

# Adult Survival Rate and 2011 Abundance of Bering-Chukchi-Beaufort Seas Bowhead Whales From Photo-identification Data Over Three Decades

G. H. Givens<sup>1</sup>   J. A. Mocklin<sup>2</sup>   L. Vate Brattström<sup>3</sup>   B. J. Tudor<sup>4</sup>   W. R. Koski<sup>5</sup>  
J. E. Zeh<sup>6</sup>   R. Suydam<sup>4</sup>   J. C. George<sup>4</sup>

## Abstract

New photo-identification data were collected from a 2011 aerial survey of Bering-Chukchi-Beaufort Seas bowhead whales. We scored and matched these images to existing images from 1985, 1986, 2003, 2004, and 2005. Other interyear comparisons between this set of years were also conducted to generate a complete matching matrix for the 6 years. These data were used to estimate bowhead adult survival rate and population abundance using Huggins models embedded in a Robust Design capture-recapture analysis. Our estimated survival rate was 0.996 with approximate lower confidence bound 0.976, which is consistent with previous estimates and with research showing that bowhead lifetimes can be very long. Estimated 2011 abundance was 27,133 (CV=0.217, 95% CI 17,809 to 41,337). Although much less precise than the 2011 ice-based abundance estimate (16,820 with CV=0.052, 95% CI 15,176 to 18,643) of Givens et al. (2016), the 2011 photo-id estimate adds to the evidence that the stock is abundant, increasing from previous years, and unlikely to be harmed by limited subsistence hunting.

## 1 Introduction

The Bering-Chukchi-Beaufort (BCB) Seas stock of bowhead whales, *Balaena mysticetus*, is subject to subsistence hunting by Alaska natives along the whales' migratory route in the spring and fall. In addition to domestic regulation, this hunt is overseen by the International Whaling Commission (IWC), which sets annual and 6-year block hunting quotas using the BCB bowhead component of the Aboriginal Whaling Management Procedure (AWMP) being developed by its Scientific Committee (SC).

A critical component of management under the AWMP is periodic stock abundance estimation. The most recent estimate of 16,820 whales (95% CI: 15,176, 18,643) was based on data obtained

---

<sup>1</sup>Givens Statistical Solutions LLC, 4913 Hinsdale Drive, Fort Collins CO 80526 USA. geof@geofgivens.com.

<sup>2</sup>Shearwater Consultants LLC, 5829 NE 75th St., Seattle WA 98115, USA.

<sup>3</sup>Bardval, 6305 224th St. SW, Mountlake Terrace, WA 98043, USA.

<sup>4</sup>North Slope Borough, Department of Wildlife Management, Utqiagvik, Alaska, USA.

<sup>5</sup>LGL Limited environmental research associates, 22 Fisher Street, King City, ON, L7B 1A6, Canada.

<sup>6</sup>Department of Statistics, University of Washington, Seattle, WA, USA.

from an ice-based survey in 2011 (Givens et al., 2016). Previous ice-based surveys were successful in 2001 (George et al., 2004), 1993 (Raftery and Zeh, 1998), and several times in the 1980s (Zeh and Punt, 2005). These ice-based surveys used visual observers and a submerged hydrophone array to detect and locate whales passing through open leads beyond the shorefast ice.

This paper uses an altogether different sort of data. Photo-identification data are images of whales at or near the ocean surface taken from a low flying aircraft. Whales are identified, and their photos are matched within and between years on the basis of markings visible in the photographs.

Bowheads acquire markings on the rostrum, back and flukes in many ways. Although the epidermis is mainly black, it is common to see some white and gray coloring, particularly on the rostrum, peduncle and trailing fluke edges, especially in older animals. Furthermore, bowheads acquire distinct white marks or scars from other sources, including contact with the ice or seabed, entanglement in fishing gear, killer whale encounters, and other unknown sources. These marks are essentially permanent if the injury is deep enough to damage the melanocyte layer in the epidermis, which does not regrow (Haldiman and Tarpley, 1993).

Photographs of the same whale in different years can sometimes be matched on the basis of these distinct marks (Rugh et al., 1998). Some matches are obvious, e.g., from a 1.5 m long scar on the midback. Others are tricky, e.g., from a ‘constellation’ of 2 cm white specks in a unique configuration.

Independent of the ice-based surveys, aerial photo-id surveys have been conducted over the past 30 years and served as an alternate basis for abundance estimates. In this paper, we use these photo-id data from 1985, 1986, 2003, 2004, 2005 and 2011. One goal of our analysis is to derive an abundance estimate for comparison with the 2011 ice-based estimate. Another goal is to obtain an estimate of bowhead adult survival rate. Several studies have suggested that bowheads can live for a very long time, perhaps up to 200 years in some cases (George et al., 1999, 2011b; Rosa et al., 2013; Wetzal et al., 2017). Thus the annual bowhead survival rate is likely very high. Resighting animals over several photographic surveys has offered an opportunity for estimating bowhead survival using several different statistical approaches (Zeh et al., 2002; da Silva et al., 2007; Schweder et al., 2010).

These photo-id data provide the basis for mark-recapture modeling. Such analyses are commonly used in other wildlife settings to estimate abundance and survival, among other parameters. In many such cases, animals are captured, then tagged in some manner so that they may be re-identified if they are recaptured later. In the bowhead case, the initial ‘tagging’ of a whale is represented by noting its marks and, if appropriate, classifying it as being sufficiently well marked that it could be re-identified if photographed on a subsequent occasion. A recapture occurs (in a subsequent year) when a later photo can be matched to an earlier one on the basis of the markings. Thus, the abundance and survival rate *of marked animals* can be estimated. Given an estimate of the proportion of the population that is marked, we can then estimate total population abundance.

The remainder of this paper begins with a description of the aerial surveys and data collection. Following this, we present the main statistical model used for analysis: Huggins models embedded in a Robust Design. We pursue a BIC averaging approach over a large and complex space of potential models. Then we present estimates of adult survival and abundance. The paper ends with discussion of the results, including a comparison of survey methods and implications for management of this whale stock.

## 1.1 History of surveys

The 2011 survey introduced in this paper represents the latest in a series of BCB bowhead photo-id projects spanning several decades. Researchers conducting aerial surveys for bowhead whales in the Beaufort and Chukchi Seas and Baffin Bay during the mid-to-late 1970's noted that many of the whales that they saw had readily recognizable marks on them, and they photographed whales through windows in the side of the aircraft. These images were of poor quality for photo-identification because of distortion caused by the poor quality windows, and the oblique photographs did not provide a consistent view of the whales. Hence potential matches could not be unequivocally confirmed. During a 1981 aerial survey in the Beaufort Sea the first vertical bowhead whale photographs were obtained through a camera port in the floor of a Twin Otter using a Pentax  $6 \times 7$  medium format film camera (Davis et al., 1982). These photographs were shared with other bowhead researchers and it was agreed that the high quality of these photographs and the marks that could be seen could make a major contribution to the understanding of bowhead whale abundance and biology. In 1982, the National Marine Fisheries Service (NMFS) sponsored a dedicated effort to photograph bowhead whales in the Canadian Beaufort Sea and Amundsen Gulf where 453 different bowhead whales were photographed (Davis et al., 1983). During the following four years dedicated photographic surveys were conducted in the Canadian Beaufort Sea and Amundsen Gulf that were sponsored by NMFS, the Minerals Management Service (MMS, now Bureau of Ocean Energy Management, BOEM), the Canadian Government and various oil companies. These photographic surveys had varying objectives but were all related to being able to re-identify whales from day to day.

A research priority during these early years was to determine and monitor the population structure of BCB bowhead whales. Segregation in the summering areas (Cubbage and Calambokidis, 1987; Koski et al., 1988) made it difficult to obtain a representative sample of the whales in the summering areas. Thus, in 1984 NMFS investigated the potential for conducting these photographic surveys near Utqiagvik (previously, Barrow) where almost the entire population was believed to pass during the spring migration. In 1985, the first spring photographic survey was conducted using Barrow as a base (Nerini et al., 1987) and from 1985 to 1992 annual photographic surveys were conducted there (Angliss et al., 1995). Although surveys were incomplete in some years due to weather and technical problems, they made significant contributions to the BCB bowhead whale photo catalog which permitted estimation of additional life-history parameters such as calving intervals (Miller et al., 1992; Rugh et al., 1992b), growth rates of individual whales (Koski et al., 1992), survival of large whales (Zeh et al., 2002; da Silva et al., 2007; Schweder et al., 2010), first year survival (Koski et al., 2010a), better estimates of population structure (Koski et al., 2006), and estimates of population size (da Silva et al., 2000; Schweder et al., 2010; Koski et al., 2010b).

Several other studies conducted out of Kaktovik and Barrow that were funded by MMS supplemented the above studies from 1988-1994, 1998-2000 and 2006-2011. These latter studies provided only partial coverage of the spring or fall migrations but contributed significant numbers of photographs of whales to the BCB bowhead whale photo catalog that can be used for estimation of life-history parameters of bowhead whales.

In 2003-2005, the North Slope Borough (NSB) and NMFS sponsored spring surveys at Barrow (2003-2004) that covered the spring migration past Barrow as during the 1984-1992 studies there and in the Bering Sea (2005). The 2003-2004 surveys were conducted to obtain capture-recapture estimates of the bowhead whale population size (Koski et al., 2010b) that could be compared to

estimates from ice-based counts because it was feared that ice-based counts might not be possible in the future due to declining sea ice conditions.

The 2011 survey is a new addition to this history. It is presented in the next section.

## 2 Methods

A project this size draws upon methods and expertise from a variety of fields. The primary components of our study are: the collection and processing of raw image data, scoring and matching photographs, and statistical analysis of the resulting mark-recapture data. We organize our Methods section in this manner. We particularly focus on the 2011 data since many of the details about earlier years are documented elsewhere (Angliss et al., 1995; Koski et al., 1992, 2006, 2010a; Schweder et al., 2010; Zeh et al., 2002).

### 2.1 Imaging

The 2011 aerial abundance survey was conducted jointly by the North Slope Borough Department of Wildlife Management (NSB-DWM) and the US Marine Mammal Laboratory (MML). Methodology was similar to earlier studies (Koski et al., 1992; Angliss et al., 1995). Surveys were conducted in an Aero Commander 690 with bubble windows in the two forward observer positions. The aircraft flew at an average air speed of 217 km/h (117 kts) and altitude of 200 m (656 ft), and flew directly over bowhead whales during photographic passes.

Photographs were taken with a handheld Canon Mark III-1DS digital camera affixed with a Zeiss 85 mm fixed f/1.4 Planar T\* lens pointed directly downward through the aircraft's ventral camera port that was covered with optical quality glass. Shutter speed was typically 1/2500th second or faster. Aircraft altitude was recorded every 2 seconds with a portable Garmin 76CSx GPS unit and a laser altimeter. The altitude data were saved on a laptop computer and subsequently linked, by time, to the photographic data. Additionally, Robogeo software used this information to embed latitude, longitude and altitude into the Exif metadata of each photograph post-survey.

### 2.2 Scoring and matching images

The process of determining whether two photos are of the same whale is called 'matching'. 'Scoring' refers to the process of evaluating how useful an image is for matching. Scoring is completed before matching. No attempt is made to match some images with poor scores; see below.

The same methods for scoring and matching whale images were used as in past studies (Rugh et al. (1992a; 1998)). The only modifications relate to handling digital imagery and improvements to the electronic matching process.

Among the original raw photos, all photos of the same flight pass were identified. The number of whales in each pass was determined and the best image of each whale (that shows the most of the whale with the least amount of splash or glare) was cropped and labeled. When there were multiple whales in the frame, each whale was cropped to be alone in a print (if possible). Cow-calf pairs were cropped together in the same photo, assuming they were right next to each other.

Rugh et al. (1998) describe why it is important to distinguish photo quality from whale identifiability for capture-recapture estimation. For this reason, they developed a system of evaluating and scoring four zones on the bowhead whale independently for quality and identifiability in each

photograph. The entire dorsal side of a whale is rarely completely visible in a single photo and scores often vary a great deal among the zones in a photograph. Thus, the body was divided into four zones: mid-back (MB), lower back (LB), rostrum (R) and flukes (F).

All the cropped and labeled images were evaluated for how useful they would be in the matching phase. This depended on the quality of the image and whether whales had marks that would permit them to be re-identified in later photographs. Image quality was evaluated as 1+ (excellent), 1-, 2+(good), 2-, or 3 (poor). The degree of marking, i.e., identifiability, was evaluated as H+ (highly marked), H-, M+(moderately marked), M-, U+ (barely marked), U- (unmarked), or X (cannot tell). Each zone received separate quality and identifiability scores.

After whales were scored, matching was undertaken. There were three phases of matching. The first two phases (within-day and look-ahead) were done to identify multiple images of the same whale in the same year. The third phase, inter-year matching, produces the photo-id capture-recapture data analyzed below.

First, every image taken on a given day was compared to all others from that day. This helps identify whales that may otherwise never be matched, for example unmarked whales with an identifiable pattern of mud on the body or sloughing skin. This is a tedious phase, since we study every image taken, but it enables us to get the best information on the number of unmarked whales and allows us to increase the number of images per whale.

After identifying all individual whales in a day, the next phase, four-day look-ahead matching (4DLAM), could begin. The look-ahead matching was conducted to identify images of whales that linger during the migration and sometimes to include information on marks in zones that could not be seen in the best photograph of a whale. This prevents over-counting the number of unique individuals. By matching ahead (in time), whales that moved slowly through the area or that returned to Utqiagvik from areas to the east, and therefore might be re-photographed on a subsequent day, could be re-identified. The 4DLAM phase was constrained by removing all images with X in all zones for identifiability. Starting with images from 20 May, we removed the 4 day look-ahead restriction and matched all subsequent days for the rest of the season. This adheres with past matching protocol and follows the logic that the migration slows at the end of the season and the previous assumption that whales would leave the area after 4 days may no longer be valid. In particular, satellite telemetry has documented that starting in late May some whales may return to the Utqiagvik area after they migrated east as far as Amundsen Gulf and the eastern Beaufort Sea (Quakenbush et al., 2013). In previous efforts, five-day look-ahead was used, but past data supported 4DLAM as being sufficient.

When the intrayear matching effort was complete, whales with multiple images were assigned composite quality and identifiability scores for each zone. Composite scores represent the best possible ratings, and are integrative (as opposed to simple maximums) over the set of images for the whale, for each zone.

Finally, interyear matching was conducted. During this phase, the (set of) images of each whale photographed in one year was compared to the (set of) images of each whale photographed in another year. This phase was constrained by only searching images of at least moderately marked animals (M- and above in any zone).

The 1985 and 1986 images had been matched against each other previously (Koski et al., 1992). Similarly, years 2003, 2004 and 2005 had been previously matched against each other (Koski et al., 2010b). However, the 1985 and 1986 images had not been matched against 2003, 2004 or 2005. We performed that matching for this study. No past year had been matched against 2011. Therefore

we matched 2011 images against each of these previous years. This completed a full matching matrix between the years 1985, 1986, 2003, 2004, 2005 and 2011.

Consistency is very important in the scoring and matching processes, and is very hard to achieve. Lengthy training is given before anyone may score or match images. The people who have scored for quality and identifiability have been Gary Miller and Lisa Baraff for the 1985 and 1986 data and Julie Mocklin for 2003-2011. Matching has involved the following people: 1985-1986 were matched by Gary Miller, Bill Koski and Dave Rugh, 2003-2005 were matched by Julie Mocklin and Andrew Davis, and the remainder of the matching matrix was matched by Julie Mocklin, Linda Vate Brattstrom and Barbara Tudor.

Every match found was given a confidence score by three people independently and only matches with a confidence scores of at least 95% by all three people were counted as verified matches. For matches against 2011, a fourth independent match verification was made. For matched whales, integrative interyear composite identifiability and quality scores were made.

For the purposes of capture-recapture modeling, we used the interyear composite scores wherever possible, reverting to the within-year composites if necessary, and to the individual photo rating when only a single photo of a whale was available. This treatment of markedness reflects the fact that marks are essentially permanent and the identifiability scores are considered to reflect a characteristic of the whale itself. While accumulation of marks may have occurred between an initial capture and subsequent recaptures, the accumulation rates are low and we do not believe that they would prevent identification of a recaptured marked whale. There are possibly a few unmarked whales that might accumulate marks that would change their status from unmarked to marked between sampling periods but this was not incorporated into our models. This omission would introduce a slight positive bias in our abundance estimate. Finally, the interyear composite scores for years prior to 2011 did not include photographs obtained in 2011 because information from those photographs was not available when the 2011 photographs were matched to earlier years. This affects only the small number of 2011 matches and is not believed to introduce bias.

The use of composite quality reflects the integrative nature of the matching process. Specifically, composite scores represent the combined information about a zone across all available photographs. In fact, the composite quality score for a zone may be superior to the quality scores from any of the individual images used during the matching process if, for example, different portions of the zone are visible in different images. There is a weak tendency for whales photographed in multiple years to have more photos than whales seen in only one year. The effect of the number of photographs on capture probabilities is explicitly modeled to incorporate any such effect.

Our estimation of the number of marked whales in 2011, denoted  $N_{2011}^M$ , requires a precise definition of what constitutes a marked whale. First we established whether each zone for a whale is marked following the approach of Zeh et al. (2002). A zone with quality 2- or better and identifiability M- or better is called marked. In addition, a zone with quality 3 is called marked if the identifiability is M+ or better. In such cases, the detectable markings are so identifiable that they would be recognized in a subsequent image of that zone. Such cases are very rare. A zone is called unmarked if it has quality 2+ or better with identifiability U+ or U-. A zone is indeterminable if it is not called marked or unmarked, or if its identifiability is scored as X.

Second, we defined whether a whale is marked based on its zones. We use a modified version of the screen process of Koski et al. (2010b) which was designed to provide a larger sample size of marked whales, compared to considering only a single zone, without the possibility of including a whale more than once. Without such a precaution, duplication could happen if, for example, the

midback was not visible in one photo while the lower back was marked and the lower back was not visible in another but the midback was marked.

Initially, a whale is defined as marked if it is marked in the midback zone. From the remaining whales, we considered only those with unmarked midbacks. Any such whales that were marked in the lower back were added to the sample of marked animals. Next, whales that are unmarked in the midback and lower back are called marked if they are marked in the rostrum zone. Finally, whales that are unmarked in the previous three zones are called marked if they are marked in the flukes. All whales not designated as marked during this process are called unmarked.

This screening process is a source of heterogeneity in capture probabilities. Other sources include natural differences in how well marked individual whales are, individual covariates associated with images and screening, and annual survey effort. Such heterogeneity is addressed in our statistical modeling below.

Note that a different set of marked whales would be produced if the screening process examined the zones in a different order. The opinion of the matchers was that the midback zone should definitely be treated first because it is the zone most often marked and most easily evaluated since it tends to be above the water surface in the largest number of usable photos. The matchers considered flukes to be the least useful zone. Thus, we considered three zone orderings for the screening process: (MB, LB, R, F), (MB, R, LB, F), and (MB, R, F, LB). The last of these is the ordering used by Koski et al. (2010b). We chose the first of these because it produced the largest number of marked whales while remaining consistent with matcher opinion. Thus, for the purposes of our analysis, our screening process with this zone ordering was applied to all the photos from every year.

## 2.3 Statistical model

The primary goals of our analysis are the estimation of adult survival rate,  $\phi$ , and whale abundance in 2011, denoted  $N_{2011}$ . To produce the latter, we first estimate the abundance of marked whales in 2011,  $N_{2011}^M$ , and the proportion of marked whales in the population,  $p^*$ .

### 2.3.1 Marked abundance, $N_{2011}^M$

We adopted the Robust Design model (Pollock, 1982; Kendall and Nichols, 1995; Kendall et al., 1995, 1997), which combines aspects of both closed and open population models. This model assumes that there are several primary sampling occasions, or ‘periods’, over a long-term study. The population is assumed to be open (e.g., there may be births and deaths) from one period to the next. Within each primary period there are several secondary sampling occasions. The population status between secondary occasions (within the same primary period) is assumed to be closed.

If the secondary data are pooled within primary occasions, then the mark-recapture data could be analyzed using an open Jolly-Seber model (Jolly, 1965; Seber, 1965). In contrast, within any single primary period, one could apply a closed model to the secondary sampling occasions. Our approach uses the model of Huggins (1989) within each closed period. This allows for capture probabilities to depend on individual covariates.

The bowhead data are well-suited for the Robust Design model with a Jolly-Seber primary model and Huggins secondary model. There are three primary periods: 1985-86, 2003-05, and 2011. The third primary period has only a single secondary occasion, and therefore does not conform to the Robust Design. However, one can create dummy data for 2012, consisting of no

captures whatsoever, then apply the Robust Design to the augmented data while constraining the final capture probability to equal zero (P. Conn and B. McClintock, pers. commn).

Bowhead productivity and, especially, mortality are quite low. Indeed, the minimum calving interval may be 3-4 years (Miller et al., 1992; Rugh et al., 1992b; George et al., 2011a). Several analyses indicate that bowheads may routinely live over 100 years and some possibly over 200 years (George et al., 1999, 2011a; Rosa et al., 2013; Wetzal et al., 2017). It is therefore reasonable to adopt the closed population assumption for primary periods since these span only two or three years. Koski et al. (2010b) used closed population models for previous analysis of the 2003-05 period. We also assume a constant survival rate,  $\phi$  over the entire study period.

Kendall et al. (1995) present the full likelihood for the Robust Design. This can be simplified here for  $k = 3$  primary periods with 2, 3, and 2 secondary periods (with time spacing matching our surveys) as follows.

Let  $i$  index primary occasion and using standard capture-recapture terminology define:

$U_i$  = number of unmarked whales just before the  $i$ th primary period (noting that in our case this actually refers to marked whales not yet captured in a photograph)

$u_i$  = number of whales from  $U_i$  that are captured in the  $i$ th primary period

$m_{hi}$  = number of whales captured in primary period  $i$  that were last captured in primary period  $h$  (and 0 for the first primary period)

$n_i = u_i + \sum_{h=1}^{i-1} m_{hi}$  = total number of whales captured in primary period  $i$

$X_{0i}^{\omega(\ell)}$  = number of whales from  $u_i$  having capture history  $\omega(\ell)$  (of length  $\ell$ , which will vary by context) within primary period  $i$

$X_{hi}^{\omega(\ell)}$  = number of whales from  $m_{hi}$  having capture history  $\omega(\ell)$  (of length  $\ell$  which will vary by context) within primary period  $i$

$r_i$  = number of individuals from  $n_i$  that are recaptured in some subsequent primary period

$p_{ij}$  = probability that a whale is captured in secondary sample  $j$  of primary period  $i$ , given that it is alive

$q_{ij} = 1 - p_{ij}$

$p_i^*$  = probability that a whale is captured in at least one of the occasions within the  $i$ th primary period, calculated as  $p_i^* = 1 - \prod_j q_{ij}$

$q_i^* = 1 - p_i^*$

$\phi$  = annual survival rate

Since photographing (i.e., ‘capturing’) whales is non-invasive, there is no difference between the probability of capture and the probability of recapture. In standard mark-recapture notation, this means that we adopt the assumption that  $c_{ij} = p_{ij}$ . This is why the above definitions do not include  $c_{ij}$ . Also, we assume no immigration, emigration, or similar changes in availability. In the terminology of the MARK software package, this means  $\gamma' = 1$  and  $\gamma'' = 0$ .

The likelihood for our model is a product of conditional multinomials:

$$L\left(\{X_{0i}^{\omega(\ell)}\}, \{X_{hi}^{\omega(\ell)}\}\right) = L_1(\{u_i\}|\{U_i\}, \{p_i^*\}) \times L_2(\{m_{hi}\}|\{n_i\}, \{p_i^*\}, \phi) \\ \times L_3\left(\{X_{0i}^{\omega(\ell)}\}, \{X_{hi}^{\omega(\ell)}\}|\{u_i\}, \{m_{hi}\}, \{p_{ij}\}\right).$$

The first two components are exactly the Jolly-Seber model if within-primary-period capture information is ignored. Specifically,

$$\begin{aligned} L_1 &= \prod_{i=1}^3 \binom{U_i}{u_i} p_i^{*u_i} q_i^{*U_i-u_i} \\ L_2 &= L_{21} \times L_{22} \end{aligned}$$

where

$$\begin{aligned} L_{21} &= \binom{n_1}{m_{12} \ m_{13}} (\phi^{17} p_2^*)^{m_{12}} (\phi^{23} q_2^* p_3^*)^{m_{13}} ((1 - \phi^{17}) + \phi^{17} q_2^* (1 - \phi^6) + \phi^{23} q_2^* q_3^*)^{n_1 - r_1} \\ L_{22} &= \binom{n_2}{m_{23}} (\phi^6 p_3^*)^{m_{23}} ((1 - \phi^6) + \phi^6 q_3^*)^{n_2 - r_2}. \end{aligned}$$

after noting, for example, that  $\phi^{17}$  is the probability of surviving from 1986 to 2003. The third component can be written as

$$L_3 = L_{31} \times L_{32} \times L_{33}$$

where

$$\begin{aligned} L_{31} &= \begin{pmatrix} u_1 \\ X_{01}^{10} & X_{01}^{01} & X_{01}^{11} \end{pmatrix} \left( \frac{p_{11} q_{12}}{p_1^*} \right)^{X_{01}^{10}} \left( \frac{q_{11} p_{12}}{p_1^*} \right)^{X_{01}^{01}} \left( \frac{p_{11} p_{12}}{p_1^*} \right)^{X_{01}^{11}} \\ L_{32} &= \begin{pmatrix} u_2 \\ X_{02}^{100} & X_{02}^{010} & X_{02}^{001} & X_{02}^{110} & X_{02}^{011} & X_{02}^{101} & X_{02}^{111} \end{pmatrix} \\ &\quad \times \begin{pmatrix} m_{12} \\ X_{12}^{100} & X_{12}^{010} & X_{12}^{001} & X_{12}^{110} & X_{12}^{011} & X_{12}^{101} & X_{12}^{111} \end{pmatrix} \\ &\quad \times \left( \frac{p_{21} q_{22} q_{23}}{p_2^*} \right)^{X_{\bullet 2}^{100}} \left( \frac{q_{21} p_{22} q_{23}}{p_2^*} \right)^{X_{\bullet 2}^{010}} \left( \frac{q_{21} q_{22} q_{23}}{p_2^*} \right)^{X_{\bullet 2}^{001}} \\ &\quad \times \left( \frac{p_{21} p_{22} q_{23}}{p_2^*} \right)^{X_{\bullet 2}^{110}} \left( \frac{q_{21} p_{22} p_{23}}{p_2^*} \right)^{X_{\bullet 2}^{011}} \left( \frac{p_{21} q_{22} p_{23}}{p_2^*} \right)^{X_{\bullet 2}^{101}} \left( \frac{p_{21} p_{22} p_{23}}{p_2^*} \right)^{X_{\bullet 2}^{111}} \\ L_{33} &= \begin{pmatrix} u_3 \\ X_{03}^{10} & X_{03}^{01} & X_{03}^{11} \end{pmatrix} \begin{pmatrix} m_{13} \\ X_{13}^{10} & X_{13}^{01} & X_{13}^{11} \end{pmatrix} \begin{pmatrix} m_{23} \\ X_{23}^{10} & X_{23}^{01} & X_{23}^{11} \end{pmatrix} \\ &\quad \times \left( \frac{p_{31} q_{32}}{p_3^*} \right)^{X_{\bullet 3}^{10}} \left( \frac{q_{31} p_{32}}{p_3^*} \right)^{X_{\bullet 3}^{01}} \left( \frac{p_{31} p_{32}}{p_3^*} \right)^{X_{\bullet 3}^{11}} \end{aligned}$$

and  $X_{\bullet i}^{\omega(\ell)} = X_{0i}^{\omega(\ell)} + \sum_{h=1}^{i-1} X_{hi}^{\omega(\ell)}$  for  $i = 2, 3$ . Further explanation of the general model is given by Kendall et al. (1995).

Capture probabilities are modeled to depend on covariates, as introduced by Huggins (1989). Specifically, we adopt the model

$$\text{logit}(p_{ijk}) = \alpha + \mu_{ij} + \boldsymbol{\beta}^T \mathbf{X}_k$$

where  $\alpha$  and the  $\mu_{ij}$  are parameters establishing the baseline model,  $\mathbf{X}_k$  is a vector covariates and  $\boldsymbol{\beta}$  is a parameter vector. Here, the indices indicate that the capture probability and covariates refer to the  $k$ th whale in the  $j$ th occasion of the  $i$ th primary period.

Our models were fit with the MARK software (MARK Team, 2017). The “alternate optimization method” (namely simulated annealing) was used due to convergence difficulties with the standard approach.

It is evident that in  $L_{33}$  there is no information about the  $p_{3j}$  provided by the  $X_{\bullet 3}^{\omega}$  data. This is intuitively clear because the absence of recapture events means there is no evidence about capture probabilities. Notwithstanding this, the third primary occasion does come in to play for  $L_1$  and  $L_2$ , including the dependency between capture probabilities and the covariates. Hence, when constructing a design matrix for fitting our model in MARK, we limited covariate effects to the earlier primary periods to ensure estimability.

As noted above, the annual adult survival rate  $\phi$  is modeled as constant, independent of covariates and unchanging over time. Although it is possible that changing environmental factors or population density might affect the survival rate, there is scant evidence to support this. For example, large mortality events have not been observed for bowhead whales like they have for some other mysticete whales like gray whales (Rugh et al., 2005) and southern right whales (Uhart et al., 2009).

### 2.3.2 Model selection

The simplest plausible model (“baseline model”) for capture probabilities must allow them to vary between years since survey effort, average photo-quality, rating team and various other factors varied annually. In addition, we considered several dozen possible predictors derived from the variables described in Table 1.

We used BIC to rank and compare models. Raftery (1995) provides a very useful discussion of how BIC can be used for this purpose. BIC was selected because it tends to favor parsimonious models. The weight of evidence in favor of one model compared to another can be evaluated as follows: BIC difference of 0–2 gives “weak” evidence with the posterior probability of the favored model about 0.50–0.75 (assuming equal a priori probabilities); difference of 2–6 is “positive” evidence with probability 0.75–0.95; 6–10 gives “strong” evidence with probability 0.95–0.99; and >10 gives “very strong” evidence with probability >0.99 (Raftery, 1995).

The model selection process followed the following steps.

1. We fit separate models using each possible predictor individually. We used BIC to partition the predictors into ones yielding models with BIC better than the baseline model, and the rest.

Aside from the baseline model being a natural choice for the ‘worst’ model considered, there are two additional reasons we used this standard here and in other steps below.

First, Madigan and Raftery (1994) suggest to drop all models with BIC more than 6 greater than the best model. This corresponds to models 20 times less likely than the best, under standard Bayesian modeling approximations those authors describe. Our BIC difference cutoff is about 12, corresponding to models that are up to 400 times less likely. Thus, our choice is very conservative in the sense that we entertain a very large and diverse model set.

The second reason pertains to the pattern of BIC values seen in the model set. There were three standout best models, followed by a sequence of increasingly poor models with BIC values trending steadily upward until models about as good as the baseline model were

variable	definition
bz	Zone with best identifiability, as indicator functions for each zone. Ties are broken in the zone rank order (MB, LB, R, F).
bzi	Best zone identifiability score, as indicator functions for H+, H-, M+, M-. Also I(bzi=H+ or H-) was used.
nz	Number of zones marked M- or better. This variable ranged from 1 to 4, treated continuously. I(nz≤2) was also used.
np	Maximum number of photos of a whale in one year. This ranged from 1 to 6, or 7+. Treated as continuous.
isc	Overall identifiability score. Each of the four zones was assigned a numerical value from H+=4 to M-=1 (or worse = 0). The mean of these four numbers equals the covariate value.
mz	Indicates if zone is marked (M- or better for 2- or better quality, or M+ or better for 3 quality). Coded as indicator variables for each zone separately.
tz	Zone that first triggers the decision to designate the whale as marked. Indicator functions for each zone.
tzi	Identifiability in the zone that triggered the markedness declaration. Indicator functions for H+, H-, M+, M-.
yb	Year block (85/86, 03/04/05, 11/dummy) as two indicator functions. Useful for interactions with variables above.
yr	Year. All models assume the capture probabilities vary between years.

Table 1: Key predictors considered during model selection. A variety of related variables and alternative codings were also considered during an exploratory phase prior to the formal analysis.

reached. At this point, there was a fairly abrupt BIC jump to the remaining, inferior models. Thus the baseline model was a sensible stopping point.

2. For all the predictors retained in the first step (and for some of the remaining ones), we fit a new version where the effect of the predictor varied with time block (i.e., 1985/1986 versus 2003/2004/2005). After this, the remaining variables and their year-block interactions were isc, bzi (three codings), nz (two codings), tzi, np, and mz (for lower back).
3. At this point, isc and I(bzi=H+) each yielded a model that was clearly superior to all others. Thus we fit two new sets of models. The first set included isc and, sequentially, each of the other retained predictors. The second set was analogous, but included I(bzi=H+) in each model.
4. The four predictors with main (additive) effects in at least one retained model were isc, bzi, nz, and tzi. Each of these had a best coding: for example, I(bzi=H+) was the best way to use the information in the four bzi indicator variables. We next examined all models with three additive effects from these four best codings. None of these met our BIC threshold.
5. We next tried interactions between isc and each variable having a main (additive) effect in any retained model. Then we repeated this looking for interactions with I(bzi=H+). Of all these potential interaction terms, only one merited retention.
6. The above process involved evaluating a large number of models. Models with BIC worse than the baseline were discarded. We also discarded a few models where convergence was not assured since the optimization routine in MARK led to survival rate estimates equaling 1. Most of these had very poor BIC values compared to the remaining models.

At the end of this process, 17 models plus the baseline were retained. These are given in Table 4. Note that the range of BIC values extends to the point where any inferior models have “very strong” evidence against them.

Our main results rely on model averaging with BIC weights. The weights are calculated for model  $m = 1, \dots, M$  as

$$w_m = \exp\{-BIC_m/2\} / \sum_{\ell=1}^M \exp\{-BIC_\ell/2\}.$$

These weights are conveniently calculated from BIC differences in a table of results for the set of  $M$  models. Raftery (1995) notes that these weights approximate the posterior distribution of the models in the model space under a uniform prior. Therefore, weighted averages of real (e.g.,  $\hat{\phi}$ ) and derived (e.g.  $\hat{N}_{2011}^M$ ) parameter estimates effectively represent posterior means.

For estimating the standard error of the 2011 abundance estimate, we apply the ‘unconditional’ variance calculation using BIC weights so for a parameter  $\theta$

$$\widehat{\text{var}}\{\hat{\theta}^*\} = \sum_{m=1}^M w_m \left( \widehat{\text{var}}\{\hat{\theta}_m | \text{Model}_m\} + (\hat{\theta}_m - \hat{\theta}^*)^2 \right)$$

where

$$\hat{\theta}^* = \sum_{m=1}^M w_m \hat{\theta}_m.$$

Here,  $\hat{\theta}_m$  and  $\widehat{\text{var}}\{\hat{\theta}_m|\text{Model}_m\}$  are the point estimate and estimated variance from the fit of the  $m$ th model. This yields the variance estimate recommended by Burnham and Anderson (2004) and the MARK software manual. In other words, it accounts for model uncertainty.

### 2.3.3 Proportion marked, $p^*$

The proportion of the bowhead whale population that is marked is  $p^*$ . The estimate  $\hat{p}^*$  is based on all images with midback quality better than 3 and midback identifiability better than X. The data screening procedure described above is used to determine which whales to use in the calculation of  $\hat{p}^*$ . After screening images for inclusion, each image is given a weight reflecting the amount of effort dedicated to photographing the whale. Then whales, and the proportion marked, are treated as weighted sums of counts, or equivalently, sums of weights. The weights, which we denote as  $a_i$  for the  $i$ th whale, were 1.0 for the vast majority of the images. However, following Koski et al. (2006), images of cows accompanied by calves were given less weight because of increased effort to photograph cow-calf pairs and the greater amount of time spent at the surface by calves. Cows and yearlings travelling together were given intermediate weight because, like cows with calves, increased effort was made to photograph them, but their surface times are similar to other non-calves.

The methods of Koski et al. (2004) were used to estimate the increased probability of photographing a cow-calf pair relative to other whales. They estimated time at the surface from data on surfacing, respiration and diving (SRD) behavior during the spring migrations of 1989-91 and 1994. Durations of surfacings and dives were recorded for 248 calf SRD cycles and 302 SRD cycles of other whales. Calves were found to spend 1.71 times as long at the surface as other whales, with SE = 0.14. The ratio of images per whale for cows with calves to images per whale for whales not accompanied by a calf or yearling during the part of the migration when calves were seen was 1.304 in 2011. Thus, the weights were  $a_i = 1/1.304 = 0.767$  for cows and yearlings seen together, and  $a_i = 1/1.304 \times 1/1.71 = 0.448$  for cows seen with calves.

Furthermore, the migration was divided into ‘weeks’ (actually, short discrete periods) and the weeks’ proportions of marked whales combined to obtain the overall proportion. This approach avoids positive or negative bias in  $\hat{p}^*$  that could result if a week with unusually large numbers of marked whales was oversampled or under sampled, respectively. The hours of effort for each week were summed over the relevant survey years to obtain  $e_w$ , the total hours of effort for week  $w$ . It was assumed that if each week had the same amount of effort, the number of images per week would be related to the fraction of the bowhead population migrating past Point Barrow during that week.

Next, we estimated the passage rate of whales per hour during week  $w$  as  $R_w = A_w/e_w$  where

$$A_w = \sum_{i \text{ in week } w} a_i.$$

Each week was given a weight  $W_w = R_w/\sum_w R_w$ . The  $W_w$  ensure that weekly marked proportions are weighted by the relative passage of whales during each week. The proportion marked in week  $w$  was estimated as  $p_w = M_w/A_w$  where

$$M_w = \sum_{i \text{ in week } w} a_i z_i$$

and  $z_i = 1$  if the whale is marked and 0 otherwise. Finally,  $p^*$  was estimated as  $\hat{p}^* = \sum_w W_w p_w$ .

A rough estimate  $V(\hat{p}^*)$  of the variance of  $\hat{p}^*$  was computed under the assumption that  $M_w$  follows a binomial distribution with parameters  $A_w$  and  $p_w$  as:

$$V(\hat{p}^*) = \sum_w W_w^2 p_w (1 - p_w) / A_w.$$

This is an approximation since  $A_w$  and  $M_w$  are not integers. Further details are given in previous analyses (Koski et al., 2006, 2008, 2010b).

We used the 2011 estimate of  $p^*$  to calculate our overall population abundance estimate. In Section 3.5 we present an alternate abundance estimate that uses the inverse variance weighted mean of  $p^*$  estimates from three periods: 1984-87; 1989-92, 1994, 2003 and 2004; and 2011.

### 2.3.4 Total abundance, $N_{2011}$

Total abundance is estimated as

$$\hat{N}_{2011} = \hat{N}_{2011}^M / \hat{p}^*.$$

Under the assumption that  $\hat{N}_{2011}^M$  and  $\hat{p}^*$  are independent, we can apply the delta method to derive an approximate variance for  $\hat{N}_{2011}$ :

$$\text{var}\{\hat{N}_{2011}\} = \text{var}\{\hat{N}_{2011}^M\} / (\hat{p}^*)^2 + (\hat{N}_{2011}^M)^2 \text{var}\{\hat{p}^*\} / (\hat{p}^*)^4.$$

Following Buckland (1992), an approximate 95% confidence interval for  $N_{2011}$  is

$$(\hat{N}_{2011}/c, c\hat{N}_{2011})$$

where

$$c = \exp \left\{ 1.96 \sqrt{\log \left( 1 + \text{var}\{\hat{N}_{2011}\} / \hat{N}_{2011}^2 \right)} \right\}.$$

### 2.3.5 Adult survival rate

MARK performs maximum likelihood estimation of the survival rate,  $\phi$ , under the specified Robust Design model. Our overall estimate is a BIC weighted average of the results from the selected models.

Unfortunately, interpreting estimates of  $\phi$  is hindered by limitations of the MARK software. Confidence intervals employing the Wald approach and delta method are not useful in the present case because the estimated survival rate is so near its upper boundary (namely, 1). Instead, we take the approach of examining the set of profile likelihood lower confidence bounds individually, and computing their BIC weighted average lower bound. This cannot be interpreted as a confidence limit but could be interpreted in the Bayesian sense of Raftery et al. (1995) mentioned above.

## 3 Results

The dataset consisted of 11,070 images of 8,159 whales, of which 1,095 were designated as marked. The number of whales having 1, 2, 3, or 4 marked zones was 622, 357, 177, and 35, respectively.

trigger zone	Identifiability				Totals
	H+	H-	M+	M-	
midback	93	146	103	265	607
lower back	75	88	85	160	408
rostrum	3	12	12	29	56
flukes	3	1	1	19	24
Totals	174	247	201	473	

Table 2: Counts of the zone triggering the designation of a whale as marked and the identifiability score for that zone.

captured	initial	recaptured				
		1986	2003	2004	2005	2011
1985	172	14	8	5	1	8
1986	102		7	4	0	8
2003	130			9	2	16
2004	219				2	14
2005	66					3
2011	315					

Table 3: Counts of marked whales in each year, and subsequent recaptures. Whales recaptured more than once are counted several times in the table.

Table 2 shows how whales were marked. Entries are counts of whales. The ‘trigger zone’ (tz using the nomenclature of Table 1). refers to the zone that caused the whale to be designated as marked according to the screening process previously described. This is cross-tabulated with the identifiability score for that zone (tzi).

Table 3 counts the captures and recaptures of marked whales. Represented in this table are eight whales that were captured more than once. Such instances are counted in all relevant cells of the table. Altogether, there were 101 recaptures of 86 whales.

Table 4 lists the set of models used in our final analysis. Also listed are the number of model parameters, the BIC difference (relative to the baseline model), the BIC weight, estimates of  $N_{2011}$ , estimates of  $\phi$ , and the lower bounds of 95% confidence intervals for  $\phi$  using the likelihood profile approach.

### 3.1 Adult survival

The BIC weighted average survival estimate is 0.996. We interpret the very high estimates as indicating that there is little information about survival in our data aside from the survival rate being near one. This is not too surprising since the span of the study is only a fraction of a bowhead lifetime. Since we are unable to compute a model-averaged precision estimate, we have presented the full list of likelihood profile based confidence bounds. We have not generated a confidence interval using model averaging. However, informally, we can note that the limits are pretty similar, with a weighted average lower bound of 0.976.

Model	pars	$\Delta\text{BIC}$	BIC wt	$\hat{N}_{2011}^M$	$\hat{\phi}$	LB $\phi$
isc	8	0.00	0.372	5,529	0.994	0.981
I(bzi=H+)	8	0.51	0.288	5,934	0.999	0.974
I(bzi=H+ or H-)	8	3.19	0.076	5,810	0.997	0.972
isc + I(bzi=H+)	9	3.69	0.059	5,877	0.998	0.979
I(nz $\leq$ 2)	8	4.17	0.046	4,938	0.984	0.961
isc + nz	9	5.00	0.031	5,845	0.998	0.978
isc + I(bzi=H+ or H-)	9	5.13	0.029	5,744	0.997	0.974
isc + np	9	5.55	0.023	5,524	0.993	0.976
isc + I(nz $\leq$ 2)	9	5.75	0.021	5,348	0.993	0.973
isc:yb	9	6.14	0.017	5,592	0.991	0.976
I(nz $\leq$ 2) + I(bzi=H+):yb	10	6.45	0.015	5,745	0.997	0.971
isc + I(mz='L'):yb	9	6.68	0.013	5,589	0.994	0.970
I(bzi=H+ or H-):yb	9	8.91	0.004	5,951	0.999	0.975
isc + I(bzi=H+):yb	10	10.18	0.002	5,980	1.000	0.979
np + I(bzi=H+):yb	10	11.13	0.001	5,983	1.000	0.977
I(nz $\leq$ 2):yb	9	11.19	0.001	4,944	0.984	0.960
nz	8	11.64	0.001	4,995	0.985	0.976
baseline	7	11.65	0.001	5,019	0.986	0.961

Table 4: Set of models used for model averaging estimation, listing additive terms. A colon indicates that both additive and interaction terms have been fit. All models include time (yr) as a predictor of capture probabilities. Also shown are the number of parameters, the BIC difference compared to the best model, the BIC weight for model averaging, the estimated number of marked whales in 2011 ( $\hat{N}_{2011}^M$ ), the estimate for survival rate ( $\hat{\phi}$ ) and the lower bound (LB) of a profile likelihood 95% confidence interval for  $\phi$ .

### 3.2 Number marked

For the 18 models in Table 4, estimates of  $N_{2011}$  are provided. Using BIC weights, the weighted average estimate of marked abundance is  $\hat{N}_{2011}^M = 5,679$  with an unconditional standard error of 1,213 and a 95% confidence interval (3,301, 8,056).

### 3.3 Proportion of whales marked

Koski et al. (2010b) report two  $p^*$  estimates: 0.28968 (s.e. 0.00707) based on data from 1989-1992, 1994, 2003 and 2004; and 0.33937 (s.e. 0.01225) based on 1984-1987 data. Our estimate for 2011 was 0.2093 (s.e. 0.0084). The inverse variance weighted mean estimator used in our alternative abundance estimate (see Section 3.5) is  $\hat{p}^* = 0.2700$  (s.e. 0.0050).

### 3.4 Total abundance

Using the methods described above, total abundance is estimated to be 27,133 with 95% confidence interval (17,809, 41,337). The CV of this estimate is 0.217.

### 3.5 Alternative estimates

The photo-id dataset is rich with information and could be analyzed in many different ways. For exploratory and comparison purposes, we completed three additional analysis.

The first two alternatives involve splitting the 2011 data into spring and fall observations (there were no flights for most of June-August). The sampling effort was highly lopsided toward spring: there were 348 marked whales photographed in spring 2011 and 16 in fall 2011.

The first of these analyses repeated the main analysis, using a Robust Design with a Huggins model for capture probabilities within each primary period. The difference was that spring and fall 2011 were treated as separate secondary capture opportunities during the third primary block. Therefore, a final dummy opportunity was not necessary.

Due to the scant fall 2011 data, we did not repeat the model selection process. Instead, we used the same set of models as for the main analysis, with terms fit to accommodate the extra (real) data. All remaining parts of the original analysis were replicated.

We believe that this analysis was a failure. Model fits seemed sensitive and unreliable. Abundance estimates were unrealistically high. We did not pursue a full post-mortem of this effort.

The second alternative analysis uses only the 2011 data for marked whales. For this analysis, we fit a simple Chapman (1951) estimator, namely

$$\hat{N}_{Chapman}^M = (n_f + 1)(n_s + 1)/(m + 1) - 1 \tag{1}$$

where  $n_s = 348$  is the number of marked whales in the spring 2011 dataset,  $n_f = 16$  is the number of marked whales in the fall 2011 dataset, and  $m = 2$  is the number of marked whales matched for these two seasons (Amstrup et al., 2010). The variance of this estimate was estimated using a parametric bootstrap. Alternative two-occasion capture histories were generated from the 2011 data using a multinomial model as suggested by Buckland and Garthwaite (1991). Independently, values for  $\hat{p}^*$  were sampled from a normal distribution with mean and variance equal to the estimated values given above. Bootstrap values of  $\hat{N}_{Chapman}^M$  were calculated applying equation (1). A 95% confidence interval was derived using the percentile method.

The uncorrected Chapman estimate of total abundance is 9,942 with bootstrap 95% confidence interval (4,800, 31,946). The proportion of fall 2011 marked whales that were recaptures of spring 2011 marked whales is 50% greater than the proportion of other whales that were recaptures. This unusually high recapture rate for fall 2011 is an obvious explanation for why the Chapman estimate is considerably lower than our main result. Alternatively, one may view the overall fall 2011 sample size as too small to be representative. Moreover, Robson and Regier (1964) have shown that the Chapman estimator is negatively biased when the number of recaptures is small. In our case, we estimate that  $\hat{N}_{Chapman}^M$  is biased downward by about 5%, based on the results of those authors and the discussion of Seber (1982, p. 60).

It is worth noting that the fall 2011 survey did not cover the entire migration season. Also, it was not designed specifically to estimate abundance. Instead, it was primarily designed for a feeding study.

There are also several reasons why spring and fall 2011 might not be independent samples, and why the recapture rate in fall 2011 is high. Quakenbush et al. (2013) documented that a few satellite tagged whales returned to the Utqiagvik area and other areas near Wrangel Island during late spring and summer after traveling to the eastern Beaufort Sea and Amundsen Gulf. Moreover, those whales were subjectively estimated to be adult whales when tagged, and adult whales tend to be better marked and therefore more likely to be recognized as matches.

There is also some evidence from local hunters and boat surveys that some large bowheads stay north of Pt. Barrow through the summer rather than migrating to the Eastern Beaufort Sea (George et al., 2013).

Finally, catch data from the fall hunt at Utqiagvik show this trend: larger whales are taken early and smaller whales are taken later in the season. Thus the large marked whales photographed in the early fall could be some of the same whales at Utqiagvik in late spring.

Notwithstanding this discussion, it is important to note that the Chapman estimate confidence interval contains the point estimates from both the (main) photo-id estimate and the ice-based analyses.

The third alternative estimator repeated all aspects of the main analysis but used an inverse variance weighted  $p^*$  estimate from three periods spanning several decades. For this analysis, we estimated  $\hat{p}^* = 0.2700$  and  $\hat{N}_{2011} = 21,028$  (95% CI: 13,976 to 31,638; CV=0.214). Among our main analysis and the alternatives presented here, this alternative result is the most consistent with the estimate from the 2011 ice-based survey (Givens et al., 2016), but we believe that the more appropriate scaling for  $\hat{N}_{2011}^M$  is the estimated proportion marked from that specific year, namely 2011, as done in our main analysis.

## 4 Discussion

Since photographic matching for this dataset is an enormous and painstaking process, it is worth considering how matching failures might impact our results. Missed matches would lead to a positive bias in the population estimate.

Clearly, when photos are examined many times it is increasingly likely that all or at least almost all of the matches would be identified. The earlier 1985–2005 photos have been examined more often, and rarely has a match been found during follow-up studies. No such follow-up studies have been done using the 2011 photos yet so the probability of a missed match between the 2011 and earlier images is higher than between the earlier images. Also, during the 2003–2005 study

(and all earlier years) all whales were independently examined for matches by NMML and LGL experts, with the resulting number of matches having increased when the two matching datasets were integrated. No independent search for matches was done for 2011 so it is possible that one or more missed matches could be found with an independent search. The opportunity to find missing matches will arise when the next photo survey is conducted and the new images are compared with those from the earlier studies.

There are some whales that the ice-based census doesn't count: whales migrating past Utqiagvik more than 20 km beyond the ice edge, whales that do not migrate past Utqiagvik (e.g., to the northern Chukchi Sea or around Chukotka), and whales that pass Utqiagvik before or after the survey period. Givens et al. (2016) discuss other biases.

It is therefore reasonable to expect that the ice-based estimate is a little low and the photo-id estimate may be a little high. Our results appear to reflect this, although the confidence intervals for the two estimates overlap.

The photo-id and ice-based approaches are quite different in another respect. In a successful ice-based survey, observers may see over 3,000 whales in a single season. In contrast, the number of marked whales photographed in a season is limited to a few hundred (71 to 364 in our data). The total number of recaptures in all years analyzed here is only 101. This difference in sample sizes is a primary reason why the ice-based approach yields estimates with much lower CVs than the photo-id method. For 2011, the ice-based CV was 0.052, compared to the photo-id CV of 0.217. The photo-id estimate of 2004 abundance was 12,631 with CV= 0.244 (Koski et al., 2010b), but in the 1980s when abundance was lower, CVs from the two methods were more comparable and depended on various factors related to the success of the survey.

It has been argued that, over the long run, a photo-id method should be preferred because it accumulates data each survey, eventually yielding a very large photographic library, whereas a successful ice-based survey is a single discrete occurrence (Schweder and Sadykova, 2009). Although this is true in theory, the real world outcome is rather different. Our photo-id database spans more than three decades. Although sampling may have been more sporadic than envisioned by Schweder and Sadykova (2009), it is realistically about the best that could be hoped for, given the expense and difficulties of undertaking full-scale BCB bowhead aerial surveys. Indeed, future photo-id surveys are not currently planned because of the expense and time required, and the need to complete two such surveys in years that are consecutive or nearly so. With sufficient technology advancements in the use of unmanned aerial systems and automated image matching algorithms that would permit economical acquisition and analysis of photos, bowhead photo-id surveys may again become practical.

Our estimate of adult survival is higher than previous estimates. Those estimates include 0.985 (0.958, 0.999) (da Silva et al., 2007); 0.984 (0.948, 1) (Zeh et al., 2002); and 0.983 (0.977, 0.989) (Schweder et al., 2010). Our analysis, like those, is based on marked whales, which are primarily larger whales and not representative of the population as a whole. Calf and yearling survival is likely much lower. Missed matches would lead to lower survival estimates.

It is worth emphasizing that our photo-id abundance estimate of 27,133 (17,809, 41,337) was derived from data that are entirely independent from the data used for the ice-based estimate of 16,820 (15,176, 18,643) of Givens et al. (2016). Our new estimate confirms that bowhead abundance is strongly increasing in the BCB region (Givens et al. (2016) estimate a 3.7% (2.9%, 4.6%) rate of annual increase). It also strengthens the evidence that the current subsistence harvest is not endangering or hindering the growth of this stock.

The IWC Scientific Committee should consider whether to include this photo-id abundance estimate in the set of estimates used by the SC's Bowhead Strike Limit Algorithm. Since the estimate pertains to the same year as the ice-based estimate which is considerably more precise, any impact of the photo-id estimate may be limited.

## Acknowledgments

Much of this analysis was funded with “qualified Outer Continental Shelf oil and gas revenues” through the Coastal Impact Assistance Program (CIAP) under the auspices of the Fish and Wildlife Service, US Department of the Interior. The North Slope Borough Department of Wildlife Management also made considerable contributions both in kind support and direct revenues to complete this analysis. This work was also financially supported by the North Slope Borough (Alaska) and NOAA/NMFS. The surveys were conducted with Scientific Research Permits 782-1719 and 14245 issued to MML under the provisions of the US Marine Mammal Protection Act and Endangered Species Act.

We greatly appreciate the support and patience of the Barrow Whaling Captains Association and the Alaska Eskimo Whaling Commission, as this research is conducted concurrently with their hunting activities in the same region, and can lead to disturbance.

We are grateful to all of the aerial observers in the 2011 survey: Amelia Brower, Vicki Beaver, Brenda Rone, Cynthia Christman, Becky Shea and Corey Accardo. Megan Ferguson and Dave Rugh were also crucial in making the 2011 survey a success. Andy Harcombe of Clearwater Air, Inc., and his pilots Stan Churches and Eric Seller and were extremely helpful in all aspects of the survey.

We thank Paul Conn and Brett McClintock for helpful advice on using MARK and fitting Robust Design models.

The staff at the NSB-DWM, including Director Taqulik Hepa, was particularly helpful with logistical support for the study. We also thank NSB Grants Division for their assistance administering this large program.

## References

- Amstrup, S. C., McDonald, T. L., and Manly, B. (2010). *Handbook of capture-recapture analysis*. Princeton University Press.
- Angliss, R. P., Rugh, D. J., Withrow, D. E., and Hobbs, R. C. (1995). Evaluations of photogrammetric length measurements of the Bering-Chukchi-Beaufort Seas stock of bowhead whales (*Balaena mysticetus*). *Report of the International Whaling Commission*, 45:313–324.
- Buckland, S. T. (1992). Report of the Scientific Committee, Annex H. Proposal for standard presentation of abundance estimates. *Report of the International Whaling Commission*, 42:235.
- Buckland, S. T. and Garthwaite, P. H. (1991). Quantifying precision of mark-recapture estimates using the bootstrap and related methods. *Biometrics*, 47:255–268.
- Burnham, K. P. and Anderson, D. R. (2004). Multimodel inference – understanding AIC and BIC in model selection. *Sociological Methods and Research*, 33:261–304.

- Chapman, D. G. (1951). Some properties of the hypergeometric distribution with applications to zoological censuses. *University of California Publications in Statistics*, 1:131–160.
- Cubbage, J. C. and Calambokidis, J. (1987). Size-class segregation of bowhead whales discerned through aerial stereophotogrammetry. *Marine Mammal Science*, 3:179–185.
- da Silva, C. Q., Gomes, P. V. S., and Stradioto, M. A. (2007). Bayesian estimation of survival and capture probabilities using logit link and photoidentification data. *Computational Statistics and Data Analysis*, 51:6521–6534.
- da Silva, C. Q., Zeh, J., Madigan, D., Laake, J., Rugh, D., Baraff, L., Koski, W., and Miller, G. (2000). Capture-recapture estimation of bowhead whale population size using photo-identification data. *Journal of Cetacean Research and Management*, 2:45–61.
- Davis, R. A., Koski, W. R., and Miller, G. W. (1983). Preliminary assessment of the length-frequency distribution and gross annual reproductive rate of the Western Arctic bowhead whale as determined with low-level aerial photography, with comments on life history. Paper SC/35/PS5 presented to the Scientific Committee of the International Whaling Commission, June 1983. Available from the IWC, [www.iwc.int](http://www.iwc.int).
- Davis, R. A., Koski, W. R., Richardson, W. J., Evans, C. R., and Alliston, W. G. (1982). Distribution, numbers and productivity of the Western Arctic stock of bowhead whales in the eastern Beaufort Sea and Amundsen Gulf, summer 1981. Prepared by LGL Limited, Toronto, Ontario, for Dome Petroleum Limited, Calgary, Alberta, and Sohio Alaska Petroleum Company, Anchorage, Alaska, 135pp. Available from LGL Limited, [www.lgl.com](http://www.lgl.com).
- George, J. C., Bada, J., Zeh, J., Scott, L., Brown, S., O’Hara, T. M., and Suydam, R. (1999). Age and growth estimates of bowhead whales (*Balaena mysticetus*) via aspartic acid racemization. *Canadian Journal of Zoology*, 77:571–580.
- George, J. C., Follmann, E., Zeh, J., Sousa, M., Tarpley, R., Horstmann-Dehn, L., and Suydam, R. (2011a). A new way to estimate the age of bowhead whales (*Balaena mysticetus*) using ovarian corpora counts. *Canadian Journal of Zoology*, 89:840–852.
- George, J. C., Givens, G. H., Herreman, J., DeLong, R., Tudor, B., Suydam, R., and Kendall, L. (2011b). Report of the 2010 bowhead whale survey at Barrow with emphasis on methods for matching sightings from paired independent observations. Paper SC/63/BRG3 presented to the IWC Scientific Committee, June 2011.
- George, J. C., Tudor, B., and DeLong, R. (2013). North Slope Borough Research A – Local Boat Surveys. In Shelden, K. E. W. and Mocklin, J. A., editors, *Bowhead Whale Feeding Ecology Study (BOWFEST) in the western Beaufort Sea*. Final Report, OCS Study BOEM 2013-0114. National Marine Mammal Laboratory, Alaska Fisheries Science Center, NMFS, NOAA, 7600 Sand Point Way NE, Seattle, WA 98115-6349.
- George, J. C., Zeh, J., Suydam, R., and Clark, C. (2004). Abundance and population trend (1978–2001) of western Arctic bowhead whales surveyed near Barrow, Alaska. *Marine Mammal Science*, 20:755–773.

- Givens, G. H., Edmondson, S. L., George, J. C., Suydam, R., Charif, R. A., Rahaman, A., Hawthorne, D., Tudor, B., DeLong, R. A., and Clark, C. W. (2016). Horvitz-Thompson whale abundance estimation adjusting for uncertain recapture, temporal availability variation, and intermittent effort. *Environmetrics*, 27:134–146.
- Haldiman, J. T. and Tarpley, R. T. (1993). Anatomy and physiology. In Burns, J., Montague, J., and Cowles, C., editors, *The Bowhead Whale.*, chapter 4, pages 71–156. Special Publication No. 2 of the Society for Marine Mammalogy.
- Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika*, 76:133–140.
- Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration—stochastic model. *Biometrika*, 52:225–247.
- Kendall, W. L. and Nichols, J. D. (1995). On the use of secondary capture-recapture samples to estimate temporary emigration and breeding proportions. *J. Applied Statistics*, 22:751–762.
- Kendall, W. L., Nichols, J. D., and Hines, J. E. (1997). Estimating temporary emigration using capture-recapture data with Pollock’s robust design. *Ecology*, 78:563–578.
- Kendall, W. L., Pollock, K. H., and Brownie, C. (1995). A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, 51:293–308.
- Koski, W. R., Davis, R. A., Miller, G. W., and Withrow, D. E. (1992). Growth rates of bowhead whales as determined from low-level aerial photogrammetry. *Report of the International Whaling Commission*, 42:491–499.
- Koski, W. R., George, J. C., Zeh, J., and Brandon, J. R. (2010a). Preliminary analyses on identifying yearling bowhead whales (*Balaena mysticetus*) in aerial photographs. Paper SC/62/BRG29 presented to the Scientific Committee of the International Whaling Commission, May, 2010. Available from the IWC, [www.iwc.int](http://www.iwc.int).
- Koski, W. R., Miller, G. W., and Davis, R. A. (1988). The potential effects of tanker traffic on the bowhead whale in the Beaufort Sea. Environmental Studies 58. Report from LGL Limited, King City, Ontario, for the Department of Indian Affairs and Northern Development, Hull, Quebec. 150pp. Available from [www.lgl.com](http://www.lgl.com).
- Koski, W. R., Mocklin, J., Davis, A. R., Zeh, J., Rugh, D. J., George, J. C., and Suydam, R. (2008). Preliminary estimates of 2003-2004 Bering-Chukchi-Beaufort bowhead whale (*Balaena mysticetus*) abundance from photo-identification data. Paper SC/60/BRG18 presented to the Scientific Committee of the International Whaling Commission, June 2008. Available from the IWC, [www.iwc.int](http://www.iwc.int).
- Koski, W. R., Richardson, G. W., and Würsig, B. (2004). Bowhead whale (*Balaena mysticetus*) mothers and calves during spring migration in the Alaskan Beaufort Sea: movements, behaviour, and life history data. Paper SC/56/BRG27 presented to the Scientific Committee of the International Whaling Commission, July 2004. Available from the IWC, [www.iwc.int](http://www.iwc.int).

- Koski, W. R., Rugh, D. J., Punt, A. E., and Zeh, J. (2006). An approach to minimize the bias in estimation of the length-frequency distribution of bowhead whales (*Balaena mysticetus*) from aerial photogrammetric data. *Journal of Cetacean Research and Management*, 8:45–54.
- Koski, W. R., Zeh, J., Mocklin, J., Davis, A. R., Rugh, D. J., George, J. C., and Suydam, R. (2010b). Abundance of Bering-Chukchi-Beaufort bowhead whales (*Balaena mysticetus*) in 2004 estimated from photo-identification data. *Journal of Cetacean Research and Management*, 11:89–99.
- Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using Occam’s window. *Journal of the American Statistical Association*, 89:1535–1546.
- MARK Team (2017). MARK Software Package. <http://www.phidot.org/software/mark/index.html>.
- Miller, G. W., Davis, R. A., Koski, W. R., Crone, M. J., Rugh, D. J., Withrow, D. E., and Fraker, M. A. (1992). Calving intervals of bowhead whales – an analysis of photographic data. *Report of the International Whaling Commission*, 42:501–506.
- Nerini, M. K., Withrow, D., and Strickland, K. (1987). Length structure of the bowhead whale population derived from aerial photogrammetry, with notes on recruitment, spring 1985 and 1986. Paper SC/39/PS14 presented to the Scientific Committee of the International Whaling Commission, June 1987. Available from the IWC, [www.iwc.int](http://www.iwc.int).
- Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture. *J. Wildlife Management*, 46:757–760.
- Quakenbush, L. T., Small, R. J., and Citta, J. J. (2013). Satellite tracking of bowhead whales: movements and analysis from 2006 to 2012. Report to the U.S. Department of the Interior, Bureau of Ocean Energy Management, Alaska Outer Continental Shelf Region, Anchorage, AK. OCS Study BOEM 2013-01110, 56pp.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25:111–163.
- Raftery, A. E. and Zeh, J. E. (1998). Estimating bowhead whale population size and rate of increase from the 1993 census. *Journal of the American Statistical Association*, 93:451–463.
- Robson, D. S. and Regier, H. A. (1964). Sample size in Petersen mark-recapture experiments. *Trans. Amer. Fish. Soc.*, 93:215–226.
- Rosa, C., Zeh, J., George, J. C., Botta, O., Zauscher, M., Bada, J., and O’Hara, T. M. (2013). Age estimates based on aspartic acid racemization for bowhead whales (*Balaena mysticetus*) harvested in 1998-2000 and the relationship between racemization rate and body temperature. *Marine Mammal Science*, 29:424–445.
- Rugh, D. J., Braham, H. W., and Miller, G. W. (1992a). Methods for photographic identification of bowhead whales, *Balaena mysticetus*. *Can. J. Zool.*, 70:617–624.

- Rugh, D. J., Hobbs, R. C., Lerczak, J. A., and Breiwick, J. M. (2005). Estimates of abundance of the eastern north Pacific stock of gray whales *Eschrichtius robustus* 1997-2002. *Journal of Cetacean Research and Management*, 7:1–12.
- Rugh, D. J., Miller, G. W., Withrow, D. E., and Koski, W. R. (1992b). Calving intervals of bowhead whales established through photographic identifications. *Journal of Mammalogy*, 73:487–490.
- Rugh, D. J., Zeh, J. E., Koski, W. R., Baraff, L. S., Miller, G. W., and Shelden, K. E. W. (1998). An improved system for scoring photo quality and whale identifiability in aerial photographs of bowhead whales. *Report of the International Whaling Commission*, 48:501–512.
- Schweder, T. and Sadykova, D. (2009). Information dynamics and optimal sampling in capture-recapture. *Biometrika*, 99:488–493.
- Schweder, T., Sadykova, D., Rugh, D., and Koski, W. (2010). Population estimates from aerial photographic surveys of naturally and variably marked bowhead whales. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:1–19.
- Seber, G. A. F. (1965). A note on the multiple recapture census. *Biometrika*, 52:249–259.
- Seber, G. A. F. (1982). *The Estimation of Animal Abundance and Related Parameters, 2nd Edition*. Griffin, London, UK.
- Uhart, M. M., Rowntree, V., Sironi, M., Chirife, A., Mohamed, N., Pozzi, L. M., Musmeci, L., Franco, M., McAloose, D., Doucette, G., Sastre, V., and Rowles, T. (2009). Continuing southern right whale mortality events at Peninsula Valdés, Argentina. Paper SC/61/BRG18 presented to the Scientific Committee of the International Whaling Commission, June, 2009. Available from the IWC, [www.iwc.int](http://www.iwc.int).
- Wetzel, D. L., Reynolds III, J. E., Mercurio, P., Givens, G. H., Pulster, E. L., and George, J. C. (2017). Age estimation for bowhead whales (*Balaena mysticetus*) using aspartic acid racemization with enhanced hydrolysis and derivatization procedures. *Journal of Cetacean Research and Management*. to appear.
- Zeh, J., Poole, D., Miller, G., Koski, W., Baraff, L., and Rugh, D. (2002). Survival of bowhead whales, *Balaena mysticetus*, estimated from 1981-1998 photoidentification data. *Biometrics*, 58:832–840.
- Zeh, J. E. and Punt, A. E. (2005). Updated 1978-2001 abundance estimates and their correlations for the Bering-Chukchi-Beaufort Seas stock of bowhead whales. *Journal of Cetacean Research and Management*, 7:169–175.