Review of potential methodologies for estimating abundance of dolphin stocks in the Eastern Tropical Pacific

Stephen T. Buckland[1], Cleridy E. Lennert-Cody[2], Tim Gerrodette[3], Jay Barlow[3], Jeffrey E. Moore[3], Andrew Webb[4], Peter T. Fretwell[5], Hans J. Skaug[6], Wayne L. Perryman[3]

[1]University of St Andrews
[2]Inter-American Tropical Tuna Commission
[3]Southwest Fisheries Science Center
[4]HiDef
[5]British Antarctic Survey
[6]University of Bergen

Workshop on Methods for Monitoring the Status of Eastern Tropical Pacific Ocean Dolphin Populations
18-20 October 2016, La Jolla, California

DRAFT

Review of potential methodologies for estimating abundance of dolphin stocks in the Eastern Tropical Pacific

Stephen T. Buckland[1], Cleridy E. Lennert-Cody[2], Tim Gerrodette[3], Jay Barlow[3], Jeffrey E. Moore[3], Andrew Webb[4], Peter T. Fretwell[5], Hans J. Skaug[6], Wayne L. Perryman[3]

[1]University of St Andrews
[2]Inter-American Tropical Tuna Commission
[3]Southwest Fisheries Science Center
[4]HiDef
[5]British Antarctic Survey
[6]University of Bergen

Introduction

In this review, we consider methods for estimating animal abundance, with a focus on both contemporary and potential methods suitable for surveys of dolphin species that typically occur in large schools over extensive areas of ocean. Of particular interest are methods for use in the eastern tropical Pacific Ocean, primarily targeting stocks of the offshore pantropical spotted dolphin (*Stenella attenuata*), the spinner dolphin (*S. longirostris*), and the common dolphin (*Delphinus delphis*).

The best absolute abundance estimates of the stocks are considered to be the estimates from the series of research vessel surveys conducted by NMFS (Gerrodette *et al*., 2008). To assess whether stocks are recovering from the effects of past catches in the tuna fisheries, we need estimates of trend in abundance. There have been three main approaches to estimating trends. The first is to model tuna vessel observer data to estimate trends from the annual abundance indices (Lennert-Cody *et al*., 2016), the second adopts a similar approach using the time series of research vessel estimates (Gerrodette and Forcada, 2005; Gerrodette *et al*., 2008), while the third combines research vessel abundance estimates with estimates of dolphin mortality in the tuna fisheries through a population dynamics model (Hoyle and Maunder, 2004; Wade *et al*., 2007).

In this review we focus on methodologies for fishery-independent data sources. (A brief description of tuna vessel observer data and challenges associated with their use to estimate indices of relative abundance is provided in Scott *et al*., Background Document 1). New technology means that improved field and analysis methods may now be feasible and affordable, but a change in field methods will create bias in trend estimates, unless it is possible to calibrate the new methods against the old.

Shipboard surveys

Current methods

Line transect surveys by the SWFSC in the ETP began in 1974 using a combination of aircraft and ships. Shipboard procedures were refined each year and, by 1979, were close to current procedures. Since 1986, the surveys have used a stratified random design. In general, about three times more effort per unit area has been allocated in the central core area than in the outer or peripheral area (Fig. 1). The central area includes the main dolphin stocks of interest, namely northeastern offshore spotted and eastern spinner dolphins, and the main area where tuna seiners set on dolphins. Because the ETP area is large and the research vessels have a limited range of 20-30 days, it is not possible to lay transect segments strictly at random. Instead, prior to departure, waypoints are chosen to achieve the desired allocation of effort among strata, approximately even spatial coverage within each stratum, and a length of trackline that returns the ship to port at the end of each leg. Since 1986, each survey has utilized 2 ships (3 in 1998) for 120 sea days each, with 4-5 legs per ship from late July to early December. Prior to the start of a survey, observers undergo 3 days of training in data collection procedures, species identification and group size estimation. Most observers have previous experience conducting ETP dolphin surveys, and for them this is refresher training.

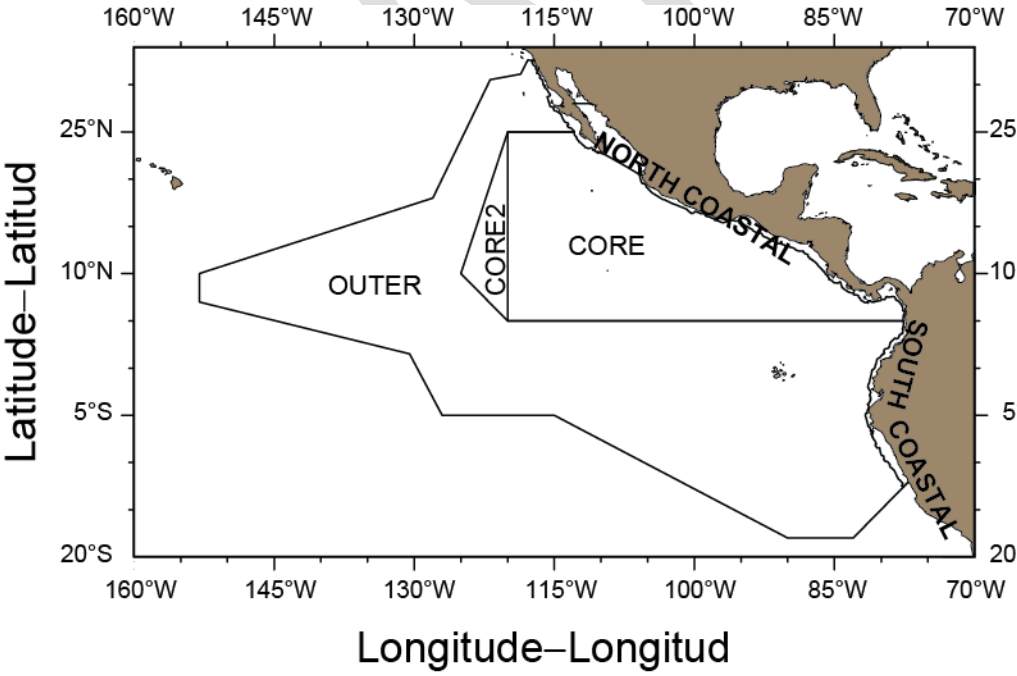

Figure 1. Strata for the STAR06 cruise (used with permission; Gerrodette *et al*., 2008). (The ‘core’ area was expanded to include the ‘core2’ area during the 2003 and 2006 surveys.)

During the survey, the ship proceeds from waypoint to waypoint at 10 knots. Waypoints are typically hundreds of miles apart. Search effort takes place when there is sufficient light for effective detection of animals (normally about 30 minutes after sunrise to about 30 minutes before sunset). Search effort is suspended if it is too windy (normally Beaufort sea state > 5), if visibility is severely limited by rain or fog, or if the horizon is not visible due to haze. At night and during such periods of suspended effort, the ship continues along the planned trackline to stay on schedule.

The line transect surveys carried out by the SWFSC in the ETP have used teams of three observers. Early experiments with helicopters established that most dolphin schools were detected before there was a significant reaction of the dolphins. For dolphin schools of the stocks targeted by the purse seine fleet, it has been assumed that the probability of detection on the trackline [$g(0)$] is 1.0, but this has recently been called into question (Barlow, 2015). While "on effort," two observers search through 25X pedestal-mounted binoculars, one on each side of the flying bridge, from 90° abeam to the centerline. (In the early years, observers searched out to 10° on the other side, to ensure some overlap of effort near the line.) The third observer searches by naked eye or with hand-held 7X binoculars over the whole 180° in front of the ship; the third observer also enters data into a laptop computer. Observers spend 40 min at each position for a total of two hours on watch, and then have two hours off duty. It has been found that this schedule allows observers to maintain their concentration during their search period on the large binoculars. There are six observers on each vessel, three on duty at a time.

The laptop computer is connected to the ship's GPS system, and automatically records date, time and position whenever data are entered, or after 10 min if no data have been entered. Data include the observers on duty and sighting conditions, such as sea state, swell height and sun angle. When a group of cetaceans is sighted, the observer measures the horizontal angle (azimuth) from centerline to sighting using an angular scale mounted on the pedestal, and the vertical angle (declination) from horizon to sighting using a reticle scale inscribed in one eyepiece of the binocular. Radial distance to the sighting is calculated from the reticle value (Kinzey and Gerrodette, 2001).

After the angle and reticle to the sighting have been recorded, the observer team typically goes "off effort" and directs the ship to leave the transect line and approach the sighting. The purpose of "closing" on the sighting is to identify the proportion of each species present in the group (often not clear at first detection) and to obtain the best possible estimates of school size. Experiments have shown that both kinds of data are compromised if the ship remains on the trackline and does not close on the sighting. Approaching the sighting also allows other kinds of data to be collected, such as biopsy samples and behavioural observations.

Estimating the size of dolphin schools is a crucial but difficult task. The SWFSC surveys have devoted considerable effort to verify the accuracy of group size estimates, particularly because the mean sizes of spotted and spinner schools recorded by research vessels are much smaller than reported from fishing vessels. It is thought that the school size differences between research and fishing vessels are due to a combination of several factors (Ward, 2005). Since large dolphin schools tend to be associated with more tuna, fishermen seek out areas with large schools. In addition, within those areas, fishermen may not report smaller groups of dolphins with few or no tuna to the observer. There may also be a tendency for fishermen and/or fishery observers to make larger estimates of dolphin group size.

School size estimates by observers on research vessels in the ETP have been extensively checked by taking vertical photographs of entire schools, and comparing observers' estimates to counts from the photographs (Gerrodette *et al*., 2002). The accuracy of group size estimates varies from observer to observer, and from group to group for a single observer. On average, over all schools and all observers, group size is estimated accurately for groups of about ten dolphins, but there is a tendency to underestimate larger groups by 20-30% (Gerrodette *et al*., in prep). It should be emphasized that estimating the size of a group of active dolphins is a very challenging task, and the variance among observers for a single school, as well as the variance for a single observer from school to school, is high.

To reduce this variance and to improve accuracy, the SWFSC has used three main strategies. First, during pre-cruise training, observers learn group size estimation techniques. They practice estimating group sizes using photographs, videos and computer simulations. Second, after the ship approaches a sighting, the three on-duty observers make independent estimates of group size. They write their numbers down in individual notebooks, and the cruise leader later enters the estimates in the electronic data file. The mean of the three independent estimates is less variable than single estimates. Third, the tendency of each observer to under- or over-estimate group size has been assessed with aerial photographs of the schools, as described above. Each observer's estimates are adjusted according to his/her individual tendency, and this improves the overall accuracy of group size estimation. These procedures also allow group size estimation error to be included in the variance of the estimate of abundance.

Abundance has been estimated from these survey data using a multivariate extension of conventional line transect methods (Gerrodette and Forcada, 2005; Gerrodette *et al*., 2008). This methodology is used to account for covariate effects on the estimated probability density evaluated at zero perpendicular distance (Buckland *et al*., 2004). Covariate effects considered in the analyses include: school size, sea state, swell height, time of day, survey ship, sighting cue, method of sighting, presence/absence of glare on the trackline, and presence/absence of seabirds.

Strengths and weaknesses

The above methods are tried and tested. The target species form large, easily detected schools, and a wide strip can be surveyed using the pedestal-mounted 25x binoculars. It is relatively easy to evaluate assumptions. Animals are likely to be detected before any significant response to the vessel occurs, at least in good conditions.

Movement of animals (independent of the vessel) will generate some upward bias in estimates, but this bias should be modest. (If average animal speed is half the speed of the vessel, then bias is of the order of 5%, Glennie *et al*., 2015.) It can be difficult to estimate group size and species proportions (mixed groups), but as noted above, aerial photographs of a sample of schools are used to quantify and correct for bias. Precision is rather poor, given the resources that have been devoted to these surveys. Barlow (2015) has conducted analyses that indicate that g(0) might be appreciably below one in all but the best sighting conditions, which may be linked to a reduced window in which a school is available for detection in poorer sighting conditions together with responsive movement.

It is also costly to conduct effective shipboard surveys over such a large study area.

Suggestions for addressing the weaknesses

Changes in field methods might improve abundance estimates, but also risk compromising having a time series of comparable estimates. If g(0) is less than one, using a double-platform approach may allow its estimation (Borchers *et al*., 1998). The apparent effect on g(0) of sea state may arise due to responsive movement of schools prior to detection. Attempts to study responsive movement in 1998 used an observer searching with 25X binoculars from a higher "tracker" platform but failed to detect groups appreciably before the primary search team. However, a drone or helicopter might provide a more effective tracker platform, operating ahead of the survey vessel, and setting up trials for the main observation platform. This allows estimates to be corrected for both responsive movement and g(0) (Buckland and Turnock, 1992). We discuss these issues further in the next section.

Correlations among sea state, location, extent of evasive behavior and group size may partially explain the results of Barlow (2015). Model-based analysis methods may help to resolve this, and perhaps provide estimates of abundance with greater precision. Model-based methods are useful both for modeling encounter rate and for modeling the detection function. In the latter case, using multiple covariate distance sampling methods, it is possible to jointly model data from different species, with species as a factor in the detection function model, to improve precision (e.g. Barlow *et al*., 2011). However, the larger source of variance is encounter rate, and so encounter rate modeling perhaps merits more attention, especially in light of recent developments in spatial distance sampling methods (e.g. Yuan *et al*., submitted).

Improved designs based on oceanographic conditions and adaptive sampling may be able to contribute to higher precision, although we would expect gains to be rather modest.

Use of commercial vessels

The most important principle of survey design for design-based estimation of abundance (e.g., distance sampling methods, Buckland *et al*., 2001, 2004) is that units of survey effort are placed randomly with respect to the distribution of animals or groups of animals (Buckland *et al*., 2001). Violation of this principle can lead to an unrepresentative sample and hence biased estimates of abundance. This is one of the primary disadvantages of opportunistically collected survey data (e.g., fisheries observer data), and it has been shown that the non-random search of tuna purse-seine vessels during fishing operations is problematic with respect to estimation of dolphin indices of relative abundance (Lennert-

Cody *et al*., 2015 and references therein). Therefore, if data collected aboard tuna vessels were to supplement data collected by research vessels, or were to be the primary data source for abundance estimation, it is critical that effort allocation be determined by a designed randomized survey.

The best option for a commercial vessel survey to produce abundance estimates of similar quality and precision as those of the previous NMFS surveys would be for the commercial vessel survey to replicate all aspects of the NMFS survey methods and design (with the obvious exception of use of the same vessels), including: the number of vessels, the amount of search effort, the set-up of the observation platforms, the use of specially trained observers, and the calibration of observers' estimates of group size by aerial photogrammetry. Because vessels, and perhaps observers, will be different from those used in the NMFS surveys, biases may differ, which could compromise the time series of estimates. (Ideally, we would seek to estimate absolute abundance without bias for both, in which case we avoid this issue.) Whether a commercial vessel survey would be advantageous depends in part on vessel cost, which is the largest component of the survey budget. Examples of current cost per day of research vessels are: US$25K-US$29K for the NMFS/NOAA vessels (US$6.65M/266 ship days - US$7.2M/266 ship days, from SWFSC 15 July 2016 survey budget in 2017 US dollars; the 266 ship days corresponds to 240 actual sea-days) and about US$18K for the Universidad Nacional Autónoma de México vessel El Puma (although this vessel would have to be outfitted with marine mammal survey equipment, at an additional cost). Although the cost of a commercial vessel will vary with the size and age of the vessel, and country-specific costs of fuel and insurance, among other factors, a recent example of the possible commercial vessel cost range (Lennert-Cody *et al*., 2016) was US$16K-US$25K per day. There likely would be additional costs to outfit commercial vessels with marine mammal survey equipment.

If commercial vessel time for surveys were to be provided (e.g., donated) by vessel owners, it might be anticipated that more vessels would be involved in the survey, each for a shorter period of time than 120 days. Three aspects of a many-vessel survey are worthy of further discussion. First, if each commercial vessel were to survey during a portion of a fishing trip, it would be important to determine the optimal allocation of survey segments from a random design to each vessel so as to minimize transit time from the fishing location to the survey location, taking into consideration spatial gradients in dolphin abundance (Reilly, 1990; Redfern et al., 2008; Forney et al., 2012) and the constraint that vessels should all operate at the same time of year to avoid any potential biases due to dolphin population movement. A simulation using historical commercial vessel fishing trip trajectories, in combination with historical NMFS survey tracklines, or new survey tracklines, and information on dolphin spatial distributions, could be conducted to determine how best to allocate blocks of survey segments from an ETP-wide design to individual tuna vessels in time and space.

Second, in a many-vessel survey, it will be critical that vessels, survey platforms, observers, and data collection procedures, are as similar as possible so as to reduce heterogeneity in the data. Even for the research vessel surveys, ship effects on the detection function were identified in some years (Gerrodette and Forcada, 2005). With certain assumptions, including perfect detection on the trackline (i.e., $g(0)=1$), the property of “pooling robustness” (Buckland *et al*., 2004) implies that heterogeneity in the data should have little effect on estimation. However, recent analyses (Barlow 2015) suggest that the assumption that $g(0)=1$ during all survey conditions may not be as likely to hold as previously assumed, and therefore minimizing heterogeneity in the data as much as is practical through survey design is important. An additional important aspect of the survey design would be to determine the optimal number of vessels that should participate in the survey, based on target CVs and logistical constraints. If the data collected are of sufficient detail, and with appropriate covariates, other covariate effects on detection, such as Beaufort and dolphin group size, can be modeled (Buckland *et al*., 2004). One way to ensure a similar level of skill among observers would be to conduct a training cruise aboard one of the commercial vessels participating in the survey, prior to survey season, with all observers. Such a cruise might be run by senior NMFS observers because of their extensive field experience. A training cruise might provide an opportunity for observer school size calibration at the end of the cruise, either with a helicopter (carried by many tuna vessels that fish for tunas associated with dolphins) or a remotely-operated drone.

Finally, for a commercial vessel survey it may be important to have a survey design and data collection protocol that will allow for testing of the assumption of perfect detection on the trackline (i.e., the assumption that $g(0, z_j)=1$ for covariates z_j), and if necessary, estimation of $g(0, z_j)$. For research vessel surveys, there is evidence that Beaufort affects $g(0)$ (Barlow, 2015). If environmental factors affect detection on the trackline for research vessels, the same would be anticipated for tuna vessels. In addition, detection on the trackline for tuna vessels may be problematic because of an evasive response of ETP dolphins to commercial vessels, which varies spatially across the area occupied by the fishery (e.g., Pryor and Norris, 1978; Lennert-Cody and Scott, 2005 and references therein). Estimation of $g(0, z_j)$ would require a survey design that is different from the previous NMFS surveys. Double-count survey designs, from which $g(0, z_j)$ can be estimated, involve two teams of observers; if one of these teams searches at greater distance (possibly from a helicopter or using video from a drone flying ahead of the ship), while the other carries out normal search, it is possible to correct for dolphin movement in response to the survey vessel. Abundance estimation is based on mark-recapture distance methods (e.g., Borchers *et al*., 1998; Buckland *et al*., 2004, Chapter 6). Such survey designs would require additional observers and equipment. Although commercial vessels that are suitable for a survey tend to carry a helicopter, it would first need to be ascertained whether the helicopters elicit an evasive response in dolphins, given that helicopters are used during fishing operations. In addition, observer safety may be a concern with respect to helicopter use; if helicopter use is restricted to good conditions, it would be of limited use for quantifying $g(0, z_j)$, which is expected to be lowest in poor conditions. Drones might provide a less disruptive vehicle for distant observations in a double-count survey design.

Aerial surveys

Manned aircraft with observers

Aircraft surveys with observers have been widely used for estimating abundance of marine mammal populations. The effective width of the strip searched is appreciably lower than for shipboard surveys, but a much greater distance is covered in a given time. Aside from practical issues, aircraft surveys would be very effective for surveying dolphin species that occur in large schools. Given a suitable choice of altitude and bubble and/or belly windows so that observers can see below the aircraft, the probability of detection of such schools on the line is likely to be at or close to one. School size is easier to estimate from the air than from a ship, and the aircraft can fly over a detected school, taking photos for later analysis of school size and species identity. However, survey aircraft do not have the range to operate throughout the ETP, and there are safety issues to consider. Safety and limited coverage of the study area were the main reasons the SWFSC discontinued aerial line transect surveys in the ETP in the 1970s. Helicopters can operate from a ship, and can ensure that a wider area is searched in a given time than is possible for ship-based observers, but the costs are much higher than for aircraft operating from land, and the risk to scientists associated with operating a small helicopter at sea is also a consideration.

Commercial digital aerial surveys

Commercial manned digital aerial survey methods were developed in the United Kingdom to provide survey data on potential impacts of developing offshore wind farms. Early tests demonstrated their effectiveness for census of seaduck at one coastal site, with abundance estimates that exceeded those of traditional visual aerial survey methods (Buckland *et al*., 2012). The same authors predicted that digital survey methods would replace visual aerial methods for seabirds in offshore waters, which has proved to be the case in the UK, Germany, Denmark, and increasingly in the eastern United States with thousands of sorties now flown. Digital video aerial survey methods have been found to give comparable results to dedicated visual aerial survey methods for harbour porpoise *Phocoena phocoena* in the UK (Williamson *et al*., 2016), and for other marine megafauna (Williams *et al*., 2015; Gordon *et al*., 2013).

Two technologies have emerged for commercial census by digital aerial survey: high resolution video; and high resolution digital stills. In general, the video methods use bespoke camera rigs to scan a strip transect using four cameras in a comb pattern over the sea. Stills methods usually use medium-format photogrammetry cameras to sample plots (or quadrats) or transects at sea. For seabird surveys, cameras ideally collect images at a ground sample distance (GSD) of 2-3cm, and this allows species identification rates of at least 80% of all seabird species in the UK, and considerably higher rates for cetaceans. Lower resolutions of 3 – 5cm also achieve high identification rates for cetaceans. The higher camera resolutions are achieved while flying at 550m above sea level (a.s.l.) for digital video methods and 270 – 400m a.s.l. for digital stills, depending on the GSD used.

Both methods in the UK use a two-phase method for analysing digital data generated. The first phase requires a review of all material, with 10% or 20% of all material subjected to a random blind audit, and robust procedures for handling failed audits. The second phase requires all objects to be assigned to the lowest order taxon possible. Again 10 – 20% of all objects are subjected to a randomised blind audit, with procedures for handling failed audits. Digital stills and digital video methods have attempted to use automated methods for detection and identification of objects using machine learning methods, with varying success. While detection methods are reasonably successful in calm sea conditions, they have much poorer accuracy at higher sea states, particularly for marine mammals. Automated methods for identification of objects require considerable human intervention and oversight, negating the potential efficiency benefits of such methods. While success so far has been low in these methods, it is likely that more sophisticated artificial intelligence methods will ultimately be able to replicate the undoubted accuracy of experienced human operators.

Digital aerial methods offer a number of advantages over conventional aerial survey methods:

- Because the aircraft operate at greater altitudes and have fewer crew, digital methods are considerably safer than visual aerial surveys.
- Detection rates are uniform across the whole image for digital methods, meaning that it is not necessary to account for missed detections using distance methods, and double-review methods are simpler.
- All individual animals can be counted and group sizes do not influence detection rates in digital methods, removing the need to estimate group size and to account for detection bias when estimating abundance.
- A permanent record of the survey can be kept for subsequent analysis should the need arise with digital aerial methods.
- Bespoke rigs are used to angle digital cameras away from sun glare and avoid detection problems of fixed camera systems and visual aerial survey methods.
- Digital video methods are still effective at higher sea states, when compared to digital stills and visual methods, although there is some unpublished evidence for lower detection rates in video methods, mainly for sub-surface marine mammals at higher sea states.
- Identification rates are higher for digital methods.

Some issues remain for digital aerial methods when compared to other methods and the survey requirements of the eastern Tropical Pacific surveys:

- Digital aerial methods are more expensive than visual aerial methods (but typically cheaper than dedicated boat-based surveys).
- While automated review methods are available, they are still not sufficiently efficient compared to manual review. Considerable investment is required to develop methods that will provide significant time and cost savings.
- Availability bias for diving seabirds and cetaceans in digital survey methods is acute, but difficult to account for. There exists a theoretical method for measuring this bias *in situ* using digital video methods which so far is untested. This is most likely to be effective for cetacean species with relatively short dive cycles (typically 2-3 minutes or less). No method exists for measuring this bias *in situ* for digital stills methods. Generic methods can be used, based upon known dive rates where these exist, for estimating availability bias for digital survey methods (Webb *et al.*, 2015).
- As with visual aerial survey methods, the endurance of the aircraft used for these surveys is limited and insufficient for reaching the furthest limits of the ETP survey

> area safely from suitable airports. (While deploying helicopters from boats offshore is possible, helicopters have been found not to provide a sufficiently stable platform for digital transect-based surveys.) Some aircraft are able to re-fuel mid-air, and one of the aircraft used for digital video aerial surveys has a pilot-less version which increases the endurance and safety significantly.

Commercial aircraft could potentially be fitted with cameras to gather high-resolution images. This has the advantage of low cost relative to dedicated aerial surveys, although potentially costly certification of aircraft for installing cameras might negate this advantage. The main disadvantages are: commercial aircraft fly at a much greater altitude than dedicated survey aircraft, resulting in low-quality images; commercial aircraft usually fly much faster which would compromise the number of images or frames that can be captured; and commercial aircraft routes do not sample the ETP evenly, so that spatial modelling methods will be required to extrapolate across the whole region.

Drone surveys with high-resolution imagery

The use of Unmanned Aerial Systems (UAS), also known as drones, has burgeoned in the ecological survey sector in the last decade and, much like in manned digital aerial surveys, can be used for transect or plot sampling of marine mammal distribution and abundance. They are available in many forms, from small multi-copter systems that carry small video cameras that have high definition or ultra-high definition (4X) resolution and save images to flash memory cards, up to military-grade fixed wing UASs that are capable of carrying much larger payloads with higher resolution cameras and server-based image storage systems. At the smallest end of the size spectrum, the camera systems are unlikely to deliver images of sufficient quality. Most attention in the use of UASs for marine census has been given to small- to medium-sized systems, such as the AH22, that are able to carry sensors of sufficient payload to capture higher end quality images or video material.

Some small- to medium-sized UASs are designed to be recovered at sea, but most need to be deployed and recovered from the deck of a ship if they were to be used to census in the entire ETP. This would elevate the cost considerably by the addition of the price of a mother ship that is able to reach the more distant parts of the proposed study area. Part of this restriction is imposed by limited access to airspace; in Europe and USA, aviation regulations require that UASs are flown within line of sight of an operator. A further limitation on the use of such systems is their endurance, both in the number of hours that can be surveyed in a single mission and in the storage capacity for the images. The endurance of even medium-sized systems is limited to about five hours at about 100 km per hour, which would mean that a survey of the ETP would be slow, unless carried out by multiple UASs. Storage capacity also limits the duration of sorties to a few hours and also means that raw image formats cannot be stored, thus reducing image quality slightly.

Military-grade systems are able to take much larger payloads and would be able to carry the payload of a commercial digital aerial survey system on board, including multiple cameras and server-based data storage systems. This gives them considerably greater endurance. Such systems would also need to take off and land at commercial or military airstrips and cannot be recovered at sea at present. The Diamond Aviation DA42, used by HiDef for its digital video aerial surveys, has a pilotless version used for military purposes. It would have an endurance of about 15 hours and would be licensed to carry cameras and increased data storage capacity for a wider-area survey such as the ETP. To use such as system would require negotiated access to airspace of the ETP study area.

The same issues exist with a lack of automated detection method – this is possible in near calm conditions, but becomes problematic in the likely sea states typical of the ETP. There may also be reliability issues with UASs being lost and not re-located at sea.

Satellite surveys

Very High Resolution (VHR) satellites now have the ability to capture large areas of ocean (>1000 km^2 per image) at a spatial resolution of 30 cm per pixel. Recent work on cetaceans using lower resolution imagery (50 cm) has shown the utility for counting baleen whales in optimal conditions and initial tests using 30cm imagery on humpback whales have shown a clear improvement in detection, both on the surface and beneath it. With 30 cm satellite imagery it should be possible to identify the pattern of breaching small cetaceans in relatively calm seas, although species identification is unlikely. In calm conditions the signature of the splashes will be very bright relative to the surrounding waters, and due to the radiometric resolution of the satellites, it may also be possible to automate or semi-automate the process of finding these patches for large pods of dolphins. If agreements could be made with the satellite provider, this could be a very cost-effective way to survey large expanses of ocean to give first order abundance or presence estimates, or estimated indices of abundance (i.e., relative abundance). Other advantages are the ease of use of satellites, the ability to capture extremely large amounts of imagery in any area of ocean, the non-invasive nature and the lack of logistical set-up or flight planning effort for satellites.

However, this use of satellite technology is still developing and much testing would be needed before a workable system using satellite data could be incorporated into other survey methods. There are some comparisons to be drawn between satellites and the use of high-resolution aerial survey using digital stills; each has similar drawbacks – the need for favourable sea conditions, the problem of single instantaneous image acquisition and potential problems, and the need for automation over large areas. The main differences between the two systems are the higher resolution of the aerial imagery and the greater potential coverage from satellites.

The potential cost of the highest resolution imagery could be high for large area studies unless an agreement can be gained from the satellite provider; this is more likely either over areas with less demand for imagery (open ocean) or areas where high profile research could be conducted. As the use of this technology is unproven for small cetacean survey, the algorithms needed for automated or semi-automated identification still need to be constructed and proven. Manual checking over 1000’s of square kilometres is difficult, although crowd-sourcing the imagery might solve this in the longer term. Species identification will be impossible with satellites as the resolution is too coarse and estimating school size could be difficult without ground truthing.

Capture-recapture and close-kin methods

The following draws heavily on Lennert-Cody *et al*. (2016).

Although mark-recapture methods can be used to estimate abundance, potentially insurmountable problems arise when applying these methods to ETP dolphin stocks. These problems arise from several sources: large population size; heterogeneous and non-independent probabilities of capture; possible errors in matching marked animals; tag loss; difficulty in defining the population that is being estimated, given the potential for movement in and out of the ETP. Buckland and Duff (1989) summarized the problems of estimating numbers of Antarctic minke whales by mark-recapture methods; their population size is similar to that of the main ETP stocks of dolphins.

Marking individuals with visual tags requires an initial capture to attach the tag to the animal, and subsequent visual 'recaptures' of tagged individuals by observers or vessel crew. Tagging and tracking studies of dolphins using radio tags and visual tags have been conducted previously in the ETP (e.g., Perrin *et al*., 1979; Butler and Jennings, 1980; Scott and Chivers, 2009), and while much was learned about behavior, movements and migration patterns, the number of tag returns was not large enough to estimate abundance. Tagging methodology has improved, but tagging would need to be on a very large scale to make abundance estimation feasible.

Passive Integrated Transponder (PIT) tags have been proposed for studying ETP dolphin movement (Inter-American Tropical Tuna Commission, 2005). An animal is 'marked' when the PIT tag is implanted internally (thus reducing tag loss) and 'resighted' when a receiver detects the presence of the PIT tag during a purse-seine set. PIT tags have a potential advantage because the detector could be located in the cork line of the purse-seine net, over which dolphins pass when they are released during the 'backdown' procedure, if detectors could be built to be sufficiently rugged to withstand the purse-seine net retrieval process. However, reliable detection at a distance of several metres would be required. A failure to detect any tagged animals amongst those encircled will bias abundance estimates. Assuming these problems can be overcome, the recapture rate should be very high, given the large number of animals encircled per year. However, marking a sufficiently large and representative sample of animals from the entire range of the stocks may be a challenge. Note that, provided the sample of animals marked is a random sample from the population, those 'recaptured' during back-down do not need to be, provided animals do not change behavior after being marked in a way that alters their probability of encirclement. However, marking a large random sample would be a challenge, and require considerable resources. It would not be cost-effective for example to mark just a single animal from a group, but if several from a group are marked, they cannot be considered to be randomly and independently sampled; any analyses would need to take account of this.

Genetic mark-recapture techniques (e.g., Pearse *et al*., 2001) and 'close-kin analysis' methods (e.g., Skaug, 2001; Bravington *et al*., 2016), use genetic information to estimate abundance. Mark-recapture methods require an initial sample for marking and a subsequent recapture sample. Based on the 2006 estimates of dolphin abundance of about 1.3 million offshore spotted dolphins (Gerrodette *et al*., 2008), tissue samples from roughly 17,000 offshore spotted dolphins would need to be collected during each phase for the mark-recapture study. For parent-offspring based close-kin analysis samples from about 9000 animals would be required in the first year, but perhaps fewer samples in subsequent years (Lennert-Cody *et al*., 2016). Collecting such numbers of biopsies may require more than the 240 sea-days estimated for the NMFS line transect surveys. For close-kin methods, and for recaptures for the mark-recapture method, it may be possible to use samples from dead animals only, but at the current annual fisheries mortality level of about 1000 animals (IATTC, 2013), it would take a decade or more to obtain sufficient samples. However, with close-kin methods, if more distant relationships can realistically be determined from genetics (down to first cousins), five years of such data may be sufficient (taking into account that dead animals may be aged). A possible alternative (perhaps unrealistic) is that vessel crew or fisheries observers collect biopsies from individuals at the time that the dolphins are released from the purse-seine net in the backdown procedure during normal fishing operations. If even a small fraction of those animals could be successfully sampled, at the current level of about 10,000 dolphin sets per year (IATTC, 2015), genetic mark-recapture will be feasible with a few years of data. Close-kin methods will also be applicable, providing additional information from the same set of biopsy samples.

Mark-recapture methods that take advantage of the fishing process, such as PIT tags and genetics mark-recapture, would seem most promising from the point of view of maximizing data collection with minimal interruption to daily fishing activities. For all these mark-recapture methods, however, there are research costs that have yet to be determined, as well as costs for logistics and technical support, and analysis, and so it is not presently clear how mark-recapture methods compare to line transect methods in terms of cost. Moreover, genetic markers for dolphin species would need to be developed.

For all mark-recapture methods, if implemented during normal fishing operations, it would be essential to establish that animals encountered by the fishery are representative of the population; i.e., that it is not just a specific subset of the dolphin population that gets encircled by purse-seines. Because the purpose of fishing is to catch fish, fishermen attempt to encircle dolphins that are associated with tuna, which may not be the entire dolphin herd. Prior to encirclement with the purse-seine net, fishermen attempt to exclude dolphins not associated with tunas and some dolphins actively evade capture. Spatial structure in dolphin evasive response, for example, implies that recapture probabilities would not be expected a priori to be spatially invariant.

Passive acoustics

The following draws heavily on Heinemann *et al*. (2015).

Distance sampling, adapted for acoustic data, is the most commonly used approach to estimation of abundance from passive acoustic data (Heinemann *et al*., 2015). Assuming a species can be unequivocally identified by its vocalization repertoire, to estimate trends in relative abundance from acoustic data, there are several key requirements for acoustic distance sampling methods:

- Detection probability can be estimated as a function of horizontal distance from the "cue" (e.g., vocalization count) to the acoustic instrument. To obtain detection as a function of horizontal distance, the depth of the cue (i.e., animal or group of animals) is often assumed, and this may bias the estimated detection function. In addition, the detection probability has to be corrected for the false detection rate (i.e., vocalizations that were incorrectly assigned to the target species during data processing). Although sound-propagation modeling has been used to estimate detection range in order to estimate distance to the cue, accurate estimation of range from these models is currently challenging. This is especially true for highly directional signals like echo-location clicks. A drifting vertical hydrophone array can be used to estimate range empirically which holds more promise than model-based range estimation.
- Density estimation methods can be based on individual-count methods, group-count methods, or cue-count methods. Individual-count methods are typically not practical because individuals within a group cannot be discriminated acoustically. Group-count methods require an estimate of group size, and methods to estimate group size from only acoustic data currently do not exist. (Group size is often obtained from concurrently-collected visual survey data.) Methods to convert cue counts to individual density require estimates of the cue production rate (vocalizations per unit time) under environmental and social conditions that are likely to be encountered during the survey.

Acoustic methods may be most valuable for estimating trends in relative abundance rather than absolute abundance. The accuracy of estimated trends in abundance will depend on the number and location of acoustic platforms used in the survey, and whether parameters such as detection probability, vocalization rates, and area effectively surveyed can be estimated or assumed to be constant. The number of surface drifters can be increased at relatively little cost to obtain the number of detections to achieve a desired power to detect changes in abundance. At present, statistical methodological challenges exist for estimating abundance from acoustic data collected with slow-moving autonomous platforms. Although passive acoustics have been used to estimate trends in both relative abundance and absolute abundance, in the near future this type of data is not likely to replace visual line transect data for monitoring trends in ETP dolphins. Assuming that detection distances can be measured for each acoustic detection, the remaining key uncertainties are the degree to which acoustic behavior and group size vary over time.

Composite methods

There are many ways in which methodologies could be combined, some of which have already been discussed. For example, a helicopter or a drone could operate ahead of a ship, providing a second platform, and data from which corrections for responsive movement and for g(0) may be estimated. Helicopters have been used in the past to estimate group sizes and species proportions, and hence estimate bias in observer group-size estimates by species. Surface drifters or gliders might be used to gather acoustic data, and jointly modeled with sightings survey or high-resolution imagery data, using a model-based approach. The acoustic data might have good spatial coverage, and spatially-varying calibration against relatively sparse sightings survey data would in principle allow conversion to absolute density. A similar strategy may allow us to utilize tuna vessel observer data together with research vessel or aerial survey data. Exploratory analyses using existing tuna vessel observer data and research vessel data may be useful.

Discussion

The safe option is to replicate past research vessel surveys as far as possible. However, these surveys are costly, and precision on abundance estimates is not high. Other approaches would incur development costs, and unforeseen problems may arise. Further, unless both research vessel surveys and any new methods provide unbiased estimates of abundance, estimates from a new approach are unlikely to be directly comparable with past estimates. Implementing a new approach together with a research vessel survey would allow the two approaches to be calibrated, but the cost of the exercise would be high, and unless it was repeated over several years, the calibration factor would be imprecisely estimated. A possible option is to implement a less costly approach (e.g., using drones or satellite images) with the aim of obtaining an annual index of relative abundance, together with an occasional full survey (perhaps using methods closely comparable with past research vessel surveys) to attempt to estimate absolute abundance.

Of the potential new approaches, perhaps the most promising in terms of cost, practicality and precision is the use of high-resolution video taken from long-range drones. Suitable drones have until recently been the preserve of the armed forces, but are now becoming commercially available. A pilot survey followed by annual surveys for perhaps four or five years would allow a new time series of abundance estimates to be generated fairly quickly. If the drones can be flown from land rather than from a ship (which is feasible, given their range), after initial development costs, it seems likely that this option will have appreciably lower cost than do research vessel surveys.

Options that involve capturing a large number of animals are likely to incur high cost and/or high bias. Acoustic surveys may be able to generate data throughout the ETP, but question marks remain over the reliability of abundance estimates (relative or absolute) from such data. Satellite surveys may be a serious contender, especially if resolution improves to the point that species identification becomes fairly reliable. They would be dependent on obtaining images when sighting conditions are good over a large region, and effective software would be needed for reliable automated search of dolphin schools in vast images.

References

Barlow, J. 2015. Inferring trackline detection probabilities, *g*(0), for cetaceans from apparent densities in different survey conditions. *Marine Mammal Science* **31**, 923-943.

Barlow, J., Ballance, L.T. and Forney, K.A. 2011. Effective strip widths for ship-based line-transect surveys of cetaceans. NOAA-TM-NMFS-SWFSC-484.

Borchers, D.L., Buckland, S.T., Goedhart, P.W., Clarke, E.D. and Hedley, S.L. 1998. Horvitz-Thompson estimators for double-platform line transect surveys. *Biometrics* **54**, 1221-1237.

Bravington, M.V., Skaug, H.J. and Anderson, E.C. 2016. Close-kin mark-recapture. *Statistical Science* **31**, 259-274.

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

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., Thomas, L. 2004. Advanced Distance Sampling. Oxford University Press, Oxford.

Buckland, S.T., Burt, M.L., Rexstad, E.A., Mellor, M., Williams, A.E. and Woodward, R. 2012. Aerial surveys of seabirds: the advent of digital methods. *J. App. Ecol.* **49**, 960-967.

Buckland, S.T. and Duff, E.I. 1989. Analysis of the Southern Hemisphere minke whale mark-recovery data. In *The Comprehensive Assessment of Whale Stocks: the early years*, pp. 121-143. G.P. Donovan (ed.). International Whaling Commission, Cambridge.

Buckland, S.T. and Turnock, B.J. 1992. A robust line transect method. *Biometrics* **48**, 901-909.

Butler, R.W. and Jennings, J.G. 1980. Radio tracking of dolphins in the eastern tropical Pacific using VHF and HF equipment. Pp. 757-759 in *A Handbook on Biotelemetry and Radio Tracking* (eds C.J. Amlaner Jr. and D.W. MacDonald). Pergamon Press, Oxford, pp. 757–759.

Forney, K.A., Ferguson, M.C., Becker, E.A., Fiedler, P.C., Redfern, J.V., Barlow, J. Vilchis, I.L. and Balance, L.T. 2012. Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. *Endangered Species Research* **16**, 113-133.

Fretwell, P.T., Staniland, I.J. and Forcada. J. 2014. Whales from space: counting southern right whales by satellite. *PLOS ONE* **9**, e88655.

Gerrodette, T., Forcada, J. 2005. Non-recovery of two spotted and spinner dolphin populations in the eastern tropical Pacific Ocean. *Marine Ecology Progress Series* **291**, 1-21.

Gerrodette, T., Watters, G., Perryman, W. and Ballance, L. 2008. Estimates of 2006 dolphin abundance in the eastern tropical Pacific, with revised estimates from 1986-2003. NOAA Technical Memorandum NMFS-SWSFC **421**, 1-45.

Glennie, R., Buckland, S.T. and Thomas, L. 2015. The effect of animal movement on line transect estimates of abundance. *PLoS ONE* **10**, e0121333.

Gordon, C., Kujawa, M., Luttrell, J., MacArthur, D., Robinson-Willmott, J. and Thaxter, C. 2013. *High-resolution Aerial Imaging Surveys of Marine Birds, Mammals, and Turtles on the US Atlantic Outer Continental Shelf — Utility Assessment, Methodology Recommendations, and Implementation Tools*. Final Report by Normandeau Associates to Bureau of Ocean Energy Management, OCS Study BOEM 2013-01130.

Heinemann, D., Gedamke, J., Oleson, E., Barlow, J., Crance, J., Holt, M., Soldevilla, M. and Van Parijs, S. 2015. Report of the Joint Marine Mammal Commission – National Marine Fisheries Service Passive Acoustic Surveying Workshop, 16-17 April 2015, La Jolla, California.

Hoyle, S.D. and Maunder, M.N. 2004. A Bayesian integrated population dynamics model to analyze data for protected species. *Anim. Biodivers. Conserv.* **27**, 247-266.

Inter-American Tropical Tuna Commission, 2005. IATTC Scientific Advisory Board Document SAB-2-08a. 2nd Meeting of the Scientific Advisory Board, 19 June, 2005. Lanzarote, Spain. http://www.iattc.org/PDFFiles2/SAB-2-08a-Dolphintagging.pdf.

Inter-American Tropical Tuna Commission, 2013. IATTC Annual Report for 2009. http://www.iattc.org/PDFFiles2/AnnualReports/IATTC-Annual-Report-2009.pdf.

Inter-American Tropical Tuna Commission, 2015. IATTC Fishery Status Report 13. http://www.iattc.org/FisheryStatusReportsENG.htm.

Lennert-Cody, C.E. and Scott, M.D. 2005. Spotted dolphin evasive response in relation to fishing effort. *Marine Mammal Science* **21**, 13-28.

Lennert-Cody, C.E., Maunder, M.N., Fiedler, P.C., Minami, M., Gerrodette, T., Rusin, J., Minte-Vera, C.V., Scott, M. and Buckland, S.T. 2016. Purse-seine vessels as platforms for monitoring the population status of dolphin species in the eastern tropical Pacific Ocean: the use of fishing vessels as scientific platforms. *Fisheries Research* **178**, 101-113.

Pearse, D.E., Eckerman, C.M., Janzen, F.J. and Avise, J.C. 2001. A genetic analogue of ‘mark-recapture’ methods for estimating population size: an approach based on molecular parentage assessments. *Molecular Ecology* **10**, 2711-2718.

Perrin, W.F., Evans, W.E. and Holts, D.B. 1979. *Movements of Pelagic Dolphins (Stenella spp.) in the Eastern Tropical Pacific as indicated by Results of Tagging, with Summary of Tagging Operations*. NOAA Technical Report NMFS SSRF-737.

Platonov, N.G., Mordvintsev, I.N. and Rozhnov, V.V. 2013. The possibility of using high resolution satellite imagery for detection of marine mammals. *Biology Bulletin* **40**, 197-205.

Pryor, K., Norris, K.S. 1978. The tuna-porpoise problem: Behavioral aspects. Oceanus 21: 31-37.Scott, M.D. and Chivers, S.J. 2009. Movements and diving behavior of pelagic spotted dolphins. *Marine Mammal Science* **25**, 137-160.

Redfern, J.V., J. Barlow, L.T. Ballance, T. Gerrodette, and E.A. Becker. 2008. Absence of scale dependence in dolphin-habitat models for the eastern tropical Pacific Ocean. *Marine Ecology Progress Series* **363**, 1-14.

Reilly, S.B. 1990. Seasonal changes in distribution and habitat differences among dolphins in the eastern tropical Pacific. *Marine Ecology Progress Series* 66, 1-11.

Skaug, H.J. 2001. Allele-sharing methods for estimation of population size. *Biometrics* **57**, 750-756.

Stapleton, S., LaRue, M., Lecomte, N., Atkinson, S., Garshelis, D., Porter, C. and Atwood, T. 2014. Polar bears from space: assessing satellite imagery as a tool to track Arctic wildlife. *PLoS ONE* **9**, e101513.

Wade, P.R., Watters, G.M., Gerrodette, T. and Reilly, S.B. 2007. Depletion of spotted and spinner dolphins in the eastern tropical Pacific: modelling hypotheses for their lack of recovery. *Marine Ecology Progress Series* **343**, 1-14.

Webb, A., Irwin, C. and Elgie, M. 2015. *Kincardine Offshore Wind Farm: Final Report on Aerial Surveys from April 2013 to September 2014*. Report by HiDef Aerial surveying Limited to Pilot Offshore Wind Limited. http://pilot-renewables.com/pdf_docs/KOWL_ES_AppendixB_FinalAerialSurveysReport_Issued.pdf.

Williams, K.A., Connelly, E.E., Johnson, S.M. and Stenhouse, I.J. (eds) 2015. *Baseline Wildlife Studies in Atlantic Waters Offshore of Maryland: Final Report to the Maryland Department of Natural Resources and the Maryland Energy Administration.* Report BRI 2015-17, Biodiversity Research Institute, Portland, Maine. 437pp.

Williamson, L.D., Brookes, K.L., Scott, B.E., Graham, I.M., Bradbury, G., Hammond, P.S. and Thompson, P.M. 2016. Echolocation detections and digital video surveys provide reliable estimates of the relative density of harbour porpoises. *Methods in Ecology and Evolution* **7**, 762-769.

Yuan, Y., Bachl, F.E., Lindgren, F., Borchers, D.L., Illian, J.B., Buckland, S.T., Rue, H. and Gerrodette, T. submitted. Point process models for spatio-temporal distance sampling data.