

Estimation of detection probabilities from the 2010 ice-based independent observer survey of bowhead whales near Barrow, Alaska

G. H. GIVENS¹, S. L. EDMONDSON¹, J. C. GEORGE², B. TUDOR², R. A. DeLONG³ AND R. SUYDAM²

Contact email: geof@lamar.colostate.edu

ABSTRACT

We investigate independent observer data from the spring, 2010, ice-based survey of bowhead whales near Barrow, Alaska. Our goal is to estimate detection probabilities to use in a future abundance estimate, replacing the 25-year-old detection probability estimates summarized by Zeh and Punt (2005). Standard capture-recapture methods are not directly applicable to the 2010 data because of ambiguities and group size inconsistencies introduced by the data collection protocol. We present a variety of analysis methods within the general framework of Huggins (1989) to address such issues, including a novel approach for weighted data analysis in this context. Results from this method can be calculated from available output of the MARK program without substantial new software coding. For our recommended analysis approach, the mean estimated detection probability is 0.468. Detailed detection probability estimates depend on the effects of visibility, distance, group size, lead condition, whale passage rate, and some two-way interactions of these factors. Estimates range from about 0.70 to less than 0.10, with standard errors of roughly 0.03. Sensitivity analysis to investigate the potential impact of excluding periods with anomalous, temporary events hindering the sighting or matching of whales showed no substantial effects. We anticipate that our detection probability estimates will be used to produce an overall estimate of absolute abundance for this bowhead population from 2011 survey counts.

INTRODUCTION

In April and May of 2009, 2010, and 2011, scientists from the North Slope Borough Department of Wildlife Management attempted ice-based counts of bowhead whales (*Balaena mysticetus*) from the Bering-Chukchi-Beaufort Seas population as the animals migrate northward past Barrow, Alaska. A detailed description of the survey is given by George et al. (2011).

These surveys employ a two-perch independent observer protocol. Specifically, two teams of observers stand at fixed survey sites ('perches') situated atop pressure ridges near leads and open water. The two perches are sufficiently distant that the teams cannot hear each other or incidentally cue each other about their sightings. All sightings and other data are recorded independently at each perch. Each recorded sighting may be of a single animal or a group. These groups are not whale 'pods' in the conventional sense. Except for cow-calf pairs, migrating bowheads appear to have only weak and probably brief allegiance to any aggregation (Zeh et al., 1993). When convenient hereafter, we may refer to single whales as groups of size 1.

Data recorded for each sighting include location (via theodolite readings), time, group size, calf presence, and a wide variety of covariates related to the whales, environmental conditions, and survey factors (e.g., observed scars, ice coverage and observer team membership). Many of these covariates are discussed later. One of the most important variables recorded for a sighting is a 'link code', which indicates whether the observer team believes that it has previously detected the sighted whale(s). The link codes and their meanings are given in Table 1. Some of these codes explicitly specify *which* previously sighted group the current sighting corresponds to. Like all our data, these link codes are assigned independently at each perch and pertain only to the sightings from that perch. For example, a New whale seen at one perch might never be seen at the other; or perhaps this New whale is a Y Duplicate within a sequence of previous/future sightings at the other perch, such as N-X-Y-R-Z.

¹ Department of Statistics, Colorado State University, Fort Collins CO 80523, USA.

² North Slope Borough Department of Wildlife Management, Barrow, AK.

³ DeLongView Enterprises, Box 85044, Fairbanks, AK 99708, USA.

Table 1: Linking codes for sightings of whale groups. Every sighting is classified using one of these codes. The three fundamental classifications are New, Duplicate, and Conditional.

Code	Meaning
N	New whale or group. Observer team is confident that it is seen for the first time.
R	Roll. The sighting is part of a sequence of surface dives or ‘roll series’ of a previously sighted whale or group. A link is assigned to indicate the associated previous sighting. A Roll is a type of Duplicate.
X	Duplicate. The observer team is 100% confident that the whale or group can be linked to a specific previous sighting. A link is assigned to indicate the associated previous sighting.
Y	Duplicate. The observer team is about 90% confident that the whale or group can be linked to a specific previous sighting. A link is assigned to indicate the associated previous sighting.
Z	Duplicate. The observer team is quite sure that the whale or group has been previously sighted but the team cannot link it back to a specific previous sighting with 90% confidence. Rarely a link is assigned.
C	Conditional. The observer team cannot determine whether this whale or group is New or a Duplicate of some previous sighting. Links to earlier sightings are forbidden.

The main purpose of the surveys is to obtain a population abundance estimate. Previous abundance estimates from ice-based surveys include those of Braham et al., (1979); Zeh et al. (1986); Zeh et al. (1991); Raftery and Zeh (1998); George et al. (2004); Zeh and Punt (2005). The most recent abundance estimate, derived from an aerial photo-identification survey, is 12,631 (95% confidence interval 7,900 to 19,700) by Koski et al. (2010). The key requirements for estimating abundance from the ice-based survey data are the count of whales seen, the proportion of the population available to be seen as whales migrate past the perches, and an estimate of the detection probability, i.e., the probability of detecting a whale given that it passes the perch within viewing range. Our paper is focused solely on detection probabilities. The perches may have separate detection probabilities, and these probabilities may each depend on covariates.

For some early ice-based surveys a two-perch removal method survey protocol was used (Zeh et al., 1986). All ice-based abundance estimates since 1986 have relied on detection probability estimates from 1985 and earlier. Those detection probability estimates are now 25 years old and should be updated to reflect current methods, conditions, and characteristics of the bowhead population and its migration.

Due to weather conditions, the 2009 survey effort failed. Only a few hundred whales were observed—far too few to estimate detection probabilities or abundance. In 2010, the survey was partially successful. There were 397 hours when at least one perch was operating, during which 304 hours of two-perch independent observer effort was maintained. The total number of New and Conditional whales, respectively, reported by each perch was 1060 and 275 for South Perch, and 1216 and 255 for North Perch (after correcting a few entries in the raw data logs). Note that many of these were instances where the same whale or group of whales was seen from both perches, and some Conditional whales were likely resightings of New whales. Despite the partial success of the 2010 survey, it would be very difficult to produce a reliable abundance estimate from these data. Ice and weather conditions prevented all survey effort for much of April and May 4-6, and survey effort ended on May 28 before the migration ended. As many as a third to one half of the bowhead population probably passed Barrow during those unmonitored times. However, the 2010 data are ample for detection probability estimation.

The 2011 visual survey is ongoing at the time of writing and is expected to end approximately 1 June. Conditions have generally been very favourable. Like 2010, the migration has exhibited a surprisingly large number of sightings in early April and high passage rates at certain times thereafter. As of May 12, over 3,000 New whales had been observed. Based on past migration patterns and assuming that good weather continues, we expect that the 2011 survey will yield at least the second-highest number of sighted whales compared to all past surveys. These numbers are even more impressive if the proportion of whales passing Barrow beyond viewing range in 2011 is higher than historical levels.

This paper describes estimation of detection probabilities from the 2010 survey data. First we describe how the raw survey data were converted into sight/resight data appropriate for detection probability estimation. After a description of the estimation approach, we present the results and sensitivity analyses. The paper ends with a discussion of our findings and implications for future use of the 2010 and 2011 survey data.



intermediate
chases, links

DATA

To understand key data issues, we must begin by previewing the plan for analysis. In the Methods section we describe the statistical model used to estimate detection probabilities (Huggins, 1989). The basis for the analysis is the capture-recapture principle. In a generic case, sighted groups are classified into three categories: those seen by team 1, those seen by team 2, and those seen by both teams. The detection probability (for whale groups) for a team is estimated as the ratio of the number of groups seen by both observers to the total number of animals seen by the other observer team.

In other settings, these counts are usually known exactly. For example, animals may be marked with bands or matched via photo-identification. In the bowhead case, the best way to count captures and recaptures is less clear. Below we describe how we address this issue and develop a suitable dataset for analysis.

Link Data

The first stage of our data treatment is to distill the sightings data into ‘captures’. As described in the introduction and Table 1, each sighting from a single perch is labelled as New, Duplicate or Conditional, where the latter two categories represent a possible re-sighting of a previously seen group. Duplicate and Conditional sightings are explicitly perch-specific (as are New sightings) and hence are not recaptures. Due to the presence of Duplicate (R, X, Y, Z) and Conditional (C) sightings, the total number of sightings reported by a perch exceeds the total number of distinct groups sighted because sightings at a perch are logically connected on the basis of the link codes. We call these connections ‘links’. Links are established by the observer team at the time of sighting, so links can only refer backwards in time to previously seen whales. It is required that a New or Conditional whale be identified as the originating sighting for any subsequent sighting coded as a R, X, or Y Duplicate. By definition, C sightings are not linked back to a previous sighting and Z sightings rarely are, but both types may be connected to future sightings via those future sightings’ backward links. Critically, sightings and links at one perch are totally independent of those from the other perch: a group sighted at one perch may never have been sighted from the other. If the group was seen by both perches, the links (if any) assigned by each perch are unrelated.

Since time is recorded with each sighting, a sequence of sightings connected by links has a unique link ordering. We call a sequence of linked sightings a ‘link chain’. By definition, a link chain must begin with a N or C (or rarely but improperly a Z) and contain no subsequent N or C. Since the perches operate independently, link chains are associated with a single perch. Let an arrow denote a link, with superscripts representing the number of whales reported in the group and subscripts representing perch. Then the link chain $N_1^1 \rightarrow X_1^1 \rightarrow Y_1^2$ represents a link chain created by perch 1, where a new sighting of one whale was linked to an X Duplicate of a single whale, and then to a Y Duplicate where two whales were seen in the group. Note that group size may not be consistent within a chain.

Table 2: Counts of sighting chains by type and perch. Group size counts use the assignment method described in the Covariate Data section, except not counting calves. These counts pertain to the subset of data used for analysis.

Chain Type	South	North	Total	Group Size =1	Group Size = 2	Group Size = 3
Single	611	648	1259	1148	101	10
Link Only	44	38	82	64	17	1
Match			589	506	74	9

We will see later that certain other types of chains may involve loops or other features that preclude the simple notation adopted here. Finally, note that link chains are permitted to be length one, i.e., a single New, Conditional, or Z Duplicate whale.

Match Data

In order to determine the number of ‘recaptures’ we must identify which whales were seen from both perches. The observer teams do not communicate, so the process of matching a link chain from one perch to a link chain from the other perch must be done by a third party.

In 2010, matching was attempted both in real time and retrospectively. For the period 1-14 May, observer teams radioed the sighting time, location, swim direction and speed to a ‘command center’. The two teams used different radio frequencies. The ‘master matchers’ in the command center plotted these data using software adapted specifically for this task and tracked sightings approximately in real time, trying to identify between-perch matches. The software was for data display only—matches were determined using human judgment integrating all relevant information. An example of the match display is shown in Figure 1. Sightings are displayed with Link Code and time. Links and matches are shown with line segments in three different colours. One motivation to attempt real time matching was that the master matchers could radio the perch for additional information if the matchers had a question or needed clarification about a sighting. George et al. (2011) provide much more detail about matching.

At high whale passage rates, it became evident that the real time protocol was not sustainable. Observers on the perches did not have enough time to record sightings, make links, and also radio back the data to the command center. For the rest of the season, sightings data were collected without real time matching. Several months later, the entire dataset was scrutinized to identify matches. The master matchers used the same software and essentially simulated what would have happened during real time matching, except that the software allowed time to be slowed as much as needed to allow full and careful identification of matches. Match decisions were twice re-assessed and validated through comprehensive reviews later in the year. Table 2 shows counts of match chains, link chains and single sightings. In this table, link chains are chains that involve no matches. Counts are shown separately for each perch as well as in total, and split by group size.

It is critical to understand that a match connects sightings *between* perches rather than *within* a perch. Moreover, matches explicitly connect *sightings*, not link chains. However, since every sighting is a member of (exactly one) perch-specific link chain, a match to any sighting implicitly connects a link chain from one perch to a link chain from the other perch. The entire set of sightings connected by a sequence of matches and links is called a ‘match chain’. These match chains may connect a variety of link chains at various perches, thereby containing multiple matches and multiple links. Note that the matches within a match chain often do not directly connect the earliest sightings within each constituent link chain. Adopting the notation that matches are indicated by double arrows, an example match chain might be $N_1^1 \rightarrow X_1^1 \leftrightarrow N_2^1 \rightarrow Y_2^1 \leftrightarrow C_1^1$. Quickly it becomes clear that our notation is insufficient. If the first link chain was extended so that the X Duplicate at perch 1 was linked onward to a Y Duplicate at perch 1 then the overall match chain would have an additional ‘loose end’. If the New whale from perch 1 was matched to the Y Duplicate from perch 2 then the match chain would include a ‘loop’. ‘Forks’ occur when one sighting is matched to several others. In principle a match chain may have quite a few loose ends, loops, and/or forks.

Covariate data

At the time of sighting, a variety of covariate data were collected in addition to the time, sighting location, group size and link code. We list some of these in Table 3. Other covariates including whale passage rate can be calculated

Table 3: Some covariates recorded at the time of sighting.

Visibility	Unacceptable, Poor, Fair, Good, Very Good, Excellent
Weather	Ten categories describing features of cloud cover, precipitation, and fog
Wind Speed and Direction	Miles per hour; Degrees clockwise from magnetic north
Ice Coverage	Percent
Lead Condition	Unknown, Closed, Not Continuous ('Patchy'), Continuous, Wide Open
Lead Width	Meters
Ocean Current Speed and Direction	Knots; Degrees clockwise from magnetic north
Swim Speed and Direction	Kilometers per hour; Degrees clockwise from magnetic north
Group Size and Calf Presence	Counts of whales; zero or one calf
Behaviour	Fifteen categories, e.g., migrating, lingering, breach
Number of Observers	Usually 3 or 4
Observer Name	Only for the theodolite operator
Markings	E.g., scarring

from the observed information. Passage rate was calculated using the running mean number of whales sighted within one hour of the sighting, expressed in whales per hour.

Since link and match chains contain several sightings, there is ambiguity about how to assign covariate values to chains. For link chains, the assigned covariate value is taken to be the last value observed within the chain. For match chains, the assigned covariate value is taken to be the value recorded at the first recapture event within the chain. Assignment of group sizes to chains is a special case discussed later. The sighting data are imbalanced, with more sightings of single whales and intermediate distances; see Table 4.

During data analysis, we investigated a variety of ways to treat certain covariates. The relevant ones for our analyses are as follows. Distance from the perch is binned into 1km bands, with sightings beyond 4km discarded. Visibility is consolidated into Fine (Excellent, Very Good, or Good), and Limited (Fair or Poor). Passage rate is binned into Heavy (more than 10 whales per hour) and Light. Group sizes are consolidated into '1' and '≥1', with calves counting towards the total. The rationale for this choice (aside from a good model fit and very few sizes exceeding 2) is that observers felt strongly that the presence of a calf increased detectability *at least* as much as a second adult whale due to the surface behaviour of the cow-calf pair.

Summary of chain data

Some chains are removed from the raw data before analysis, namely: undeterminable lead conditions (8 sightings); theodolite operators with fewer than 10 records (19); sighting distances exceeding 4 km (80); unlinked X/R/Y Duplicates (12); unlinked Z Duplicates (136). The isolated Z Duplicates are resightings of previously seen groups or single animals, but with limited assurance of which previous sighting the resighting corresponds to; therefore removing them does not reduce the number of 'captures' for that perch. Z Duplicates involved in match chains are retained. After this data cleaning, approximately 94% of the data remain, and all results and tables in this paper pertain to the trimmed dataset.

For the 589 match chains, the median length is 2, the mean is 2.7, and the longest is 16. For the 82 link chains, the median, mean, and maximum are 2, 2.2, and 4, respectively. To illustrate the potential complexity of chains, the longest match chain has 9 links, 11 matches, 4 forks (one of which is 3-way), 7 loose ends and several loops.

Recall the earlier example link chain $N_1^1 \rightarrow X_1^1 \rightarrow Y_1^2$. Note that the group sizes are not consistent. Such instances are relatively common among match chains. For link chains, 79 chains had no group size inconsistencies and 3 chains had group size inconsistencies of 1 whale. For match chains, 482, 96, 10, and 1 chains had (maximal) group size inconsistencies of 0, 1, 2, and 3 whales, respectively. This corresponds to an inconsistency rate of 3.7% for link chains and 18.2% for match chains. Of the 107 inconsistencies for match chains, 48 were chains of length 2 with group sizes (1,2) or (2,1); 5 had group sizes (2,1,1); 4 had group sizes of (2,2,1), 4 had group sizes of (1,2,1); and the remainder were other configurations of three or more sightings having fewer than 4 instances.

ANALYTICAL METHODS

Level of analysis and group size consistency

The records of single whales, link chains and match chains contain all the information about sighting and resighting. These are the data to be used in the statistical capture-recapture analysis. However, the chains pertain to groups, not individual whales. There are several ways one might analyse sighting chains.

First, the data could be analysed at the group-level. Detection probabilities would refer to the chances of detecting a group when it is present. If estimated detection probabilities were used to estimate population abundance, the abundance estimator would need to include a scaling factor to correct for average group size. In the present case, this group-level approach has a potentially non-ignorable bias. Except for cow-calf pairs, bowheads have loose group allegiance that cannot be characterized as 'pods' in the normal sense. When group sizes are seen to vary along a chain, this may be attributable to whales joining or leaving a chain. Thus treating a match chain as having a consistent group size systematically categorizes these fleeting group allegiances as recaptures rather than allocating them between recaptures and New whales. The same effect occurs for link chains, but since 88% of non-single chains are match chains rather than link chains the net bias is towards increased recaptures, which introduces an upward bias in detection probability estimates corresponding to a downward bias in abundance estimates. A poll of observer opinions after the survey finds divided views, with some observers tending to attribute group size inconsistencies on failure to detect group members, and others tending to attribute whales joining or leaving groups. An optimal choice for treating inconsistencies is not clear.

Alternatively, the data could be analysed at the level of individual whales. In this case, detection probabilities would pertain directly to single whale detection. However, (re)detection of whales within the same group is more correlated than (re)detection of whales in different groups. Thus, sightings are not fully independent within groups.

Table 5: Numbers of non-recaptures (n) and recaptures (r) resulting from various group size reconciliation methods for six example match chains. Example match chains are denoted in the column headings by a sequence of numbers representing group sizes at distinct sightings along the chain, with ‘m’ indicating the position of the match. The entries (counts) within the table represent whales, not groups. In the 131m212 case, the asterisked entries are instances where the count depends on the linking pattern; full linkage is assumed here. See the text for further explanation.

Chain:	1m1		2m2		1m2		32m3		111m21		131m212	
Method:	n	r	n	r	n	r	n	r	n	r	n	r
W_{\max}	0	1	0	2	0	2	0	3	0	2	0	3
W_{mode}	0	1	0	2	0	2	0	3	1	1	3*	1
W_{mean}	0	1	0	2	0.5	1.5	0.7	2.7	0.8	1.2	1.7*	1.7
W_{decon}	0	1	0	2	1	1	2	2	1	1	3*	1

However, bowheads’ relatively loose group affiliation may mitigate concern in two ways. First, because group associations are weak, the correlation may be limited. Second, the theodolites have very narrow field of view, which reduces the ability to detect a second whale while sighting the first through the theodolite. To the extent that there is correlation between whales within groups, this creates a source of heterogeneity in detection probabilities, and this generally leads to a downward bias in abundance estimates using standard capture-recapture abundance estimation models. See Carothers (1973, 1979); Otis et al. (1978); Seber (1982); Pollock et al. (1990); Hwang and Chao (1995); Pledger and Efford (1998); Pledger and Phillpot (2008). If the effect of potential correlation is believed to be substantial enough to render standard error estimates inadequate, bootstrapping could be used to mitigate concern. We did not bootstrap here.

An attractive aspect of a whale-level analysis is its greater fidelity to the original group size data. Field observations are taken at face value rather than second-guessing the trained observers after the fact. The extreme rarity of group size inconsistencies within link chains (3%) supports our faith in the observers’ original records. A group-level analysis requires greater departure from the raw data in the sense that internally inconsistent data within a chain must be reconciled artificially.

Given the uncertainty about which is the best method, we apply seven different approaches: three group-level analyses and four whale-level analyses. The first group-level analysis forces group sizes to be consistent within the chain by defining the group size to equal the group size for the first recapture sighting, as defined previously. Actually there are two variants of this approach. Later we discuss how sightings should be properly weighted for analysis. Our baseline group-level variant employs an appropriate weighting and is denoted G_{1st} . A crude approach ignoring these weights, denoted G_0 , is also presented merely as a ‘reality check’. The third group-level analysis sets group size equal to the most common value within the chain, with ties favouring the larger value. We label this approach G_{mode} .

The whale-level analyses employ different methods for breaking down a chain into sub-chains corresponding to single whales. Define a sub-chain to be any chain of connected sightings with consistent group size that can be obtained by reducing the counts at all sightings in the original chain by the same amount, with the additional constraint that the sub-chain must be as long as possible for any specified group size. The set of sub-chains of a given chain is therefore unique. For example, the sub-chains of a chain having group sizes (2,1,1,2,2) are the chains (1,1,1,1,1), (1,x,x,x,x), and (x,x,x,1,1) where ‘x’ indicates absence of chain membership. Let g_c denote the group size of any chain or sub-chain, c , if the group size is consistent within the chain. Then the whale-level options we examine are as follows:

1. *Deconstruction.* The longest *match* sub-chain with $g_c = 1$ is extracted. (If the match chain includes a calf, the calf is also assigned to this first sub-chain.) If no such chains remain (see below) then the longest *link* sub-chain with $g_c = 1$ is taken. After extraction, the group sizes at each sighting along the chain are decremented by 1. Any links or matches connecting a sighting whose group size reaches 0 are broken.

What remains may include individual sightings, link sub-chains, and match sub-chains. These may include remnant group sizes that remain inconsistent and/or greater than 1. This deconstruction process is then iterated. When finished, the process will have produced a collection of individual sightings, link chains, and match chains, each with $g_c = 1$. This approach is labelled W_{decon} .

2. *Mode*. All group sizes along the chain smaller than the modal size are changed to equal the mode within this chain. The result is then deconstructed as for W_{decon} . Note that this produces a number of full-length sub-chains equal to the modal group size reported, plus (potentially) some remnant shorter sub-chains. When there is a tie for the mode size, the larger value is taken. We label this method W_{mode} .
3. *Mean*. All group sizes along the chain smaller than the mean size for the chain are changed to equal the mean. The result is then deconstructed. However, deconstruction steps (after the first one) can produce sub-chains with fractional whales. Therefore, deconstruction removes sub-chains with the largest possible $g_c \leq 1$ even if g_c is a fraction. Note that this method, denoted W_{mean} , produces at least one full-length chain with $g_c = 1$.
4. *Maximum*. All group sizes along the chain are changed to equal the largest value recorded in that chain. The result is deconstructed. This produces only full-length chains with $g_c = 1$ and the number of resultant chains equals the maximum group size within the original chain. This method is denoted W_{max} .

Table 5 presents several examples of these methods. In this table, a variety of example match chains are listed as column headings. All these match chains are examples where one link chain from one perch is matched to one link chain from the other perch. For simplicity we have omitted the link codes. Each number in the chain sequence represents the number of whales at a distinct sighting within the chain. The ‘m’ represents the match and therefore divides the two link chains. In a few instances, there is not a unique solution without specifying further details about the hypothetical link chains. In those cases, we assume ‘full linkage’, which means that every sighting in a link chain is linked to the N or C sighting that initiates that chain. For comparison, note that the group-level approaches will produce 0 group non-recaptures and 1 group recapture for all the example chains in Table 5.

There are several rationales for these whale-level options. First, we adopt the assumption that observers never report more whales in a group than are actually present. Although group sizes may be over-counted occasionally at large distances, this error is believed to be very rare. Second, group affiliation is believed to be weak, so whales may join and exit groups. Group-level approaches ignore this. The W_{max} approach will clearly over-count recaptures. On the other hand, W_{decon} probably undercounts recaptures. The W_{mode} and W_{mean} approaches attempt to balance errors in both directions. Finally, note that our tie-breaking philosophy slightly inflates the number of recaptures, thereby decreasing abundance estimates.

In any of these whale-level methods, when deconstruction occurs the covariate values associated with the new sub-chains are determined in the same manner as the group-level analyses. The exception is group size. For this variable, each single-whale sub-chain inherits the group size attribute of the original chain. We take this approach because the event of detecting that sub-chain occurred during the sighting of the larger group, not the smaller sub-chain.

Detection probability estimation model and variable selection

We adopt the model of Huggins (1989) for capture-recapture estimation for closed populations. This model assumes that captures at each perch are independent and that the catch history is therefore multinomial for each individual. To form the likelihood, the model conditions on the total number of groups detected. Separate capture probabilities are allowed for each perch, although our results will show that this is not critical and consequently we constrain our analysis to assume that probabilities are equal among the perches. The general model also allows for capture response. This is impossible in the bowhead case since the captures—i.e., sightings—are not invasive. Finally, the detection probability for a group is allowed to depend on covariate observations for that group. The effect of covariates is the primary focus of our analysis.

Models are fit using the MARK software (White and Burnham, 1999), using the RMark interface (Laake, 2011). However, all of the analyses described here (except G_0) include consideration of fractional whales, which is equivalent to weighted observations. The appendix of this paper explains why fractional whales are needed for each approach. Methods for the weighted fitting of the Huggins model and discussion of appropriate weights are also given in the appendix.

Model selection is carried out using a stepwise philosophy with AIC_c as the objective function (Akaike, 1974; Hurvich and Tsai, 1989). The simplest model within 2.0 units of AIC_c from the minimum is selected (Burnham and Anderson, 2002). Nearly always, this was also the model with the lowest AIC_c . We use a stepwise approach (generally forward selection) considering additive and two-way interactive effects, with backward elimination and tangential explorations conducted when it appears they might be informative or helpful. Due to very limited sample sizes for a few combinations of variables, two-way interactions between lead condition and either group size or distance are excluded (the latter for group-level analyses only). After tentatively selecting the model, we retested single-term and two-way additions to the chosen model but this did not improve AIC_c or identify a superior model in any case. Automated model selection methods and model averaging are not used.

There are several reasons not to average models here in the manner of Burnham and Anderson (2002). Most importantly, the model fitting process shows that there is very little model uncertainty in the senses that i) there were few selection choices presenting ambiguous AIC_c decrements near 2.0 and ii) nearly the same model was selected for each approach. Furthermore, the dominant source of model uncertainty is associated with the choice among the seven analysis methods described above. These approaches use different datasets of different sizes, thereby yielding AIC_c values that are incomparable. Although predicted detection probabilities could be averaged in some informal sense, any advantage of this seems limited and sacrifices parsimony.

RESULTS

Main findings

Table 6 lists the model terms selected for each approach. Distance, Visibility, Lead Condition, Group Size, and Passage Rate are commonly used, along with some two-way interactions. Although our results differ somewhat between methods, several broad conclusions can be drawn. Foremost, the sample-weighted mean detection probability and corresponding 95% range of values for each method are quite similar across methods (see Table 7).

Table 7 also shows relative estimated population abundances for each method, using the Horvitz-Thompson-type method implemented in MARK. Above we discussed the limitations of the 2010 data for abundance estimation. In particular, the survey was not operational for periods during which as much as half the total population may have passed. No correction is made for this, for whales passing beyond 4km of shore, or for whales not available to be sighted (e.g., failing to surface) within the region of detectability. For this reason, all relative abundance estimates in Table 7 have been arbitrarily scaled by a common factor so that the estimate for the W_{mode} method is 10,000. All these abundance estimates use weighting not only for detection probability estimation but also when counting whales for the abundance result (e.g., Conditional whales count half).

Figure 2 graphs estimated detection probabilities as a function of distance for all the methods for four sets of covariate values. From left to right and top to bottom, the covariate values are a) fine visibility, wide open lead, single whales, heavy passage rate, b) limited visibility, continuous lead, group sizes exceeding 1, heavy passage rate, c) limited visibility, continuous lead, single whales, light passage rate, and d) limited visibility, group sizes

exceeding 1, wide open lead, light passage rate. These graphs typify our results. The modal values of these covariates in the dataset are limited visibility, continuous lead, single whales, and light passage rate; this corresponds to case c) above.

Complete lists of estimated detection probabilities for each method are given in Tables 8-14. From Figure 2 and these tables, we can draw some broad conclusions. First, the effect of increasing distance from the perch has a very strong negative effect on detection probabilities. At nearly 4km from the perch, detection probabilities can fall drastically. The effect of patchy ice conditions amplifies this effect. For example, when the ice is patchy the odds of detection decreases by a multiplicative factor of 0.23 and 0.11 for the G_{mode} and W_{mode} strategies respectively. (Changes in odds are the natural units to express such model results; the odds of an event equal the probability that it happens divided by the probability that it doesn't.) Thus, at nearly 4km in the worst conditions of ice and the other covariates, estimated detection probabilities can plunge to nearly zero. Fortunately, patchy ice is reported for only 4.2% of our data records.

Second, increased passage rate has a positive impact on detection probabilities. This may seem counterintuitive if one believes that observers sight whales more effectively when the whales are more sparse and the observers less stressed. We find the opposite result, and there are some plausible explanations. At high passage rates, our experience on the perch has been that—if anything—observers may tend to be more vigilant and over-link whales. To the extent that this occurs, the effect is to reduce the number of N and C whales, thereby increasing the proportion of recaptures and producing higher apparent detection probabilities. This effect could be compounded by the behaviour of the ‘master matchers’. When passage rates are high, the options for matching whales are much greater. To the extent that this encourages additional matches, the effect is the same as over-linking.

The effect of visibility on detection probability is primarily through a key interaction. Marginally, changes in visibility have only weak effects on detection. However, visibility and distance from the perch interact strongly. Detection probabilities for more distant whales tend to be low, but the reduction is exacerbated when visibility is Limited. Conversely, Limited visibility actually increases detection probabilities in the 1-2km range substantially, compared to Fine visibility. This interaction is clearly seen, for example, comparing the top two panels of Figure 2.

For group size, our results are somewhat equivocal. For many models, larger groups are associated with higher estimated detection probabilities. The straightforward interpretation is that larger groups display more visual cues of their presence. However, in a few of our results larger groups exhibit the opposite effect. This finding is clearly apparent for the deconstruction approach. The explanation is equally direct here, but less immediately intuitive. More than the other methods, deconstruction tends to produce single individuals. This is intentional because the method adopts the philosophy that group size discrepancies are often caused by separate whales joining/leaving groups. The deconstruction method ‘extracts’ these extra, inconsistent whales as new individuals. This favours lower detection probabilities. Now, the key step in the argument is to recall that these new individuals inherit their group size covariate value from their original chain—it is *not* set to 1. Then, since the only opportunities for such

extractions are for chains (hence whales) labelled with group sizes greater than one, a negative correlation between group size and detection probability tends to be induced.

It is important to note what variables do not appear to affect detection probability. These include: weather conditions, number of observers on the perch, lead width, percentage ice cover, wind speed, whale swim direction, whale behaviour, and calf presence (except as this contributes to group size). Some of these factors are among the ones that the observers consider when assigning an overall visibility score.

Ancillary results

None of our group-level analyses detect a significant difference in estimated detection probabilities between perches. On the other hand, the whale-level analyses do detect small differences. Although these effects exceeded our AIC_c threshold of 2.0, the differences are very small in a practical sense. For example, for the W_{mode} approach, the estimated detection probabilities for single whales at 2km in a continuous lead under fine visibility and light passage rate are 0.616 and 0.592 for the north and south perches, respectively. On the ice, the north perch was higher than the south perch: 7.3 m and 5.4m, respectively. Observers on the perches felt that the different perch heights did not greatly affect their ability to detect whales, and we believe that the small differences between perch detection probabilities we found using the whale-level analyses has minimal implications for abundance estimation.

Observer effects are more dramatic. Our results show that detection probabilities vary significantly between observer groups. Such extra heterogeneity will tend to cause a downward bias in abundance estimates using standard capture-recapture abundance estimation models. Although it is common to find observer effects in studies like ours, there is one surprising aspect here. Given the survey protocol, search success is a team effort, yet it is only the operator of the theodolite whose identity is recorded for the sighting. The theodolite operator is not the only—nor the primary—person discovering whales, particularly considering the device's narrow field of view. Moreover, team memberships were continuously remixed so that any particular theodolite operator used the device with many different combinations of teammates. For these reasons, it is difficult to explain the observer effects. Despite an extensive analysis not reported here, we have found no correlation between observer effects and the level of observer experience (e.g., total hours of effort during the survey) or mean whale passage rate during the observer's effort. Finally, note that accounting for observer effects in detection probabilities will be difficult if our estimates are applied to counts from 2011 because the 2011 observers were not all present in 2010, and vice versa. Our current plan is to ignore observer effects in any future abundance estimate and consider this to be a source of negative bias.

We conducted two sensitivity analyses to investigate special concerns that arose during the survey. First, the height and location of the south perch sometimes made it difficult for observers to see quite to 4km except for sighting blows, due to the nature of ice pans on the water and obstruction from the north perch. Thus, we repeated the W_{mode} analysis after limiting the data to sightings not exceeding 3km from the perch, using the same model chosen for the original data. The population-weighted mean detection probability for the sensitivity analysis is 0.493, compared to 0.468 for the main analysis. The respective scaled abundances (comparable to Table 7) are 10,032 and 10,000. The probabilities and abundances are ordered differently because the sensitivity analysis abundance calculation incorporates an additional correction (the ratio of weighted sightings) that accounts for the sample size difference between the two datasets. Fitting a new model from scratch to the reduced dataset produces the same final model form.

The second sensitivity analysis is motivated by the fact that there were a few days when the survey effort appeared potentially unreliable. We refit the model after removing data from the following periods: April 30 to May 3 (new observers), May 7, 11, 12, 15 and 20 (markedly different viewing conditions at the two perches due to ice, but only 7:15-16:30 on the 15th), May 8, 10:00-14:00 (minor equipment miscalibration), and all periods with patchy lead conditions. Adopting the W_{mode} analysis framework, the population weighted mean detection probability for the sensitivity analysis is 0.510 and the scaled abundance (comparable to Table 7) is 10,595. Although removing days where comparatively poor viewing conditions at (only) one perch should yield a higher estimated detection probability due to the reduced number of recaptures, this effect is dominated by an additional new correction factor to account for missing days to maintain comparative relative abundances because the remaining days have slightly larger numbers of sightings than the deleted days.

DISCUSSION

The 2010 survey was not successful for the purpose of estimating abundance, but it was extremely successful as an independent observer experiment. It clearly presents the best opportunity in 26 years—perhaps ever—to estimate ice-based detection probabilities.

It may be tempting for readers to discount our estimates of detection probabilities from the 2010 data and await the data from the 2011 survey. However, this is a mistake that would reflect a misunderstanding of our planned approach. Early in the 2010 survey it became apparent that sufficient coverage of the migration would be unlikely, so effort was overwhelmingly dedicated to ensuring that we obtained a large amount of high quality independent observer data. For the 2011 survey, it has been apparent that the critical need now is to obtain equally good count data. The forces that prevent or enable comprehensive, high quality survey counts (mainly visibility and ice conditions) are largely outside our control. However, allocation of effort is controllable and in 2011 we have prioritized continuous effort from one perch instead of more limited effort from both perches.

This decision makes it more likely that a final abundance estimate will apply 2010 detection probabilities to 2011 count data. In an ideal world, it would be preferable to use 2011 probabilities or a pooled estimate from both years, but the realities of environmental challenges, availability of workers, and funding may render such a goal unrealistic. Consider the alternative to using 2010 detection probabilities and 2011 counts: one would need to apply pre-1986 detection probabilities to the 2011 counts. Clearly the modern data will offer a major improvement.

Looking ahead toward the next estimation of total abundance, our findings here suggest that detecting whales during patchy ice conditions (specifically, “open water visible; lead not continuous; consists of one or a series of polynyas” in the observers’ training manual) can be very difficult. However, we have few data during these conditions from 2010, and this effect may differ between years. For example, in 2011 observers are reporting that such ice conditions have sometimes led to larger numbers of sightings when the whales surface in an open patch of water that has sometimes formed within easy view of the perches. It is worth considering whether periods with very patchy ice conditions should be excluded from analysis of the 2011 survey data (applying the standard approach to extrapolate for periods without data). A firm decision, however, must await the findings from the 2011 survey. There is also a reasonable question as to whether abundance estimation should be based on the traditional 4km boundary (as a distance from the perch), or perhaps 3km. Zeh et al. (1987) considered this question previously and found that the two choices yielded statistically indistinguishable abundance estimates. Finally, implications of the fact that the distribution of group sizes appears to have changed from the earliest surveys also warrant consideration. For group sizes of 1, 2, and ≥ 3 , the observed frequencies are 89%, 10%, and 1%, respectively for 2010, but 72%, 23%, and 5% respectively for the combined years of 1975-1978, 1987, and 1988 (Zeh et al., 1993).

To the best of our knowledge, our approach to weighting the data to address bowhead survey protocol issues such as the distinction between New and Conditional whales and group size inconsistencies is novel. Capture-recapture abundance estimation relies on two fundamental quantities: counts (i.e., catch histories) and detection probabilities. For the bowhead case, it is clear that the counts must partially discount Conditional whales. That is *not* the type of weighting we describe in the appendix or use in our analysis. Our methods weight the data for the purpose of estimating detection probabilities. This is sensible because, for example, a solitary Conditional animal does not constitute as much of a non-recapture as does a New whale. Thus, we should expect that the unweighted analysis—which is inappropriate in our opinion—would produce lower estimates of detection probabilities and higher abundance estimates. Indeed, this is confirmed in Table 7.

When comparing the detection probability estimates from our various methods given in Tables 8-14, it is important to remember that some types of observations are quite rare. This can create instances where estimates of the same quantity differ markedly between methods. Since there are very few sightings to which such detection probabilities apply, the impact on key estimates like those in Table 7 is minimal.

The issue of balancing bias and variance in final estimates of detection probabilities (and abundance) merits further consideration. Many of the terms included in our models on the basis of the AIC_c criterion exert small magnitude effects on estimated detection probabilities. Despite the statistical significance of such terms, their practical importance is minimal for the sake of estimating abundance. Thus, there is an argument in favour of selecting a more parsimonious model at the expense of minor biases (which might even partially cancel each other out). A method for doing this is to apply a stricter inclusion rule for model terms. For example, for a term to be added to the model one might require that the AIC_c improvement be at least 2.0 *and* the effect on detection probabilities exceeds 5%. Another option would be to penalize model complexity more severely, e.g., using the BIC criterion (Schwarz, 1978).

Considering analysis alternatives more broadly, it is possible that mark-recapture distance sampling approaches could be adapted to our problem (Buckland et al., 2010). However, an assumption that the distribution of animals is uniform over distances from the perch or from the lead edge would not be reasonable in the bowhead case. Moreover, in some ice and visibility conditions, the probability of detection ‘along the line’ in distance sampling parlance can be rather low.

Previous estimates of detection probabilities (most recently by Zeh and Punt, 2005) have been used in past abundance estimates. Some of the Zeh and Punt (2005) detection probability estimates (expressed as percentages) are 72 ± 6 (Very good to Excellent vis., $\leq 2\text{km}$), 60 ± 8 (Fair vis., $\leq 2\text{km}$), 40 ± 11 (Good vis., $> 2\text{km}$), and 33 ± 12 (Fair vis., $> 2\text{km}$). Broadly speaking, we estimate lower detection probabilities for single whales at low passage rates (the most common situation) and higher values for larger groups and/or higher passage rates. Our estimates are modestly lower on average. Our standard errors are considerably smaller. One can only speculate about the reasons for the difference between 2010 and the historical estimates. Annual variation in weather conditions and environmental conditions may be important. Clearly observers have also changed over the years, although there is no indication that the distribution of observer abilities is much different recently compared to past years. Although scientists have tried to maintain protocol and equipment constant throughout the history of these surveys, there have been inevitable changes and improvements. Finally, the previous detection probabilities are nearly 30 years old, and during the interim there have been marked changes in environmental conditions, particularly ice, over a much longer time span than annual variation. Changes in whale swim behaviour associated with changes in ice conditions may affect the ease of detection.

Another key aspect of the 2010 survey is that it is the first one that has used an independent observer design rather than the removal method. This change in methodology might contribute to the differences in detection probabilities in the following manner. The past (removal) method was extremely difficult for observers to carry out effectively at high whale passage rates because the north perch was required to receive detailed data radioed from the south perch and attempt to re-detect those whales while simultaneously maintaining its own separate sighting and data recording effort. One can imagine that, under busy conditions, the demands of receiving and confirming the south perch's sighting data diminished the north team's ability to detect New whales. This phenomenon would lead to an increased number of recaptures at the expense of new sightings, thereby pushing detection probability estimates upward. The 2010 survey design does not suffer from this problem and therefore might be expected to produce lower detection probability estimates.

We have presented a variety of analysis methods, and although the results are similar, the differences are not negligible. It is reasonable to ask which method should be preferred. Among the group-level analyses, G_0 is clearly biased. We suggest the G_{mode} approach is the best option. Although the modal group size does not characterize the moment of sighting as accurately as does the immediate instances used in $G_{1\text{st}}$, we feel that the mode is a better indication of the true group size. Among the whale-level analyses, it is not entirely unreasonable to support W_{max} and W_{decon} , but we believe that they tend to bracket the most realistic scenarios by overestimating and underestimating detection probabilities respectively. We suggest that the W_{mode} approach should be preferred. There is no apparent advantage to using W_{mean} since it is considerably more complex and it appears to underestimate standard errors.

Finally, we face a choice between group-level or whale-level analysis. For many animals including whales (e.g., eastern North Pacific gray whales (Rugh et al 2008), detection probabilities are estimated at the group level, with the final abundance estimate incorporating a correction for group size. However, bowheads' loose group affiliation argues against treating groups as inviolable: some adjustment for group infidelity should be incorporated. Moreover, the linking and matching protocols particular to the ice-based survey are sufficiently unique to consider a whale-level approach because, among other things, they produce data with more information about group size variation than would normally be available. In our opinion, the W_{mode} approach yields the best compromise between forcing consistent group sizes and failing to recognize groups altogether. This approach adopts reasonable, consistent group sizes accounting for whales that are unseen, while potentially extracting a small number of other individuals to reflect whales joining and leaving groups. Thus, we recommend W_{mode} as the primary approach, with G_{mode} as a backup. This suggestion is tentative, of course, pending input from interested scientists and the particulars of the 2011 survey.

Our goal in presenting these results is to provide readers with an introduction to the types of challenges inherent in independent observer data analysis for Bering-Chukchi-Beaufort Seas bowhead ice-based surveys. The methods discussed here—and likely the estimated detection probabilities themselves—will be used to produce a bowhead population abundance estimate from ice-based survey data, hopefully from 2011. We welcome suggestions and queries about this work since our goal is to produce an estimate that is as accurate and defensible as possible, and as conservative as is reasonable given the complexities of the survey and data.

ACKNOWLEDGMENTS

This work was supported by the North Slope Borough (Alaska) and the National Oceanic and Atmospheric Administration (through the Alaska Eskimo Whaling Commission). We thank the many observers, matchers, and other colleagues who conducted this challenging survey with amazing tenacity and professionalism. Jeff Laake is thanked for advice on data analysis and using RMark, and Dave Rugh is thanked for his thoughts on various aspects of this analysis. Finally, Judy Zeh is thanked for her advice and wisdom about virtually every aspect of the bowhead surveys over 40 years.

REFERENCES

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. on Automatic Control* 19: 716–723.
- Braham, H., Krogman, B., Leatherwood, S., Marquette, W., Rugh D., Tillman, M., Johnson, J., and Carroll, G. 1979. Preliminary report of the 1978 spring bowhead whale research program results. *Rep Int Whal. Commn* 29:291-306.
- Buckland, S.T., Laake, J.L., and Borchers, D.L. 2010. Double-observer line transect methods: levels of independence. *Biometrics* 66: 169-177.
- Burnham, K. P. and Anderson, D.R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edition*. Springer-Verlag. New York.
- Carothers, A. D. 1973. The effects of unequal catchability on Jolly-Seber estimates. *Biometrics* 29: 79–100.
- Carothers, A. D. 1979. Quantifying unequal catchability and its effect on survival estimates in an actual population. *Journal of Animal Ecology* 48: 863–869.
- George, J.C., Givens, G.H., Herreman, J., DeLong, R., Tudor, B., Suydam, R. and Kendell, L. 2011. Report of the 2010 bowhead whale survey at Barrow with emphasis on methods for matching bowhead whale sightings from paired independent observations. Paper SC/63/BRG 3 submitted to the Scientific Committee of the International Whaling Commission, June 2011.
- George, J.C., Zeh, J.E., 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.
- Huggins, R.M. 1989. On the statistical analysis of capture experiments. *Biometrika* 76: 133-140.
- Hurvich, C. M. and Tsai, C.-L. 1989. Regression and time series model selection in small samples. *Biometrika*, 76: 297–307.
- Hwang, W.-D. and Chao, A. 1995. Quantifying the effects of unequal catchabilities on Jolly-Seber estimates via sample coverage. *Biometrics* 51: 128–141.
- Koski, W.R., Zeh, J., Mocklin, J., Davis, A.R., Rugh, D.J., George, J.C., and Suydam, R. 2010. Abundance of Bering-Chukchi-Beaufort bowhead whales (*Balaena mysticetus*) in 2004 estimated from photo-identification data. *J. Cetacean Res. Manage.* 11: 89-99.
- Laake, J. 2011. RMark package for R. <http://www.phidot.org/software/mark/rmark/>.
- Otis, D. L., Burnham K. P., White, G. C., and Anderson, D. R. 1978. Statistical inference from capture data on closed animal populations. *Wildlife Monographs* No. 62. 135pp.
- Pledger, S. and Efford, M. 1998. Correction of bias due to heterogeneous capture probability in capture-recapture studies of open populations. *Biometrics* 54: 888-898.
- Pledger, S. and Phillpot, P. 2008. Using mixtures to model heterogeneity in ecological capture-recapture studies. *Biometrical Journal* 50: 1022-1034.
- Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. D. 1990. Statistical inference for capture-recapture experiments. *Wildlife Monographs* No. 107. 97 pp.

- Raftery, A.E. and Zeh, J.E. 1998. Estimating bowhead whale population size and rate of increase from the 1993 census. *J. American Statistical Assoc.* 93: 451-463.
- Rugh, D.J., M. M. Muto, R. C. Hobbs, J. A. Lerczak. 2008. An assessment of shore-based counts of gray whales. *Mar Mamm Sci*, 24(4): 864-880.
- Schwarz, G.E. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461-464.
- Seber, G.A.F. 1982. *The Estimation of Animal Abundance and Related Parameters*, 2nd Edition. Griffin, London, UK.
- White, G.C. and Burnham, K.P. 1999. Program MARK: Survival estimation from populations of marked animals. *Bird Study* 46, Suppl.: 120-138.
- Zeh, J.E. 1988. Effects of changing 1978-1983 visual census parameters on bowhead whale, *Balaena mysticetus*, population size estimates. *Rep. Int. Whal. Commn* 38:371-376.
- Zeh, J.E. 1999. Correcting for missed observations in a shore-based count of whales. *Marine Mammal Survey and Assessment Methods*. Garner et al. (eds). Balkema. Rotterdam. p. 127-136.
- Zeh, J.E., Clark, C.W., George, J.C., Withrow, D., Carroll, G.M., and Koski, W.R. 1993. Current population size and dynamics. In: J.J. Burns et al. (Eds) *The Bowhead Whale*, p. 408-489.
- Zeh, J.E., George, J.C., Raftery, A.E., and Carroll, G.M. 1991. Rate of increase, 1978-1988, of bowhead whales, *Balaena mysticetus*, estimated from ice-based census data. *Marine Mammal Science* 7: 105-122.
- Zeh, J.E., Ko, D., Krogman, B.D., and Sonntag, R.M. 1986. A Multinomial Model for Estimating the Size of a Whale Population from Incomplete Census Data. *Biometrics* 42: 1-14.
- 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. *J. Cetacean Res. Manage.* 7: 169-175.

APPENDIX: FITTING WEIGHTED CAPTURE-RECAPTURE MODELS WITH MARK

Some of our methods involve fractional whales. Our analyses address this by weighting each observation in proportion to the corresponding fraction. This appendix provides more details and describes a method for obtaining the results of the weighted analysis using the MARK program, thereby circumventing the need to develop a customized estimation procedure. The key results are given in equations (1), (2), and (3). Although we refer to MARK results below, it is easiest to implement these ideas using RMark, from which the necessary quantities can easily be extracted and manipulated.

Weighted observations

Suppose we wish to assign the i th outcome (i.e., the catch history for a single, link, or match chain) a weight w_i for $i = 1, \dots, n$ where n is the sample size and, without loss of generality, $0 < w_i \leq 1$. Our method temporarily replaces the original, unweighted dataset with a larger dataset that replicates the i th observation r_i times, where $r_i = R w_i$ and R is chosen to be the smallest integer such that r_i is an integer for all i . Thus the total sample size in the expanded

dataset is $n_r = \sum_{i=1}^n r_i = R \sum_{i=1}^n w_i$. If the desired weights are rounded to the first decimal place, then $R \leq 10$. In an

analysis where the presence of conditional whales requires $w_i \in \{1, 0.5, 0.75\}$ for all i (see below) then $R = 2$ or $R = 4$ depending on whether any $w_i = 0.75$.

There are several instances where weighting may be used. Consider Conditional sightings, i.e., sightings where the observer cannot determine whether the whale is New or a Duplicate. Zeh et al. (1991) and subsequent analyses have treated a single Conditional whale as half a New whale when estimating detection probabilities and abundance. Thus we assign weights of 0.5 to conditional sightings here.

Conditional whales in link and match chains must also be addressed. For example, consider a sequence of sightings like $N_1^1 \quad C_1^1 \rightarrow X_1^1$ where, importantly, there is no link between the initial New whale and the subsequent Conditional sighting. The absence of a preceding link is compelled by the definition of the term ‘Conditional’, and indeed the earlier New whale is merely implicit in the sense that every Conditional whale has at least one New whale at some earlier time. Assigning equal odds for a Conditional being new or duplicate, the above sequence of sightings can be interpreted in two ways. With probability 0.5 the Conditional has not been previously seen so the sequence is equivalent to $N_1^1 \quad N_1^1 \rightarrow X_1^1$. On the other hand, with probability 0.5 the Conditional is actually a resighting of some previous whale, in which case the sequence is equivalent to $N_1^1 \rightarrow X_1^1 \rightarrow X_1^1$ for some preceding N_1^1 .

Assume the simplest possible capture-recapture model with the probability of capture at each perch and the probability of recapture all equal to p . Then the first possibility represents two whales, thereby contributing p^2 to the likelihood. The second possibility represents a single whale contributing p to the likelihood. In the next section we develop a method allowing one to assign a weight of 0.5 to a Conditional whale by letting it contribute $p \times p^{1/2} = p^{3/2}$ to the likelihood. This approach is also consistent with the standard principle of weighting observations in statistical models: that a random variable having half the weight is equivalent to having twice the variance.

The weight of 0.5 is used for Conditional sightings that are unconnected or which lead a link chain. Treating match chains with leading Conditional sightings at one or both perches is more complex—particularly because the chain must be weighted as a single unit rather than weighting the sightings from each perch separately. Nevertheless, we can apply analogous reasoning to show that we should assign such chains a weight of 0.75 if one perch reported Conditional and 0.5 if both perches reported Conditional.

Weighting and MLEs for exponential families

We now discuss our weighting approach in the context of exponential families. For exponential family distributions with weighted observations, there is a useful relationship between certain likelihoods. In the simplest and most generic case, a density in the exponential family can be written as

$$f(x_i | \theta) = \exp \left\{ \frac{x_i \theta - b(\theta)}{\phi / w_i} + c(x_i, \phi / w_i) \right\}$$

where ϕ is a fixed dispersion parameter and w_i is a known weight. The exponential family includes many familiar distributions including the Gaussian, binomial, and multinomial distributions; the latter two are directly relevant for capture-recapture models.

Consider the simplest (i.e., null) Huggins (1989) model with one recapture opportunity. Each weighted observation in the dataset (e.g., each catch history) contributes a term to the likelihood function used to estimate the model. Denote the corresponding log likelihood contribution as

$$\ell_i(\theta | x_i) = w_i A_i + c(x_i, \phi / w_i)$$

where $A_i = (x_i \theta - b(\theta)) / \phi$, $\theta = -\log\{p / (1 - p)\}$, $b(\theta) = -\log\{p / (2 - p)\}$, $\phi = 1$, and x_i equals 1 if the catch history does not include a recapture and 0 otherwise. For an i.i.d. sample, the overall log likelihood is the sum of such contributions, namely

$$\ell(\theta | \chi) = \sum_{i=1}^n \ell_i(\theta | x_i) = \sum_{i=1}^n w_i A_i + q$$

where $\chi = \{x_1, \dots, x_n\}$ represents the entire dataset and q is a constant that does not depend on θ .

Let χ_R represent the dataset where each x_i is replicated r_i times and *there is no weighting* of the observations within χ_R . Let x_{i,j_i} for $j_i = 1, \dots, r_i$ denote these replicates of the i th case in χ . Then the contribution to the overall *unweighted* log likelihood $\ell_R(\theta | \chi_R)$ for x_{i,j_i} is

$$\ell_{R,i,j_i}(\theta | x_{i,j_i}) = A_i + c(x_{i,j_i}, \phi)$$

and the total contribution associated with the i th catch history is

$$\sum_{j_i=1}^{r_i} (A_i + c(x_{i,j_i}, \phi)) = r_i A_i + q_{R,i}$$

where $q_{R,i}$ is a constant that doesn't depend on θ . It follows that the overall log likelihood given by the unweighted, replicated dataset is

$$l_R(\theta | \chi_R) = R \sum_{i=1}^n w_i A_i + q_R^*$$

where q_R^* doesn't depend on θ .

Our key results follows from the fact that terms not involving θ are irrelevant when computing the score functions from $\ell(\theta | \chi)$ and $\ell_R(\theta | \chi_R)$. Specifically,

$$\frac{d\ell(\theta | \chi)}{d\theta} = \frac{d\ell_R(\theta | \chi_R) / R}{d\theta}.$$

Thus the MLE for the weighted likelihood equals the maximizer of the replicated likelihood, i.e.,

$$\hat{\theta}_{MLE} = \hat{\theta}_R. \quad (1)$$

Although this discussion is presented for unidimensional θ , the analogous results for vector parameters are obvious. Because it is based on the multinomial distribution, the results above pertain to the general Huggins (1989) model, including the case when detection probabilities are modelled to depend on a collection of covariates. The covariates inherent in the model lead one to express the parameter vector θ in the conditional likelihood as a function of the coefficients in the linear predictor portion of the model. The standard maximum likelihood assumptions are also implicit above and in what follows.

Computing AIC_c differences

Consider the comparison of two fitted models, which we will represent as $\hat{\theta}_1$ and $\hat{\theta}_2$, recognizing that the two models may have different parameter sets. Define

$$\Delta = \ell(\hat{\theta}_1 | \chi) - \ell(\hat{\theta}_2 | \chi) \quad \Delta_R = \ell_R(\hat{\theta}_1 | \chi_R) - \ell_R(\hat{\theta}_2 | \chi_R).$$

Then

$$\Delta = \Delta_R / R.$$

In other words, the log likelihood difference between the two models fit to the weighted data can be calculated from the log likelihood difference between the two models fit to the replicated data. Therefore, the AIC_c difference (Akaike, 1974; Burnham and Anderson, 2002) between two weighted models can be expressed as

$$AIC_c(\hat{\theta}_1, \hat{\theta}_2) = -2\Delta_R / R + K_1 - K_2 - K_{R,1} + K_{R,2} \quad (2)$$

where $K_i = 2k_i n / (n - k_i - 1)$ and $K_{R,i} = 2k_i \sum_j r_j / \left(\sum_j r_j - k_i - 1 \right)$ for $i = 1, 2$ indexing the two models and k_i

denoting the numbers of parameters therein. Thus, $AIC_c(\hat{\theta}_1, \hat{\theta}_2)$ for the weighted analysis of the original data χ can be calculated using the results of applying MARK to the unweighted replicated dataset χ_R , and then adjusting the results using a collection of known constants.

Estimated standard errors and confidence intervals

The asymptotic variance of a maximum likelihood estimator is $\text{var}\{\theta\} = I^{-1}(\theta_0) / n$ where

$$I(\theta_0) = E \left(\frac{d\ell(\theta)}{d\theta} \right)^2 \Bigg|_{\theta=\theta_0} = -E \left(\frac{d^2\ell(\theta)}{d\theta^2} \right) \Bigg|_{\theta=\theta_0}$$

and θ_0 denotes the true value of θ . Consider the likelihood corresponding to χ , namely $\ell(\theta)$. In this case,

$$I(\theta) = -E \sum_{i=1}^n w_i A_i''(\theta).$$

Next consider $\ell_R(\theta)$ originating from the unweighted, replicated dataset χ_R . Then

$$I_R(\theta) = -E \left(\frac{d^2 \ell_R(\theta)}{d\theta^2} \right) = -E \sum_{i=1}^n R w_i A_i''(\theta).$$

The sample sizes for χ and χ_R are n and $n_R = \sum_{i=1}^n r_i$ respectively. Thus it is straightforward to show that

$$\hat{v}\{\hat{\theta}\} = \frac{nR}{n_R} \hat{v}\{\hat{\theta}_R\}.$$

where \hat{v} denotes an estimated variance obtained by plugging the maximum likelihood estimate into the asymptotic expression. Further, $nR/n_R = \bar{w}^{-1}$ where \bar{w} is the mean weight for the observations in χ . Thus

$$\hat{v}\{\hat{\theta}\} = \hat{v}\{\hat{\theta}_R\} / \bar{w}. \quad (3)$$

Since \bar{w} is known and $\hat{v}\{\hat{\theta}_R\}$ is an output of the MARK analysis of χ_R the desired $\hat{v}\{\hat{\theta}\}$ can be calculated. Standard errors and confidence intervals for $\hat{\theta}$, for the coefficients of the linear predictor in the Huggins model, and for detection probabilities follow directly.

Table 8: Detection probability estimates and standard errors for the G_0 analysis, expressed as percentages. These results are for Wide Open or Continuous lead conditions. The detection odds for Patchy leads are much lower.

Distance	Light Passage Rate				Heavy Passage Rate			
	Group Size 1		Group Size >1		Group Size 1		Group Size >1	
	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.
0 – 1 km	53 ± 5	57 ± 4	63 ± 6	67 ± 5	53 ± 6	46 ± 6	63 ± 6	57 ± 6
1 – 2 km	47 ± 5	62 ± 3	60 ± 6	72 ± 4	54 ± 5	56 ± 5	64 ± 5	66 ± 5
2 – 3 km	28 ± 4	33 ± 3	38 ± 6	42 ± 5	50 ± 4	44 ± 5	60 ± 5	55 ± 6
3 – 4 km	34 ± 6	20 ± 4	44 ± 7	27 ± 6	51 ± 6	24 ± 7	62 ± 6	33 ± 8

Table 9: Detection probability estimates and standard errors for the G_{1st} analysis, expressed as percentages. These results are for Continuous leads; the results for Wide Open leads are virtually identical (42 ± 3 and 15 ± 5 would be replaced by 42 ± 4 and 16 ± 5 , respectively). The detection odds for Patchy leads are much lower.

Distance	Light Passage Rate				Heavy Passage Rate			
	Group Size 1		Group Size >1		Group Size 1		Group Size >1	
	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.
0 – 1 km	57 ± 3	61 ± 2	66 ± 5	69 ± 4	58 ± 4	49 ± 3	67 ± 5	58 ± 5
1 – 2 km	50 ± 3	63 ± 2	74 ± 4	83 ± 2	58 ± 3	58 ± 3	80 ± 3	79 ± 3
2 – 3 km	31 ± 2	34 ± 2	39 ± 4	43 ± 4	51 ± 3	42 ± 3	60 ± 4	51 ± 5
3 – 4 km	37 ± 3	22 ± 3	22 ± 6	12 ± 4	56 ± 4	27 ± 5	39 ± 8	15 ± 5

Table 10: Detection probability estimates and standard errors for the G_{mode} analysis, expressed as percentages. These results are for Wide Open or Continuous lead conditions. The detection odds for Patchy leads are much lower.

Distance	Light Passage Rate				Heavy Passage Rate			
	Group Size 1		Group Size >1		Group Size 1		Group Size >1	
	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.
0 – 1 km	56 ± 3	60 ± 2	54 ± 6	72 ± 4	62 ± 3	49 ± 3	60 ± 6	63 ± 5
1 – 2 km	50 ± 3	63 ± 2	67 ± 5	86 ± 2	59 ± 3	55 ± 3	74 ± 4	82 ± 3
2 – 3 km	30 ± 2	34 ± 2	32 ± 4	50 ± 5	53 ± 2	40 ± 3	55 ± 4	57 ± 5
3 – 4 km	37 ± 3	22 ± 3	18 ± 5	17 ± 5	57 ± 4	24 ± 4	34 ± 7	19 ± 6

Table 14: Detection probability estimates and standard errors for the W_{mean} analysis, expressed as percentages. These results are for Continuous and Wide open leads. The detection odds for Patchy leads are much lower.

Distance	Light Passage Rate				Heavy Passage Rate			
	Group Size 1		Group Size >1		Group Size 1		Group Size >1	
	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.	Fine Vis.	Limited Vis.
0 – 1 km	57 ± 1	60 ± 1	50 ± 2	65 ± 1	61 ± 1	46 ± 1	54 ± 2	52 ± 1
1 – 2 km	52 ± 1	64 ± 1	61 ± 1	82 ± 1	59 ± 1	54 ± 1	67 ± 1	75 ± 1
2 – 3 km	29 ± 1	34 ± 1	27 ± 1	44 ± 1	53 ± 1	41 ± 1	49 ± 1	51 ± 1
3 – 4 km	38 ± 1	22 ± 1	15 ± 1	13 ± 1	57 ± 1	24 ± 1	29 ± 1	14 ± 1