Methods for investigating measurement error in cetacean line-transect surveys

Rob Williams*⌠§, Russell Leaper†, Alexandre N. Zerbini‡ and Philip S. Hammond*

*Sea Mammal Research Unit, Gatty Marine Laboratory, University of St Andrews, St Andrews, Fife, KY16 8LB, Scotland, UK. ⌠Present address: Pearse Island, Box 193, Alert Bay, BC, V0N 1A0, Canada. †International Fund for Animal Welfare, (IFAW), 87–90 Albert Embankment, London, SE1 7UD, UK. ‡National Marine Mammal Laboratory, Alaska Fisheries Science Center, NOAA Fisheries, 7600 Sand Point Way NE, Seattle, WA, 98115-6349, USA. §Corresponding author, e-mail: rmcw@st-andrews.ac.uk

Line-transect estimates of abundance generally assume that distances and angles to animals are measured without error. Biased distance measurements will lead to corresponding bias in abundance estimates. Efforts to address this have been made, but measuring distance to cetaceans at sea remains problematic. Four distance-estimation experiments were conducted to explore relationships between estimated and measured distance. Heteroscedasticity was found in all four cases. Preliminary evidence suggested that an observer differed in the ability to judge distance to fixed, continuously-visible cues and ephemeral, cetacean cues, which calls into the question the common practice of using marker buoys as cetacean proxies in distance-estimation experiments. Two studies found visual estimates to be positively biased, and two studies found reticule measurements to be negatively biased. Suitable correction factors were developed to address systematic bias from visual estimates, but these varied widely among observers and were sometimes non-linear. Errors in three studies showed positive skew, suggesting that ranges were overestimated to a larger degree than they were underestimated. If reticule and photogrammetric measurements yield log-normally distributed errors generally, then a least-squares regression will always overestimate the correction factor, underestimate range, and overestimate abundance. Photogrammetric methods to measure range to cetaceans performed well, and their use is encouraged. When measurements cannot be made to all sightings, however, it is recommended that experiments be conducted that generate sufficient sample size (of the target species, across typical survey conditions, and beyond the maximum range that most estimates will be made during the survey) to assess error distributions, examine evidence for non-linearity, and to consider inter-observer differences. Distance experiments, and training on survey protocols, can be conducted using ships of opportunity prior to the beginning of a dedicated survey, which would improve observer efficiency while reducing the costs of dedicated vessel time.

INTRODUCTION

A fundamental assumption underlying distance sampling techniques is that the relative locations of animals can be determined without error (Chen, 1998; Buckland et al., 2001; Palka & Hammond, 2001). In general, line-transect methods have been considered to be robust to small random errors in distance estimates, as long as no systematic bias is present (Chen, 1998; Chen & Cowling, 2001). However, more recent work has challenged this relaxed assumption. Marques & Buckland (2004) found that random error in perpendicular distance measurements caused overestimation of density in both simulated and real datasets.

Addressing measurement error in sightings surveys for cetaceans is critical since observers are generally poor at judging distance accurately at sea (Schweder, 1997; Baird & Burkhart, 2000) but distance estimates are a key data item in analysis. The probability of detection as a function of distance must be estimated from the sightings data. For analysis based on perpendicular distances, bias in distance estimates leads to a proportional bias in the resulting estimate of effective strip width. Errors in distance data also result in increased variance and may lead to the fitting of an inappropriate detection function. Researchers have compared true and estimated distances to wooden stakes, and found systematic bias that took the form of underestimating distance (Alpízar-Jara et al., 2001). This problem with distance estimation caused much greater bias in the resulting abundance estimates than did another common source of error in line-transect surveys, namely the tendency for observers to introduce random errors by rounding estimates to preferred values. For radial distance methods based on counts within an area, bias in the final abundance is approximately inversely proportional to the square of any bias in distance estimates. Such methods are also biased due to distance estimation error even if the distance estimates themselves are unbiased. Borchers et al. (2003) found that bias in conventional cue counting estimators of abundance becomes substantial when the error cross-validation is larger than about 10%.

Thus, measurement error clearly has the potential to introduce large bias in abundance estimates from line-transect sightings surveys. However, it has received

Table 1. *Summary of four distance-estimation experiments conducted during this study.*

Study	Cue type	Survey type	Eye height (m)	Region	Measurement method	Estimation method	Observers
1	Fixed (ice, rocks, zodiacs etc.) and transient (whales, penguins etc.)	Opportunistic	13.8	Antarctica	Video	Naked eye	1
2	Fixed (vegetation and life preservers)	River dolphin survey	5	Amazon	Range-finder	Naked eye	6
3	Transient (whales)	Cetacean survey	9.2	Alaska	Video	Reticule	1
4	Transient (whales)	Cetacean survey	18	Antarctica	Video	Reticule	1

surprisingly little attention at the data collection stage. Some analytical techniques have been developed to address inaccurate distance estimates (Schweder, 1997), but these *post-hoc* analytical approaches will always be less robust than actually obtaining accurate measurements in the first place. Most shipboard cetacean sightings surveys either rely on estimates of distance using the naked eye or from binoculars marked with reticules. (Aerial surveys often use clinometers to measure range, and are beyond the scope of this paper.) Reticule marks allow researchers on shipboard sightings surveys to measure the angle of dip from the horizon to an animal. This angle can then be used to estimate distance based on the height of the platform. Trained observers can collect unbiased distance estimates to stationary targets using reticule binoculars under good conditions (Lerczak & Hobbs, 1998). Without adequate training, however, these measurements can introduce a systematic bias that may vary among observers (Thompson & Hiby, 1985).

Sightings surveys frequently devote substantial time to training and testing observers in distance estimation (e.g. Schweder, 1997; Hammond et al., 2002). However, observer training is limited by the expense of ship time and the difficulties of creating realistic experiments. Such experiments generally use fixed targets of known size that are not necessarily representative of a brief surfacing of a real cetacean under survey conditions. Cetacean sightings surveys may differ in several respects from the conditions under which experimental data typically are collected. There may be qualitative differences in the way that the human eye perceives distance to fixed, continuously-visible targets of constant size such as floats, and transient cues such as a whale's blow or body. Similarly, observers using reticule binoculars may have more time to measure range to fixed targets than to cetaceans, thereby improving apparent performance. Environmental conditions such as sea state may remain constant while distance experiments are conducted, but vary throughout the survey.

Correcting for bias at an analysis stage based on such experiments is not straightforward. Simple linear correction factors may be overly simplistic if underlying relationships are non-linear. Errors may be observer specific and also related to unmeasured covariates, such as sea state, target size, aspect or bearing. Addressing these problems statistically will require a large number of observations. However, the expense of ship time on systematic surveys may make it difficult to collect sufficient sample sizes.

One recent technological advance is the development of photogrammetric methods for measuring range (Leaper & Gordon, 2001). This technique involves mounting a video camera to binoculars, and allows unbiased distance measurement to free-moving cetaceans based on the same trigonometric relationships as used for reticule binoculars. Laser range-finders were also used for surveys of river dolphins where no horizon was visible. Our paper uses these new techniques to compare measured and estimated distances from sightings surveys of a wide range of targets under a wide range of conditions. These data were used to investigate implications of distance estimation error and some issues inherent in using fixed-target experiments to correct radial distance estimates.

MATERIALS AND METHODS

Data from four studies comparing distances estimated by observers with measured values were examined. Distance estimates to a variety of targets were made using naked eye or reticule binoculars and measurements were made using laser range-finding and photogrammetric methods. A summary of the four sets of experiments is shown in Table 1. The basic principles behind distance estimation using photogrammetric measurement of distance are described in detail in Leaper & Gordon (2001). Measurements of the angle of dip from the horizon to the whale can be obtained from images that include both the horizon and the whale taken with a lens of known focal length. Provided the observation height is known the angle of dip can be used to calculate distance using simple trigonometry plus a correction for refraction. Still images were obtained from sequences using digital video cameras. Observers were given no feedback on their performance during the trials and so the data can be considered as distance-estimation experiments, rather than distance-training exercises.

Study 1 was carried out in conjunction with a cetacean line-transect survey conducted from a ship of opportunity in the South Atlantic sector of the Southern Ocean (Williams et al., 2006). The calibration experiment was conducted between 5 December 2001 and 12 February 2002. One observer (R.W.) conducted the experiment. While the ship was moving at cruising speed, the observer scanned from the primary platform with the naked eye. When an object was detected, the observer began recording on a hand-held video recorder and making simultaneous voice recording using an external microphone. The observer decided whether the object in question formed a fixed cue (i.e. it remained visible at the surface for the entire time it took to find it, judge distance, and record the object and the distance estimate voice note on the video cassette). Cues were deemed to be transient cues if the object submerged at

least once during this procedure. Transient cues targeted for this exercise included cetaceans, pinnipeds and penguins. Fixed cues included fur seals and penguins resting on the surface, inflatable boats and chunks of ice that were judged approximately whale-sized or smaller.

When a fixed target came into view, or when an animal surfaced presenting a transient cue, the observer made a voice note of the estimated range in metres. Immediately following each video sequence of a surfacing, the observer captured a digital still image of one of the two calibration points on the ship's bow at the same level of magnification in order to measure the focal length of the lens. At the end of the experiment, the videotape was reviewed by linking the camcorder to a personal computer equipped with WinCoder software (InterVideo, Inc.). When the voice note indicated a surfacing where a simultaneous naked-eye estimate was made, a digital still image was captured and stored as a bitmap file. The program LENRAN was used to convert these bitmaps, given the camera height and magnification calibration, into range estimates (Leaper & Gordon, 2001).

Study 2 was conducted from a small (approximately 15 m length, 5 m eye height) boat during off-effort legs of a survey for boto (*Inia geoffrensis*) and tucuxi (*Sotalia fluviatilis*) in the Amazon River (S. Hedley, F. Marques, F. Trujillo and R. Williams, unpublished data). These data were different from the other three studies in being collected in restricted areas of water rather than the open sea. Thus distance-estimation and measurement methods requiring a horizon were not possible. Six forward-facing observers conducted the experiment from the main sightings platform on 1 April 2002. A data recorder chose fixed targets only to which observers judged range, because dolphins proved to offer insufficient target strength for range-finders. Each observer wrote down the trial number (22 trials), target position (i.e. ahead, port, starboard, abeam), and estimated distance, in order to retain observer independence. The data recorder measured true distance using Bushnell range-finders. Sea state was not recorded, because the river remained calm throughout the experiment.

Study 3 was conducted from a 53 m ship at the discretion of the cruise leader during off-effort legs of a dedicated sightings survey for killer whales (*Orcinus orca*: Zerbini et al., 2006) and balaenopterid whales in the Gulf of Alaska (Zerbini et al., in press). One observer (A.Z.) conducted the study, and the experiment was conducted from a small cargo ship named 'Coastal Pilot' (length 53 m and camera height 9.18 m). The observer scanned using 7×50 binoculars to which a digital camcorder was mounted. At first sight of a cetacean school, a voice note was made onto the videotape to record species and reticle reading. Next, a calibration shot of the ship's bow was taken. Reticle readings were converted to range using the same trigonometric relationships and refraction calculations as for photogrammetric measurements, based on a US Standard Atmosphere (Leaper & Gordon, 2001).

Study 4 was conducted during off-effort segments of a sightings survey from the 99 m survey vessel RRS 'James Clark Ross' in the Scotia Sea between 5 January and 10 February 2003. One observer (R.L.) collected the data. Reticle and photogrammetric measurements were made to cetaceans using methods described for Studies 1 and 3.

A dataset was compiled from all paired observations (those estimated by eye versus those calculated using photogrammetric methods) of both fixed and transient cues. These data were analysed in program *R* (R Development Core Team 2003) using a variety of linear- and non-linear modelling techniques (Krzanowski, 1998) to regress estimated range on measured range. The decision to include cue type (fixed versus transient) as a covariate was driven by a variation on Akaike's information criterion (AIC; Akaike, 1973), in which models carry a penalty for additional parameters. The variant used in this study adjusted the penalty to account for small sample size (AIC_c; Burnham & Anderson, 2003).

Model selection was guided by the desire for an explanatory model that best fitted the data, rather than a mechanistic model to explain plausibly how the errors were generated. The objective for distance sampling is a model that produces corrected radial distance estimates that are unbiased on average (Chen & Cowling, 2001). Model selection was aided by visual methods that assessed which model fit the data best across the entire range of observations (i.e. models were favoured when diagnostic plots of their residuals showed uniform spread across the x- and y-axes).

A three-stage model selection protocol was used. First, a least-squares regression was made of estimated range on measured range. If residuals indicated a relationship between range and variance, then the data were log-transformed and the model refitted. If residuals of the second model showed evidence of non-linearity, then a generalized additive model (GAM) was fitted using package *mgcv for R* (Wood, 2001), with variance proportional to range. Methods for GAM-fitting in mgcv use cross-validation to select automatically the optimum degree of smoothing and impose penalties for unwarranted smoothing. If the estimated degrees of freedom (edf) optimized by the smoothing process in mgcv for the measured range term was greater than one, then the model was selected. If the edf was near 1, then the log-transformed linear regression was selected.

In Study 1, the selected model was used to transform estimated radial distances from a survey (Williams et al., 2006). These and measured angles were used to calculate perpendicular distances, which were modelled in DISTANCE 4.0 Release 2 (Thomas et al., 2002). A half-normal model (Buckland et al., 2001) was fitted to both sets of perpendicular distances (i.e. uncorrected and corrected) to explore how the probability of sighting an animal decreased with increasing perpendicular distance from the trackline. This estimated the extent to which correcting an observer's ability to judge radial distance altered the estimates of strip half-widths effectively searched for whales.

In Study 2, a dataset was compiled from paired observations (visual estimates versus range-finder measurement) for each observer. The decision to include the observer as a covariate was aided by AICc. In Study 3, a dataset was compiled from all paired observations (those measured using the reticle binoculars versus photogrammetric methods) to transient cues for each of three cetacean species. Species was introduced as a candidate covariate, and the decision to include the term in the model was driven by AICc. Study 4, was treated as for Study 1, where photogrammetric measurements were treated as the independent variable.

RESULTS

Study 1

A total of 222 paired comparisons was made between estimates of distance by naked-eye and photogrammetric measurements to transient cues across a range of animal sizes. Of these, the majority (188) were observations of the target species of the sightings surveys (minke, humpback and fin whales). The remaining observations included small cetaceans (16), penguins (12) and fur seals (6) that were detected at closer distances. The experiment was conducted across a range of sea states from Beaufort 1–5, but with 92% of effort in Beaufort 2–4. An additional 59 paired comparisons were made to fixed cues. These included objects of known size (inflatable boats, fur seals or birds) that were continuously visible while resting at the surface of the water, and chunks of ice that were continuously visible but of unknown size.

The variance of estimated distance increased with true range showing a heteroscedastic relationship (Zar, 1996). This observed heteroscedasticity indicated that least-squares regression was not appropriate and that either a transformation or alternative model such as a GAM (Wood, 2001) was required.

Plotting log(estimated distance) against log(measured distance) revealed a linear relationship with constant variance (Figure 1). The data also lent support for stratifying by cue type, and when cue type was introduced as a covariate, ΔAIC=2.07. The intercept term was not significant at the (P=0.05) level. Thus the selected model was:

log(estimate)=1.0282×log(measured)

to transient cues, and

log(estimate)=1.0422×log(measured)
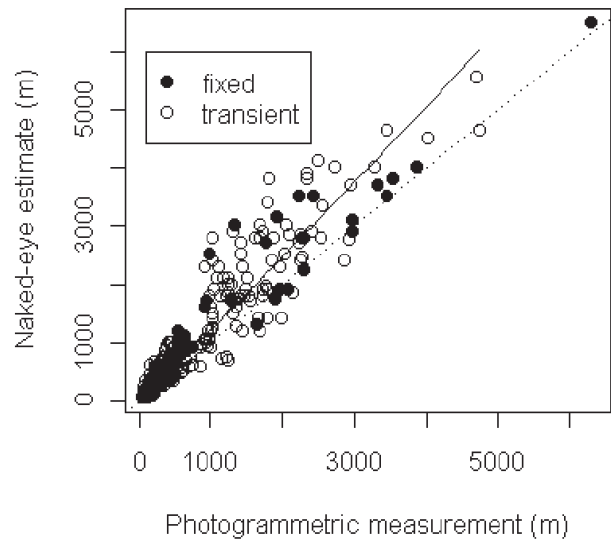
to fixed cues.

Figure 1. Scatterplot showing the relationship between photogrammetric measurements and naked eye estimates to fixed (closed circles) and transient (open circles) cues. A dotted line through the origin is shown with slope=1 to show an unbiased relationship. The solid line shows the fitted values of a least-squares regression applied only to transient cues at each observed value of measured range. The selected model, log(estimate)~log(measured), produced standardized residuals that were unbiased across the range of values observed.

The log transformation appeared to adequately address the problem of heteroscedasticity in both cases.

Comparing the regression coefficients for fixed and transient cues provided some evidence that the two slopes
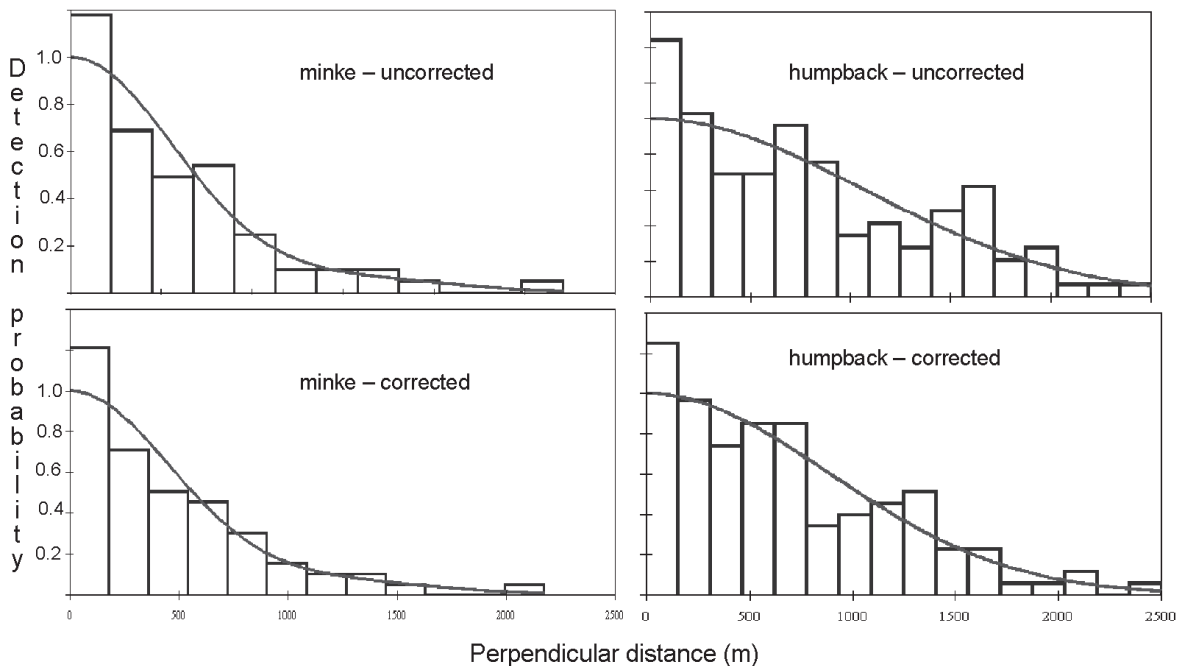

Figure 2. Detection functions for uncorrected (upper) and corrected (lower) radial distance estimates. The detection functions were based on 72 sightings of minke whales and on 121 sightings of humpback whales. The correction factor had a large effect on estimates of effective strip width (ESW) and school density (D_s in schools.km^{-2}) for minke and humpback whales. Correcting radial distances reduced ESW for minke whales from 799 (CV 0.10) to 658 m (CV 0.10), which increased the estimate of school density from 0.0046 (CV 0.26) to 0.0055 (CV 0.26). Correcting radial distances reduced ESW for humpback whales from 1314 (CV 0.07) to 1069 m (CV 0.07), which increased the estimate of school density from 0.0046 (CV 0.27) to 0.0057 (CV 0.27).
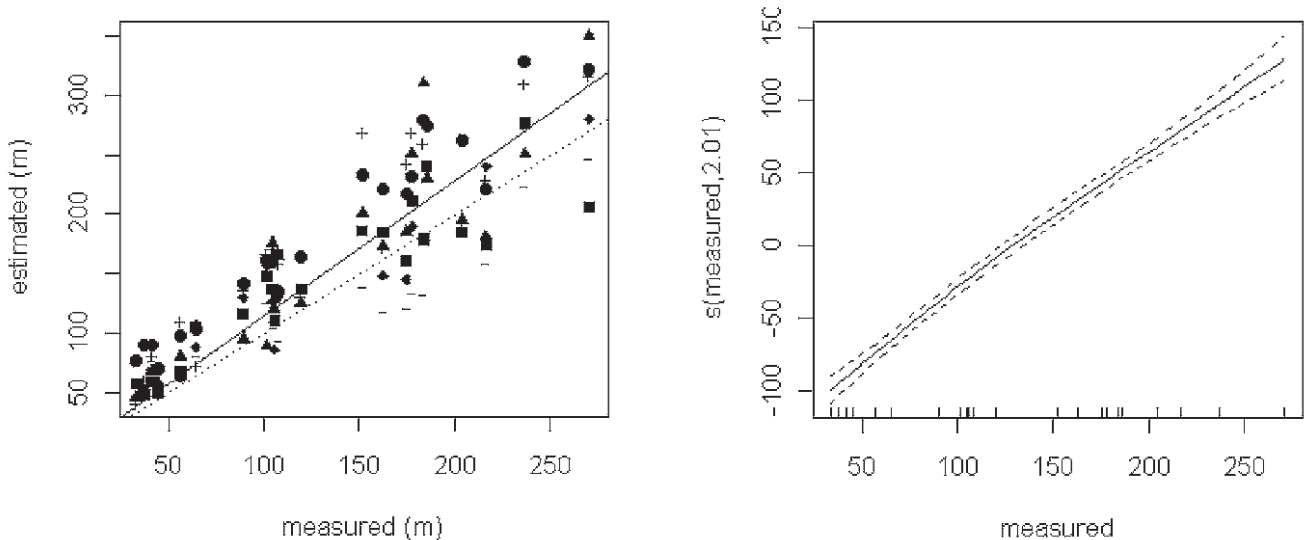
Figure 3. Scatterplot (left) showing heteroscedastic relationship between laser range-finder measurements and naked eye estimates to fixed cues. Plot characters represent values for each of six observers. A dotted line through the origin with slope=1 is shown to illustrate an unbiased relationship. The solid line shows the fitted values of a least-squares regression averaged among all observers. A GAM fitted to the data with variance proportional to the mean (right; observer effect held constant as a random factor) indicated that a smooth spline of measured distance with approximately 2 df described the estimated distances best.

differed from one another, however, this was not significant at the 5% level (t_{277}=1.752, 0.05<P<0.10). Range estimates to fixed cues were omitted from subsequent analyses because the primary aim of the study was to develop a suitable factor to correct range estimates to cetaceans.

Further efforts to explore non-linearity or other mean–variance relationships failed to produce a better-fitting model than this one. Similarly, the additional explanatory power gained by including information about sea state was insufficient to justify including additional variables in the model in terms of AIC.

The detection functions fitted to observed and corrected estimates of perpendicular distances to minke and humpback whales are shown in Figure 2. Estimated strip widths based on uncorrected radial distances were approximately 20% larger than those using corrected distances.

Study 2

Study 2 indicated that the six observers varied widely in the way that they estimated distance to 22 fixed objects. A linear regression gave the relationship estimated distance=1.137×measured. However, the data also showed evidence of heteroscedasticity but although model fit was improved by a log-transformation, evidence still remained to suggest a non-linear relationship in the transformed data. A GAM fitted to the data revealed that the best smoothing spline included approximately 2 df (Figure 3, right). Specifying a model with variance proportional to range provided a fit whose residuals were spread uniformly along the x- and y-axes, suggesting that this model would provide corrected radial distance estimates that were unbiased on average. The selected GAM modelled visual distance estimates as a smooth, non-linear function of measured distance, with observer as a covariate factor.
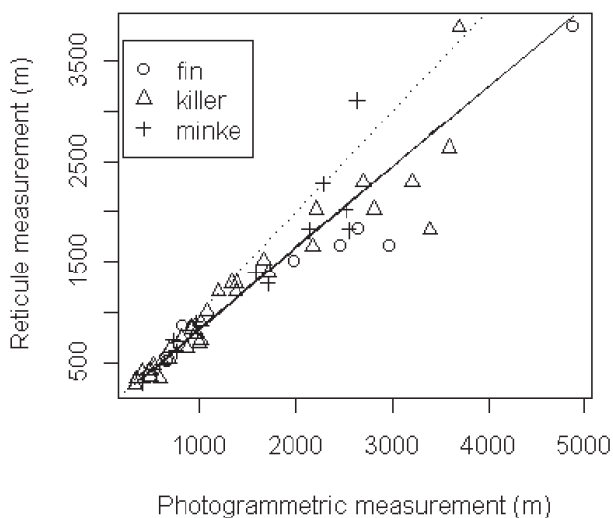

Figure 4. Scatterplot showing heteroscedastic relationship between reticule and photogrammetric measurements to whales. Plot characters represent values for each of three cetacean species. A dotted line through the origin with slope=1 is shown to illustrate an unbiased relationship. The solid line shows fitted values of a regression on the log-transformed data; log transformation had a minor effect on the slope, but improved the fit.
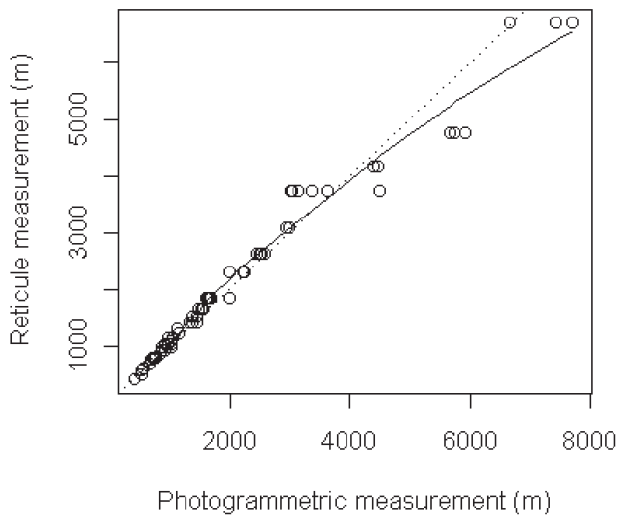
Study 3

A plot of distances estimated from reticule readings against video measurements for 54 paired observations (Figure 4) shows that the observer's use of reticule binoculars would have underestimated distance by approximately 20% (distance based on reticule measurement=0.802×photogrammetric measurement. However, the relationship was also heteroscedastic (as would be expected from a non-

Figure 5. Scatterplot showing heteroscedastic relationship between reticule and photogrammetric measurements to cetaceans. A dotted line through the origin with slope=1 is shown to illustrate an unbiased relationship. The solid line shows the fitted values of a GAM fitted to the data with variance proportional to the mean; a smooth spline of photogrammetric measurements with approximately 3 df described the reticule measurements best. This GAM specification produced residuals that were unbiased across the range of values observed in Study 4, suggesting that this calibration would yield corrected distances that were unbiased on average.

linear transformation of measured angles). Nevertheless, an attempt to fit a GAM to the data in mgcv resulted in a smoothing spline with one degree of freedom (i.e. a linear term). Log-transforming the data indicated a superior fit to the data than the least-squares regression although the slope only differed slightly. Including species as a covariate improved the model fit, but resulted in a higher AIC score, so the term was dropped.

Thus, the selected model was:

log(reticule range)=0.9747×log(photogrammetric range).

Study 4

A linear regression of 61 pairs of observations from reticule binoculars and photogrammetric measurements gave the relationship: distance based on reticule=0.95×photogrammetric measurement. However, heteroscedasticity and some evidence of non-linearity in the residuals ruled out a least-squares regression, so a smoothing spline was explored, specifying a model with variance proportional to range. A smoothing spline with approximately 3 df (Figure 5) was selected as the best fit by mgcv. The residuals from the model indicated that the model fit the data reasonably well across the range of observed values, suggesting that if this GAM were used to correct radial distances, then the data would be unbiased on average.

Synopsis

Bias was found to be a problem to varying degrees in all four experiments. Bias was positive for both studies using visual estimates and negative in both studies using reticule binoculars. All four studies found that the variance of distance estimates increased with range. Two of the four studies indicated that the relationship between estimated and true distance might be slightly non-linear. A synopsis of the key results from the four sets of experiments is given in Table 2.

DISCUSSION

Measurement error was found to be a problem in all four experiments, ranging from very minor (Study 4) to substantial (Studies 1 & 3). If ignored, these larger measurement errors would have biased distance estimates downward by as much as 20% (Study 3) or upward by as much as 25% (Study 1). These levels of bias emphasize the need for close attention to collection of distance data.

Although the use of reticule binoculars might be expected to give much more accurate estimates of distance than by relying on the naked eye, the results of Study 3 indicate that this is not necessarily always the case. Study 4 showed that reticules may give reasonable accuracy but sighting conditions were generally excellent during this study and much better than those of Study 3. The results of Studies 3 and 4 might be interpreted as evidence that observers use reticule binoculars idiosyncratically. If the observer records in Studies 3 and 4 represent typical use of reticule binoculars on sightings surveys, then their use would generally underestimate range and overestimate abundance. However, it is likely that not all observers will use reticule binoculars in the same way.

Although we have treated photogrammetric measurements as 'true' values because these allow measurement under controlled conditions, there is nevertheless scope for error in these measurements. Reticule and photogrammetric measurements are based on the same trigonometric relationships and therefore, should yield the same results. However, ship and whale movement make reticule binoculars more difficult to use than photogrammetric methods when measuring range to moving objects. By providing a permanent record in the form of a digital still image (in which

Table 2. *A summary of key findings from the four distance-estimation experiments.*

Study	Bias	Heteroscedasticity	Sample size	No. observers	Non-linearity
1	+25%	Yes	222 fixed, 58 transient cues	1	No
2	+14%	Yes	22 fixed cues	6	Yes
3	−20%	Yes	54 transient cues	1	No
4	−5%	Yes	61 transient cues	1	Yes

both the target and horizon are frozen), photogrammetric methods provide a more reliable measurement than reticle binoculars. Under typical survey conditions from a large oceanographic vessel (length 99 m, platform height 18 m), Leaper & Gordon (2001) demonstrated that photogrammetric measurements were unbiased to distances of 5 km, with a root mean square error of 3.5%.

It is interesting to note that the relationships between naked eye estimates and measured values in Study 1 were best described by a log-transformation. This relationship, termed 'compression' in the psychophysical literature, may be rooted in the manner in which humans generally perceive distance. When asked to judge the sensory tasks such as brightness and loudness, human subjects often perform according to Steven's law, an allometric relationship between perceived sensation and stimulus intensity in which the smallest difference that observers can detect fits a logarithmic scale better than a linear one (Stevens, 1970).

Information about how humans perceive distance may yield statistical models with better explanatory power than the descriptive models presented here. There is evidence to suggest that humans perceive range in much the same way as it is measured photogrammetrically and by reticle binoculars: by judging the angle below the horizon to the object of interest (Ooi et al., 2001). This may help to explain the heteroscedastic error structure observed in both visual (Studies 1 & 2) and reticle (Studies 3 & 4) experiments. Very small errors in judging declination near the horizon cause much larger errors in range estimates than similar errors do near the observer.

An additional point to emerge from this study was that the data from Study 1 suggested different parameters for fixed cues than for transient, cetacean cues. This finding calls into question the common practice of using fixed cues as cetacean proxies in distance-estimation experiments (e.g., Study 2), and warrants further attention. Distance-estimation experiments using fixed cues may provide the wrong correction factor because they provide observers with a longer opportunity to judge range than observers receive from transient, whale cues. Future research should identify whether this apparent difference might represent systematic bias in fixed-cue distance-estimation experiments that is large enough to matter in most surveys. Wu et al. (2004) present the results of distance-estimation experiments in terrestrial situations which suggest a possible mechanism due to visual scanning patterns that would result in less accurate estimation of distances to transient cues. Their results showed greater accuracy if the observer was allowed to scan from near to far, than from far to near. In the case of a transient cue, the observer is forced to scan from far to near.

Detection function fitting is especially sensitive to observations very near the trackline, and some models are strongly influenced by observations in the tail of the distribution of perpendicular distances (Buckland et al., 2001). While distance-sampling theory appears to be robust to random errors in distance estimates (Chen & Cowling, 2001), fitting the detection function may not be robust to a source of bias that causes different measurement errors at different ranges. While truncation of distant sightings is recommended to prevent distant observations from unduly

influencing the fit of the detection function in conventional distance sampling (Buckland et al., 2001), model-based abundance estimation methods (Hedley et al., 1999) benefit from including as many sightings as possible. A correction factor that fits the data well at small and large ranges may improve detection function fit and require less truncation. The extent to which bias in radial distance affects bias in estimated effective strip half-width will also be influenced by the distribution and accuracy in angle measurements. However, this paper only addresses distance-estimation.

This study serves as an example of the value of so-called ships of opportunity not only for collecting cetacean sightings data, but also as inexpensive platforms for conducting experiments that inform researchers about the sighting process itself. Distance-estimation experiments do not require randomized sampling designs. Ships of opportunity could be used for collecting data to assist existing error modelling efforts, thereby providing analysts with the opportunity to explore error in perpendicular distance measurements to cetaceans rather than potentially unsuitable proxies such as golf tees (Marques & Buckland, 2004). In addition, there is a role for these ships as cost-effective platforms for training observers and fine-tuning protocols prior to systematic sightings surveys. Although no substitute for attempting to measure all distances to sightings, even measuring a small proportion is of great value for estimating correction factors and likely to be more representative than distance experiments using static targets.

However, the importance of accurate measurements, rather than those that are merely unbiased on average, must be stressed. Indeed, for some surveys, a correction factor may be simply not good enough. The best that one could do with a correction factor is to calculate ranges that are unbiased on average. Random error that remains after applying a correction factor may still cause error in detection function fitting that leads to bias in abundance estimates (Borchers et al., 2003; Marques & Buckland 2004). Therefore, accurate distances will improve detection function fit, thereby reducing the variance of abundance estimates, which can be as important as bias when detecting population trends (Taylor & Gerrodette, 1993). Where possible, it is always preferable to measure range accurately in the field than to develop analytical techniques for correcting data at a later stage.

Further work is required to simulate how much bias is associated with the random errors remaining after applying a correction factor to remove bias in range estimates to cetaceans, rather than fixed cues. Additional work should explore how random error might influence detection function fitting in surveys with small sample size, or surveys that use advanced distance sampling techniques, such as duplicate sightings surveys to estimate g(0), the probability of sighting an animal on the trackline, and multiple covariate distance sampling (MCDS), where detection may vary with factors in addition to perpendicular distance (Buckland et al., 2001).

REFERENCES

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In *2nd international symposium on information theory* (ed. B.N. Petrov and F. Csaki), pp. 267–281. Budapest, Hungary: AkadŠemiai Kiadi.

Alpízar-Jara, R., Stefanski, L.A., Pollock, K.H. & Laake, J.L., 2001. An additive error model in line transect sampling using multiple observers. *Proceedings VII Congreso Annual da Sociedade Portuguesa de Estatística*, 539–555.

Baird, R.W. & Burkhart, S.M., 2000. Bias and variability in distance estimation on the water: implications for the management of whale watching. *Paper SC/52/WW1 presented to IWC Scientific Committee in Adelaide, Australia*. Available from www.iwcoffice.org

Borchers, D.L., Pike, D., Gunnlaugsson, T. & Víkingsson, G.A., 2003. Analyses of the NASS 1987 and 2001 minke whale cue counting surveys taking account of distance estimation errors. *Paper SC/55/NAM3 presented to the IWC Scientific Committee*. Available from www.iwcoffice.org

Buckland, S., Anderson, D.R., Burnham, K.P., Laake, J., Borchers, D. & Thomas, L., 2001. *Introduction to distance sampling: estimating abundance of biological populations*. Oxford: Oxford University Press.

Burnham, K.P. & Anderson, D.R., 2003. *Model selection and multimodel inference: a practical information-theoretic approach* 2nd ed. New York: Springer-Verlag.

Chen, S.X., 1998. Measurement errors in line transect surveys. *Biometrics*, **54**, 899–908.

Chen, S.X. & Cowling, A., 2001. Measurement errors in line transect surveys where detectability varies with distance and size. *Biometrics*, **57**, 732–742.

Hammond, P.S. et al., 2002. Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology*, **39**, 361–376.

Hedley, S.L., Buckland, S.T. & Borchers, D.L., 1999. Spatial modelling from line transect data. *Journal of Cetacean Research and Management*, **1**, 255–264.

Krzanowski, W.J., 1998. *An introduction to statistical modelling*. London: Arnold Publishers.

Leaper, R. & Gordon, J., 2001. Application of photogrammetric methods for locating and tracking cetacean movements at sea. *Journal of Cetacean Research and Management*, **3**, 131–141.

Lerczak, J.A. & Hobbs, R.C., 1998. Calculating sighting distances from angular readings during shipboard, aerial, and shore-based marine mammal surveys. *Marine Mammal Science*, **14**, 590–598.

Marques, T.A. & Buckland, S.T., 2004. Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models. *Biometrics*, **59**, 924–935.

Ooi, T.L., Wu, B. & He, Z.J., 2001. Distance determined by the angular declination below the horizon. *Nature, London*, **414**, 197–200.

Palka, D.L. & Hammond, P.S., 2001. Accounting for responsive movement in line transect estimates of abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, **58**, 777–787.

Schweder, T., 1997. Measurement error models for the Norwegian minke whale survey in 1995. *Reports of the International Whaling Commission*, **47**, 485–488.

Stevens, S.S., 1970. Neural events and the psychophysical law. *Science, New York*, **170**, 1043–1050.

Taylor, B.L. & Gerrodette, T., 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conservation Biology*, **7**, 489–500.

Thomas, L. et al., 2002. *Distance 4.0. Release 2. Research Unit for Wildlife Population Assessment, University of St Andrews, UK*. http://www.ruwpa.st-and.ac.uk/distance/

Thompson, D. & Hiby, A.R., 1985. The use of scale binoculars for distance estimation and a time-lapse camera for angle estimation during the 1983/84 IDCR minke whale assessment cruise. *Reports of the International Whaling Commission*, **35**, 309–314.

Williams, R., Hedley, S. & Hammond, P.S., 2006. Modelling distribution and abundance of Antarctic baleen whales using ships of opportunity. *Ecology and Society*, **11**, 1. Available from http://ecologyandsociety.org/vol11/iss1/art1/

Wood, S.N., 2001. mgcv: GAMs and Generalized Ridge Regression for R. *R News*, **1**, 20–25.

Wu, B., Ooi, T.L. & He, Z.J., 2004. Perceiving distance accurately by a directional process of integrating ground information. *Nature, London*, **428**, 73–77.

Zar, J.H., 1996. *Biostatistical analysis*. New Jersey: Prentice Hall.

Zerbini, A.N., Waite, J.M., Durban, J.W., LeDuc, R., Dahlheim, M.E. & Wade, P.R., 2006. Estimating abundance of killer whales in the nearshore waters of the Gulf of Alaska and the Aleutian Islands using line transect sampling. *Marine Biology*. DOI 10.1007/s00227-006-0347-8

Zerbini, A.N., Waite, J.M., Laake, J.L. & Wade, P.R., in press. Abundance, trends and distribution of large baleen whales in western Alaska and the Central Aleutian Islands. *Deep Sea Research Part I*.