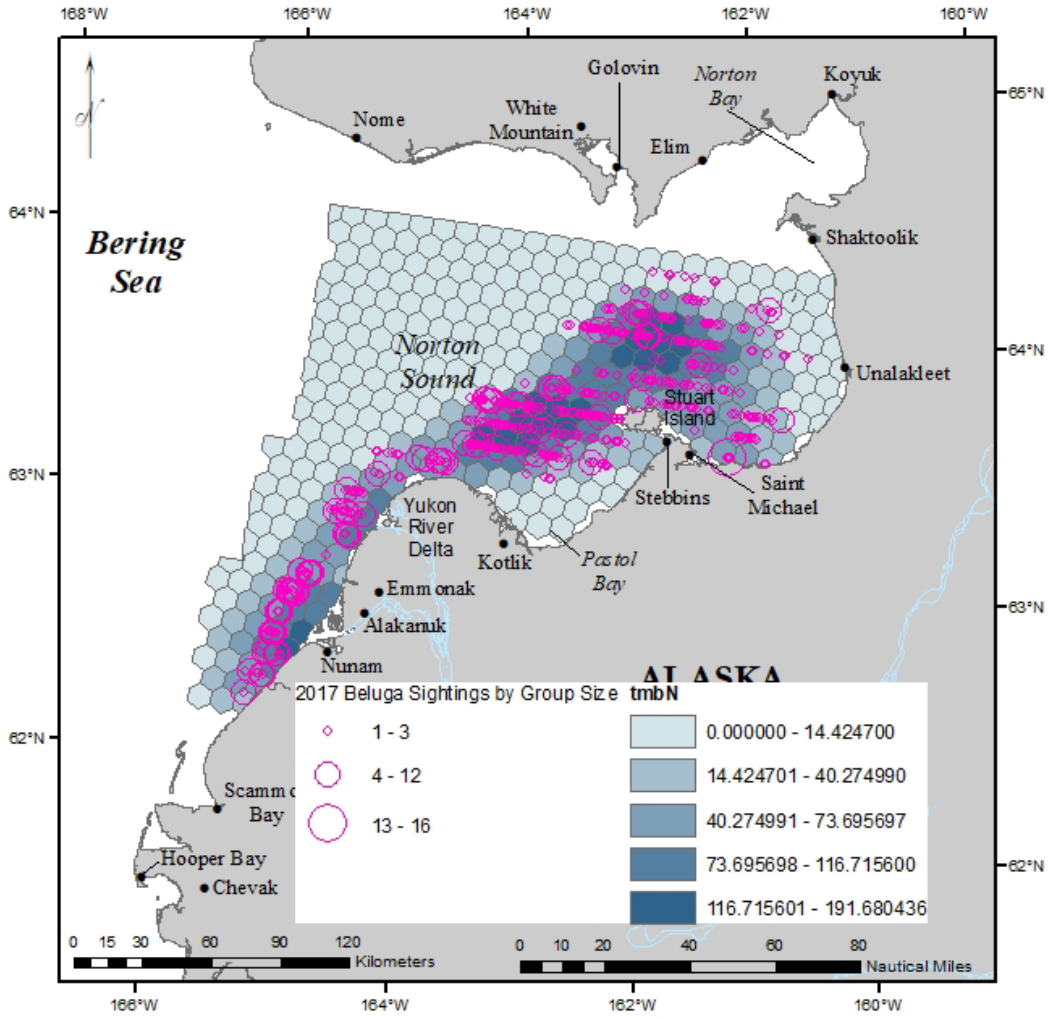


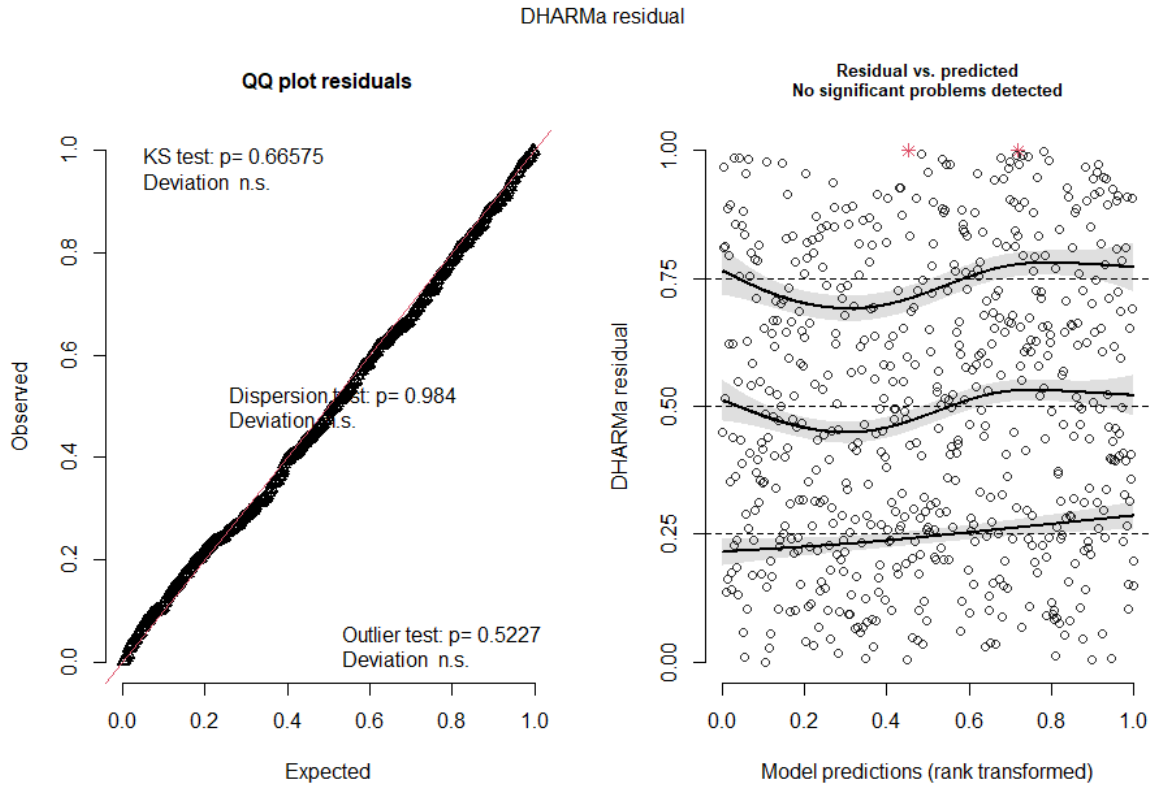
2017 TMB te(easting,northing)





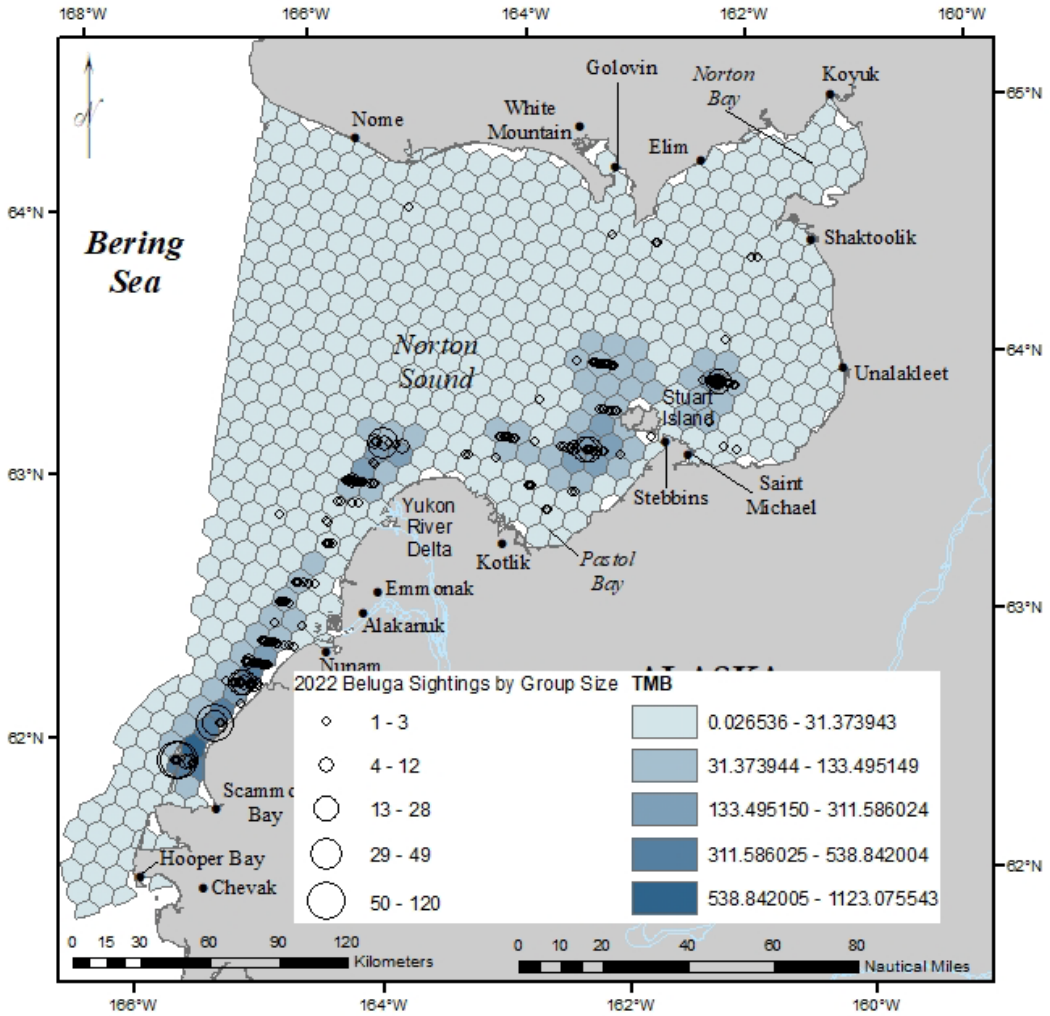
2017 TMB s(easting,northing,bs="ts")

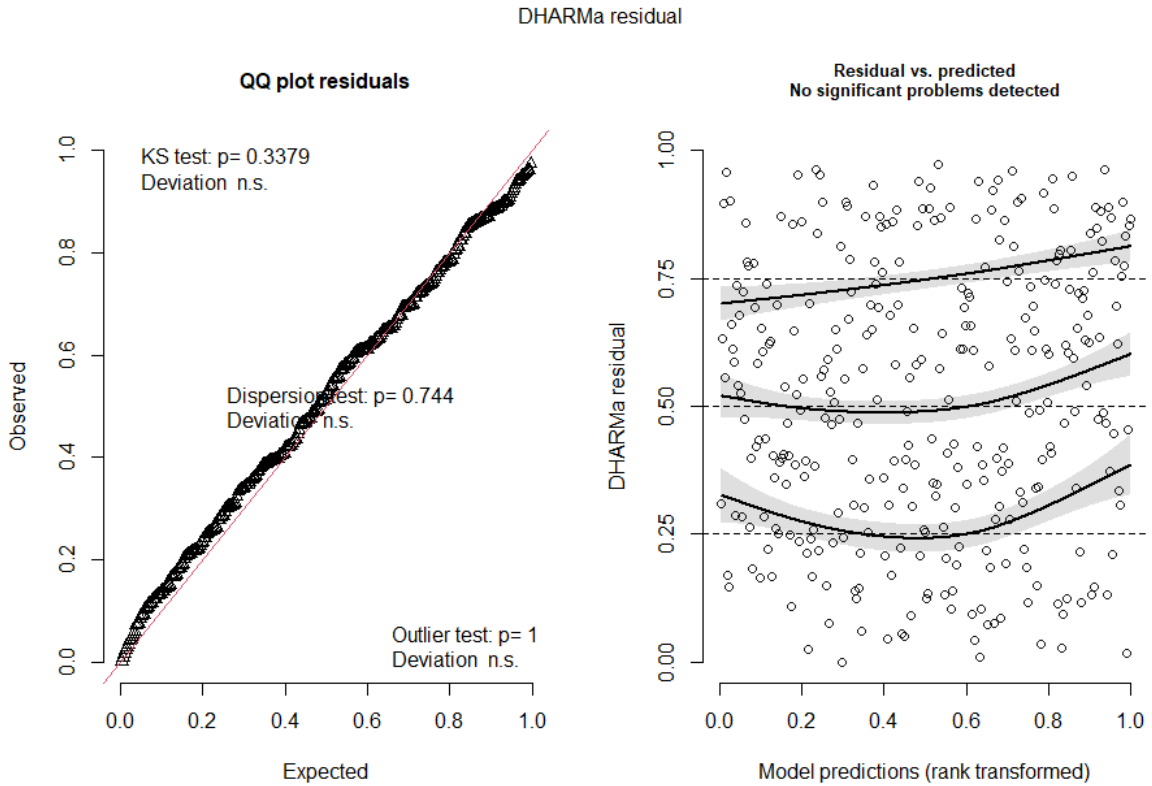




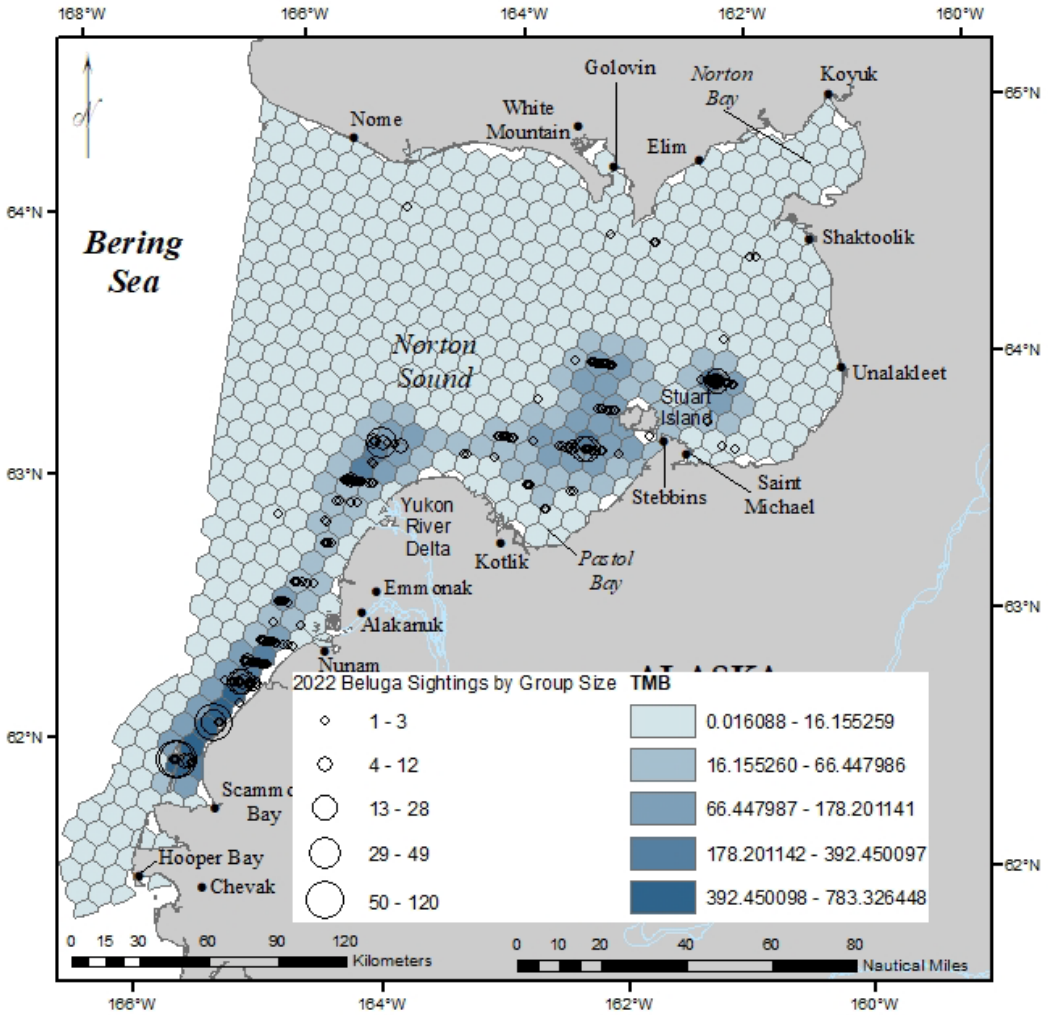
2022 candidate models in ensemble

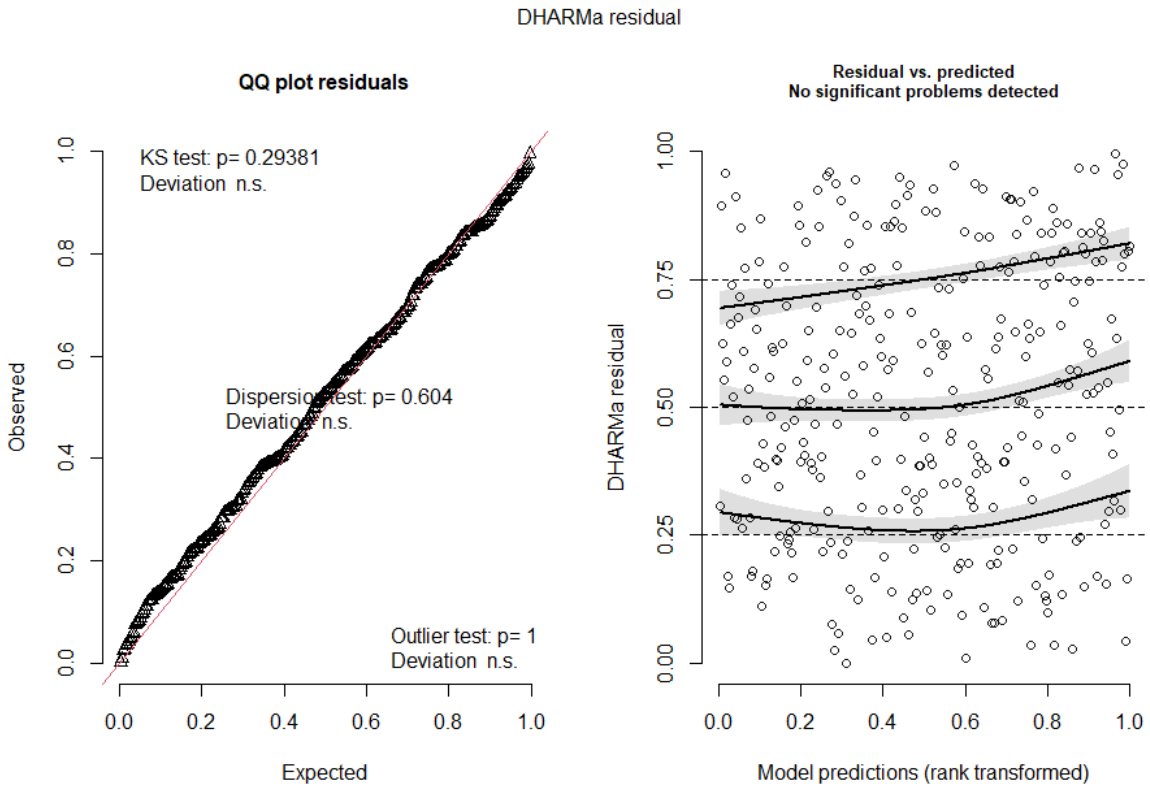
2022 TMB SPDE mesh5 (max.edge = 60 km)



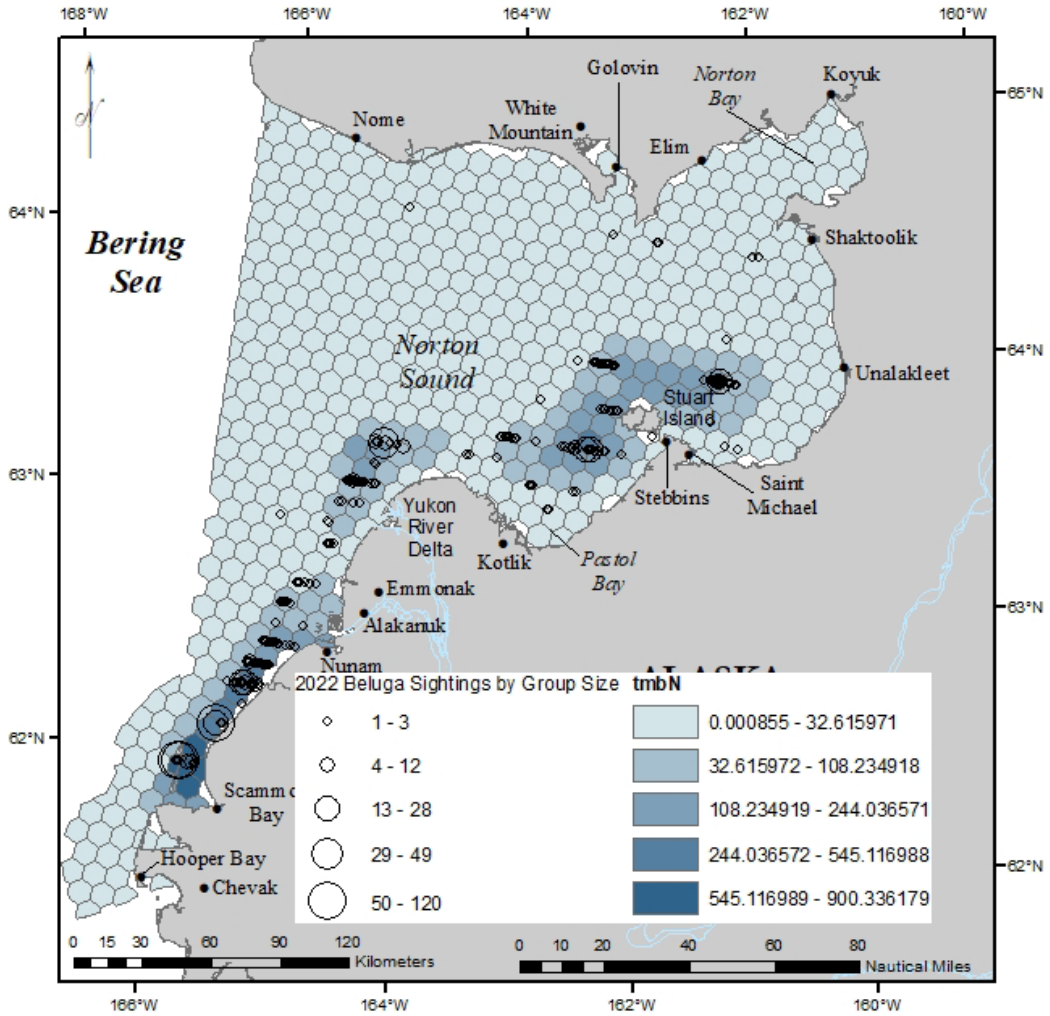


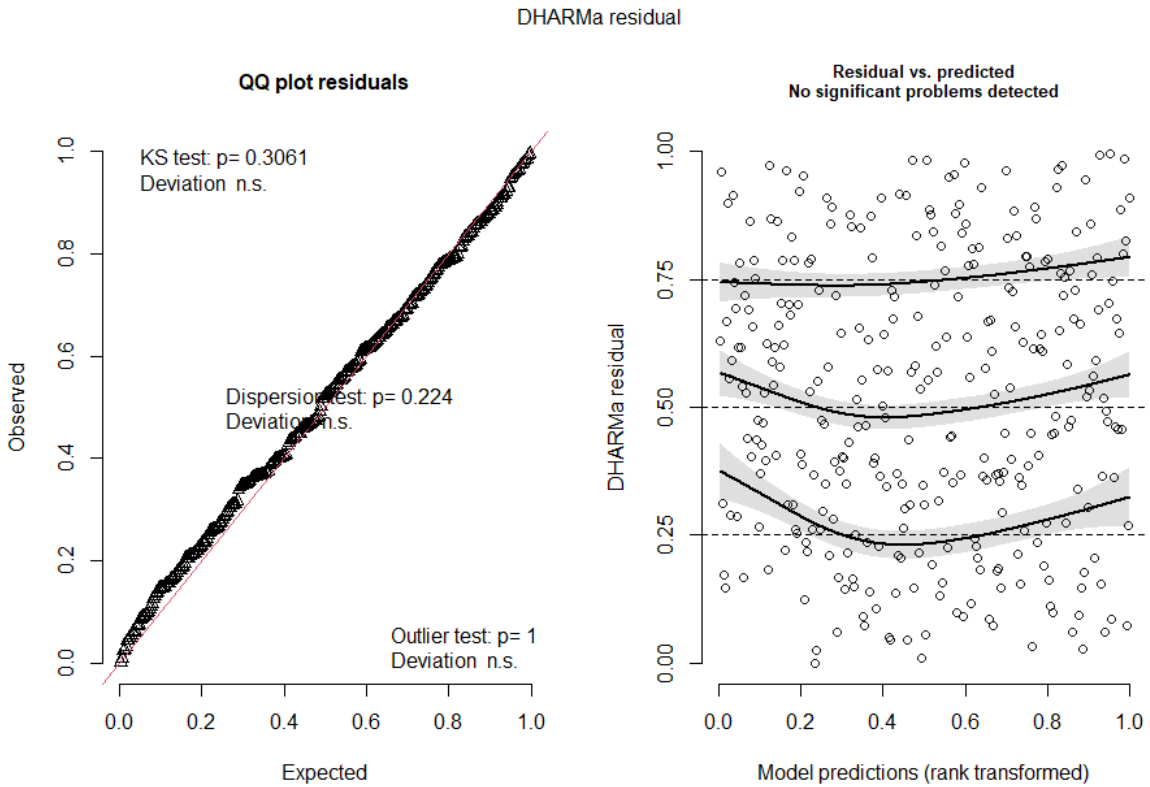
2022 TMB SPDE w/barriers mesh2 (max.edge = 60 km)



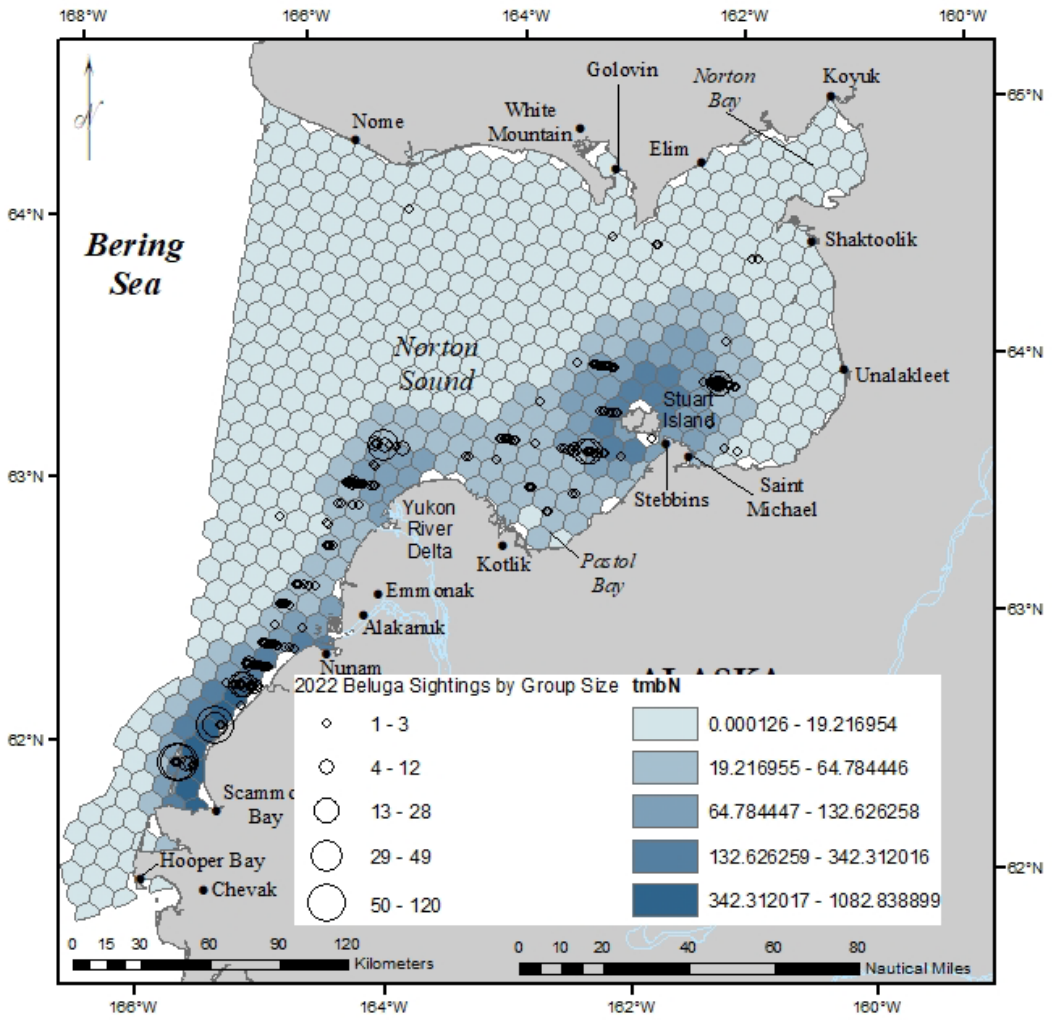


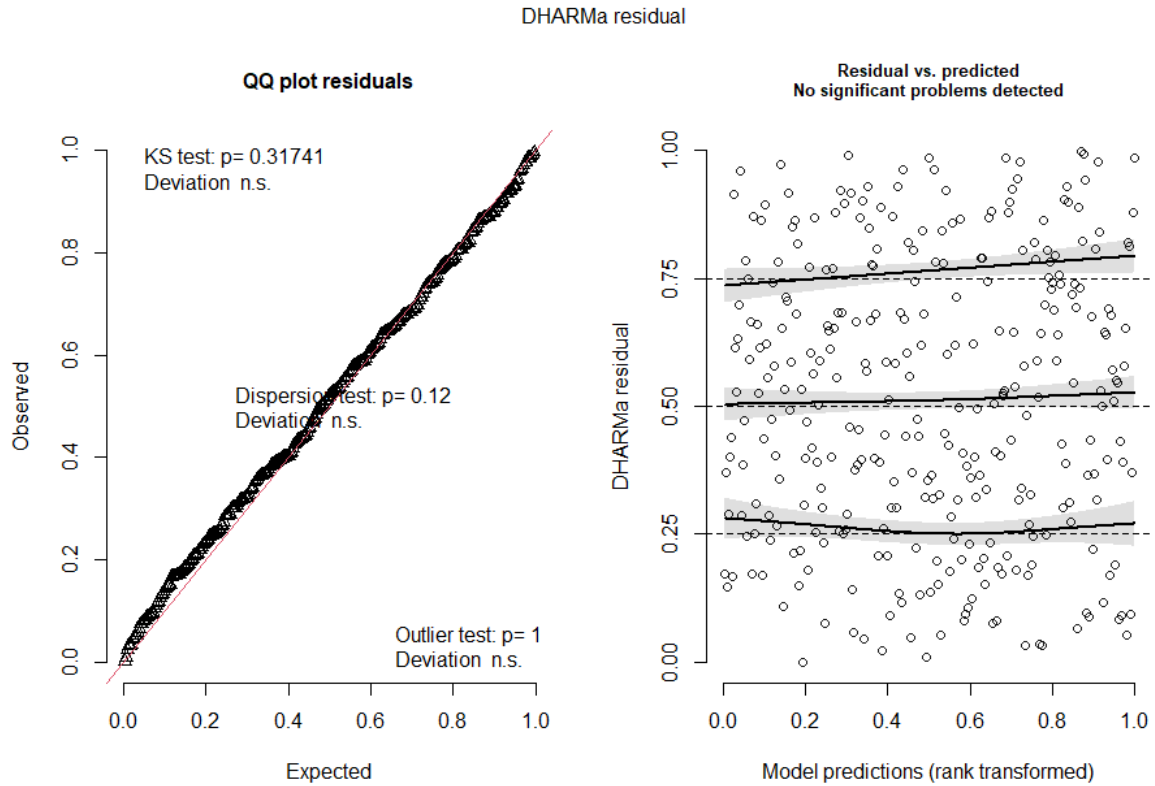
2022 TMB soap(easting,northing)





2022 TMB s(easting,northing,bs="ts")





What to do when you can't choose just one: