Fisheries and Oceans Canada
Pêches et Océans Canada
Ecosystems and Oceans Science
Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2015/034

Central and Arctic, and Quebec Regions

Identification of duplicate sightings from the 2013 double-platform High Arctic Cetacean Survey

D. Pike[1] and T. Doniol-Valcroze[2,3]

[1] Esox Associates
1210 Ski Club Road
North Bay, Ontario, P1B 8E5

[2] Maurice-Lamontagne Institute
Fisheries and Oceans Canada
P.O. Box 1000
Mont-Joli, Quebec, G5H 3Z4

[3] Freshwater Institute
Fisheries and Oceans Canada
501 University Crescent
Winnipeg, Manitoba, R3T 2N6

July 2015

Foreword

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1919-5044

Correct citation for this publication:

Pike, D. and Doniol-Valcroze, T. 2015. Identification of duplicate sightings from the 2013 double-platform High Arctic Cetacean Survey. DFO Can. Sci. Advis. Sec. Res. Doc. 2015/034. v + 22 p.

TABLE OF CONTENTS

ABSTRACT

One of the key assumptions of distance sampling is that all animals on line are detected by observers. Double-platform methods have been developed to address situations of incomplete detection at the track line, but they require the identification of sightings seen by both observers. However, there is no means to independently and unequivocally determine whether or not a given pair of sightings is in fact a duplicate pair, or to select the most likely duplicate among a set of candidate sightings observed in close proximity. Most previous studies have used ad-hoc methods and arbitrary thresholds. Here, we develop a data-driven approach to identify single and duplicate sightings made during the 2013 High Arctic Cetacean Survey (HACS). We make use of four covariates to compare sightings made by front and rear observers: difference in time of sighting, difference in declination angle, difference in group size and difference in species identity. To estimate the relative weights of these covariates, we compared two datasets in a logistic regression framework: a set of sighting pairs that contain both duplicates and non-duplicates and a similar dataset known to contain no true duplicates (the observations made at the same time but on the other side of the plane). This allowed us to determine which combinations of factors were most successful at discriminating duplicates and to rate each candidate pair within the same-side data with an index of dissimilarity. Candidates with the lowest scores were identified as duplicates using two different methods and a range of threshold values for each covariate. Depending on the procedure used, 19% to 30% of narwhal sightings in the HACS dataset were seen by both observers, whereas 36% to 50% of bowhead whale sightings were seen by both observers. However, the aggregated nature of the sightings and particularly the relatively high proportion of missing primary data such as declination and group size made the identification of duplicates uncertain in many cases.

Identification des observations effectuées en double dans le cadre de l’inventaire à deux plateformes des cétacés de l’Extrême-Arctique de 2013

RÉSUMÉ

L'une des principales hypothèses concernant l'échantillonnage en ligne avec mesure des distances est que tous les animaux situés sur le transect sont détectés par les observateurs. Les méthodes à deux plateformes ont été créées en vue de tenir compte des situations où la détection est incomplète le long de la ligne, mais elles nécessitent l'identification des observations faites en double par les deux observateurs. Cependant, il n'existe aucun moyen de déterminer de façon indépendante et sans équivoque si une paire d'observations est en réalité une observation en double, ni de choisir parmi un ensemble d'observations candidates effectuées à proximité celles qui sont les plus susceptibles de constituer des observations en double. La plupart des études précédentes utilisaient des méthodes subjectives et des seuils arbitraires. Dans la présente étude, nous avons élaboré une approche axée sur les données afin de repérer les observations simples et en double effectuées au cours de l’inventaire des Cétacés de l’Extrême-Arctique (ICE-A) de 2013. Nous avons utilisé quatre covariables pour comparer les observations effectuées par les observateurs installés à l'avant et à l'arrière : différence du moment de l'observation, différence de l'angle de déclinaison, différence de la taille du groupe et différence de l'identité de l'espèce. Afin d'estimer la valeur relative de ces covariables, nous avons comparé deux ensembles de données dans un cadre de régression logistique : un ensemble de paires d'observations contenant des observations simples et des observations en double, et un ensemble de données similaire qu'on sait ne contenir aucun double véritable (observations effectuées au même moment, mais de l'autre côté de l'avion). Cela nous a permis de déterminer quelles combinaisons de facteurs permettaient le mieux de distinguer les doubles et d'attribuer un indice de différence à chaque paire d'observations candidates de l'ensemble de données provenant du même côté. À l'aide de deux méthodes différentes et d'un éventail de seuils pour chaque covariable, on a désigné les observations candidates ayant reçu les indices de différence les plus bas comme étant des doubles. Selon la procédure utilisée, de 19 % à 30 % des observations de narvals de l'ensemble de données de ICE-A ont été effectuées par les deux observateurs, tandis que de 36 % à 50 % des observations de baleines boréales ont été effectuées par les deux observateurs. Toutefois, la nature agrégée des observations et particulièrement la proportion relativement élevée de données primaires manquantes, comme l'angle de déclinaison et la taille du groupe, ont rendu l'identification des doubles incertaine dans de nombreux cas

INTRODUCTION

Distance sampling is widely used for estimating animal abundance. One of the key assumptions of the method is that all animals on the track line are detected by observers (Buckland et al. 2001). So-called double observer or double-platform methods have been developed to extend line transect sampling to the case that not all animals on the line are detected (e.g., Quang and Becker 1997; Innes et al. 2002). The double-observer data can be regarded as two-sample mark-recapture in which detections are not necessarily independent between observers (Laake and Borchers 2004). In addition to collecting perpendicular distance for each sighting, observers collect additional data that allow classification of sightings into those seen by either observer 1 or observer 2 only (single sightings) and those by both observers (duplicate sightings). Classification can occur in the field or at the analytical stage. In the case of the 2013 High Arctic Cetacean Survey (HACS), this means duplicates between the front and rear platforms on the right and left sides.

Pollock and Kendall (1987) identified one of the major practical problems of mark-recapture distance sampling (MRDS) in aerial surveys as the difficulty of identifying duplicate sightings from single sightings. The central problem of duplicate determination is that there is no means to independently and unequivocally determine whether or not a given pair of sightings is in fact a duplicate pair. All observations are measured with error, so even if two sightings have equal covariate (beam time, declination, species, group size) values it is still possible that they are not duplicates. If we are to use covariate values to determine duplicates, it would be preferable to have a population of duplicate sightings that were identified independently of these covariates; otherwise the determination becomes tautological. Duplicate determination is in essence probabilistic, and a methodology that resulted in a probability that any two sightings were duplicates would be appropriate.

A second issue is the determination of *threshold* covariate values that identify duplicates. Most studies have relied on such thresholds to identify duplicates; for example, beam time difference of 5 seconds or less, declination difference of 10 degrees or less. However there is no obvious way to identify what thresholds are appropriate. Indeed, given that duplicate determination is probabilistic, we should not expect any particular threshold levels to be unbiased determinants of duplicate sightings. We reviewed the field and analytical techniques used in a collection of 49 aerial cetacean surveys conducted in the past 10 years. Of these 18 were double-platform surveys requiring duplicate matching comparable to that required in the HACS (not including surveys using circle-back techniques, which use a different matching procedure). Table 1 shows that all these surveys used simple rules for determining whether or not between-platform pairs of sightings were duplicates. Timing of the sighting (beam time) was most often used and sightings within 3–5 seconds of one another were considered as at least possible duplicates. Declination abeam was also used in several studies, and a difference of 10 degrees or less was considered a duplicate. Finally, a few studies used difference in group size as an indicator, with a difference of three or less considered a duplicate. A few studies used all three of these factors. In most cases little detail was provided beyond simple dichotomous rules (duplicate or not duplicate), and no studies provided any analysis of the actual data to justify these rules. This contrasts sharply with ship surveys, where potential time-in-view is much greater and elaborate duplicate matching procedures are often used.

A third issue arises where there are more than one candidate duplicate for a given sighting (i.e., several sightings made by the other observer within the time, angle and count thresholds). What combination of criteria should be used to select the best candidate among several plausible pairs of sightings? Although this situation is a common occurrence in highly aggregated species, quantitative methods to discriminate among potential duplicates are usually not given in the literature.

One example of an aerial survey analysis using a data-based, somewhat probabilistic method of assigning duplicates is provided by Southwell et al. (2002). This analysis of an Antarctic seal survey uses differences in beam time, beam declination, group size and species identity to classify potential duplicates. This allows the sensitivity of duplicate determination to each of these factors and to determination parameters to be assessed. For example, in this particular survey, duplicates were determined mainly by time abeam and beam declination, whereas group size and species identity had little effect because group size was strongly clustered for each species and it was rare to have more than one species close together. Beam time difference ≤ 5 seconds and declination difference ≤ 10 degrees captured most potential duplicates. While this study is a step forward in duplicate identification, the methodology still relies heavily on threshold covariate values that are not determined completely objectively. Also, it weights all covariates equally, even though not all covariates may provide the same level of information to discriminate between single and duplicate sightings.

The HACS dataset presents numerous problems for duplicate determination:

1. It is large, with 1,553 primary sightings and 2,565 potential duplicates with beam time differences of 10 seconds or less. It would have been time-consuming and difficult to individually examine every potential duplicate pair. An algorithm based method of duplicate determination was therefore required.
2. The distributions of the two most common species seen, narwhal and beluga, were highly aggregated, meaning that many sightings were often made in quick succession, often with beam time differences of as little as one second. Observer performance becomes degraded at such high encounter rates, and we can expect the accuracy of measurement of beam time, declination, group size and species identity to be affected.
3. For the same reason, there is a great deal of missing covariate data. Of the 2,630 sightings, 470 are missing valid declinations, 122 are missing group sizes and 18 are missing both. This problem is amplified because in any potential duplicate pair, the covariate is unusable if one member of the pair lacks a valid covariate value. Thus, of the 2,565 potential duplicates, 20% lack a valid declination difference.
4. The method used to mark time abeam is subject to error from at least three sources: watch calibration, recorder calibration and uncertainty in determining the time at which the sighting comes abeam. As beam time is an important factor in determining whether or not a pair of sightings are duplicates, this increases uncertainty by an unknown amount.
5. Other data that might have been useful in duplicate determination, such as direction of travel and the number of tusked vs not tusked narwhal, were rarely and inconsistently recorded and are therefore of little use in duplicate determination.
6. Mixed species aggregations were rarely seen; therefore species identity is likely to be of limited use in duplicate identification.
7. Group size for narwhal (the main target species) is tightly clustered in the range of 1–4. Therefore we can expect, on average, little difference in absolute group size between two sightings, even if they are not duplicates.
8. While observers did report if they were uncertain about any measurement, this is a subjective judgment and likely to vary between observers. Therefore it is not clear how this parameter can be integrated into duplicate determination.

In this analysis, we adapted the methodology developed by Southwell et al. (2002) to the HACS data and used it to identify duplicates from the visual survey data. In addition, we developed a novel method based on comparing covariate levels in a population of sightings pairs that contained duplicates, to one that did not. Duplicate determinations using both methods were then compared.

METHODS

DATA PREPARATION

In some cases declinations were recorded as ranges (e.g., 30–60), especially when large aggregations were encountered. These declinations were set to “missing”. Similarly group sizes recorded as ranges or some other descriptor (e.g., ”1 pod”) were set to “missing”. A few observations had extreme group sizes of between 30 and 600. As such sightings may involve the combining of many groups into one, sighting pairs including group size counts greater than 30 were analyzed as if group size was missing.

Some species identities were recorded as an uncertain choice between two species. A frequent example was “Mm/Dl”, which can be interpreted as “probably narwhal maybe beluga”. However this type of recording was specific only to certain observers. To facilitate comparisons such codings were revised to the probable species with low certainty; e.g., in the above case to “Mm certainty 0”. Certainty levels for sightings recorded as “Unknown” were revised to 1, meaning “certain” (i.e., the observer was certain he/she didn’t know what it was).

USE OF PHOTOGRAPHS

The HACS incorporated digital photography of the trackline in addition to the visual observers, and it was expected that these photos could be used to fill in data missed by observers and enhance the accuracy of observer observations, in cases where a photo-identified sighting could be clearly linked to an observer sighting. Sightings where angles of declinations had not been recorded or were coded as “uncertain” were compared to the photographic records. However, the process of linking sightings identified in photos to observer sightings is very similar in concept to matching front-rear observer sightings and therefore subject to the same uncertainty. Therefore, we used photographic data only if a visual sighting could be identified without ambiguity on the corresponding photograph (usually when there were only one or few sightings on the frame). In those cases, the perpendicular distance was retrieved from the pixel position of the sighting on the photo. If the sighting was not made within the swath width of the picture, could not be found, or could not be told apart from other sightings unambiguously, it was coded as missing distance. Similarly, sightings where group size had not been recorded or was coded as uncertain were also compared to the photographic records, and group size was retrieved if a match could be made based on perpendicular distance. Otherwise, sightings with missing group size were coded as missing count for the duplicate identification process.

Unlike Southwell et al. (2002), who performed calibration tests with known objects, we have no prior information on the distribution of errors around the measurements made by HACS visual observers. To gain some insights, we compared the visual and photographic measurements (angle, count) for all cases where photographic observations could be unambiguously matched to a visual sighting.

IDENTIFICATION OF POTENTIAL DUPLICATES

The dataset was segregated by side of the plane (R or L) and all sightings for each side were sorted by plane, date and beam time. For each sighting, the dataset was scanned forward 10 seconds, and all sightings made within this time range by the other station on the same side (rear for front sightings, front for rear sightings) were identified and paired with the initial sighting as a *potential duplicate*. Therefore a sighting could have several potential duplicate matches within the 10 second interval. Potential duplicate pairs for the R and L sides were then combined into a single dataset incorporating all relevant data for each sighting.

For each potential duplicate pair, the following covariates, based on variables directly recorded by the observers, were derived:

T: Difference in beam time in seconds;

D: Difference in declination angle in degrees;

C: Difference in group size.

S: Difference in species identity (Table 2).

Other potential covariates, such as direction of travel, number with tusks and number with no tusks, were recorded too infrequently to be of use in a statistical analysis.

DUPLICATE ASSIGNMENT USING SOUTHWELL SCORES

Dissimilarity indices, hereafter referred to as Southwell scores, were derived based on the methods used by Southwell et al*.* (2002).

1. Beam time $D_t = ABS(T)/T_{max}$ where D_t is the beam time index and T_{max} is the threshold time difference, in seconds.
2. Declination: $D_d = ABS(D)/D_{max}$ where D_d is the declination index and D_{max} is the threshold declination difference, in degrees.
3. Group Size: $D_c = ABS(C)/C_{max}$ where D_c is the group count index and C_{max} is the threshold count difference. For the purpose of this analysis, *C*>30 was considered implausible and *C* was considered missing for these pairs.
4. Species identity: Scores were assigned based on recorded species identity and observer certainty (1=certain, 0=uncertain) as described in Table 2.

An overall Southwell Score was the sum of the dissimilarity indices described above, divided by four, with a score of one being the maximum value for a *valid* duplicate pair candidate. For pairs for which some data were lacking, the score was the sum of the available indices.

Thresholds under the Southwell method were considered “hard” caps; i.e., candidate duplicates had to score below the threshold values on *all* available indices (and thus have an overall score below one) to be considered valid. For any one sighting there could be several valid candidate pairs that scored below this threshold; obviously any one sighting can have only one true duplicate. For these cases the candidate with the lowest Southwell Score was chosen as the true duplicate pair. For cases where one or more of the candidates had missing data, candidates were ranked according to the best Southwell Score (two, three or four indices) available for all candidates. In the case of a tied score, the candidate with the lowest time difference *T* was chosen.

Choice of threshold levels for covariates is usually subjective. Here, we tried to estimate threshold levels by examining graphs showing the number of duplicates as covariate levels changed. It was expected that such curves would show a sharp initial increase followed by a levelling-off, with the inflection being roughly equivalent to the value below which most duplicates could be found (Southwell et al*.* 2002).

DUPLICATE ASSIGNMENT USING A LOGISTIC REGRESSION METHOD

The dataset described above contains sightings pairs that are duplicates and others that are not, in unknown proportion. Duplicate sightings pairs are expected to have a particular range and mix of covariate levels that differentiate them from non-duplicate sightings pairs. Some covariates may be more important than others in differentiating duplicates.

In an effort to determine what factors were most useful to identify duplicates, we created a dataset that contained sighting pairs between front and rear observers that occurred close together in time but were certainly not duplicates. This was done in exactly the same way as the dataset described above, except that observer sightings were paired with stations on the *opposite* side of the plane. For example, a Front Right sighting would be paired with a Back Left sighting that occurred within the 10 second interval. This *opposite-side* dataset) should be similar to the *same-side* dataset except that it cannot possibly contain duplicate sightings pairs.

Logistic regression was used to determine which covariates were important in identifying data that contained duplicate pairs. The response variable was same-side (1) vs opposite-side (0). Candidate logistic regression models were fit using all combinations of individual covariates *T, D, C* (if *C*>30 it was considered missing) and *S*. For each case, the model with the highest Area Under Curve (AUC), representing the best reclassification performance, was chosen.

Separate models were created for:

1. Pairs for which all four covariates were available;
2. Pairs with missing declination angle differences *D*.
3. Pairs with missing count differences *C.*
4. Pairs with missing both *D* and *C*.

Using the coefficients from the best model in each of these situations, regressions produced *p* values corresponding to the probability that a particular sightings pair belonged to the same-side dataset, as opposed to the opposite-side dataset. Because the same-side dataset contains a mix of duplicate and non-duplicate pairs, these *p* values do not correspond directly to the probability that a pair is truly a duplicate. However, since the main difference between these two datasets was the presence of duplicate pairs in the same-side dataset, these scores were interpreted as a *relative* index of the probability that a particular sightings pair is a duplicate.

For each of these models, we calculated *p(0)* the value of *p* when all covariates are at 0 (i.e., the maximum value of *p*). To be able to pool scores from the four models, we scaled *p* values by dividing them by *p(0)* and used *1-p* to obtain a dissimilarity index ranging between 0 (most likely to be a duplicate) and 1.

This dissimilarity index obtained with the logistic method was used in the same manner as described for the Southwell Scores to rank and identify duplicate sightings pairs, except that for the logistic method, we used a single threshold *p(T)* obtained by substituting the covariate thresholds in the logistic regressions equations. This means that a duplicate can exceed a threshold for a given covariate if the other covariate values are low e.g., a candidate pair with a declination difference of 12 degrees might be considered a duplicate even if the threshold is 10 degrees, if its time, count and species differences were low enough to result in a score below *p(T)*. In other words, contrary to the Southwell approach, here the T_{max}, $D_{\max}$, C_{max} and S_{max} thesholds are “soft” caps. To avoid extremely unlikely values, however, we placed a hard cap of 20 degrees on *D*.

RESULTS

VISUAL AND PHOTOGRAPHIC DATA

The HACS dataset contains 2,553 on-effort visual sightings, of which 2,422 were made in double-platform conditions (1,428 made by the primary platform and 994 by the secondary platform). The observations include 1,600 narwhal, 224 bowhead and 314 beluga sightings.

Of these, 431 sightings were missing declination angle measurements and 114 were missing group size counts. Using photographic data, we were able to recover 299 missing angles and 39 missing group sizes.

The differences between angles measured by visual observers and those determined from photographs were mostly distributed between 0 and 5°, although about 8% were between 5 and 10°, and 4% were beyond 10° (Fig. 1). The majority of differences>10° corresponded to sightings that the visual observers had identified as “uncertain” measurements.

The difference between visual and photographic data sets for group size counts were tightly clustered around zero and one (Fig. 1), with few differences beyond two and almost none beyond five. Measurements deemed “uncertain” by visual observers accounted for the majority of the cases where the difference was >3.

COVARIATES FOR SAME-SIDE AND OPPOSITE-SIDE DATASETS

There were 2,506 pairs of sightings made by same-side front and rear observers separated by 10 seconds or less, which are therefore potential duplicate candidates. Of these, after recovery of missing data from photographs, 2,140 pairs had data for all covariates for both sightings, but 222 were missing a declination value for at least one sighting in the pair, 89 were missing a count value, and 55 were missing both declination and count.

The opposite-side dataset contained 1,526 pairs of sightings made within 10 seconds of one another, including 1,364 with all covariate data; 80 were missing a declination angle covariate, 75 were missing a count covariate and 7 were missing both declination and count.

Examination of the frequency distributions of the covariates (*T, D, C* and *S*, Fig. 2) showed that the distributions of declination difference *D* differed strongly between the two datasets, with the same side data showing an rapid decrease in frequency from 0–5° to 5–10° and continuing to decrease steadily to ~25°, whereas the opposite side distribution remained relatively flat out to declination differences of 25°.

There seemed to be less discriminatory information in time and count differences *T* and *C*. The same-side distribution showed a slow steady decrease from 1–3 seconds to ~10 seconds (with 10-second values about half as frequent as 1-second values). In the opposite-side dataset, the distribution remained relatively flat from 1–10 seconds. Frequency of count differences decreased steadily from 0–5 in the same-side dataset, but showed no substantial decrease in the opposite-side data at values <2 and then decreased more slowly than same-side differences. There was no clear difference in the distribution of species differences *S* between the two datasets.

The distribution of differences in perpendicular distance rather than declination angle is shown in Fig. 3. The frequency of distance differences of 50–100 m is reduced to less than 50% of the frequency of 0–50 m differences, while the distribution on the opposite side remains flat up to 200 m (with the 150–200 m frequency remaining at 85% of that of the 0–50 m bin).

SOUTHWELL THRESHOLDS

Southwell et al. (2002) used graphical methods to identify thresholds for duplicate identification. Graphs are provided for the HACS dataset in Fig. 4. The inflection point for time differences *T* was not clear but the rise in the number of duplicates (*Dup*) did slow at *T* greater than 5 seconds. The sensitivity of *Dup* to angle differences *D* appeared to lessen at *D* greater than 5° and slowed down further at 10°. The rise in the number of duplicates slowed rapidly at levels of count differences *C* greater than three or four. As expected from Fig. 2, there was little sensitivity of *Dup* to species differences *S*, especially at levels above 0.5. For the most part, these thresholds were not apparent in the opposite-side dataset (Fig. 4, red lines), which confirmed that these inflexion points were due to the presence of true duplicates in the same-side dataset. Based on this sensitivity analysis and on the distribution of measurement errors shown in Fig. 1, conservative thresholds of T_{max}=5 seconds, D_{max}=10°, C_{max}=3 and S_{max}=0.50 were chosen.

LOGISTIC REGRESSIONS

Table 3 describes the equations of the best logistic regression models for sighting pairs with no missing data as well as each case of missing data. The magnitude of the standardized coefficients corresponds to the relative amount of variance accounted for by that coefficient, while the AUC represents the success of each model at correctly classifying pairs into each data set (same side vs opposite). As expected, all coefficients for main effects were negative; i.e., an increase in covariate differences *T, D, C* or *S* caused a decrease in the probability that the pair was from the same side. Difference in declination (-0.56) was the most important factor in discriminating the same-side from the opposite-side cases when all covariates were available, with count differences *C* being slightly less useful (-0.37), and *T* and *S* of less importance. This was also the case when *C* was missing. When declination difference *D* was missing, the other three covariates became equally important, each explaining one third of the variance.

Models were also evaluated using perpendicular distance *P* rather than declination angle *D*, but resulted in lower reclassification rates and therefore only models with *D* were used to identify duplicates.

COMPARISON OF SOUTHWELL AND LOGISTIC SCORES

The distributions of Southwell scores (using the thresholds defined above) and dissimilarity indices based on the logistic regression are shown in Fig. 5. Differences between the distributions of the same-side and opposite-side datasets suggest that both criteria are somewhat successful at discriminating data that contain duplicates from data that do not.

Southwell scores performed well at penalizing opposite-side pairs (i.e., a majority of opposite-side pairs receive large scores), but a large proportion of same-side pairs also received high scores, suggesting that the Southwell approach may create false-negatives by penalizing a number of true duplicates. In contrast, logistic regressions gave lower scores to same-side pairs (distribution more skewed to the left), suggesting that this method may be superior at identifying true duplicates within the same-side dataset. However, the large degree of overlap between same- and opposite-side scores may imply that this approach generates a substantial amount of false-positive duplicate candidates.

IDENTIFICATION OF DUPLICATE SIGHTINGS AND SENSITIVITY ANALYSIS

Using the thresholds identified above (T_{max}=5 sec, D_{max}=10°, C_{max}=3 and S_{max}=0.5), the Southwell method identified a total of 379 duplicate sightings, i.e., 17% of 2,251 unique sightings, whereas the logistic method identified 558 duplicates (27% of 2,072 unique sightings).

Table 4 shows a breakdown for the main target species. In all cases the logistic method resulted in a higher proportion of duplicates, but the difference was least important for bowhead whale (which had the highest overall duplicate proportion) and much larger for narwhal and beluga.

Figure 6 shows the sensitivity of the thresholds used in the logistic method on the numbers of duplicate and unique sightings for all data as well as broken down by species. Only the thresholds for time differences *T* and declination angle differences *D* are shown, because they account for most of the variability (*C* and *S* have less effect). The minimum and maximum numbers of unique sightings, using extreme values of T_{max} and D_{max}, were 146–148 for bowhead whale, 253–257 for beluga, and 1316–1,335 for narwhal.

DISCUSSION AND CONCLUSIONS

IDENTIFICATION OF DUPLICATE SIGHTINGS

The identification of duplicates in the HACS dataset is not a trivial task. The aggregated nature of the sightings and particularly the relatively high proportion of missing primary data such as declination and group size made the identification of duplicates highly uncertain in many cases. Any rule or model which purports to identify duplicates is, in actuality, identifying *nominal* duplicates, which result from the interplay between true duplicate identifications, false positive identifications and false negative classifications. Thus:

$$D_T = \widehat{D} - D_{FP} + D_{FN}$$

Where $\widehat{D}$ is the number of nominal duplicates, D_T is the number of true duplicates, D_{FP} is the number of false positive duplicates, and D_{FN} is the number of false negative duplicates. Unfortunately we have no independent, verifiable method of quantifying D_T, D_{FP} or D_{FN} so the identification of duplicates will always be probabilistic and subject to error.

Ideally, the establishment of covariate thresholds for identifying duplicates should be finding the covariate level or levels where $D_{FP} = D_{FN}$ (or both equal zero), thus $\widehat{D} = D_T$. We attempted to identify covariate thresholds using the graphical methods described by Southwell et al. (2002). Unfortunately, inflection points in our data were not nearly as clear-cut as in the dataset used by Southwell et al. (2002). This is probably because our sightings were much more aggregated than those used in Southwell et al. (2002), which leads to greater uncertainty in measures of beam time, declination, group size and species identification. Southwell et al. (2002) also used instrumentation to measure beam time and declination, probably more accurately and precisely than we did. Therefore our data are almost certainly more “noisy” than that used by Southwell et al. (2002).

We compared opposite-side sightings to same-side sightings, all made within a 10 second time window, to try to determine which covariates contributed most to the identification of duplicates. These two datasets should have been similar in every respect except that the same-side data contained a high proportion of duplicates while the opposite-side data contained none. Obvious differences were expected in the covariate distributions of the two datasets. The distributions of declination differences (*D*) differed strongly between same-side and opposite-side, the distributions of time (*T)* and count *(C)* differences did not differ as much, while the distributions of species differences (*S*) were similar This can probably be explained by three features which are particular to the HACS dataset:

1. Sightings of narwhal and beluga (and to a lesser extent, bowhead whale) were highly aggregated and they make up the majority of the sightings in the dataset. Therefore front and rear observers tended to make observations close together in time, even if they were on opposite sides of the plane. This explains why T was not as important as expected in

discriminating between same side and opposite side sightings pairs, and hence in the detection of duplicates. The data also showed that differences of 0 seconds were not as frequent as differences of 1 or 2 seconds, suggesting that it is unrealistic to expect most duplicated sightings to be made at exactly the same time in this type of survey.

2. Almost all sightings tended to occur in single-species aggregations, separated in space and time from other aggregations. Mixed-species aggregations were rare. Therefore species identity was a weak discriminator of duplicates since nearly all sightings close together in time were of the same species.
3. The group size of narwhal and beluga sightings was tightly clustered between one and five. Thus the difference in group size between two randomly chosen sightings is likely to be less than four. This likely explains why C was not a strong discriminator between same-side and opposite-side sightings pairs, and thus between duplicates and non-duplicates.

Thus the difference in declination *D* remains the strongest indicator of duplicate sightings. Declination is not a biological or ecological attribute and is not, on average, affected by the contagious distribution of the sightings. It is, however, affected by the detection curve of observers, i.e., low values will be more frequent than large ones, which will have an effect on the distribution of angle differences between platforms. Similar to group count, it is sampled from a clustered distribution, but the range is substantially wider than for group count. Therefore opposite-side (non-duplicate) sightings pairs have a strong tendency to differ more in declination than do same-side (possibly duplicate) sighting pairs, which is problematic because 13% of the pairs of sightings made within 10 seconds or less lacked at least one valid declination measurement, even after photo verification. However, one appealing feature of these methodologies is that weaker discrimination due to missing covariates results in more, not fewer duplicate identifications, which will lead to a higher estimate of *p(0)* and hence a more conservative abundance estimate.

The HACS incorporated digital photography of the trackline in addition to the visual observers, and these photos were used to replace data missed or considered uncertain by observers in cases where a photo-identified sighting could be clearly linked to an observer sighting. However, the process of linking sightings identified in photos to observer sightings is very similar in concept to matching front-rear observer sightings and therefore subject to the same uncertainty. Declination has proven to be the most important datum for identifying duplicates, and if it was missing from an observer sighting, it was nearly impossible to match it to a photo sighting unless that sighting was isolated in time or had other unique attributes such as an unusually large group size.

EFFECT ON ABUNDANCE ESTIMATION AND RECOMMENDATIONS FOR ABUNDANCE ANALYSES

The two methodologies shown here provide a range of estimates of the number of duplicates in the dataset, with the Southwell method producing the fewest and the Logistic producing the most. There are also qualitative differences in the duplicate pairs chosen by the methods, because the Logistic method gives different weights to the available covariates, while the Southwell method weights them equally.

One of the main differences between the two approaches is that the Southwell places a “hard” threshold on each covariate. If a sighting pair exceeds this threshold for any of the four covariates, it is ruled out as a valid duplicate candidate. The logistic approaches, however, uses a single threshold level obtained by injecting the four covariate thresholds in the logistic equation. Therefore, if a sighting pair slightly exceeds a threshold for one covariate, it can still be considered a valid candidate provided it scores low enough on the other covariates. This allows for additional flexibility and considerably lessens the impact of choosing arbitrary thresholds. Fig. 6 shows that the resulting number of duplicate and unique sightings under the logistic method is relatively robust to the choice of thresholds. It also shows that the choice of thresholds has almost no impact on the number of bowhead whale unique sightings that enter the abundance analysis.

Differences in the number of duplicate sightings will have impacts on abundance estimates at three levels:

a) the number of unique sightings will determine the raw number of individuals seen on transect from which the total population abundance will be estimated;

b) the identity and characteristics of unique sightings (distance, count size) will have an effect on estimation processes such as fitting a detection curve and calculating expected group size; and

c) the number, distance and size of unique and duplicate sightings will have an effect on the mark-recapture distance model, and thus on *p(0)*, which serves as a multiplier on the abundance estimate.

Effect a) was quantified in our sensitivity analysis and testing ranges of 2–10 seconds and 4–16° resulted in maximum changes of around 1–2%, i.e., relatively minor effects that can implemented in the final abundance estimate if desired, for instance by using a multiplier factor of 1 with a CV of 1–2%, thus leaving the point estimate unchanged but increasing the range of uncertainty slightly.

Effect b) is only expected if changes in the identity of duplicate and unique sightings are not homogeneous in relation with characteristics such as distance and group size. Since distance and group size are part of the covariates used to identify duplicates, this possibility cannot be excluded. However, the effect of b) on abundance estimates is much harder to predict without actually running the analyses multiple times with multiple sets of unique sightings.

The same applies to effect c) because distance and group size both play a role in the MRDS model. However, we can at least illustrate its potential impact by considering a simple case where *p(0)* is calculated by assuming independence between the probabilities of detection by observer 1 and 2 (p_1 and p_2). In that case, the proportion of duplicate sightings is equal to the product p_1p_2 and the combined *p(0)*$=1-(1-p_1)(1-p_2)$. In the case of narwhal sightings from our sensitivity analysis, the resulting *p(0)* would vary from 0.73 with 1,316 unique sightings to 0.68 with 1,335 unique sightings. One can see that, in this simplified example, the total abundance would vary from 1,316/0.73=1,803 to 1,355/0.68=1,963, a range that represents about 8% of the mean estimate.

For the HACS analysis, we recommend using the logistic method because it is less sensitive to threshold choices and provides more conservative estimates. We also recommend testing the sensitivity of the final abundance and PBR estimates to the choices made at the duplicate identification stage, by performing multiple analyses on the datasets of unique sightings obtained for threshold values of 3–7 seconds for time differences *T* and 5 –15° for declination angle differences *D*. This would allow the inclusion of a correction factor that encompasses all three effects described above simultaneously.

RECOMMENDATIONS FOR FUTURE SURVEYS

The degree of aggregation encountered on this survey was problematic for observers, leading to degradation in the accuracy and precision of their measurements, more missed measurements, and probably more missed sightings. Given that a photographic record is available, it might be better to rely on the photos in a strip survey mode when local encounter rate rises above a threshold level. However, photographic surveys do not cover the same strip width, and are analyzed differently (no detection function, instantaneous availability correction factor, no perception bias). Therefore, combination of visual and photographic data would likely require development of an appropriate ad-hoc statistical framework.

In this survey, the observer manual instructed observers as follows: “For all sightings the first three variables are required for abundance estimation (in order of importance below) and if you have time, and have no other sightings while observing, then record in order of importance the remaining information: 1) Species; 2) Group size [if overwhelmed by numbers of groups, you should track group size (adding small groups as described above) rather than trying to collect clinometers angles]; 3) Clinometer angle (to the nearest degree)…” These instructions are somewhat ambiguous in that they say on the one hand that the recording of species, group size and declination are required for abundance estimation, but on the other hand order them in a priority which implies that declination can be sacrificed. As a result, most observers went into a mode of recording only species and group size when encounter rate was high. This is problematic for at least two reasons. First, it makes duplicate assignment uncertain, which in turns creates uncertainty about perception bias. Second, sightings for which declination is not available can be incorporated into a line transect estimate, but this requires the assumption that the detection function is the same under these circumstances (i.e., within aggregations) as in areas with lower encounter rates, where observers did not miss declinations. This assumption is almost certainly false, as observers do change their observation patterns in areas of high density. For example, they cease doing *D*-scans and do not scan far ahead of the plane.

One approach would be to instruct observers to obtain species, group size AND declination for all observations under all circumstances, and to narrow their field of view in areas of high density, so that they have time to record all data. However, this would result in a narrower effective strip width and probably more missed sightings (i.e. greater perception bias) in the relatively few areas with high densities, perhaps giving them a disproportionate weight. It is not clear how well these biases could be accounted for in the statistical analysis. Species with high aggregation rates such as narwhal will always be problematic by overwhelming visual observers and may benefit from a more extensive use of photographs.

REFERENCES

Asselin, N.C. and Richard, P.R. 2011. Results of narwhal (*Monodon monoceros*) aerial surveys in Admiralty Inlet, August 2010. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/065. iv + 26 p.

Asselin, N.C., Ferguson, S.H., Richard, P.R. and Barber, D.G. 2012. Results of narwhal (*Monodon monoceros*) aerial surveys in northern Hudson Bay, August 2011. DFO Can. Sci. Advis. Sec. Res. Doc. 2012/037. iii + 23 p.

Borchers, D.L., Pike, D.G., Gunnlaugsson, Th. and Vikingsson, G.A. 2009. Minke whale abundance estimation from the NASS 1987 and 2001 aerial cue-counting surveys taking appropriate account of distance estimation errors. NAMMCO Sci. Publ.7:95–110.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. 2001. Introduction to Distance Sampling. Oxford University Press, Oxford.

Heide-Jørgensen, M.P. and Acquarone, M. 2002. Size and trends of bowhead whale, beluga and narwhal stocks wintering off West Greenland. NAMMCO Sci. Publ. 4: 191–210.

Heide-Jørgensen, M.P., Laidre, K., Borchers, D., Samarra, F. and Stern, H. 2007. Increasing abundance of bowhead whales in West Greenland. Biol. Lett. 3: 577–580.

Heide-Jørgensen, M.P., Borchers, D.L., Witting, L., Laidre, K.L., Simon, M.J., Rosing-Asvid, A. and Pike, D.G. 2008. Estimates of large whale abundance in West Greenland waters from an aerial survey in 2005. J.Cetacean Res. Manag. 10(2): 119–129.

Heide-Jørgensen, M.P., Laidre, K., Borchers, D., Marques, T.A., Stern, H. and Simon, M. 2009. The effect of sea-ice loss on beluga whales (*Delphinapterus leucas*) in West Greenland. J.Cetacean Res. Manag. 29: 198–208.

Heide-Jørgensen, M.P., Laidre, K.L., Burt, M.L., Borchers, D.L., Marques, T.A., Hansen, R.G., Rasmussen, M. and Fossette, S. 2010a. Abundance of narwhals (*Monodon monoceros*) on the hunting grounds in Greenland. J. Mammal. 91: 1135–1151.

Heide-Jørgensen, M.P., Witting, L., Laidre, K.L., Hansen, R.G. and Rasmussen, M. 2010b. Fully corrected estimates of common minke whale abundance in West Greenland in 2007. J.Cetacean Res. Manag. 11: 75–82.

Hobbs, R.C. and Waite, J.M. 2010. Abundance of harbor porpoise (*Phocoena phocoena*) in three Alaska regions, corrected for observer errors due to perception bias and species misidentification, and corrected for animals submerged from view. Fish. Bull.108: 251–267.

Innes, S., Heide-Jørgensen, M.P., Laake, J.L., Laidre, K.L., Cleator, H., Richard, P. and Stewart, R.E.A. 2002. Surveys of belugas and narwhals in the Canadian High Arctic in 1996. NAMMCO Sci. Publ. 4: 169–190.

Kelly, N. Peel, D., Bravington, M.V. and Gales, N. 2010. Analysis of minke whale sighting data from aerial surveys over pack ice in east Antarctica. IWC SC/63/IA3

Laake, J.L., and Borchers, D.L. 2004. Methods for incomplete detection at distance zero. In Advanced Distance Sampling. Edited by S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L Laake, D.L. Borchers and L. Thomas. Oxford University Press, Oxford. p. 108–189.

Laidre, K.L. and Heide-Jorgensen, M.P. 2011. Life in the lead: extreme densities of narwhals *Monodon monoceros* in the offshore pack ice. Mar. Ecol. Progr. Ser.423: 269–278.

Lawson. J.W. and Gosselin, J.-F. 2009. Distribution and preliminary abundance estimates for cetaceans seen during Canada's marine megafauna survey - A component of the 2007 TNASS. DFO Can. Sci. Advis. Sec. Res. Doc. 2009/031. vi + 28 p.

Pike, D.G. 2009. Icelandic aerial survey 2009: Preliminary survey report. NAMMCO SC/17/AE/10. Available on request from the NAMMCO Secretariat.

Pike, D.G. and Gunnlaugsson, Th. 2008. T-NASS Icelandic aerial survey: Survey report and a preliminary abundance estimate for minke whales. NAMMCO SC/15/AE/04. Available on request from the NAMMCO Secretariat.

Pike, D.G., Gunnlaugsson, Th. and Vikingsson, G.A. 2011. Icelandic aerial survey 2009: Survey report and a preliminary abundance estimate for minke whales. NAMMCO SC/18/AESP/15. Available on request from the NAMMCO Secretariat.

Pollock, K.H. and Kendall, W.L. 1987. Visibility bias in aerial surveys: a review of estimation procedures. J. Wildl. Manag. 51: 502–510.

Quang, P. X. and Becker, E. F. 1997. Combining line transect and double count sampling techniques for aerial surveys. J. Agric. Biol. Environ. Stat. 2: 230–242.

Richard, P. 2010. Survey index of the Northern Hudson Bay narwhals, August 2008. DFO Can. Sci. Adv. Sec. Res. Doc. 2010/021.

Southwell, C., de la Mare, W., Underwood, M., Quartararo, F. and Cope, K. 2002. An automated system to log and process distance sight-resight aerial survey data. Wildlife Soc. Bull. 30: 394–404.

TABLES

Table 1. Duplicate matching procedures used in some recent aerial cetacean surveys.

REFERENCE	DUPLICATE MATCHING PROCEDURE
Asselin and Richard 2011	Timing ≤ 5 sec, declination ≤ 10 deg.
Asselin et al. 2012	Timing ≤ 10 sec, declination ≤ 15 deg.
Borchers et al. 2009	Timing ≤ 3 sec.
Heide-Jørgensen and Acquarone 2002	No detail
Heide-Jørgensen et al. 2007	No detail
Heide-Jørgensen et al. 2008	No detail
Heide-Jørgensen et al. 2009	No detail
Heide-Jørgensen et al. 2010a	Time <3 sec, Group ± 3, Perp Dist ± 200m
Heide-Jørgensen et al. 2010b	No detail
Hobbs and Waite 2010	Timing ≤ 5 sec, declination ≤ 10 deg, ancillary data
Innes et al. 2002	Primarily timing, also group size and declination. No detail.
Kelly et al. 2010	Time and declination. No detail
Laidre and Heide-Jørgensen 2011	Time <3 sec, Group ±3, Perp Dist ±200m
Lawson and Gosselin 2008	Timing and declination, no details given
Pike and Gunnlaugsson 2008	Timing ≤ 3 sec, declination ≤ 10 degrees
Pike 2009	Timing ≤ 3 sec, declination ≤ 10 degrees
Pike et al. 2011	Timing ≤ 3 sec, declination ≤ 10 degrees
Richard et al. 2010	Timing ≤ 5 sec, declination ≤ 10 deg., group size ±3.

Table 2. Species certainty D_s *for hypothetical species A and B, where U = Unknown. Certainty: 1 = certain, 0 = uncertain. Adapted from Southwell et al. (2002).*

OBSERVER 1		**OBSERVER 2**		***Ds***
ID	**CERT**	**ID**	**CERT**	
A	1	A	1	0
A	1	A	0	0.25
A	0	A	0	0.25
A	0	U	1	0.25
U	1	U	1	0.25
A	0	B	0	0.5
A	1	U	1	0.5
A	1	B	0	0.75
A	1	B	1	1

Table 3. Standardized coefficients of logistic regression equations with "same" and "opposite" sides as response variable, with no missing data (full) and sets missing D, C *or* D *and* C. p(T)*: Probability of belonging to the same side dataset at threshold levels of* T, D, C *and* S. p(0)*: Probability of belonging to the same side dataset with all covariates set at 0. ROC(AUC): "Receiver Operating Characteristic, Area Under Curve", it corresponds to the probability that a randomly chosen datum will be classified correctly.*

MODEL	***T***	***D***	***C***	***S***	***p(T)***	***p(0)***	**ROC(AUC)**
Full	-0.24	-0.56	-0.37	-0.20	0.58	0.75	0.63
Missing *D*	-0.32	-	-0.33	-0.35	0.52	0.71	0.59
Missing *C*	-0.23	-0.54	-	-0.09	0.62	0.72	0.60
Missing *D* + *C*	-0.31	-	-	-0.25	0.56	0.68	0.56

Table 4. Duplicate assignments using the logistic (LOG) and Southwell methods, for bowhead whale (Bm), beluga (Dl) and narwhal (Mm). N_p *– number of primary sightings from double platform effort;* N_s *– number of secondary sightings from double platform effort; N - total primary and secondary sightings from double platform effort; Dup – number of duplicate sightings;* N_U *– number of unique sightings, i.e.* N_U *= N - Dup.*

				LOG			**SOUTHWELL**		
SPECIES	**N_P**	**N_S**	**N**	**Dup**	**N_U**	**Dup %**	**Dup**	**N_U**	**D up%**
Bm	132	93	225	75	150	50%	59	166	36%
Dl	187	174	361	77	284	27%	40	321	12%
Mm	1000	714	1714	395	1322	30%	274	1440	19%

FIGURES

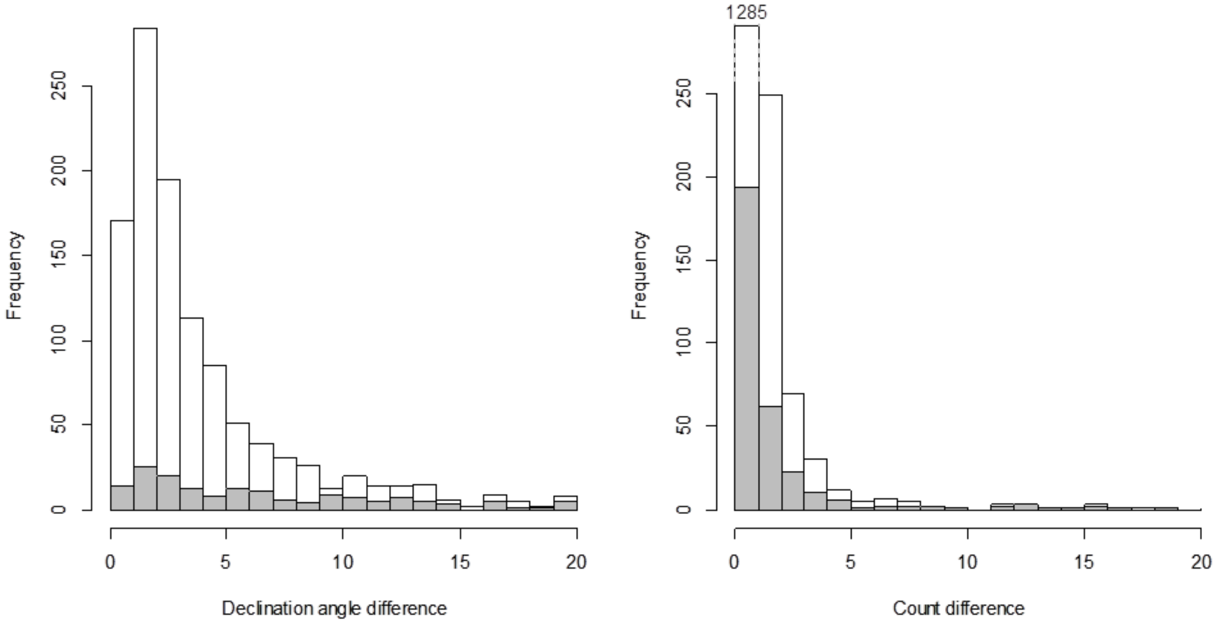

Figure 1. Distributions of the absolute differences in declination and group count between visual observations and the corresponding sightings found on photographs. White bars: all sightings. Grey bars: sightings coded as “uncertain” by visual observers for declination angle (left) or count (right).

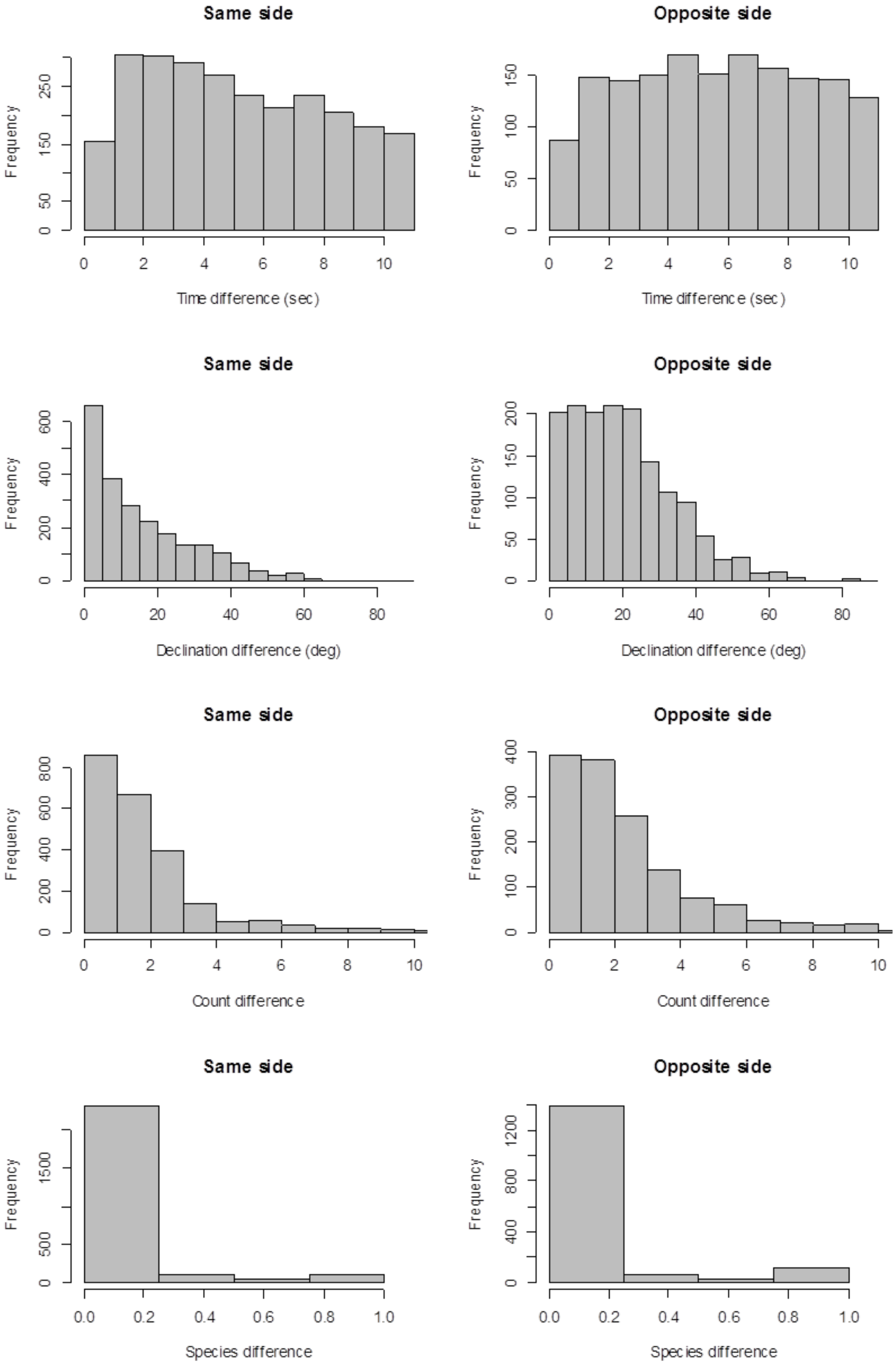

Figure 2. Distributions of the differences in beam time, declination, group count and species identity for the same side and opposite side data.

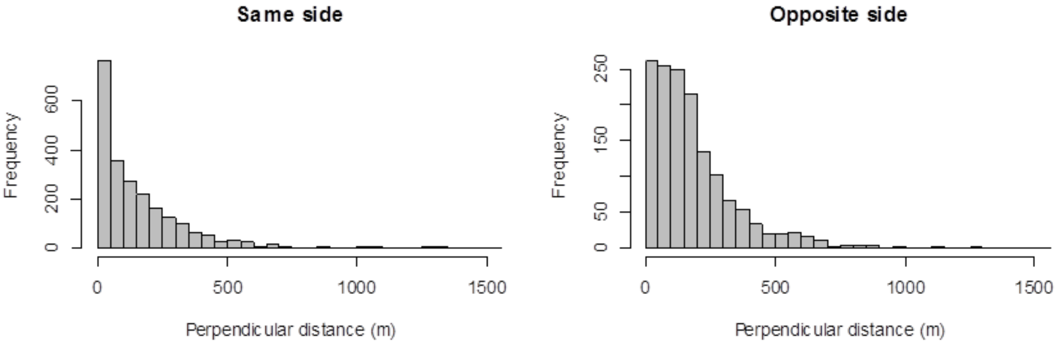

Figure 3. Distributions of the differences in perpendicular distance for the same side and opposite side data.

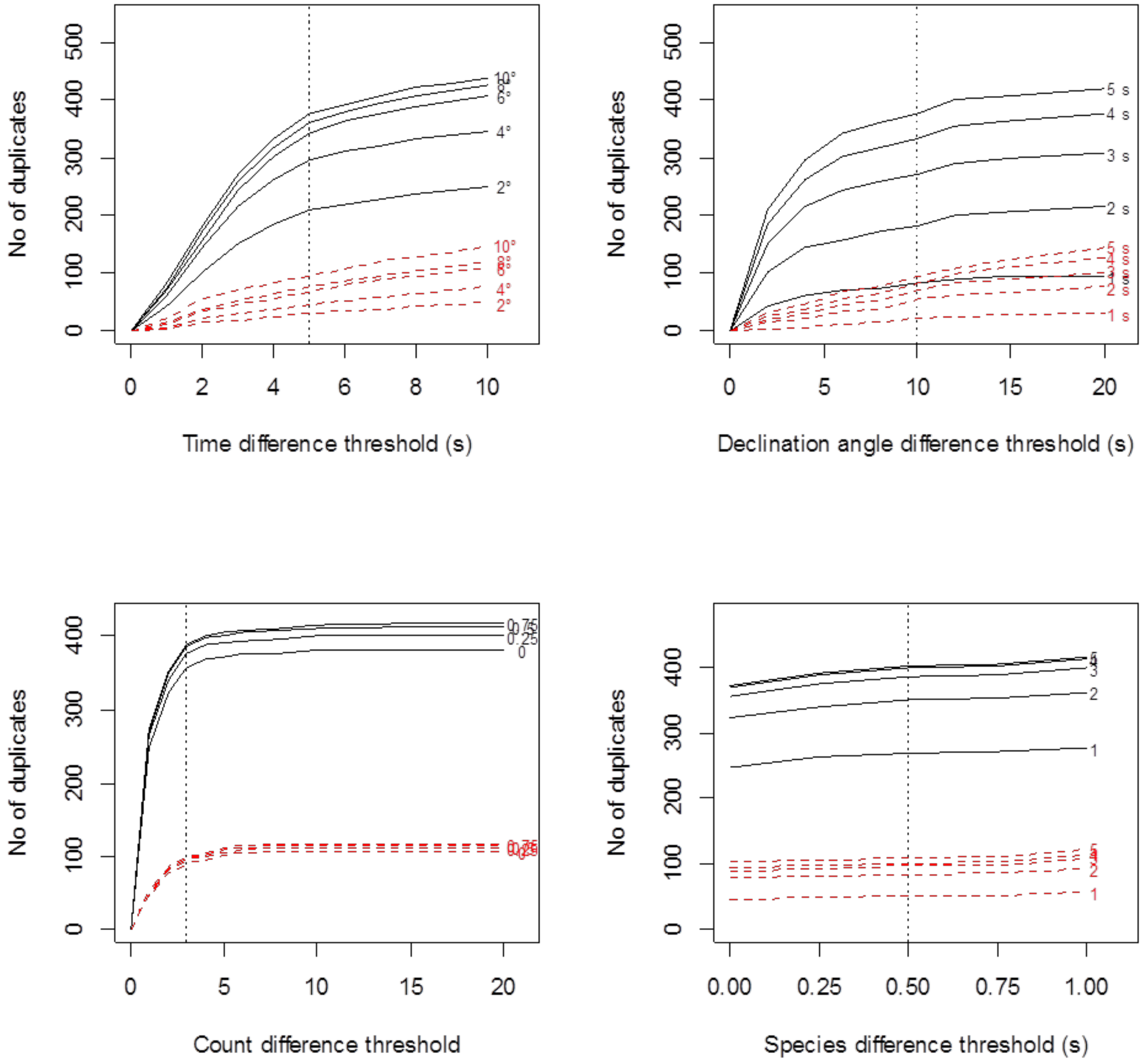

Figure 4. Sensitivity of the number of duplicate sightings identified with the Southwell method to variation in the between-observer difference in covariates. Solid black lines: same side dataset. Dashed red lines: opposite side dataset. Top: as a function of beam time difference T *and declination* D *(each line corresponds to a different threshold value), while holding* C_{max}*=3 and* S_{max}*=0.50. Bottom: as a function of group count difference* C *(x-axis) and species identification* S *(each line corresponds to a different threshold), while holding* T_{max}*=5 seconds and* D_{max}*=10°. Vertical dotted lines: proposed threshold value for each covariate.*

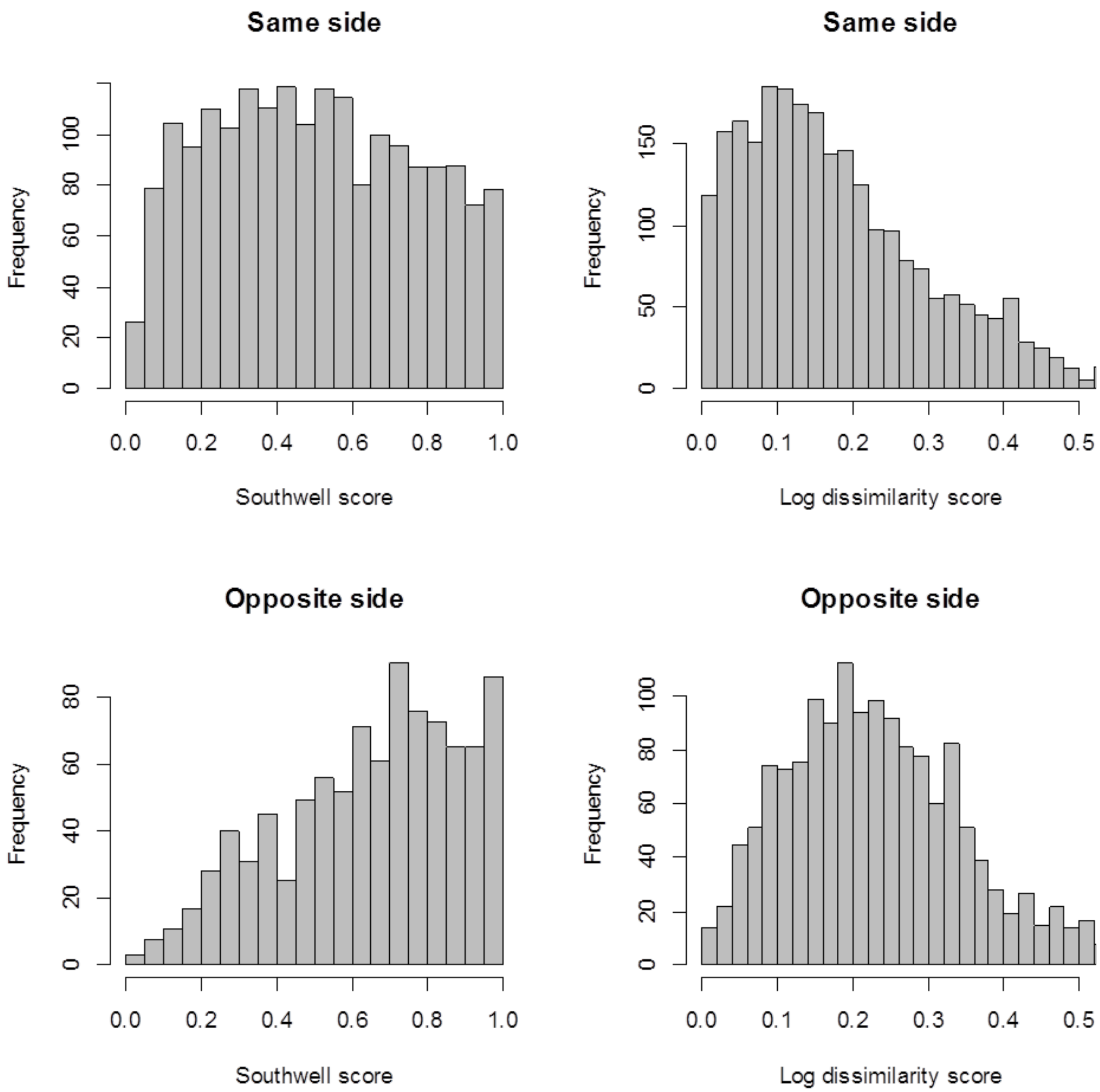

Figure 5. Left: Southwell scores for the same-side and opposite-side datasets. Right: Dissimilarity scores based on logistic regressions, for the same-side and opposite-side datasets.

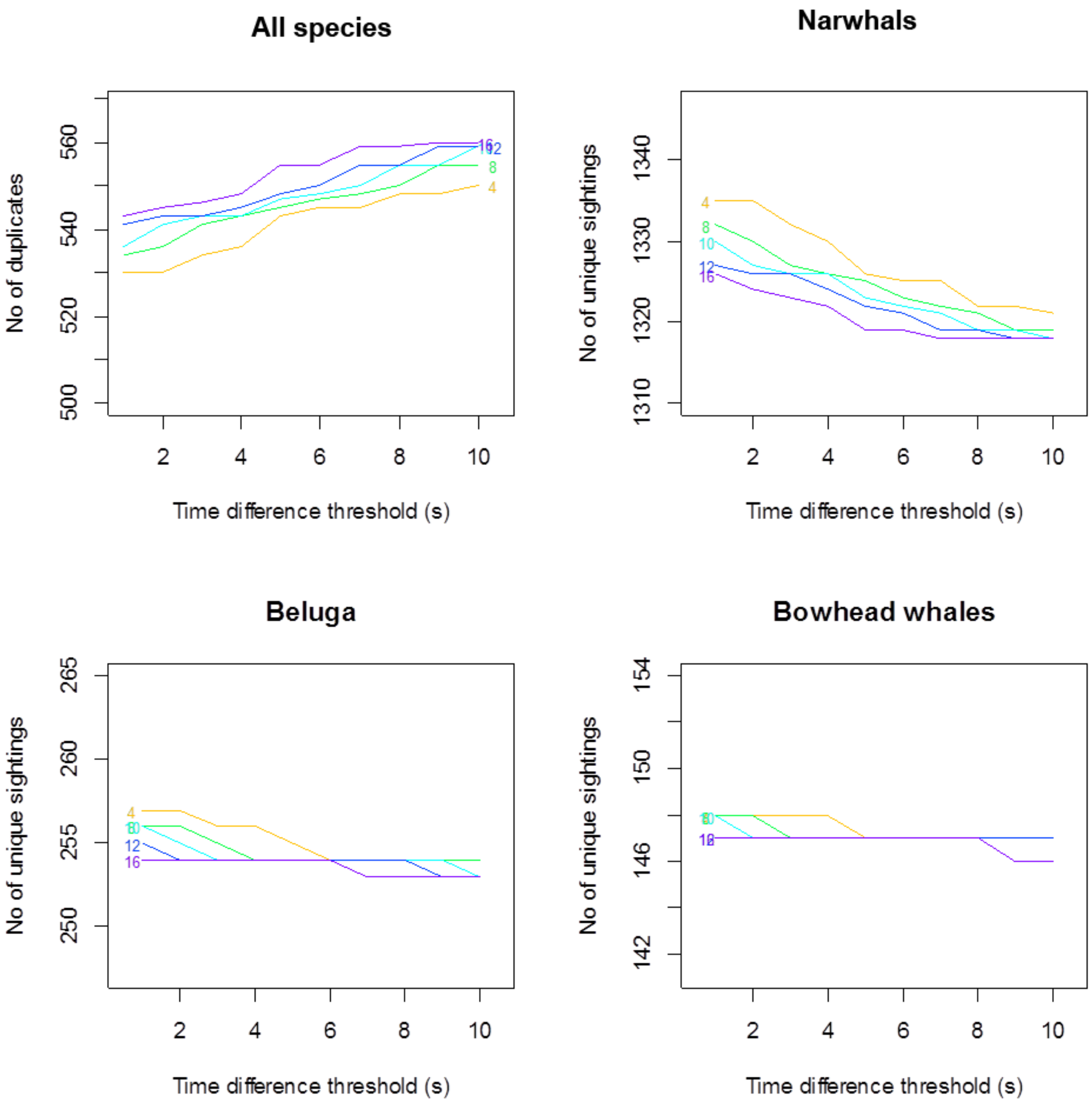

Figure 6. Sensitivity of the number of duplicate and unique sightings identified with the logistic regression method to variation in the between-observer difference in covariates beam time T *(x-axis) and declination angle* D *(each line corresponds to a different threshold value of* D*), keeping constant values of thresholds* C_{max}*=3 and* S_{max}*=0.50.*