Comparing two methods of shore-based counts of eastern North Pacific gray whales

DURBAN, J.[1], LANG, A.[1], WELLER, D.[1], RUGH, D.[2], HOBBS, R.[2], PERRYMAN, W.[1]

Contact email: john.durban@noaa.gov

ABSTRACT

Counts of southbound migrating whales at Granite Canyon, California, form the basis of abundance estimation for the eastern North Pacific stock of gray whales (*Eschrichtius robustus*). In 23 years, between 1967 and 2007, counts of the number of observed pods have been rescaled by a series of correction factors to provide abundance estimates. The "traditional" counting approach involved single observers independently searching for whales and hand-recording entries onto a data form. However, a new observation approach has now been adopted wherein a paired team of observers work together, using a computer to log data and map whale sightings. We evaluated the performance of the traditional and new counting approaches by comparing the pooled number of pods, whales and pod size distributions during simultaneous and independent trials conducted during both the 2006/2007 and 2007/2008 southbound migrations. In general, the number of pods counted showed a high degree of similarity between stations, with a coefficient of variation of only 7 and 8% in each of the two years. However, there was a tendency for the new paired-observer teams to count fewer pods but estimate relatively higher numbers of whales, reflecting differences in the distribution of estimated pod sizes. The single-observer stations generally counted more pods of size one, with the paired team recording a higher proportion of larger-sized pods. This difference was particularly apparent for a single-observer station staffed by observers with extensive counting experience, which recorded a pod size distribution that was significantly different to that of the paired-observers (p = 0.95 and p = 0.70 for the 2006/2007 and 2007/2009, respectively), suggesting that a different counting process underlay the collection of these data. These differences likely represented the tendency for the paired-observers to lump rather than split whales into recorded pods because the tracking software facilitated the repeated relocation of whales in close proximity to each other. However, because paired-observers typically counted more total whales from fewer sightings, there may also have been a differential pod size estimation bias, and we suggest that teamwork and computer-assisted tracking may have reduced underestimation of pod sizes by enabling the paired team to observe a greater number of re-sightings of individual pods. Notably, a second single-observer station, staffed by less-experienced observers, recorded an average pod size and pod size distribution that was more similar to the paired-observers than to the more experienced single-observers. This highlights the need for new calibration data to evaluate the different pod size estimation biases of new counting methods and new observers before count data can be reliably rescaled to estimate abundance.

KEYWORDS: Whale counts, Survey shore-based, migration, monitoring, pod size

INTRODUCTION

Abundance estimates for the eastern North Pacific gray whale (*Eschrichtius robustus*) have been made for 23 years, between 1967 and 2007, derived from data from shore-based counts of the southbound migration past Yankee Point or Granite Canyon, near Monterey, California (Reilly et al., 1980; Reilly, 1984; Laake et al., 1994; Buckland and Breiwick, 2002; Buckland et al., 1993; Hobbs et al., 2004; Rugh et al., 2005; Rugh et al., 2008, Laake et al., 2009). Gray whales pass closer to shore in the Granite Canyon area than along much of their migratory route, enabling shore-based observers to see across most of the migratory corridor (Shelden and Laake, 2002).

The "traditional" counting approach has involved single observers independently searching for whales and hand-recording sightings onto a data form (Rugh et al., 2008). To produce abundance estimates, the observed number of pods has been multiplied by correction factors for pods missed during watch periods, pods passing outside watch periods, night travel rate and bias in pod size estimation (Laake et al., 2009). Pods missed during watch periods have been estimated using a mark-recapture approach by matching sightings between two independent stations of single observers in each estimation year. Pods passing outside of watch periods have been estimated by extrapolating from an estimated distribution for the rate of whales passing the counting site through each estimation period. However, the remaining correction factors have not been estimated in each survey. Corrections for night travel rates were

[1] *National Marine Fisheries Service, Southwest Fisheries Science Center, 8605 La Jolla Shores Dr., La Jolla, CA 92037, U.S.A.*
[2] *National Marine Fisheries Service, Alaska Fisheries Science Center, 7600 Sand Point Way NE, Seattle, WA 98115, U.S.A.*

established first using radio tracking of tagged gray whales (Swartz et al. 1987) and then through thermal imagery conducted at the Granite Canyon site (Perryman *et al*. 1999); pod size bias corrections were calculated when aircraft (Laake *et al*., 1994), video imagery (Perryman *et al*. 1999) or tracking teams (Rugh *et al*., 2008) were available. These correction factors are therefore assumed constant - an assumption that needs testing to ensure validation, particularly with changes in the counting methodology and with new, uncalibrated observers.

Due to concern over the ability of a single observer to simultaneously track and accurately record multiple pods of whales, a new observation approach was adopted in 2007, involving a paired team of observers working together using a computer data-logging and tracking program. The rationale behind this change was that a team of observers working together would be better able to track pods of whales, especially with the help of real-time data logging and tracking software. The advantages of the new system are several fold: 1) communication between observers reduces biases involved in any observer working alone; 2) collaborative tracking of pods allows for more repeated observations of each whale pod, enabling pod size estimates to be re-assessed and refined; 3)tracking software provides a visualization that allows distinct pods of whales to be more easily distinguished; and 4) having a dedicated data recorder frees primary observers from needing to look away from the search area to record data.

In this document, we have evaluated the performance of the traditional and new counting approaches during simultaneous and independent trials conducted during the 2006/2007 and 2007/2008 southbound migrations. We based our comparison on the pooled number of pods, whales and pod size distributions in watch periods with simultaneous effort. This avoids the inherent biases and assumptions required to match sightings of individual pods of whales. Notably, we adopted a novel Bayesian hierarchical modeling approach to compare the observed distributions of pod sizes, to make inference about the similarity of the underlying observation processes.

METHODS

Data Samples

Counts of gray whales were conducted from three different watch stations at Granite Canyon during the 2006 /2007 southbound migration and from two different watch stations during the 2007/2008 migration. In January 2007, counts from the new paired-observer station were assessed from 26 three-hour watch periods with simultaneous effort by two single-observer watch stations (N = North and S = South), each counting independently of the other stations. In January 2008, counts from the paired-observer team were compared to counts from a single-observer station during 16 three-hour periods (Table 1).

Table 1: Summary of the independent counts made by each single-observer watch stations and the paired-observer team during the 2006/2007 and 2007/2008 southbound migrations. Presented are the dates during which the stations were operating simultaneously, the number of watch periods (over # days) when all stations were operating in good weather conditions (visibility code <5), the total number of pods counted, the total number of whales calculated from estimated pod sizes and the mean (standard deviation) pod size. A paired-observer station was compared to two single-observer stations (N = North and S = South) in 2006/2007 and to just one single-observer station in 2007/2008.

	2006/2007 migration			2007/2008 migration	
Dates	6th Jan – 27th Jan, 2007			7th Jan – 18th Jan, 2008	
Watches (days)	26 (18)			16 (9)	
Station	Single_S	Single_N	Paired	Single	Paired
# Pods / Whales	482 / 1043	503 / 907	439 / 999	220 / 338	197 / 387
Mean Pod (sd)	2.2 (1.8)	1.8 (1.2)	2.3 (1.6)	1.5 (1.0)	2.0 (1.3)

The single-observer stations operated as in previous years. Up to three 3-hour watch shifts were used to cover daylight hours from 0730 to 1630, with a rotating observer team using 7X50 binoculars to detect passing pods, record pod size estimates and note environmental conditions, specifically visibility (subjectively categorized from 1 to 6 for excellent to useless) and sea state (Beaufort scale). Magnetic compasses in the binoculars provided horizontal bearings, and reticle marks in the binoculars provided vertical angles relative to the horizon. Observation methods were described in Rugh et al. (2008). Notably, each single-observer operated independently and hand-recorded entries onto a data form. Observers tried to keep track of each pod travelling through the viewing area in order to record a “north sighting” when the whale was first seen and a “south sighting” as close to the standard azimuth (a line perpendicular to the coastline at 241^0 magnetic that intersects the survey site) as possible. This was achieved by using a table based on average swimming speeds to predict the time and vertical angle where a pod might be expected to cross the azimuth. The single-observer station in 2007 / 2008 was staffed by observers with extensive experience of the counting process. In 2006 / 2007, these observers staffed one of the single stations (N station), and a set of observers with less experience staffed a second station (S station) that operated independently with a near-identical field of view, altitude and observation process.

A notably different approach was employed by the new paired-observer station, which used a rotating team of observer pairs. From a similar field of view and altitude to the single-observer stations, one observer in the pair kept continual visual watch using naked eye aided by 7X50 binoculars, while the second served as a data recorder, but also watched with naked eye and binoculars whenever possible. Sightings were entered into a real-time data logging PC program, which had a mapping screen to help track repeated sightings of the same pod. The map projected the movement patterns of the pods using predicted swimming speeds (same speeds as used by single-observers), allowing re-sightings and new sightings to be queried. A key feature of the approach was that the primary observer need not break visual watch to record sightings or re-sightings, but rather just communicate with the data recorder.

For comparison between stations, each sighting was assigned to the three-hour effort period into which it fell as a function of the calculated time that it crossed the standard azimuth. The calculation was performed by a Visual Basic program (see Rugh et al., 2008) which converted the recorded sighting time and location closest to the standard azimuth to a time and distance offshore at which each pod crossed this line. Whale sightings were eliminated from the analysis if they crossed this line prior to the start of an effort period or if they had not crossed the line by the end of an effort period. To control for weather conditions, we eliminated any of the matched watch periods where one or more of the stations estimated visibility to be unacceptable (visibility code >5) at any time.

Comparing pod size distributions

We developed a hierarchical model to formally compare the distributions of pod size estimates between stations. Specifically, the estimated number of pods n_{jk} counted by each watch station j for each pod size class k = 1,…,10+ was modeled as a multinomial choice out of total of N_j pod observations by each station. The multinomial choice probabilities, p_{jk}, therefore represented the proportions of pods estimated by each station to be in each size class. Heterogeneity between watch stations was investigated using a Bayesian mixture model (West, 1992), with each $p_{jk} = p_{(z_j=c)k}$ dependent on an unknown component mixture c = 1…, C. Instead of a separate probability for each watch station for each pod size, the model therefore specified a separate probability of each pod size for each mixture (i.e. cluster) of watch stations, if they existed. Cluster identities were based on estimates of stochastic indicator variables z_j, one for each watch station, each indicating which component from a ceiling of mixtures (as many potential mixtures as watch stations) had generated the probability distribution for each station (Neal, 2000). This mixture formulation therefore allowed the identification of clusters representing watch stations with similar probability distributions for the estimated pod sizes.

To complete specification of this mixture model, we adopted a random effects model for the logit transforms of the choice probabilities $\pi_{ck} = \log\{p_{ck}/(1-p_{ck})\}$, assuming them to be distributed according to a multivariate Normal distribution, with mean μ_k and covariance matrix Σ. The hierarchical distribution for π was thus stratified into 11 dimensions, to represent the number of size classes, and we adopted a Normal hyperprior on μ_k and a Wishart hyperprior on Σ^{-1} with T=11 degrees of freedom. The covariance matrix Σ was of the order TxT, where the principal (left to right) diagonal element was the variance across the cluster effects π_{ck} for each of the k in 1,..,T size classes, and the off-diagonal values represented covariance between pairs of size classes. The Wishart prior distribution for the inverse covariance matrix Σ^{-1} (Fienberg et al. 1999) was specified in terms of a scale matrix B and a degrees of freedom parameter v. We set diagonal values of B = 1 for the prior variance of effects for each size class k, and the off-diagonals were assigned B = 0 for a prior assumption of no covariance between pairs of size classes. Larger values of v represent stronger belief, and we therefore adopted a value of $v = T = 11$ to represent a vague prior and allow non-zero covariance values to emerge. To allow differences between the means to emerge, the μ_k hyper-parameters were drawn from a Normal prior with a mean of zero and large prior variance of 100.

We used the Gibbs sampling Markov Chain Monte Carlo (MCMC) method (Smith and Roberts, 1993) implemented using the WinBUGS software (Lunn et al. 2000) to update these prior distributions conditional on the data and sample 50,000 values from the posterior distributions. This Bayesian sampling approach provided an intuitive framework for the analysis and communication of uncertainty. In this case, we directly compared the full distribution of each cluster indicator z_j. Similarities between pod size distributions were evaluated from the proportion of MCMC iterations, where the sampled values of $z_1 = z_2$ for the two station trial in 2007/2008, and $z_1 = z_2$, $z_1 = z_3$ or $z_2 = z_3$ for the three-station trial in 2006/2007. With repeated samples from each posterior distribution, these proportions represented estimates of the probability that the pod size distributions from pairs of watch stations were the same.

RESULTS

There was a high degree of similarity in the number of pods counted and the number of whales estimated from these counts during simultaneous counting trials in both 2006/2007 and 2007/2008 (Table 1). The coefficient of variation (CV = standard deviation / mean) for the pod counts was only 0.07 and 0.08 for 2006/007 and 20007/2008, respectively. The CV's for the counts of whales, after rescaling for estimated pod sizes, were slightly larger at 0.7 and 0.1.

In both years, the number of pods counted by the paired-observer team was lower than the single-observer stations. The highest number of pods in both years was recorded by a single-observer station staffed by the most experienced observers ("Single N" in 2006/2007 and "Single" in 2007/2008). Notably, the paired-observers recorded relatively high numbers of whales, despite counting the fewest number of pods, with higher whale counts than the experienced single observers in both years. Counts of the numbers of pods recorded by the less experienced single observers ("Single_S") were relatively high compared to the paired observers, and more similar to the experienced single observers, but counts of the number of whales were more similar to the higher estimates of the paired observers.

The different inference resulting from comparing the number of pods and the estimated number of whales counted by each watch station suggested key differences in the pod size estimates made by the different watch stations. The average estimated pod size was highest from the paired-observer station and lowest from the experienced single observers. These differences were driven by notable differences in the distribution of estimated pod sizes from each station (Figure 1). In particular, the single observers had a higher proportion of pods of estimated size = 1 compared to the paired-observer station. This difference is most noticeable for the experienced single observers. In contrast, the paired-observer station recorded a higher number of pods of larger estimated sizes.

Inference from the cluster indicators highlighted a clear difference between the pod size distributions of the paired and single-observer stations. There was a high probability ($p = 0.95$) that the pod size distributions from the paired and experienced single-observer stations were different in the 2006/2007 comparison, along with strong evidence of a difference ($p = 0.70$) in 2007/2008, despite the smaller sample of simultaneous count data. This indicated that a different process likely underlay collection of these data sets. However, the distinction between pod size distributions from the paired-observer and less experienced single-observer stations in 2006/2007 was not as great, with a probability of only 0.37 that they were described by different models. With the exception of the relatively high proportion of pods of size one recorded by the less-experienced single observers, these distributions were very similar. It is notable that there was actually less similarity between the pod size distributions from the two single-observer stations operating in 2006/2007, with a probability of 0.66 that they were described by different models.

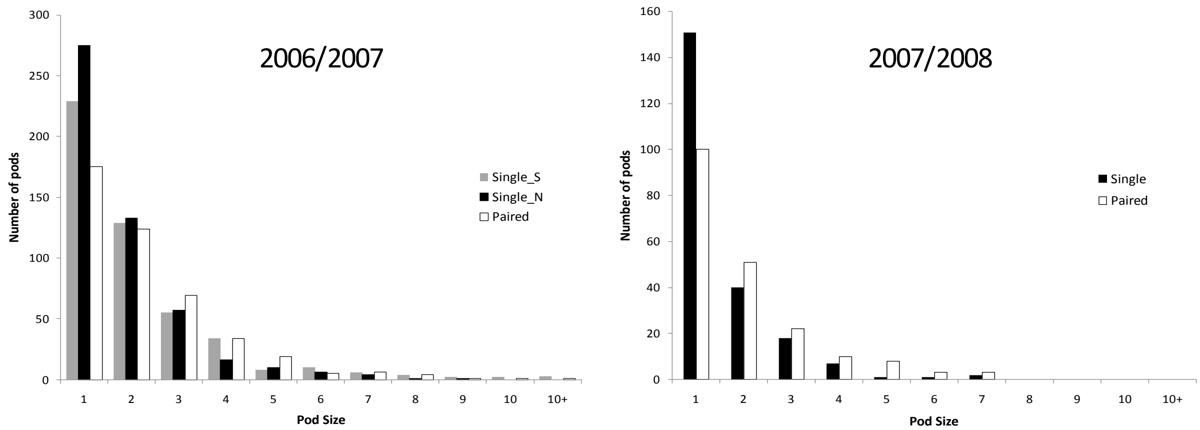

Figure 1: Frequency distributions of estimated pod sizes made by each of the three watch stations over 26 three-hour watch periods in during the 2006/2007 migration and 16 three-hour watch periods in 2007/2008. A paired-observer station was compared with two single-observer stations (N = North and S = South) in 2006/2007, and with just one single-observer station in 2007/2008.

DISCUSSION

Precise and unbiased abundance estimates of the eastern North Pacific gray whale population require shore-based counts to be as complete as possible, with minimal but quantified observation biases. To this aim, our comparison attempted to assess the efficacy of a new counting approach that has replaced a single-observer method in which the observer hand-recorded counts onto a data form. The new approach incorporates efforts from a paired team of observers who work together and enter data on a computer that maps sighting records, providing a visualization of where whales are. Simultaneous trials over two migrations (2006/2007 and 2007/2008) demonstrated a high degree of similarity between the traditional and new counting approaches, but there was a tendency for the new paired-observer teams to count fewer pods but estimate relatively high numbers of whales. Therefore, differences are not likely to be due to variable detection probabilities of the two systems but rather reflect differences in the distribution of estimated pod sizes.

The single-observer stations generally counted more pods of size one, and paired teams recorded a higher proportion of larger-sized pods. Notably, there was a significant difference between the pod size distributions estimated by the paired-observer team and those estimated by the single-observer station staffed by observers with extensive counting experience, suggesting that a different counting method underlay the collection of these data. This disparity may have represented a tendency for the paired observers to lump rather than split whales into recorded pods, likely because the tracking software facilitated the repeated relocation of whales in close proximity to each other. However, because the paired observers typically estimated more total whales from fewer sightings, and because the CV of whale counts was greater than the CV of pod counts, there may also have been a differential pod size estimation bias which introduced additional variability in estimated numbers of whales. We suggest that teamwork and computer-assisted tracking may have reduced underestimation of pod sizes by enabling the paired team to observe a greater number of re-sightings of individual pods. This would be a beneficial effect of the new counting approach, but this needs to be tested in new trails of pod size calibrations.

Notably, a second single-observer station, staffed by less-experienced observers, recorded an average pod size and pod size distribution that was more similar to the paired observers than to the more experienced single observers. This puts into question the utility of matching sightings between independent single-observer stations to estimate detection probability, particularly when pod size is used as a covariate for detectability, and highlights the need to assess and allow for individual observer effects (Laake et al. 2009). Inexperienced single observers produced the highest pod-counts in the three-station trial and differed from the paired observers by recording a relatively high proportion of pods of size one. We hypothesize that this represented over-counting of small pods by less experienced single observers, with false positives emerging due to the difficulties of hand-recording and tracking repeat sightings of multiple single whales.

The new tracking and visualization software may help to limit false positives, and working as part of a paired team is intended to buffer individual effects, such as inexperience. Data collected in a recent comparative trial of two independent paired-observer teams in 2009/2010 will allow us to assess the robustness of the new counting approach to individual effects. We plan to extend these trials in 2010/2011 by conducting a calibration of pod size estimates made using the new counting method and by new observers: a key task before count data can be reliably rescaled to estimate abundance.

REFERENCES

Buckland, S.T. and Breiwick, J.M. 2002. Estimated trends in abundance of eastern Pacific gray whales from shore counts, 1967/68 to 1995/96. *J. Cetacean Res. Manage*. 4(1):41-48.

Buckland, S.T., Breiwick, J.M., Cattanach, K.L. and Laake, J.L. 1993. Estimated population size of the California gray whale. *Mar. Mammal Sci.* 9(3):235-249.

Dey D., Muller P. and Sinha D. 1999. *Practical nonparametric and semiparametric Bayesian statistics*. Springer, New York

Escobar M.D. and West M. 1995. Bayesian density estimation and inference using Mixtures. *J. Am. Stat. Assoc.* 90: 577-588

Hobbs, R.C., Rugh, D.J., Waite, J.M., Breiwick, J.M. and DeMaster, D.P. 2004. Abundance of North Pacific gray whales on the 1995/96 southbound migration. *J. Cetacean Res. Manage*. 6(2):115-120.

Laake, J.L., Rugh, D.J., Lerczak, J.A. and Buckland, S.T. 1994. Preliminary estimates of population size of gray whales from the 1992/93 and 1993/94 shore-based surveys. Paper SC/46/AS7 presented to the IWC Scientific Committee, May 1994 (unpublished). 16 p.

Laake, J. L.,Punt, A., Hobbs. R.,, Ferguson, M., Rugh, D. and Breiwick, J. 2009. Re-analysis of gray whale southbound migration surveys, 1967-2006. NOAA Technical Memorandum NMFS-AFSC-203. 55 p.

Lunn, D.J, Thomas, A., Best, N. and Spieglehalter, D. 2000. WinBUGS – A Bayesian modelling framework: concepts, structure and extensibility. *Statist Comput.* 10:325-337.

Neal R.M. 2000. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *J. Comput. Graph. Stat.* 9: 249-26.

Perryman, W.L., Donahue, M.A., Laake, J.L. and Martin, T.E. 1999. Diel variation in migration rates of eastern Pacific fray whales measured with thermal imaging sensors. *Mar. Mammal Sci.* 15(2):426-445.

Reilly, S.B., Rice, D.W. and Wolman, A.A. 1980. Preliminary population estimate for the California gray whale based upon Monterey shore censuses, 1967-68 to 1978-79. *Rep. Int. Whal. Commn*. 30:359-368.

Reilly, S.B., Rice, D.W. and Wolman, A.A. 1983. Population assessment of the gray whales, Eschrichtius robustus, from California shore censuses, 1967-80. *Fish. Bull., US* 8(2):267-281.

Rugh, D.J., Hobbs, R.C., Lerczak, J.A. and Breiwick, J.M. 2005. Estimates of abundance of the eastern North Pacific stock of gray whales (Eschrichtius robustus) 1997-2002. *J. Cetacean Res. Manage*. 7(1):1-12.

Rugh, D., Breiwick, J., Muto, M., Hobbs, R., Shelden, K., D'Vincent, C., Laursen, I.M., Reif, S., Maher, S. and Nilson, S. 2008. Report of the 2006-2007 census of the eastern North Pacific stock of gray whales. AFSC Processed Rep. 2008-03. 157 p., Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way NE, Seattle WA 98115.

Rugh, D.J., Muto, M.M., Hobbs, R.C. and Lerczak, J.A. 2008. An assessment of shore-based counts of gray whales. *Mar. Mammal Sci.* 24(4):864-880.

Shelden, K.E.W. and Laake, J.L. 2002. Comparison of the offshore distribution of southbound migrating gray whales from aerial survey data collected off Granite Canyon, California, 1979-96. *J. Cetacean Res. Manage*. 4(1):53-56.

Smith, A.F.M and Roberts, G.O. 1993 Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *J. Roy. Stat. Soc. B* 55: 3-23.

Swartz, S.L., Jones, M.L., Goodyear, J., Withrow, D.E. and Miller, R.V. 1987. Radio-telemetric studies of gray whale migration along the California coast: a preliminary comparison of day and night migration rates. *J. Cetacean Res. Manage*. 37:295-9.

West, M. 1992. Mixture models, Monte Carlo, Bayesian updating and dynamic models. *J. Stat. Plan. Infer.* 62: 325-333