Report of the Inter-American Tropical Tuna Commission Workshop on Methods for Monitoring the status of Eastern Tropical Pacific Ocean Dolphin Populations

Edited by
Kelli Faye Johnson, André Punt, Cleridy E. Lennert-Cody

1. Overview and background

On 18-20 October 2016, the Inter-American Tropical Tuna Commission (IATTC) held a workshop on methods for monitoring the status of eastern tropical Pacific Ocean (ETP) dolphin populations at the Southwest Fisheries Science Center, La Jolla, California. Dolphins in the ETP, particularly pantropical spotted dolphins (*Stenella attenuata*), spinner dolphins (*S. longirostris*), and common dolphins (*Delphinus delphis*), co-occur with yellowfin tuna (*Thunnus albacares*). The Antigua Convention of the IATTC mandates that the status of tuna, and other species impacted by ETP tuna fisheries, be monitored (IATTC, 2003).

Since at least the 1940s, tuna purse-seine vessels have used the co-occurrence of dolphins and tuna to locate tuna (Silva, 1941; NRC, 1992). In the late 1950s vessels began encircling dolphins as a means to catch tuna (McNeely, 1961; NRC, 1992), which resulted in substantial dolphin bycatch (Perrin, 1968; Lo and Smith, 1986; NRC, 1992; Wade, 1995). Bycatch has been significantly reduced through fishermen's ingenuity and the implementation of national and international management measures (NRC, 1992; Joseph, 1994; Hall, 1998; IATTC, 2016). Nevertheless, the status of these dolphins is still in question given high levels of historical mortality (Wade, 1995) and low estimated rates of population increase (Gerrodette *et al*., 2008).

Historically, estimates of dolphin status have been based on population dynamics models (Alvarez, 2002; Wade *et al*., 2002, 2007; Hoyle and Maunder, 2004; Reilly *et al*., 2005; IATTC, 2006) that used estimates of abundance based on data collected during fishery-independent, ship-based surveys (Gerrodette and Forcada, 2005; Gerrodette *et al*., 2008) conducted by the United States National Marine Fisheries Service (NMFS). During most years, surveys were not conducted annually, and, therefore, the use of tuna vessel observer data to estimate trends in relative abundance was investigated (Hammond and Laake, 1983; Buckland and Anganuzzi, 1988). Recent analyses suggest that the tuna vessel observer data are unlikely to provide a reliable estimate of dolphin status because of time-varying biases present in the data resulting from changes in vessel search methods, for which the details remain unknown, and because estimated trends may reflect changes in the tuna-dolphin association rather than changes in the absolute abundance of dolphins (Lennert-Cody *et al*., 2001, 2016). Since the last NMFS survey in 2006, no reliable indicators are available to assess the current status of ETP dolphins. Therefore, the goal of the workshop was to identify data types and methods of analysis, both conventional and novel, for monitoring and assessing ETP dolphin status. To accomplish the meeting goal, the following questions were to be addressed: if another fishery-independent, ship-based survey could not be conducted, what other methods could be used that would produce an estimate of abundance with a CV comparable to that from previous line-transect methods; are there new methods that could provide future abundance estimates at lower costs than previously used methods; are there methods that should be used in tandem to provide complementary information; and if another fishery-independent, ship-based survey could be conducted, could the methodology be improved without reducing the comparability with the historical time series of population estimates?

Dr. André Punt chaired the workshop and Mrs. Kelli Johnson acted as lead rapporteur. Invited participants included experts in line-transect and mark-recapture (M-R) surveys, abundance estimation, population modelling, imagery, tagging, genetics, and life-history data (Appendix A). The workshop was also attended by observers (Appendix A). This report summarizes discussions among the invited participants regarding the meeting goal and proposed short- and long-term plans necessary for its achievement. Appendix B lists the meeting agenda, Appendix C lists the background documents developed for the workshop, and Appendix D provides abstracts of the presentations made at the workshop. Key aspects of the discussions related to data collection are summarized in Section 2, analysis in Section 3, and modelling population dynamics in Section 4. Section 5 summarizes discussions regarding ways to conduct future research to better understand the abundance and population dynamics of ETP dolphins, with a focus on the following three areas: ship-board line-transect surveys; unmanned aircraft to conduct strip- or line-transect surveys of the ETP; and mark-recapture-based monitoring.

2. Data

Past studies have provided abundance estimates from fishery-independent, ship-based surveys; relative abundance estimates from tuna vessel observer data; life-history data from sampling conducted by observers aboard tuna vessels and researchers aboard fishery-independent, ship-based surveys; and movement and stock-structure data from tagging studies. It has become increasingly difficult, however, to collect data on ETP dolphin populations due in part to changes in funding and availability of infrastructure for ETP research and because reductions in dolphin bycatch have limited the opportunities for life-history sampling. Many methods were used to collect historical data, and technological advances offer several new ways to collect data. The workshop discussed the feasibility of collecting survey, tagging, genetics, and life-history data to estimate absolute or relative indexes of abundance for ETP dolphins. It is particularly advantageous to continue research that builds on the time series of existing data or that could be used in a model in conjunction with previously collected data. It is important that future surveys be designed to include all components of a given stock because estimates of abundance and population parameters will be biased if stocks are outside of the study area during times of sampling.

2.1 Fishery-independent, ship-based survey data

Historically, fishery-independent, ship-based line-transect surveys (Gerrodette *et al*., 2008) were used as the primary source of information for estimating the abundance of ETP dolphins. These surveys were initiated in 1974 by the NMFS, but only data from surveys conducted during 1986-1990, 1998-2000, 2003, and 2006 were used for the most recently available estimates of abundance because of a lack of standardized stratification and sampling procedures during previous years. The time series could be continued to provide comparable estimates of absolute abundance, even if different vessels were used. If tuna vessels were to be outfitted as research vessels, research on the behavioural responses of dolphins to tuna vessels (e.g., Pryor and Norris, 1978; Lennert-Cody and Scott, 2005) would need to be explored.

Estimating group size (i.e., the number of dolphins present in a school or group) from a ship-based survey is difficult, but critical to obtaining accurate estimates of abundance. Social groups of ETP dolphins are extremely ephemeral and are known to break up daily. Group sizes range from just a few to thousands of individuals, and group size fluctuates throughout the day (Scott and Cattanach, 1998; Scott and Chivers, 2009). On average, NMFS marine mammal observers aboard fishery-independent, ship-based surveys (hereafter referred to as NMFS observers) tend to underestimate group size (Gerrodette *et al*., 2002, in prep.). NMFS observers should therefore continue to independently provide not only their "best" estimate, but also estimates of the maximum and minimum number of dolphins present for each group. Providing ranges allows for the estimation of variance within and among observers, which tend to be high. Group size estimates from NMFS observers have been validated/calibrated by comparing them with counts from vertical aerial photographs of the schools taken from helicopters. In the future, unmanned aircraft that can easily be deployed from the ship (i.e., short-range "drones") offer a means to capture digital still images of groups to corroborate species identifications and group size estimates made by ship-based observers. Additionally, images can be informative about "availability bias" ($g(0)$; see below) when captured in tandem with human observers because humans can capture behavioural information, which is known to contribute to imagery availability bias.

Previously, it was assumed that ship-based surveys detected all dolphin groups >20 individuals on the trackline (i.e., $g(0)$ was assumed to be unity; Barlow, 1995). However, this assumption has recently been called into question (Barlow, 2015). In theory, issues with respect to estimation of $g(0)$ could be informed by operating with independent observers at multiple heights on the vessel (Okamura *et al*., 2003), but NMFS attempted such a procedure during the 1998 survey for dolphins in the ETP and concluded that observers located higher on the vessel than traditional observers failed to detect groups appreciably sooner. To avoid biased estimates of abundance (e.g., Barlow, 2015), it is critical to evaluate whether $g(0) < 1$ for ETP dolphins, and this might be done during future surveys using helicopters or drones operated from the survey ship. In the future, ship-based survey design changes could be made so that $g(0)$ could be estimated, but this may lead to incompatibility with the historical abundance time series.

Ship-based, line-transect surveys are not limited to collecting dolphin sightings, and can collect environmental data and sightings of non-target species. Variability in the ability to detect subsurface animals is more of an issue for aerial visual surveys than for ship-based visual surveys because aerial platforms may allow observers to see into the water column, whereas ship-based observers typically detect animals (or other cues) above the surface of the water. Furthermore, the depth at which an aerial observer can detect an animal depends on many factors, including presence, location, and type of glare; turbidity; sea state; animal coloration; animal size; animal orientation; group size; etc. The ability to collect information on sea state, turbidity, water temperature, and other factors known to affect sighting rates, currently, is easier from a ship-based survey than from an aerial-based survey. Remote sensing datasets can be used to augment environmental data collected from surveys.

Given that the estimated error of encounter rate is larger than the estimated error for detection, $f(0)$, and group size (Gerrodette *et al*. 2008), adaptive sampling designs might be considered in the future to try to reduce the overall error associated with abundance estimates. Oceanographic information might be useful in this regard. However, the return with respect to reduced CVs of abundance estimates is likely to be modest (Pollard *et al*., 2002), especially given the increased work required to develop and implement adaptive designs.

In principle, tuna vessels, properly outfitted with marine mammal survey equipment, could operate as research vessels to collect ship-based line-transect data using a randomized survey design (see Section 5 and Background Document 2). Although it might be possible that tuna vessels could operate in multiple modes during a fishing trip (e.g., fishing and survey), it is most likely that in the future any line-transect survey data collection from fishing vessels should be limited to when fishing vessels are operating solely as research vessels because of logistical constraints. For instance, the challenge of transferring specially trained marine mammal observers among multiple tuna vessels for short survey sections could be considerable and would be avoided if tuna vessels operated in only survey mode for an entire trip.

2.2 Tuna vessel observer data

Use of tuna vessel observer data to assess dolphin stock status will be problematic. Recent analyses of these data (see Background Document 1) have revealed multiple problems with using fishery-dependent data to estimate dolphin abundance (Lennert-Cody *et al*., 2001, 2016; Ward, 2005), including changes in the data consistent with a temporal evolution of searching methods and effects of tuna vessels targeting dolphin groups associated with tunas on estimating both encounter rate and dolphin group size. Therefore, any future investigations with tuna vessel observer data should be limited to the use of resulting data as a relative index within a population dynamics model, not as a standalone index of dolphin stock abundance. If these data were to be used in population dynamics models in the future, an extensive survey of tuna vessel fishing captains should be conducted to provide information with which to try to model temporal changes in tuna vessel search behaviour, although this would not alleviate problems caused by tuna vessels targeting tuna-associated dolphin groups.

In the future, presence-absence information collected from tuna vessel observers could be informative for calibrating satellite images given that at this current time it remains unknown if dolphins can be detected from satellite images and species identification is not possible.

2.3 Acoustic data

Passive acoustic systems, specifically towed hydrophone arrays and drifting vertical hydrophone arrays, offer a potential means to estimate abundance of ETP dolphins. Acoustic systems have the advantage that detection is largely independent of sea state and weather. However, at present, these methods cannot be applied to ETP dolphins because of limitations in correctly identifying species and estimating group size from dolphin vocalizations. Furthermore, dolphin call rates are affected by social behaviour, and, therefore, the social aspects of call rates must be taken into account, but largely remain unknown. In theory, a group-based line-transect method could be applied to acoustic data to estimate absolute abundance, if a separate platform could provide estimates of group size. Towed arrays, for which range estimation has been well tested, have the complication of causing changes in behaviour when individuals respond to the approaching vessel, thereby limiting their use for abundance estimation. Drifting vertical arrays (for which range estimation has not yet been tested) could potentially be used to provide precise estimates of relative abundance using the density of acoustic cues, given assumptions about calling rates, but this method is still in the proof-of-concept (PoC) stage for dolphins.

Acoustics can help provide information about biases inherent in visual survey methods. In principle, acoustic data may provide information on $g(0)$, particularly when sighting conditions are less than ideal (e.g., Beaufort sea states > 4) because acoustic methods can detect calls independent of sea state and depth.

2.4 Aerial-based survey data

Manned aerial surveys with high-resolution imagery have provided data used to obtain abundance estimates for other marine mammal species and, theoretically, are feasible in the ETP. Methods for conducting aerial-based surveys are established, but study designs must tradeoff between maximizing sighting effort and minimizing availability biases (see additional comments below). Manned aerial surveys with observers are not discussed herein in detail because they are considered to be impractical and may be dangerous for a survey area as large as the ETP.

Aerial surveys can collect high-resolution imagery data as digital video footage or digital still images. Digital imagery offers the benefits of providing a permanent sighting record and equal detection efficiency across the imagery. Digital technology is rapidly developing, where the use of the best equipment is limited only by funding. For instance, high resolution video cameras, currently being used in the United Kingdom, can sample the ground at a resolution of 2 cm from an altitude of ~550 m (Webb *et al*., 2015).

Combinations of camera type, camera placement, camera resolution, and flight altitude offer the ability to change the strip width and ground resolution. The best combination of these parameters will differ among species and would depend on typical weather conditions, and will require investigation in the ETP. For instance, it was found that using an oblique versus horizontal camera angle allows for increased precision in measurements used for species identification (Webb *et al*., 2015), but a decreased ability to distinguish colors (Chabot and Francis, 2016), which is necessary for automated-detection algorithms to identify animals. Nevertheless, the effective strip width of an aerial-based survey will generally be smaller than that from a ship-based survey, though aerial-based surveys travel faster and can cover a longer trackline in a given day than ship-based surveys.

Digitally recorded sightings may lead to more accurate species identification and estimation of group size, but suffer from long post-processing times. As many as twenty human hours may be needed to post-process a single flight hour, which includes time to identify sightings, identify species, and estimate group size. Post-processing times can be reduced by using automated-detection software. Currently, such software is inaccurate and detection capabilities are inversely related to post-processing speed. Future software development should focus on more accurate identification of potential sightings to reduce the amount of digital content that needs human review and development of software to post-process video footage, for which software currently does not exist. Using automated-detection software necessitates the need for an additional correction factor for missed sightings because no post-processing software can match human detection efficiencies, but data to inform this correction factor are typically not collected.

Availability bias is more complicated for aerial-based than ship-based surveys because some unknown and variable proportion of subsurface individuals will be undetectable from the air (Marsh and Sinclair, 1989). *In situ* methods for measuring availability biases from video-captured sightings exist (Teilmann *et al*., 2013), but are untested, and no *in situ* method exists for digital still sightings. Data on dive profiles from telemetry studies (e.g., Scott and Chivers, 2009) can be used to calculate the amount of time individuals spend close enough to the surface to be detected in the clear waters of the ETP, allowing availability biases to be calculated given assumptions about behavior (Webb *et al*., 2015). Further research is needed on how biases vary with platform, species, lighting conditions, water turbidity, sea state, and observation angle. Additionally, the question of how to sample environmental conditions that affect availability bias during aerial-based surveys will need to be investigated

Many trade-offs exist between manned and unmanned aerial surveys. Some unmanned aircraft can cover more territory than manned aircraft before needing to refuel. Impediments due to flight duration may not be an issue for aircraft that can be refueled at sea (e.g., helicopters). If unmanned aircraft are sent out over multiple days, footage collected during sea states greater than Beaufort four and poor lighting conditions will likely be unusable and the study design and analytical methods should accommodate hours when the aircraft will not be in survey mode due to weather. The use of unmanned aircraft necessitates a prior air traffic study, which, in the case of the ETP, will involve the air spaces of multiple countries and the use of technology to avoid collisions with other aircraft (e.g., most of the international purse-seine fleet search with helicopters). Some unmanned aircraft will be capable of adaptive sampling, i.e., those that can carry automated-detection software and are able to change course in real time, but the efficiency of automated-detection software may limit the ability to perform adaptive sampling. Uncertainty about the abundance estimates has been found to be higher for unmanned than manned aircraft in arctic surveys due to differences in sample sizes (Ferguson, pers. comm.). Finally, safety concerns and costs differ between the platforms.

The ability to sight animals in inclement weather will be limited from any type of aircraft, and extreme conditions may prevent aircraft from flying at all. For instance, thunderstorms, which produce known safety threats, are common in parts of the ETP. Aircraft will be vulnerable to turbulence, and, currently, no small- to medium-sized unmanned aircraft can operate in sea states higher than Beaufort four. Conversely, data from ship-based surveys operating in Beaufort sea-state five have historically been included in abundance estimates. Ultimately, detections will always be limited by the quality of the images, which is known to be affected by camera angle, weather, turbidity, and glare.

2.5 Satellite imagery

Satellite imagery provides the benefits of covering large areas, having access to almost all areas, being cost effective when agreements with providers are pre-arranged, being non-invasive, and not requiring permits for data collection, which can be logistically challenging to obtain for other data types. However, to date, the method has been tested only for large cetaceans (Fretwell *et al*., 2014). Satellite images by themselves cannot provide the data needed to estimate absolute, or relative, abundance of ETP dolphins given its currently limited resolution (greater than or equal to 31 cm), inability to provide reliable data, particularly in less than ideal sea states, and inability to see through clouds.

Satellite images may offer information on presence/absence, and could be used to fill in the gaps regarding the stochastic, non-homogeneous distribution of ETP dolphins. Images could be requested year-round to help design ship-based or aerial surveys. Multiple days would be needed to obtain the images and process them (i.e., convert to true colours) before they could be assessed for the presence or absence of marine mammals. Satellite images also offer the potential to provide information on missed detections from other platforms should future images be obtained or catalogued. Using satellite images for calibration would be applicable only for future studies because presently images are not acquired unless requested.

Infrared satellite images are also available, but are not a viable tool because of their low resolution. Furthermore, the ability to detect marine mammals in infrared imagery depends on temperature differences between the environment and the target animal.

2.6 Tagging data

Multiple tagging methods have the potential to provide data necessary to inform recapture rates used in M-R estimation methods. Each tag type has associated advantages and disadvantages. The longer the life of the tag the more information that can be gained for M-R-based inference into movement, survivorship, or abundance. Fortunately, in the case of physical tags, manufacturers have been willing to develop solutions to specific problems such as the need to modify attachment hardware or trade-offs between battery life and battery size. In theory, externally placed tags could be monitored using Argos satellites, VHF, acoustic receivers located on the purse-seine vessel or net, or physical retrieval. All monitoring systems proposed to be placed on fishing nets must be rugged enough to withstand the purse-seine net retrieval process (e.g., passing through the power block). Furthermore, both the numbers of unmarked individuals and tagged animals must be counted if monitoring for tagged animals were to occur during the back-down dolphin release procedure.

The main impediments to using physical tags to estimate recapture rates and subsequently the abundance of ETP dolphins are the large sample sizes required, potentially high, but unknown, tag loss rates, and the difficulty of tagging a representative sample of the population. Prior to conducting any tagging experiment, simulations should be used to calculate the sample sizes needed to provide estimates of abundance with CVs similar to those obtained from previous ship-based line-transect surveys. Necessary sample sizes will more than likely be large and tag-specific. Consequently, to obtain the sample sizes needed, tags would more than likely be deployed from fishing vessels, and, therefore, the sampling design of any tagging study would need to take into account the fact that tagged individuals, not just the tag recoveries, may not represent a random sample of the population.

Genetic samples can also be used for M-R analyses to estimate abundance. As with physical tags, the required sample sizes to estimate abundance would be large for populations the size of those in the ETP. Close-kin genetics, however, would require fewer samples than standard M-R genetics because the close-kin analysis can take advantage of information on relationships among individuals (e.g., parent-offspring, half siblings, and grandparent-grandchild). It would take approximately twenty years to collect a sufficient number of samples, given the current bycatch rate if tissue samples came only from dead animals from the fishery bycatch. A sufficient number of samples may be obtained in a much shorter time period, perhaps five years or so, if live animals could be sampled using "biopsy poles" by researchers or fishing crew.

Genetic samples can provide information on biological and ecological characteristics of ETP dolphins. However, information gained from genetic samples on life-history characteristics depends on the amount of sample tissue collected and the processing method used. Consequently, if samples are collected by crew members, which necessitates using a "simple" method for on-board processing (e.g., a formalin solution), more individuals could be sampled than if samples were only taken by trained observers. However, trained observers using complex at-sea processing methods may increase the utility of the samples for future studies. Sampling may add more time to the fishing operations, but the data would have the added benefit of being informative about life-history characteristics and stock structure. Research and funding are needed to design an archival system such that samples of adequate mass would be available for future analysis should they become part of other studies or new genetic M-R methods be developed.

The logistical effort required to tag or sample tens of thousands of dolphins in the ETP is daunting. Physical tags were considered less feasible for large-scale sampling than genetic M-R methods. Should a tagging study be initiated, it will require tags that are easy and rapid to apply, with high probabilities of detection (either visually or electronically), and low and known rates of tag loss.

2.7 Life-history data

Life-history data were collected from more than 43,000 individual dolphins killed in the ETP yellowfin tuna purse-seine fishery between 1966 and 1994, providing information on biological parameters such as somatic growth and reproductive rates. These data are fishery-dependent and were collected year-round mirroring the fishery, which exhibits noted spatial shifts in distribution of effort within a year. When the collection of these data ended in 1994, additional biological data were collected during NMFS research surveys using non-lethal techniques such as biopsies and photogrammetric methods. If bycatch sampling were to be reinstated ~350 samples, including ~50 from mature females, could be collected annually. Life-history data, if data collection were to be reinstated, could be used to evaluate whether estimates of population growth rates from newly developed population dynamics models (see Section 4) are reasonable given the currently available information on reproductive rates.

2.8 Permitting prior to data collection

Regardless of the method used to collect data, scientists must be aware of the, sometimes lengthy, permitting process that must be undertaken prior to initiating data collection because the ETP contains many countries. Data collected from manned or unmanned aircraft would require the necessary research permits for collecting data on marine mammals and authorizations to enter airspaces. The collection of data using satellite imagery was the only method considered during the workshop that would not require a permit.

2.9 Evaluation of data collection methods

Tables 1-3 provide an overview of advantages and disadvantages of each data type considered during the workshop. These tables summarize the purpose for the collection of each data type, the status of the methods that could lead to estimates of abundance or trend, and the advantages and disadvantages of each data type. The categories of status range from "Established," where data collection and analysis procedures exist and have been applied to ETP dolphins, to "Proof of Concept" ("PoC"), where at least some aspect of data collection and analysis would require research and development prior to implementation. It is noted that new estimates of abundance, e.g., based on incorporating corrections for $g(0) < 1$, may result in higher abundances, which would mean that M-R sample sizes larger than those projected in these tables from existing abundance estimates would be required.

3. Statistical methodology

Previously, trends in abundance have been assessed using data collected by fishery-independent ship-based line-transect surveys (Gerrodette and Forcada, 2005; Gerrodette *et al*., 2008), observers aboard tuna vessels during normal fishing operations (Lennert-Cody *et al*., 2016 and references therein), and a combination of the previously mentioned data sources, along with estimates of incidental fishing mortality (Hoyle and Maunder, 2004; Wade *et al*., 2007). In general, new approaches could lead to improved field and analysis methods, which may lead to benefits in terms of more accurate and precise estimates of abundance. However, substantial changes in field methods could introduce time-varying bias into any abundance time series that includes the historical estimates, unless the new field methods are calibrated against the old. In contrast to the previous section that focussed on data collection and field methods, this section focuses on methods to analyse the data that can be utilized to minimize bias and variance by accounting for various factors in the analysis or improving the statistical design of data collection.

Although the fishery-independent, ship-based line-transect surveys are costly, continuing these surveys for some period of time would ensure a means for evaluating existing assumptions, as well as validation for any new methods under development. This would be particularly valuable for research and development of methods in the PoC stage (Table 2) that may prove successful. Regardless, the likelihood of any method providing an estimate of true absolute abundance is questionable. For this reason, proposed methods should be designed to produce estimates of abundance that are as close to absolute as possible and with a CV equivalent to or less than previously used methods (e.g., for the northeastern stock of offshore spotted dolphin, the most recent five surveys had CVs around ~0.15-0.20; Gerrodette *et al*., 2008).

3.1 Line-transect methods

Line-transect methods for estimating abundance can accommodate distance sampling data collected by a variety of platforms, including observers aboard research or tuna vessels, manned aircraft with observers, and various types of unmanned aircraft with high-resolution imagery (although for the latter, these are technically strip transects). Discussions on reducing bias and variance of estimates from line-transect data focused on methods for data collected using research-vessel surveys because this is the source of historical absolute abundance estimates.

One of the primary sources of bias discussed was that which can arise from an invalid assumption of perfect detection on the trackline (i.e., incorrectly assuming $g(0) = 1$). Previous survey estimates of abundance assumed $g(0) = 1$ (Barlow, 1995; Gerrodette *et al*., 2008). However, recently that assumption has been called into question based on analyses of Barlow (2015), which indicate that $g(0)$ might be appreciably below one except for times during the best sighting conditions; i.e., there may be a reduced window during which a dolphin group is available for detection in poorer sighting conditions, especially when taking into consideration responsive movement with respect to the survey vessel. With the existing survey data, bias corrections might be achieved following the methods of Barlow (1999, 2015). In the future, modifications to Horvitz-Thompson-type estimators for double-platform data (e.g., Buckland and Turnock, 1992) offer one way to address imperfect detection on the trackline. An example of such a modification is provided by Borchers *et al*. (1998), who extended the approach of Buckland and Turnock and used a logistic regression model to estimate the probability of detection as a function of covariates. Double-platform data could be collected in the ETP in the future with a sampling design that included a drone or helicopter operating ahead of the survey vessel, or acoustic data coupled with the ship-based visual survey data, although for the latter responsive movement may become a much greater issue.

Another issue is that the precision of group size estimates varies with specific covariates, yet these covariate effects on precision are not taken into account in the estimation of abundance. Observer estimates of group size have been shown to be highly variable (Gerrodette *et al*., 2002), and some of this variability might be attributable to specific covariates that have already been measured as part of the survey data collection process. Whether the use of "uncorrected" group size could lead to a large amount of bias in the estimates of abundance depends on the magnitude of the error in group size and the extent to which the effective strip width depends on the true group size. This source of bias might be minimized by taking into consideration the distribution of uncertainty about observed group size, as a function of covariates, when computing the Horvitz-Thompson-like estimator of abundance (Borchers *et al.*, 1998). In other words, using the expectation of group size in the numerator of the Horvitz-Thompson estimator and using the conditional expectation of effective strip width in the denominator, where in both cases the expectation is taken with respect to the estimated distribution of group size for each covariate combination. Another option for adjusting the estimate of effective strip width for uncertainty in group size would be to estimate the detection function using an errors-in-variables type of model.

The estimate of error associated with the existing abundance estimates might be improved in several ways. First, the precision of the estimate of $f(0)$ might be increased by pooling data from multiple species to estimate the shape of the detection function. This can be done by using multiple covariate distance-sampling methods (Buckland *et al*., 2004) to jointly model data from different species with species as a factor in the detection function model (e.g., Barlow *et al*., 2011). However, the largest source of variance in estimates of abundance is due to encounter rate, not $f(0)$ (Gerrodette *et al*. 2008).

Furthermore, the current estimates of variance about the estimated abundances could be improved if the variance components could be further decomposed based on their source. In addition to the variance components attributable to encounter rate, effective strip width (including $g(0)$ uncertainty), and group size, there is uncertainty due to the following sources: measurement error, calibration factors, and process error. Estimating these other sources of error and incorporating these estimates into the estimated abundance error would lead to more realistic estimates of overall uncertainty. It could also improve understanding of the main causes of uncertainty and provide information relevant to the design phase of future surveys, potentially reducing future uncertainty.

Finally, encounter rate modelling perhaps merits more attention, especially in light of recent developments in spatial distance sampling methods (e.g., Yuan *et al*., submitted) because encounter rate is currently the greatest source of variability in the estimates of abundance. In the future, spatial modelling of survey data collected from adaptive sampling designs may result in greater precision because survey effort could be directed to areas of better dolphin habitat, perhaps reducing the variance associated with estimated encounter rate (if such areas can be detected and tracked over time). This might be achieved using adaptive sampling designs informed by near real-time oceanographic conditions, for example. However, improvements in precision with adaptive sampling designs are expected to be modest.

3.2 Mark-recapture methods

Statistical methods for mark-recapture data that account for non-random recaptures would need to be developed for ETP dolphins. A large number of individuals would need to be marked for M-R methods to be of use for ETP dolphins (i.e., produce an estimate of abundance with a CV comparable to that from line-transect methods). Realistically, sufficient recaptures may only be possible through the identification of individuals during the back-down procedure performed by tuna vessels during fishing on tunas associated with dolphins. Any tagging study that relies on fishing vessels for recaptures may have a non-random sample of recaptures, and, therefore, animals must be marked randomly. Analytical methods to account for the non-random recaptures have been developed for other species, but have yet to be developed for ETP dolphins.

3.3 Composite methods

Ship-based surveys have a high cost, and, therefore, statistical methods for estimating abundance that can combine data from different platforms into an estimate of absolute abundance should be investigated. For instance, spatial modelling with sightings data from multiple platforms, as well as other covariates, may reduce estimation uncertainty compared to estimates of abundance from a single data source. Several hypothetical examples for future consideration include annual satellite surveys with occasional ship-based, fishery-independent surveys; ship-based, fishery-independent surveys with a drone as a tracker platform; acoustic surveys with good spatial coverage combined with high-resolution imagery in a model-based spatial analysis; and, tuna vessel observer data combined with ship-based, fishery-independent survey data in a model-based approach.

Genetics and life history data can help to improve population modelling if they were collected. Genetic data can estimate mixing proportions to inform stock structure assumptions and design-based survey protocols. In addition, life history and genetic information regarding stock structure could be used in M-R abundance models. Life-history data can provide age and reproductive inputs for population modelling. Finally, although it remains to be proven, genetic data have the potential to provide information on ages.

Statistically rigorous designs to collect data for composite estimation methods, including sample size requirements, and methods to appropriately summarise the data for use in the population dynamics models, need to be developed. The PoC field trials (Section 5) could provide “pilot study” data sets with which to develop sampling designs.

4. Cetacean stock assessment models

Population dynamics models are required as filters of the available data to yield inferences about quantities or questions of management or scientific interest (Table 5). The required features of the model depend on the data to be used and on the questions of interest. For example, a population model needs to include individual life-history and movement processes to use M-R data. Model complexity ranges from simple exponential trend models that ignore density-dependence to complex multi-stock age-sex- and stage-structured models that form the basis for management strategy evaluations.

Highly significant and complicated patterns of heterogeneity in the sampling process used to collect data available for population dynamics models makes it challenging to identify the relatively weak signals from population processes against the background of strong heterogeneity effects. In the case of survey data, most of the pre-analysis to cope with heterogeneous detection rates can be performed external to the population dynamics model, generating “cleaned up” abundance estimates or indices that can be used as input into a population dynamics model. These abundance estimates will be the primary source of data for modelling population dynamics, although a variety of other data types, including relative abundance indices and M-R data can be included. In general, at least one estimate of absolute abundance is needed for parameter estimation because there is a lack of catch-induced declines in abundance captured by indices of relative abundance. Data on fleet-based catches also represent an important source of information.

Most models are deterministic, but variation in cohort strength must be accounted for with species that are relatively short-lived. Additionally, variation in cohort strength must also be accounted for if age- or length-composition data are included in the model, although such data are rarely available. Most analyses assume density-dependence impacts on calf survival (which implicitly includes maturity and pregnancy rate), but it could also impact the survival-rate of adults or age-at-maturity. The models differ in terms of whether the population projections start when substantial catches first occurred or whether allowance is made for time-varying carrying capacity by starting the model in a more recent year. Female cetaceans seldom have more than one calf per year, which limits the variation in calf numbers and places an upper (but not lower) limit on the recruitment rate.

It is important to include both demographic and environmental variability for stocks that are at low abundance. Interactions between environmental variability and density-dependent effects can lead to populations that are more variable when they have recovered from past depletion, and, therefore, constant-K models will eventually show a lack of fit given a long enough time series. Simulation studies show that fitting constant-K models when K is time-varying can seriously bias estimates of mean productivity (r) and K. Consequently, it is most appropriate to allow parameters such as K to vary through time as a stochastic (and potentially auto-correlated) process. Assuming such parameters are constant will lead to biases, and relationships between measureable environmental variables and biological parameters are likely to break down with time.

The future for population dynamics models for dolphins will likely involve multi-stock models that include age-, sex-, and spatial-structure fitted as state-space formulations. At present, such models are often too computationally intensive to be feasibly implemented or there is insufficient information in the data to estimate the parameters representing all the processes. Consequently, models must be simplified, with the result that the performance of some methods need to be better understood, including through simulation testing. Uncertainty about the results can be quantified using Bayesian methods, which allow information on biological parameters, particularly r and K, to be included in the analyses. Alternatively, bootstrap or asymptotic methods could be used. For most models, leave-one-out validation processes are limited by a lack of yearly data (on for example abundance), such as the case for ETP dolphins.

It was recommended that the available data for ETP dolphins be re-analysed to provide updated estimates of abundance and trend, even though fishery-independent surveys have not been conducted since 2006. An updated assessment model could include model-based, instead of design-based, estimates of absolute abundance that include a correction for imperfect detection on the trackline and estimates of pregnancy rates from photogrammetric data. The incorporation of corrections for $g(0) < 1$ should lead to higher estimates of abundance. Furthermore, results from age-structured models with stock structure could be compared to results from simpler model formulations to determine the benefit of added model complexity. Most importantly, all available data should be included in a single, updated population dynamics model ensuring that population estimates are based on all available data sources.

5. Proposed research

The workshop focused on the following three methods for estimating abundance: ship-based line-transect surveys, M-R studies, and aerial-based survey approaches. Of the three projects, the ship-based line-transect survey is the method most based on established methods, while research and development would be needed to implement the remaining projects. Section 5.4 outlines a project to estimate tag-loss rates that could help assess the viability of M-R studies and a project to re-initiate the collection of life-history data that could be used in population modelling, but these were not discussed in detail during the workshop. Costs are provided for the all projects, but these are rough and would need refining. In addition, the costs are related to obtaining estimates of abundance with CVs of ~0.15-0.20. The workshop did not assess whether such CVs were sufficient for fully addressing questions of management importance.

5.1 Ship-based line-transect surveys

For reasons of comparability, future ship-based line-transect surveys used to estimate dolphin abundance in the ETP should use the same field methods as the NMFS surveys carried out prior to 2007, i.e., two ships for 120 sea days each, or a total of 240 sea days, with a rotating team of three observers using 25X binoculars at an eye height of approximately 10 m. Surveys carried out in this manner can produce estimates of abundance with CVs of ~0.15-0.20 for all ETP dolphin stocks of interest. It was suggested that radar capable of detecting seabird flocks (as used on purse-seine vessels) might assist in studies of responsive movement of dolphin groups, but a person with experience using radar in this way would be required. Care would have to be taken to ensure that the survey design would be comparable with that of previous surveys.

Several survey-design issues were identified that should be addressed before the initiation of a ship-based survey. The area to be covered by the survey, and the stratification of effort within that area, should be reviewed. Neither the area nor the stratification need be identical to previous surveys, but the benefits of any changes should be carefully weighed against the costs of decreased comparability. Adaptive sampling, possibly aided by satellite imagery, could also be considered, but, again, the potential benefits should be weighed against costs of decreased comparability. In light of Barlow (2015), which estimated that an appreciable fraction of dolphin schools are missed on the trackline, a future cruise should be conducted to better understand the factors underlying $g(0)$. Acoustics, bird radar, and drones might all contribute to a better understanding. The parameter $g(0)$ is central to unbiased estimation, and, therefore, dedicated experiments during the cruise, or even a separate cruise with a helicopter, might be needed.

Other valuable scientific data not directly related to estimating dolphin abundance could be collected during ship-based surveys. For example, data could be collected on turtles, seabirds, other cetaceans (using line-transect methods and passive acoustics), and marine debris and drifting acoustic buoys could be deployed and/or retrieved. Except for line-transect data on other cetacean species, some of these ancillary projects would require additional crew, for which the costs are not part of the included, rough budget.

General estimates of the costs of a ship-based survey were given in the workshop Background Document 1. The included budget, which is based on an estimate of NMFS surveys costs for one year in 2017 U.S. dollars (made publically available by Cisco Werner and Lisa Ballance on July 15, 2016), encompasses data collection, checking, and archiving; the budget does not factor in costs pertaining to the analysis of the data. The estimated total is \$9.4M, of which 70% is ship costs. If ship time were donated or provided at a reduced rate, the costs would be reduced substantially. The presentation of a NOAA-based budget for an ETP survey does not imply that NOAA would or should conduct future surveys, only that NOAA-based cost estimates were readily available. Similarly, the indicated levels of NOAA in-kind support for past cruises does not mean that NOAA has offered such support for future cruises. Research generated from the above proposal would provide an estimate of current abundance after the collection of one year of survey data, where the estimate could be compared to previous estimates of abundance, generating an estimate of the current trend.

5.2 Mark-recapture surveys based on genetic methods

Genetic M-R provides the least infeasible option among M-R methods because of the logistical difficulty of physically tagging tens of thousands of dolphins. A 5-year program would target a sample size of 50,000 dolphins (based on the rule-of-thumb: 20 $\sqrt{N}$ per stock). Approximately 30 animals could be sampled per set using biopsy poles inside the purse seine by sending two additional scientists aboard 10-12 fishing trips each year. However, it would be better to collect data from more trips to attempt to mark a more representative sample of the population. With about 30 sets per trip, about 10,000 samples could be collected annually. Sampled trips would need to be chosen to spread effort around the fishing grounds, seasons, and stocks. A two-stage analysis would be conducted: first, the genetic sample would be used to determine stock structure and identity; and second, genetic samples would be used to identify individuals and calculate M-R abundance estimates. Close-kin analyses could also be used to estimate population size (e.g., Bravington *et al*., 2014). Table 4 provides an approximate CV prognosis by stock (northeastern and western/southern spotted dolphins; eastern and whitebelly spinner dolphins) by year. This two-stage project would provide information on stock structure and abundance. Morever, survival and population trends could be estimated if sampling occurred over multiple years. An ancilllary benefit of the project would be the collection of biopsy samples that could be used for other studies (e.g., reproductive hormones, stress hormones, pesticides, and trophic levels from stable isotopes).

The sample size of 50,000 dolphins is predicated on random sampling, and a larger sample size may be required to ensure that geographic areas and all stocks are sampled representatively. Simulation analyses could be conducted prior to sampling to determine the representativeness of several sampling designs, though this cost was not determined. Also, there may be no way to guarantee that biopsies from animals associated with a fishing net will be representative of the population no matter how trips are selected for samples. The following additional logistical issues should also be addressed before individuals are tagged: the chosen purse-seine vessels must have space for two extra personnel; the anticipated sample size would exceed current storage capacity and analysis capabilities, requiring new infrastructure and more staff; and biopsy sampling would increase set time by about 20 minutes. Delays of releasing dolphins for tagging purposes would have to be balanced against the possibility of mortality.

Annually, it is estimated that there would be \$200K in field expenses, \$600K in laboratory expenses, and \$200K in overhead expenses. The total cost for the 5-year study would be \$5M, and an estimate of abundance would be available after the second year of data collection, though with a high CV (Table 4).

5.3 Drone-based aerial imagery

Drone technology is developing rapidly, but a number of key unknowns regarding their use in surveying dolphins would need to be addressed prior to their use. Most importantly, the probability of detecting a dolphin from aerial photographs varies with environmental conditions (sea state, cloud cover, water turbidity, sun angle, and glare), which are known fluctuate on the order of minutes to hours. Consequently, correction factors to account for covariate effects on variability in detection probability based on the target animal's depth must be developed before such data can be used for estimating abundance and trends.

It remains unknown if such correction factors can be estimated and if the precision of estimates will be sufficient for reliable estimation of trends in relative or absolute abundance. Therefore, the development of drones for the use of estimating dolphin status should be done in two phases.

The first phase (Phase I) would test the feasibility of estimating correction factors and provide estimates of their precision. A small hexacopter drone with cameras and a multi-spectral sensor, operated from a vessel, could be used to estimate the detection probabilities for dolphins (or dolphin-like objects) under a variety of environmental conditions. Hexacopters are likely to be more cost-effective than helicopters, but it would be imperative to use equipment that can be used during subsequent phases. Detection probability and dive profiles could be evaluated simultaneously if the feasibility study is performed using live dolphins. If, instead, the study were performed using a dolphin-like object deployed at known water depths, ancillary data on dive profiles for each species of interest would be needed to estimate the proportion of time dolphins spend at varying depths (e.g., Scott and Chivers, 2009). Using dolphin-like objects instead of live dolphins, which cannot precisely be controlled, would allow for a more in-depth assessment of how environmental conditions affect viewing conditions because it would omit variability in, and complications arising from, animal behaviour. Estimated costs of \$550-615K include in-kind contribution of ship time (\$0K), a study design workshop (\$40K), two hexacopters with multi-spectral camera and other primary and backup instrumentation (\$100K), two scientists for two months of field-based research and 10 months of analysis (\$400K), and the development of dolphin-like object (\$10K) or tagging study of target stocks (\$75K). Image processing time would likely contribute to a substantial amount of analysis time, unless automatic detectors could be developed.

The second phase (Phase II) would be contingent upon the success of Phase I, and would include a full-scale survey. The survey would need to be considered and designed separately from Phase I. One option might be a hired FlexRotor drone, which can fly for 40 hrs at 50 knots (~2,000 km range) and may be able to refuel aboard tuna vessels using helipads. Before its use, questions relating to air-traffic permitting and collision avoidance would need to be resolved. Costs will depend on design and technology. For instance, costs will increase proportional to the amount of area sampled. Estimated costs for a drone survey with 300 hrs of surveying plus image processing would be around \$1-1.5M to achieve the same coverage as ship-based line transect data for the ETP. However, if backwards compatibility to previous research vessel surveys is required, several years of concurrent drone and ship-based surveys would be needed, which might be prohibitively costly, depending on monitoring objectives.

5.4 Other projects and proposals

5.4.1 Tuna-vessel research surveys

Tuna vessels were suggested as an alternative to using research vessels for the collection of standard line-transect data. Prior to collecting data, tuna vessels would need to be modified to ensure their effectiveness as a survey platform. Data collection would be performed by trained observers aboard two vessels for several months. Limiting the survey to two vessels was proposed to limit the costs accrued from necessary vessel modifications and to alleviate the logistical practicalities of transferring observers among vessels while at sea. This constraint may need to be revisited if it is found that the survey design needs to be augmented to account for seasonality in dolphin distributions. For instance, an increased number of vessels could cover more area in a shorter time period (e.g., during the 2-month fishing closure), but would require modifying more vessels and training more observers. Costs would depend on contributions from the industry, which could depend on the study design.

5.4.2 Satellite imagery

Satellite images could be examined for their ability to identify dolphin groups in the ETP. Images would need to be examined in conjunction with data from comparative platforms such as tuna vessel observer data (although these estimates of dolphin group size are not calibrated) or survey data that could provide a more accurate estimate of group size. As a result, detection probabilities could be estimated for satellite images. Even though the images themselves would never provide enough information to estimate dolphin status, they might be used in conjunction with other platforms in the future to provide more accurate estimates of status. Estimated costs for a pilot study are $10K.

5.4.3 Estimation of tag-loss

Although the logistics are formidable for putting tags on tens of thousands of dolphins and keeping them on, new tag designs (e.g., Wildlife Computer Splash 10-268C satellite tag) are easier to mount than previous tags and have demonstrated longevity in the field. This study would test the ease and speed of attaching tags to dolphins encircled in a purse seine, tag longevity, and loss rate. Thirty tags would be mounted along the rear edge of the dorsal fin of spotted or spinner dolphins during 1-2 trips aboard fishing vessels. The locations of tagged dolphins and the fates of tags would be monitored remotely. Those tags that stopped transmitting prior to the estimated battery life could be assumed to be premature a tag loss. These tags would also report dive-depth information to the satellite.

In addition to estimating tag-loss rates, this project would provide information on the time it takes to tag multiple dolphins encircled in a purse seine, which would inform the practicality of a large-scale tagging program; depth profiles, which would be transferred in real time to satellites providing information relevant to $g(0)$; and habitat use, which could inform stock boundaries.

The total cost for a one-year study would be $220K ($120K for tags, $40K field operations and overhead, and $60K for the use of satellites).

5.4.4 Regular sampling of life-history data

Dolphins that have died during fishing operations can be sampled or collected by observers already aboard tuna purse-seine vessels. The IATTC and national programs presently place observers on all Class-6 vessels of the international tuna purse-seine fleet. Observers currently record body length, girth, sex, and spotted dolphin colour phase, when possible, but the re-initiation of life-history sampling of teeth (for age estimation), gonads (for reproductive analyses), and stomach contents (for food habits and trophic research) would provide added information relevant to assessing population status. This re-initiation of life-history research was approved by the Meeting of the Parties to the Agreement for the International Dolphin Conservation Program (IATTC, 2005).

The acquired life-history data would have many applications. Age distributions could complement future population dynamics models and provide information on current status if the relative vulnerabilities of different ages to capture were known (e.g., age distributions skewed towards old animals can be an indicator of future declines in population size). Information gained from gonads could provide reproductive rates, another important component of population dynamics models. Life-history data can provide information about population condition, although the data often need to be interpreted in light of other information such as current and historical mortality, environmental changes, and previous population estimates. Additionally, these data can assist in the interpretation of abundance trends. For example, life-history data can provide insights into trophic relations and environmental changes affecting population condition, as well as evidence of effects of climate change on populations through changes in food habitats.

The approximate sampling costs would be $255K per year for the first two years, with decreased costs in subsequent years. Sampling would need to be carried out over a long-term, continuous basis to gather an adequate sample size to facilitate comparisons with previously collected life-history data and to provide ongoing monitoring of the population. Additional funds of approximately $150K per year would be needed to process the samples.

Acknowledgments

Funding for this workshop was provided by the European Union and the Pacific Alliance for Sustainable Tuna. We gratefully acknowledge the Southwest Fisheries Science Center for providing the meeting facilities for the workshop.

References

Alvarez-Flores, C.M. 2002. Uncertainty in the management of activities affecting marine mammal populations. The tuna-dolphin conflict, a case study. PhD dissertation, University of Washington, Seattle, WA.

Au, D., and Perryman, W. 1982. Movement and speed of dolphin schools responding to an approaching ship. Fisheries Bulletin 80(2), 371-379.

Barlow, J. 1999. Trackline detection probability for long-diving whales. Pages 209-221 in G.W. Garner, S.C. Amstrup, J.L. Laake, B.J.F. Manly, L.L. McDonald and D.G. Robertson, eds. Marine mammal survey and assessment methods. Balkema Press, Leiden, The Netherlands.

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., Balance, 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., Grewe, P., and Davies, C. 2014. Fishery independent estimate of spawning biomass of Southern Bluefin Tuna through identification of close-kin using genetic markers. FRDC Report 2007/034, CSIRO, Australia.

Buckland, S.T., and Anganuzzi, A.A. 1988. Estimated trends in abundance of dolphins associated with tuna in the eastern tropical Pacific. Report of the International Whaling Commission 38, 411-437.

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

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

Chabot, D., and Francis, C.M. 2016. Computer-automated bird detection and counts in high-resolution aerial images: a review. Journal of Field Ornithology 00, 0-0. doi: 10.1111/jofo.12171.

Fretwell, P.T., Staniland, I.J., and Forcada, J. 2014. Whales from space: counting southern right whales by satellite. PLoS ONE 9(2), e88655.

Gerrodette, T., and 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., Perryman, W., and Barlow, J. 2002. Calibrating group size estimates of dolphins in the eastern tropical Pacific Ocean. Southwest Fisheries Science Center, Administrative Report LJ-02-08. 20 pp.

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 Tech. Memo. NMFS-SWFSC422, 1-45.

Hall, M. 1998. Ecosystem research and tuna fisheries management: some key questions. International Commission for the Conservation of Atlantic Tunas, Collective Volume Scientific Papers 50(2), 671-672.

Hammond, P.S., and Laake, J.L. 1983. Trends in estimates of abundance of dolphins (*Stenella* spp. and *Delphinus delphis*) involved in the purse-seine fishery for tunas in the eastern Pacific Ocean, 1977-1981. Report of the International Whaling Commission 33, 565-588.

Hewitt, R.P. 1985. Reactions of dolphins to a survey vessel: effects on census data. Biometrics 29, 29-42.

Hoyle, S.D., and Maunder, M.N. 2004. A Bayesian integrated population dynamics model to analyze data for protected species. Animal Biodiversity and Conservation 27, 247-266.

Inter-American Tropical Tuna Commission (IATTC). 2003. Antigua Convention. http://www.iattc.org/PDFFiles2/Antigua_Convention_Jun_2003.pdf

Inter-American Tropical Tuna Commission (IATTC). 2005. Agreement on the International Dolphin Conservation Program: 14th Meeting of the Parties. IATTC 20 October 2005 Minutes. http://www.iattc.org/PDFFiles2/MOP-14-Minutes-Oct-2005.pdf

Inter-American Tropical Tuna Commission (IATTC). 2006. Technical workshop on calculating N_{min} for the dolphin stocks of the eastern Pacific Ocean. IATTC Special Report 14, 35 pp. http://www.iattc.org/PDFFiles2/SpecialReports/IATTC-Special-Report-14ENG.pdf

Inter-American Tropical Tuna Commission (IATTC). 2016. Report on the International Dolphin Conservation Program, 34th Meeting of the Parties, La Jolla, California, USA MOP-34-05, 22 pp. https://www.iattc.org/Meetings/Meetings2016/Oct/Pdfs/MOP-34-05-Report-on-IDCP.pdf

Joseph, J. 1994. The tuna-dolphin controversy in the eastern Pacific Ocean: biological, economic, and political impacts. Ocean Development and International Law 25, 1-30.

Lennert-Cody, C.E., Buckland, S.T., and Marques, F.C. 2001. Trends in dolphin abundance estimated from fisheries data: a cautionary note. Journal of Cetacean Research and Management 3, 305-319.

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. Fisheries Research 178, 101-113.

Lo, N.C.H., and Smith, T.D. 1986. Incidental mortality of dolphins in the eastern tropical Pacific, 1959-1972. Fishery Bulletin 84, 27-34.

Marsh, H., and Sinclair, D.F. 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. Journal of Wildlife Management 53(4), 1017-1024.

McNeely, R.L. 1961. The purse seine revolution in tuna fishing. Pacific Fisherman 59(7), 27-58.

National Research Council (NRC). 1992. Dolphins and the tuna industry. National Academy Press, Washington, D.C. 176 pp.

Perrin, W.F. 1968. The porpoise and the tuna. Sea Frontiers 14, 166-174.

Pollard, J.H., Palka, D., and Buckland, S.T. 2002. Adaptive line transect sampling. Biometrics 58, 862-870.

Pryor, K., and Norris, K.S. 1978. The tuna-porpoise problem: behavioural aspects. Oceanus 21, 31-37.

Okamura, H., Kitakado, T., Hiramatsu, K., and Mori, M. 2003. Abundance estimation of diving animals by the double-platform line transect method. Biometrics 59(3), 512-520.

Reilly, S.B., Donahue, M., Gerrodette, T., Forney, K.A., Wade, P., Ballance, L., Forcada, J., Fiedler, P., Dizon, A., Perryman, W., Archer, F., and Edwards, E. 2005. Report of the scientific research program under the International Dolphin Conservation Program Act. NOAA Tech. Memo. NOAA-TM-NMFS-SWFSC-372, 100 pp.

Scott M. D., and K. L. Cattanach. 1998. Diel patterns in aggregations of pelagic dolphins and tunas in the eastern Pacific. Marine Mammal Science 14, 401-428.

Scott, M. D., and S. J. Chivers. 2009. Movements and diving behavior of pelagic spotted dolphins. Marine Mammal Science 25, 137-160.

Silva, G. 1941. Porpoise fishing. Pacific Fisherman 39(3), 36.

Teilmann, J., Christiansen, C.T., Kjellerup, S., Dietz, R., and Nachman, G. 2013. Geographic, seasonal, and diurnal surface behaviour of harbour porpoises. Marine Mammal Science 29(2), E60-E76.

Wade, P.R. 1995. Revised estimates of incidental of dolphins (Delphinidae) by the purse-seine tuna fishery in the eastern tropical Pacific, 1959-1972. Fishery Bulletin 93, 345-354.

Wade, P.R., Reilly, S.B., and Gerrodette, T. 2002. Assessment of the population dynamics of the northeastern offshore spotted dolphin and the eastern spinner dolphin populations through 2002. SWFSC Administrative Report LJ-02-13, 58 pp.

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.

Ward, E.J. 2005. Differences between fishery-dependent and fishery-independent estimates of single- and mixed-species dolphin schools: implications for single-species stock assessments. Marine Mammal Science 21(2), 189-203.

Webb, A., Irwin, C., and Elgie, M. 2015. Kincardine offshore wind farm: final report on aerial surveys from April 2013 to September 2014. HiDef Aerial Surveying Ltd, Cumbria, CA, UK.

Yuan, Y., Bachl, F.E., Borchers, D.L., Lindgren, F., Ilian, J.B., Buckland, S.T., Rue, H., and Gerrodette, T. submitted. Point process models for spatio-temporal distance sampling data. Annals of Applied Statistics 00, 00-00.

Tables

Table 1. Data types and estimation methods for mark-recapture (M-R) abundance estimation of eastern tropical pacific (ETP) dolphin populations. Permits are an issue with all types of research, but omitted from the table. The column "Status" indicates whether the method could be applied immediately ("Established") or requires additional research prior to implementation. The first row applies to all items in this table.

Data type	What does it aim to give us?	Status	Advantages	Disadvantages
Mark-recapture	Absolute / relative abundance Survival Movement / stock structure Individual identity	Established	Can be combined with other data in a population dynamics model	Heterogeneity in recapture probabilities Design impacts whether estimates are absolute or local Need large sample sizes
Conventional tag	Fishery interactions	Established	Can be applied relatively easily	Tag loss Tag reporting Tagging large numbers is difficult Tag effects
Telemetry / radio tag	Location Fishery interactions Dive depth & time Behaviour state (activity) Habitat association	Established	Argos: global coverage	Need rapid sampling to estimate dive cycle Tag loss Tag reporting Tagging large numbers is difficult Tag effects High cost per tag
Acoustic / PIT tag	Location Fishery interactions	Established	Lower tag loss rate versus telemetry	Limited tag detection range Tag loss Tag reporting Tagging large numbers is difficult Tag effects Implanting tags is a surgical procedure
Conventional genetic M-R	Genetic population structure	Applied to other species	Archive samples for later analysis Possible recaptures via fishery Possible with a short time series	Need to develop markers
Close-kin M-R	Fecundity Social structure	Applied to other species	Archive samples for later analysis Dead animals / bycatch Fewer samples than other M-R Possible with a short time series	

Table 2. Data types and estimation methods for line-transect (LT) abundance estimation of eastern tropical pacific (ETP) dolphin populations. Population dynamics (PD) model; species identification (spp ID). The column “Status” indicates whether the method could be applied immediately (“Established”) or requires additional research prior to implementation. Proof of concept (PoC) status refers to its need to be established prior to its use. The first row applies to all items in this table.

Data type	What does it aim to give us?	Status	Advantages	Disadvantages
Line transect	Absolute / relative abundance Distribution / stock structure Habitat association			Design impacts whether estimates pertain to local or total abundance
Ship-based LT survey (Visual component)	Group size	Established	Platform for other studies Existing time series Double platform possible	Possible behaviour changes before sighting Long survey time Light- and weather-dependent detection *g*(0) dependent
Acoustic Towed array	*g*(0)	Some work exists	Detection is independent of visibility conditions Detection distances > than ship-based survey Independent of *g*(0), but dependent of the fraction of animals calling.	Spp ID estimated statistically Group size estimation not possible Call rate affected by group size and behaviour Detection is dependent on physical environmental conditions
Drifting buoy		PoC – range est.	Detection is independent of environmental conditions Animals do not react Fishing vessels could recover buoy Independent of *g*(0), but dependent of the fraction of animals calling.	Spp ID estimated statistically Group size estimation not possible Call rate affected by group size and behaviour Detection depends on physical environmental conditions Track dependent on currents
Aerial (photographic; high-resolution imagery) Manned aircraft	Body condition Cow-calf association Group size Reproductive output (proportion calves in schools)	Established–mixed spp ID PoC – detection probabilities / coverage	Animals much less likely to react Sampling can be adaptive Images provide permanent record Double platform possible (observer and photographs) Independent of *g*(0), but dependent on detection probability within the water column Rapid Count individuals rather than estimate group size	Light-, weather-, and turbidity-dependent detection Range (needs ship support) Thunderstorms affect ability to fly Groups not running are less visible Long post-processing times

Unmanned aircraft	Body condition Cow-calf association Group size Reproductive output (proportion calves in schools)	PoC – mixed school spp ID PoC – detection probabilities / coverage	Animals do not react Flight duration can be > manned Images provide permanent record Technology improving rapidly Independent of *g*(0), but dependent on detection probability within the water column Rapid survey Count individuals rather than estimate group size	Light-, weather-, and turbidity-dependent detection Design to account for night flight Thunderstorms affect ability to fly Airspace access & safety concerns May need ship support Long post-processing times Groups not running are less visible
Satellite	Group size	PoC – availability bias PoC – detection	Animals do not react Cover large & unserviceable areas Images available in short time Low set-up cost Less need for permits Potentially repeat images Images provide permanent record g(0) independent, but dependent on other detection factors	Cannot get dolphin spp ID Large data sets Light-, weather-, and turbidity-dependent detection Need satellite provider agreements Need automated image processing Groups not running so less visible Long post-processing times

Table 3. Other data types applicable for methods used to estimate the abundance of eastern tropical pacific (ETP) dolphin populations. If these data are available to collect, then an emphasis should be placed on collecting them. The column “Status” indicates whether the method could be applied immediately (“Established”) or requires additional research prior to implementation.

Data type	**What does it aim to give us?**	**Status**	**Advantages**	**Disadvantages**
Life-history data	Stock structure Survival Fecundity Population growth rates (using population dynamics models) Somatic growth	Established	Obtained from various platforms Large sample sizes possible Could be compared to previous estimates of fecundity and population growth rates	Discontinuity among time series Recently, low mortality Pulsed sampling Information content dependent on knowledge of processes such as selection
Other				
Oceanographic sampling	Habitat information Stock structure	Established	Can be obtained from various platforms	Sources have different temporal spatial-temporal coverage / resolution Some products are model-based Needs to be combined with spatial abundance information
Fishery-dependent data	Relative abundance	Established	Lots of data Extensive spatial-temporal coverage	Biased sampling design Incomplete information from all search methods Observers’ estimates of group size are not calibrated

Table 4. Approximate CV prognosis by stock (northeastern and western/southern spotted dolphins; eastern and whitebelly spinner dolphins) across years.

Year	1	2	3	4	5
New marked	2,500	2,500	2,500	2,500	2,500
Surviving marked		2,350	4,559	6,635	8,587
Recaptures		8	15	22	29
Cumulative recaptures		8	23	45	74
CV – conventional genetic M-R		0.36	0.21	0.15	0.12
CV - close-kin		0.25	0.15	0.11	0.08

Table 5. Management goals and their modelling and information needs. DML: Dolphin Mortality Limit; MNPL: Maximum Net Productivity Level.

Management goal	Minimal Model	Data/information	Uncertainty	Reliability
Abundance estimates (DML)	Exponential regression	Absolute abundance	Dependent on abundance estimates	Moderate
Recent trends	Exponential regression	Relative abundance	Dependent on abundance estimates	Moderate
Depletion level	Total catch history model	Absolute (preferable) or relative abundance, and catch	Dependent on historical catch and to some extent density dependence assumptions	Low
Reference point (e.g., MNPL) evaluation	Model that includes the total catch and dynamics processes	Absolute (preferable) or relative abundance, catch, and life-history information	Dependent on historical catch and density dependence assumptions	Low

Appendix A: Participants

Invited Participants

André E. Punt (Chair; UW), Lisa T. Ballance (SWFSC), Jay Barlow (SWFSC), Steve Buckland (University of St. Andrews, Scotland), Susan J. Chivers (SWFSC), Justin Cooke (CEMS, Germany), Michel Dreyfus (Instituto Nacional de Pesca, México), Paul C. Fiedler (SWFSC), Karin A. Forney (SWFSC), Megan C. Ferguson (AFSC), Peter Fretwell (British Antarctic Survey, UK), Tim Gerrodette (SWFSC), Robert Jannarone (Brainlike, USA), Toshihide Kitakado (Tokyo University of Marine Science and Technology, Japan), Jeff Moore (SWFSC), Phil Morin (SWFSC), Bernie McConnell (Sea Mammal Research Unit, University of St. Andrews, Scotland), Wayne Perryman (SWFSC), Robert Pitman (SWFSC), Hans J. Skaug (University of Bergen, Norway), and Andy Webb (HiDef)

Workshop staff

Kelli F. Johnson (UW), Cleridy E. Lennert-Cody (IATTC), Mark N. Maunder (IATTC), and Michael D. Scott (IATTC)

Observers

Ernesto Altamiran (IATTC), Eric Archer (SWFSC), Dan Averill (Marine Stewardship Council), Guillermo A. Compeán (IATTC), Kerri Danil (SWFSC), Luis Fleischer (Comisión Nactional de Pesca y Acuacultura, México), Noressa Giangola (Pacific Alliance for Sustainable Tuna), Guillermo Gomez (Pacific Alliance for Sustainable Tuna), Shane Griffiths (SWFSC), Martín Hall (IATTC), Annette Henry (SWFSC), Al Jackson (SWFSC), Kristin Koch (SWFSC), Aimée Lang (SWFSC), Rebecca Lent (Marine Mammal Commission, USA), Carolina V. Minte-Vera (IATTC), Sarah Mesnick (SWFSC), Jaime Bolaños Jiménez (Especialista Externo Ministerio del Poder Popular para la Pesca y Acuicultura, República Bolivariana de Venezuela), Paula Olson (SWFSC), Mariana Ramos (Pacific Alliance for Sustainable Tuna), Shannon Rankin (SWFSC), Rebecca Regnery (Humane Society International), Kelly Robertson (SWFSC), Jerry Scott (International Seafood Sustainability Foundation), and Suzanne Yin (SWFSC)

AFSC - Alaska Fisheries Science Center, National Marine Fisheries Service National Ocean and Atmospheric Administration, USA
IATTC - Inter-American Tropical Tuna Commission, USA
SWFSC - Southwest Fisheries Science Center, National Marine Fisheries Service National Ocean and Atmospheric Administration, USA
UW - University of Washington, USA

Appendix B: Draft Agenda

Tuesday, October 18

09:00 Opening address (Workshop Chair, André Punt)
09:15 Background paper 1 - Data sources (Michael Scott; 15 min + 5 min questions)
09:35 NMFS survey data (Tim Gerrodette; 15 min + 10 min questions)
10:00 Life history data for ETP dolphins (Susan Chivers; 15 min + 10 min questions)
10:30-10:45 Coffee break
10:45 Tracking technology (Bernie McConnell; 15 min + 10 min questions)
11:10 High-resolution imagery (Andy Webb; 15 min + 10 min questions)
11:35 Aerial photographic techniques (Wayne Perryman; 15 min + 10 min questions)
12:00 -13:00 Lunch
13:00 Drone application in marine mammal survey (Megan Ferguson; 15 min + 10 min questions)
13:25 Satellite imagery, advantages and disadvantages (Peter Fretwell; 15 min + 10 min questions)
13:50 Genetics mark-recapture and close kin (Hans Skaug; 15 min + 10 min questions)
14:15 Acoustic surveys (Jay Barlow; 15 min + 10 min questions)
14:40-15:00 Coffee break
15:00 Automated analysis of airborne imagery (Robert Jannarone; 15 min + 10 min questions)
15:30-16:30 Group discussion – data sources (60 min)
16:30-17:30 Public comment period

Wednesday, October 19

09:00 Background paper 2 – Abundance estimation (Steve Buckland; 50 min + 10 min questions)
10:00 Background paper 2 – discussion presentation (Toshihide Kitakado; 20 min + 20 min questions)
10:40-11:00 Coffee break
11:00 Group discussion – data and abundance estimation (60 min)
12:00-13:00 Lunch
13:00 Background paper 3 – Population Modelling (André Punt; 50 min + 10 min questions)
14:00 Background paper 3 – discussion presentation (Justin Cooke; 20 min + 20 min questions)
14:40-15:00 coffee break
15:00 Group discussion – data, abundance estimation, population modelling (60 min)
16:00-17:00 Public comment period
17:00-19:30 Social

Thursday, October 20

09:00 Group discussion – research plan, short- and long-term (90 min)
10:30-10:45 Coffee break
10:45-12:00 Public comment period
12:00-13:00 Lunch
13:00-15:00 Group discussion - research plan, short- and long-term (120 min)
15:00-15:15 Coffee break
15:15-16:15 Group discussion and draft outline of workshop report
16:15 Public comment period (15 min)
16:30 Closing address (Workshop Chair, André Punt)

Appendix C: Background documents

Background document 1. Scott, M.D., Lennert-Cody, C.E., Gerrodette, T., Skaug, H.J., Minte-Vera, C.V., Hofmeister, J., Barlow, J., Chivers, S.J., Danil, K., Duffy, L.M., Olson, R.J., Fiedler, P.C., Ballance, L.T., and K.A. Forney. Data Available for Assessing Dolphin Population Status in the Eastern Tropical Pacific Ocean

Background document 2. Buckland, S.T., Lennert-Cody, C.E., Gerrodette, T., Barlow, J., Moore, J.E., Webb, A., Fretwell, P.T., Skaug, H.J. and W.L. Perryman. Review of potential methodologies for estimating abundance of dolphin stocks in the Eastern Tropical Pacific

Background document 3. Punt, A.E. Review of Contemporary Cetacean Stock Assessment Models.

Appendix D: Abstracts of presentations

Michael Scott (Available data sources for ETPO dolphin populations)
A description of the data sources available for monitoring the status of ETP dolphin was presented. Within the ETP there has been a history of tagging and tracking of dolphins. Additional information was provided on data collected during purse seine operations when setting on tuna associated with dolphins.

Tim Gerrodette (Line-transect surveys to estimate dolphin abundance in the eastern tropical Pacific Ocean)
Line-transect surveys using research vessels were carried out by the Southwest Fisheries Science Center from the late 1970s to 2006 to estimate dolphin abundance in the ETP. A team of three observers searched visually from the flying bridge of the vessel, primarily using 25X pedestal-mounted binoculars, at a height of 10-11 m. Observers' estimates of group size were checked using photographs collected from a helicopter. There is a general tendency to underestimate group size, and the tendency varies by observer and species.

Susan Chivers (ETP dolphin life-history data)
Biological data were collected by observers from more than 43,000 individual dolphins killed in the ETP yellowfin tuna purse-seine fishery between 1966 and 1994. The data and tissue samples collected were used in studies to characterize the essential elements dolphin life history (i.e., reproduction, growth and survival, and to estimate population growth rates). Since 1994, the IATTC has continued monitoring the fishery, but the comprehensive dolphin sampling program established in the early 1970s has not been continued. However, the NMFS has continued biological studies of ETP dolphins using remote technologies. For example, steroid hormones analysed from blubber biopsy samples have been used to identify pregnant females and photogrammetric count data have been used to estimate calf production. Both methods provide the ability to monitor reproduction in wild dolphin populations and continue the time series from the observer program data.

Bernie McConnell (Tracking technology)
Telemetry is a toolbox of building blocks that may be assembled to optimally answer specific questions about specific species. The combination of these blocks, and the development of new blocks, is limited purely by imagination, physics, and money. It is likely that Cetacean Tagging Guidelines will be published in 2017. Single pin satellite tags can last up to 163 days. The use of computational fluid dynamics in tag design is important in reducing drag and increasing longevity. For dolphins, the only realistic option for global relay of data is the Argos satellite system. For shorter, local studies VHF or physical retrieval is an alternative option. Numerous low-energy sensors are potentially available for answering specific questions. In summary, the user community must proactively engage with manufacturers to develop innovative telemetry solutions.

Andy Webb (High resolution digital aerial surveys)
Digital, aerial-based surveys in Europe first emerged in 2006 and were developed primarily for environmental surveys of marine megafauna around offshore wind farms in the UK. The principal driver for their development was the need to fly and survey effectively above wind turbine generators, which would be considerably safer than flying between them with better sampling, the need for an evidence trail, and the potential for improved count accuracy. Since acceptance of the validity of the method, over 1500 digital, aerial-based surveys have been flown in NW Europe and USA, mainly for characterising seabird and marine mammal abundance and their distribution around offshore wind farms, but also for monitoring post-construction effects and for monitoring at protected sites. For the most part, these surveys measure relative abundance of marine mammals, but have used generic corrections based upon average dive depth and duration to approximate absolute abundance.

HiDef's high resolution video survey method uses a bespoke camera rig either in a modified nose cone or in the standard photogrammetry hatch of various light aircrafts. HiDef's cetacean-only method uses four cameras which each survey a 187.5 m swathe separated by a 30 m gap. The cameras are angled at 30º from vertical on a plinth that rotates at the end of each transect such that cameras point permanently away from sun glare. The aircraft flies at 610 m ASL and has a ground sample distance (GSD) of 3 cm. Data are streamed continuously for storage onto hard drives with RAID for backup. After the survey, a two-stage process is used for review of video footage and identification of animals. Some 20% of all video material undergoes a blind re-review and a minimum of 90% agreement is required for data quality to be passed, but an average of about 96.7% agreement is typical. Marked objects are then identified and 20% also undergoes blind review requiring at least 90% agreement (typically 96% is achieved). All review and identification is manual; thus far, no automated system has been found by HiDef to match or improve on the performance of human review processes.

The other digital imagery systems use still cameras and have not been employed for cetacean-only surveys in Europe to date. Still cameras are all based on off-the-shelf medium format or photogrammetry systems. They sample in plain view and are either used for recording continuous transects or for plot-based sampling. These systems are flown at 400 m ASL typically and achieve 3 cm GSD resolution and have a strip width of 250-460 m, depending on the sensor size. Some of these use automated processes to detect some of the animals within the imagery with unknown success. Sun glare is an issue, and processes have been developed for cutting out affected parts of images or even the whole sample. As in the case of video surveys, relative abundance estimates are obtained for marine mammals unless generic correction factors can be obtained from dive data.

While high resolution surveys have come a long way in NW Europe, there are still some reservations, mainly because it is not yet possible to obtain in-situ measures of availability bias during surveys. A potential double-platform solution has been designed but is not yet tested by HiDef. Automation solutions exist, but cannot yet match humans for detection efficiency. Manned digital aerial surveys can cover up to 1400 km in one day, but this is unlikely to be sufficient to reach all parts of the IATTC study area. Unpiloted versions of survey aircraft exist which would increase their range to over 3000 km.

Wayne Perryman (Aerial photography: background, challenges, successes, and moving forward)
Estimates of group size by observers on tuna vessels in the late 1970s were 7-8 times higher than estimates from observers on ship-based line-transect surveys. Consequently, aerial surveys were used to calibrate estimates from tuna vessel observers. Since then, digital technology has rapidly developed and now smaller, higher resolution cameras can be placed on unmanned aerial-survey platforms (drones) that can take off and land vertically, have an endurance of ~20 min, and are capturing high-resolution images from ~300 ft. Images are helpful in estimating group size, species identification, length, and body shape, which is indicative of life history. New aircraft should be available shortly with two times the endurance. Even now, some drones can fly in a sea state of five and change flight patterns based on sighting detections.

Megan Ferguson (Comparing estimates of arctic cetacean density and associated uncertainty derived from manned and unmanned aerial surveys: Operations, methods, and preliminary results)
Manned aerial surveys have been used successfully for decades to collect data to infer cetacean distribution and density. Unmanned aerial systems (UAS) have potential to augment or replace some manned aerial surveys for cetaceans in the future. To ascertain the utility of UAS for such missions, however, it is first necessary to define the specific scientific objective(s) and then compare the cost-benefit of alternative platforms and methodologies. NOAA led and conducted such a direct comparison of aerial surveys for cetaceans near Barrow, Alaska, during fall 2015 via a collaborative effort that included the Bureau of Ocean Energy Management, US Navy, North Slope Borough Department of Wildlife Management, and Shell. We conducted a three-way comparison among visual observations made by marine mammal observers aboard a Turbo Commander operated by Clearwater Air, Inc; imagery autonomously collected by a Nikon D810 camera system mounted on the belly of the Turbo Commander; and imagery collected by a similar camera system on a remotely-controlled ScanEagle operated by the Naval Surface Warfare Center Dahlgren Division. The platforms each conducted five flights within a 16,800 km^2 study area. Surveys from manned and unmanned platforms did not directly overlap geographically and temporally to maintain safety of flight; the two platforms operated as close as safely possible. The Turbo Commander collected 44,849 images in 26.7 flight hours. The ScanEagle collected 24,600 images in 21.8 flight hours.

Manual image processing and analysis by marine mammal photo analysts required 332.5 total hours, averaging 6.9 hours to analyze one flight hour, which involved reviewing every third image. In total, eight bowhead whales (*Balaena mysticetus*) and 16 belugas (*Delphinapterus leucas*) were identified in the images from the Turbo Commander. Fifteen bowhead whales, six belugas, and three gray whales (*Eschrichtius robustus*) were identified in the UAS images. Sixty-one bowhead whales, 54 belugas, nine gray whales, and 48 unidentified cetaceans were sighted by the marine mammal observers aboard the Turbo Commander. Bowhead whale density estimates derived from the marine mammal observer data and Turbo Commander imagery were similar. Beluga density estimates derived from the marine mammal observer data were greater than estimates derived from either imagery dataset. The uncertainties in density estimates derived from the marine mammal observer data were lower than estimates derived from either imagery dataset. The cost of the UAS survey was considerably more expensive than the manned aerial survey.

Peter Fretwell (Satellite imagery: Advantages and disadvantages)

The study of cetaceans by satellite imagery is a technique that is in its infancy. Satellite sensors with the spatial, temporal, and radiometric resolution capable of pragmatically identifying cetaceans have only recently become available. Currently, only test studies on larger whale species have been conducted. Although these show potential promise and have many advantages over more traditional methods, the limited resolution of satellite imagery results in a number of drawbacks, and the technique remains unproven for smaller cetaceans. There are only two published papers that use satellite imagery to identify whales. The first, by Ron Abileah in 2005 used IKONOS imagery with a spatial resolution of 1.5 m per pixel to look for humpback whales near Maui, HI. This resolution of imagery could differentiate boats from objects in the water, but wide-scale identification was not possible. In 2014, a study using 50 cm resolution QuickBird2 imagery in optimal conditions successfully counted southern right whales at Península Valdés over an area of 115 km^2. With the relaxation of federal regulations on satellite data in 2015, higher resolution WorldView3 imagery at 30 cm per pixel has become available and ongoing preliminary studies on humpback whales in Hawaii and fin whales in the central Mediterranean both show the capability of counting large cetaceans. Advantages of satellite data include large coverage, with each image covering over 1000 km^2; the ability of repeat imagery; the low potential cost relative to other survey techniques; the low set-up costs; lack of bureaucracy; the ability to target any part of the ocean; and the safe nature and lack of disturbance from the satellite. Disadvantages include the fact the method is untried for dolphins and it is likely that only the splashes of dolphins will be countable given the relatively coarse resolution of even the best imagery. Species identification will not be possible, unless combined with other survey techniques. The method performs badly in poor sea-states and considerable analysis will need to be undertaken to understand the availability bias needed to covert counts into population estimates because of the novelty of the data. Finally, agreements with satellite providers will have to be sought before the method is cost effective.

Hans Skaug (Genetic mark-recapture and close-kin)

Genetic M-R is ordinary M-R with physical tags replaced by DNA profiles. Both abundance and survival may be estimated, but due to the large population size the required number of biopsy samples may be prohibitive for ETP dolphins. Close-kin methods exploit the fact that DNA profiles contain information about the biological relationship among individuals in the sample. It can also be viewed as a M-R method, but with “recapture” meaning the presence of a close relative in the sample. Close-kin has been successfully applied to southern bluefin tuna, which has an abundance in the same range as ETP dolphins. There does not yet exist a standard software package for analysing close-kin data, so some statistical method development must be anticipated for each new application. Close-kin methods are applicable to tissue samples collected from dead animals, such as those that are lethally bycaught in the tuna fisheries. With current bycatch levels, sufficient sample sizes will be obtained over a 20 year period for this data source alone. The price of genetic analyses continues to go down, so the limiting factor for both genetic M-R and close-kin seems to be availability of tissue samples.

Jay Barlow (Use of passive acoustics for estimation of cetacean population density: Realizing the potential)

Methods to estimate cetacean abundance using passive acoustic surveys have advanced considerably in the past decade, but applying these methods to estimate dolphin abundance is more difficult than for the other species that have been studied to date. For distance sampling methods applied to acoustic data, the unit of analysis can be an individual sound (a cue), an individual animal, or a group. The group-based method is the most feasible approach for dolphins, but group size cannot be estimated using acoustic data alone. Acoustic detection platforms could include towed horizontal hydrophone arrays, free-floating vertical hydrophone arrays, bottom-mounted hydrophones, gliders, or profiling buoys. Detection range, which is required for distance-sampling estimates, can be best estimated from towed and vertical hydrophone arrays. Dolphin movement in reaction to the towing vessel is a problem for abundance estimation with towed hydrophone arrays. Range estimation from vertical hydrophone arrays may be feasible, but this approach is new and has never been tested. The lack a group size estimates is a concern for both types of detection systems. At this point, absolute abundance of dolphins cannot be reliably estimated using any acoustic-only technology. Towed arrays might be useful in acoustically detecting groups that are not seen by observers on visual-sighting surveys. Vertical arrays might be useful in estimating relative densities of dolphins based the density of acoustic cues.

Robert (Bob) Jannarone (Automated image processing: marine mammal monitoring prospects)
Airborne sensor and unmanned aerial survey (UAS) advances are making airborne surveys of marine mammals more affordable. Thousands of maritime images may now be gathered in a single, un-piloted flight, launched from ship or land. However, one critical component is lagging - the capacity to automatically identify marine mammals from high resolution data. Without automatic identification, human observers must analyse massive amounts of data manually. Analysing images manually in real time runs the risk of missing target animals and distracting observers from other important tasks. Post-flight, manual analysis can cause expensive delays in marine mammal detection and mitigation. Either way, manual data analysis requires human intervention, takes time, and costs money. For example, a UAS may be configured with high resolution cameras to look for marine mammals to meet regulatory oil drilling or fishing requirements. Highly compressed video data may be streamed to the UAS operator in real time, allowing the operator to redirect the UAS for adaptive sampling when marine mammals are found. However, identifying marine mammals in real time from compressed data can be difficult and distracting. Alternatively, trained experts may analyse images post-flight with better chances than real-time observers of finding marine mammals. Post-flight analyses can take time, cost money, and happen too late. In this presentation, automated marine mammal detection availability for post-flight marine mammal detection will be described and demonstrated. Its operational use, potential value, and key transition enablers will be discussed.

Steve Buckland (Review of potential methodologies for estimating abundance of dolphin stocks in the Eastern Tropical Pacific)
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*). We focus on methodologies for fishery-independent data sources. 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.

We consider ship-based surveys conducted from research vessels, from tuna vessels operating as research vessels, and from tuna vessels in normal fishing mode. We also consider aerial surveys of different types: manned aircraft with observers; manned aircraft with high-resolution imagery; long-range "military-grade" drones with high-resolution imagery; and short-range drones with high-resolution imagery. Surveys using satellite imagery are also addressed, as are capture-recapture and close-kin methods. Acoustic surveys may be conducted using ships, gliders, or drifters. Finally, we consider composite methods that combine methodologies in an attempt to improve abundance estimates.

We conclude that the safe (if costly) option is to continue ship-based surveys. In any such survey, additional data should be collected to improve understanding of the apparent effect of sea state on the probability that schools on the trackline are detected. For example, a drone or helicopter might be flown ahead of the survey ship, providing a ‘tracker’ platform, allowing $g(0)$ and responsive movement to be estimated. If aerial surveys were to replace ship-based surveys, then the option that reduces risk and which is potentially achievable is drone surveys conducted using drones with a range of thousands of kilometres, together with high-resolution imagery. To use capture-recapture or close-kin methods, large numbers of dolphins must be marked. Realistically, to ensure sufficient recaptures, a method would be needed to identify marked animals during back-down by tuna vessels. The difficulty in marking a sample of dolphins that is sufficiently large and representative is considerable. Acoustic survey data may be useful for estimating trends in relative abundance, although bias might arise if acoustic behaviour or school size changes over time. All methods based on new technology will have development costs.

Line-transect surveys by the NMFS in the ETP began in 1974 using a combination of aircraft and ships. Ship-based procedures were refined each year and, by 1979, were close to current procedures. The methods are tried and tested. The target species form large, easily detected schools, and a wide strip can be surveyed using the pedestal- or tripod-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. It can be difficult to estimate group size and species proportions (mixed groups), but aerial photographs of a sample of schools are used to quantify and correct for bias. Precision of abundance estimates is rather poor, given the resources that have been devoted to these surveys. Jay Barlow 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 ship-based surveys over such a large study area.

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. A drone or helicopter might provide an effective tracker platform, operating ahead of the survey vessel, and setting up trials for the main observation platform, allowing estimates to be corrected for both responsive movement and $g(0)$.

Correlations among sea state, location, extent of evasive behaviour, and group size may partially explain the results obtained by Jay Barlow. 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 modelling encounter rate and for modelling 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. However, the larger source of variance is encounter rate, so encounter rate modelling 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 contribute to higher precision, although we would expect gains to be rather modest.

The most important principle of survey design for design-based estimation of abundance is that units of survey effort are placed randomly with respect to the distribution of animals or groups of animals. 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. 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.

Toshihide Kitakado (Discussion on “abundance estimation)

Much work exists on the abundance estimation of dolphins in the ETP. Unfortunately, many issues exist with respect to the use of fishery-independent shipboard surveys for the estimation of absolute abundance. First, it is suggested that uncertainty in the observed school size and its corrected estimate be more carefully addressed in the estimation and assessment of variance. For instance, instead of using corrected school size as a plug-in into an underlying Horvitz-Thompson like estimator, the use of the expectation of corrected school size and effective strip width using the conditional distribution of corrected school size given observed school size may be useful and contribute to producing increasingly stable and accurate abundance estimates. Second, with respect to *g*(0), which is crucial in obtaining unbiased absolute estimates of abundance, it is suggested that, among many methods, mark-recapture type methods such as Buckland-Turnock could be useful, especially with simultaneous use of other equipment such as the drones and passive acoustics, when considering large school sizes in number and space. These methods would also provide another chance for correcting for the response movement. Finally, regarding the variance estimation of the abundance estimate, decomposing information on the various sources of variance components would be useful, if possible, for understanding the main causes of uncertainty and for planning future surveys to reduce uncertainty. Furthermore, the use of spatial modelling with data from multiple platforms, as well as covariates, may reduce estimation uncertainty.

André Punt (Review of contemporary cetacean stock assessment models)

Model-based methods of analysis are widely used to conduct assessments and to provide the operating models on which management strategy evaluation is based, for cetacean stocks. This paper reviews recent assessments and management strategy evaluations for cetacean populations, with a view towards establishing best practice guidelines for such analyses. The models on which these analyses are based range from simple exponential trend models that ignore density-dependence to complex multi-stock age-sex- and stage-structured models that form the basis for management strategy evaluation. Most analyses assume that density-dependence is on calf survival (which implicitly includes maturity and pregnancy rate), but it could also impact the survival rate of adults or the age-at-maturity. Female cetaceans seldom have more than one calf per year, which limits the variation in calf numbers and places an upper limit on the effects of density-dependent calf survival. The models differ in terms of whether the population projections start when substantial catches first occurred or whether allowance is made for time-varying carrying capacity by starting the model in a more recent year. Most of the models are deterministic, but account needs to be taken of variation in cohort strength for analyses that include age-composition data or for species that are relatively short-lived. A limited number of analyses include process variability using a state-space-like modelling framework. Abundance is very low for some stocks, so both demographic and environmental variability need to be included in models for these stocks. The primary source of data for parameter estimation is time-series of estimates of absolute abundance, although the analyses reviewed made use a variety of data types, including relative abundance indices, mark-recapture data, and minimum abundance estimates based on haplotype counts. In general, at least one estimate of absolute abundance is needed for parameter estimation because there is a lack of catch-induced declines in abundance that are captured by indices of relative abundance and hence could be used to provide information on absolute abundance. Similarly, information on abundance from age- and length- composition data is limited. Most of the analyses quantify uncertainty using Bayesian methods to allow information on biological parameters, particularly the intrinsic rate of growth and the relative population at which maximum production occurs, to be included in the analyses, along with sensitivity testing. However, some analyses also quantify uncertainty using bootstrap and asymptotic methods. The future for the models on which assessments and management strategy evaluation is based will likely involve multi-stock models that include age-, sex- and spatial-structure and are fitted as state-space formulations, although at present such models are often too computationally intensive to be feasible for implementation or there is insufficient information in the data to estimate the parameters representing all the processes, leading to simplifications, with the result that the performance of some of the methods of assessment used for cetacean stocks needs to be better understood, including through simulation testing.

Justin Cooke (Discussion presentation on Background Document 3)

Background document 3 summarizes the different population and assessment models that have been applied to cetaceans. Population models are required as filters of the available data to yield inferences about quantities or questions of management or scientific interest. The required features of the population model depend both on the data to be used and on the questions of interest. For example, to be able to use individual identification (capture-recapture) data, a population model needs to include individual life history and movement processes, even if the quantities of ultimate interest are aggregate in nature (such as population size and trend).

Experience to date with whale individual identification data shows that there can be highly significant and complicated patterns of heterogeneity in the sampling process, such that it can be a challenge to identify the relatively weak signals from population processes against the background of strong heterogeneity effects. In the case of survey data, most of the pre-analysis to cope with heterogeneity in detection rates can be performed externally to the population model, such that "cleaned up" abundance estimates or indices are produced that can be used as input into the population model.

Environmental variability affects cetacean population dynamics differently from many fish species. Fish populations can be dominated by a few exceptionally strong year classes, but the limited reproductive capacity of cetaceans, specifically the odontocetes, limits their annual increase to 2-4%. However, sudden large decreases are possible (die-offs), and such events can have a major impact on the population dynamics. The interaction between environmental variability and density-dependent effects means that cetacean populations will become more variable when they have recovered from past depletion. The implications for population modelling are that constant-K models will eventually show a lack of fit given a long enough data series. Simulations studies have shown that fitting constant-K models can seriously bias estimates of the mean r and K when K is variable. Lack of fit can often be patched up by hypothesizing a discrete, one-off change in K, but simulation studies have shown that this usually exacerbates the biases in r estimates. It is more appropriate to allow parameters such as K to vary throughout time as a stochastic process.