Program Distance: Recent developments in analysis and design capabilities for cetacean population estimation

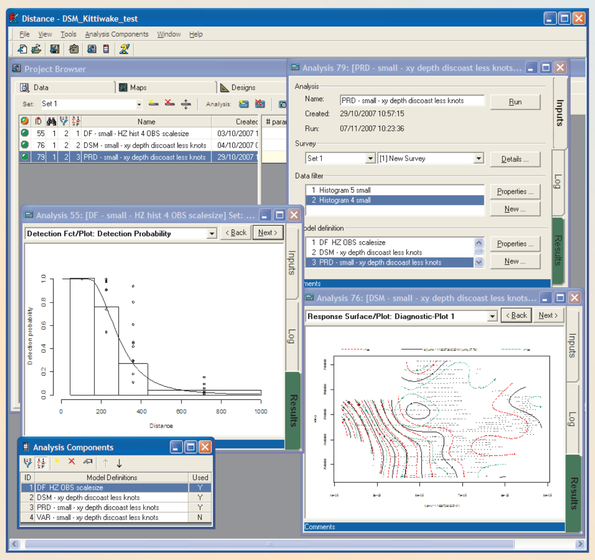
Thomas, L.[1], Borchers, D.L.[1], Buckland, S.T.[1], Laake, J.L.[2], and Rexstad, E.[1]
[1]Research Unit for Wildlife Population Assessment, University of St. Andrews, The Observatory, Buchanan Gardens, St. Andrews KY16 9LZ Scotland
[2]National Marine Mammal Laboratory, 7600 Sand Point Way NE F/AKC3, Seattle WA 98115-6349, U.S.A.

Distance sampling principles

A set of related methods suited to estimating abundance and/or density of marine mammals. The basic concept involves:

- traverse a transect searching for animals,
- record distance to all animals detected,
- use the distribution of distances to estimate the probability of detection given distance ("detection function"),
- from this the number of animals not detected can be estimated, and hence the abundance or density of the population in the study area can also be estimated.

Program Distance is purpose-built for the design and analysis of distance sampling surveys. It is intended to be easy to use, with a Windows graphic interface and voluminous help. The most recent public release version of Distance is 5.0



An example of the Distance Windows user interface.

Software availability and training opportunities

Software availability

In the past 5 years, Distance 5.0, and its predecessors have been downloaded over 10,000 times. Distance is available free of charge from our website http://www.ruwpa.st-and.ac.uk/distance/

Training opportunities

We regularly hold training workshops (both at introductory and advanced levels). Participants from nearly 50 countries have attended these workshops. Please chech this website for upcoming dates and locations http://creem.st-and.ac.uk/conferences.php

Photo: Lee Brady

Acknowledgements

This software has been under development for nearly 30 years. Besides the poster authors, the list of individuals contributing to the software include S. Strindberg, F.F.C. Marques, D.R. Anderson, K.P. Burnham, S.L. Hedley, J.H. Pollard, J.R.B. Bishop, and T.A. Marques.

Funding for software and methods development has been provided by

Colorado Cooperative Fish & Wildlife Research Unit
National Marine Mammal Laboratory (U.S.)
Wildlife Conservation Society
Office of Naval Research (U.S.)
National Park Service (U.S.)
Engineering and physical sciences research council (U.K.)
Biotechnology and biological sciences research council (U.K.)
GeoMarine Inc.
Fisheries and Oceans Canada

Covariates influencing detectability in addition to distance

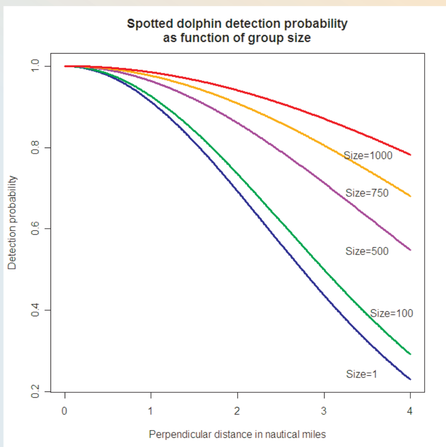
Photo: Tim Gerrodette

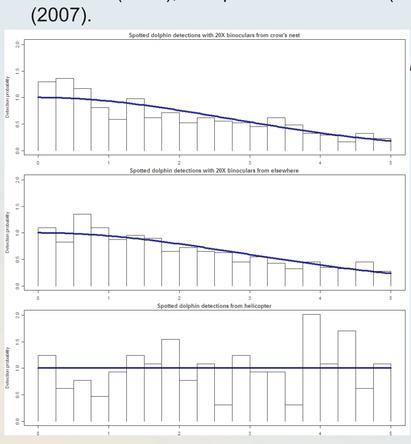
Issue

Estimation of abundance or density using a subset of the data, e.g., geographic strata, species, seasons, particularly when there are few observations in some of the subsets.

Solution

Conventional solution is stratification: i.e., estimate a separate detection function for each subset of the data for which we want to estimate density. However, when there are few observations in a subset, the estimated detection function is unreliable. An alternative is to assume that the general shape of the detection function is constant over all subsets of data, but that the scale is a function of covariates that vary by subset–effectively bending or shrinking the detection function for each subset (covariate levels). These covariates levels may include characteristics of the animals being detected (group size), the observers conducting the surveys (vessel type or observer identity), or environmental characteristics influential in the detection process (sea state, sun angle, or time of day). In each instance, the modelling of the detection function includes covariates that influence the scale of the detection function, and the models for one level of the covariate "borrows strength" (i.e., enhances the fit) from other levels of the covariates. Details and examples of this method of estimation can be found in Marques and Buckland (2003), Marques and Buckland (2004), and Marques et al. (2007).





Example

Analysis of spotted dolphins observed in the Eastern Tropical Pacific (Inter-American Tropical Tuna Commission) was conducted using the multiple covariate distance sampling (MCDS) anaysis engine in Distance 5.0. The result demonstrates the influence of group size upon detectability. All groups (regardless of size) are more difficult to detect as their distance from the observer increases. However, the detectability of large groups declines more slowly than large groups.

At left are histogram and fitted detection functions of spotted dolphin detections for 3 search methods used during the survey (binoculars from crow's nest, binoculars from another location on the ship, and from a helicopter. Note that detectability of spotted dolphins does not diminish as a function of distance from the ship's transect for the helicopter search.

The estimated detection function for a variety of cluster sizes is depicted in the graph above. Observed cluster sizes during the survey ranged from 8 individuals to over 6800 individuals estimated in a single cluster.

Uncertainty in animal detection on transects

Photo: Colin Southwell

Issue

Although classical distance sampling theory assumes detection of animals on the transect line is certain, this is not always the case.

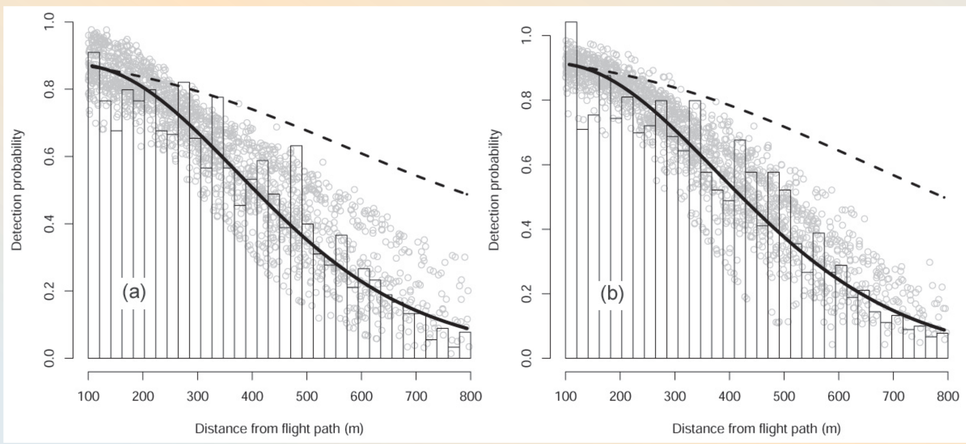
Solution

Using a double observer survey design, wherein observers acting independently, observe animals, and a tracking methods is available to ascertain which animals are "duplicates," that is, seen by both observers. The result is that there are three facets to the problem of estimating animal density, a) a component related to the probability of seeing some number of animals by both observers, b) a component related to the change in detection probability as a function of detection distance, and c) a component related to the probabilities of animals detected by observer 1, observer 2, or both observers detecting animals at distance zero. This combined method of estimation is known as MRDS or mark-recapture distance sampling. More details of this method are provided in Borchers et al. (2006). Analyses of this type can be performed in Distance 5.0 using the MRDS analysis engine.

An important aspect of the MRDS estimation process is that unmodelled heterogeneity in the detection function can give misleading results. This is because pooling robustness (an inherent property in conventional distance sampling) is not applicable because detectability is no longer assumed certain at distance zero.

Example

Borchers et al. (2006) analyzed aerial double observer data of crabeater seals in the Antarctic. With 1740 crabeater seal group detections, roughly equal numbers of groups were detected by observer 1, observer 2, and by both observers. Striking differences were found between the shapes of detection functions from conventional distance sampling methods (the solid curve shapes below) and those from analysis of the double-platform data assuming independence between observers' detections (the dashed curves below). This is a consequence of increasing 'unmodelled heterogeneity' with distance from the line: both observers tended to see only very detectable things far away, so duplicate proportions are inflated at large distances. The problem can be dealt with by using a conventional distance sampling method to estimate detection function shape and the double-observer data to estimate detection probability at the closest distance (a method called 'Point Independence'). Program Distance implements both 'Full Independence' and 'Point Independence' methods.



Estimates of the detection functions from (a) the front-seat and (b) the back-seat observers. The histograms show the distances of detected seals by each observer. The dots are the estimated detection probabilities for each detection. The solid lines are the estimated detection functions assuming no unmodelled heterogeneity at distance 100m (i.e. using the "Point Independence" method of Borchers et al (2006). The dashed lines are the estimated detection functions assuming no unmodelled heterogeneity at any distance. The dashed lines do not fall as fast as the solid lines because at greater distances both observers tend to see only more detectable seals, artificially inflating the proportion of duplicates.

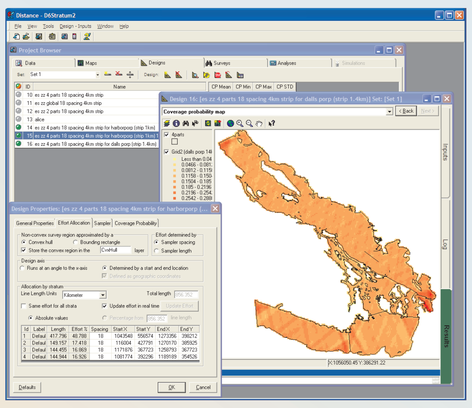
Automated survey design

Issue

Design of line transect surveys is often currently performed using paper and pencil employing "algorithms" of debatable quality, possessing unknown properties. Commonly the design is sketched without any comparison among alternative survey designs.

Solution

Distance 5.0 includes a survey design engine, which incorporates a built-in geographic information system (GIS). Users can enter or import a map of their study site, and can then specify and contrast various survey designs (such as randomly spaced parallel transects, equal angle zigzags and equal spaced zigzags). Conventional analysis methods assume that every part of the study site has an equal chance of being surveyed ("equal coverage probability"), and this assumption can be tested using a built-in simulation capability within the design engine. This generates many realizations of a design, and calculates the proportion of time each point in the study area is covered by a transect. Other design properties can also be calculated, such as the mean, maximum and minimum line length, number of lines and proportion of time spent transiting between transects ("off effort"). Details of various survey designs and exploration of the properties of the survey designs derived using Distance can be found in Strindberg and Buckland (2004 a, b).

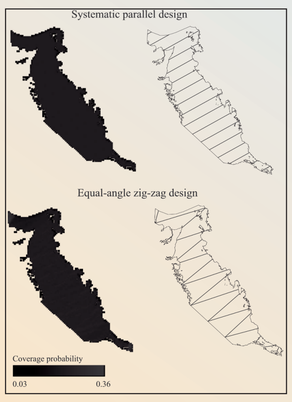


After a particular survey design has been identified, the survey design engine can produce a single realization of the chosen design. Coordinates for the transect lines can then be downloaded from the software onto navigational equipment of the survey vessel.

Example

Thomas et al. (2007) employ the survey design engine of Distance 5.0 to design a multi-species marine mammal survey of the fjord system known as British Columbia.

Panel at upper right shows Distance user interface for producing survey designs, in particular the multi-species survey along the British Columbia coastline. Panel lower right shows contrast of coverage probabilities for systematic parallel design versus equal-spaced zig-zag designs in a different stratum of the study area, the Queen Charolette Basin. The dark areas at the extreme north and south of the study region have abnormally high coverage probability in the zig-zag design due to the extreme irregularity of the study region.



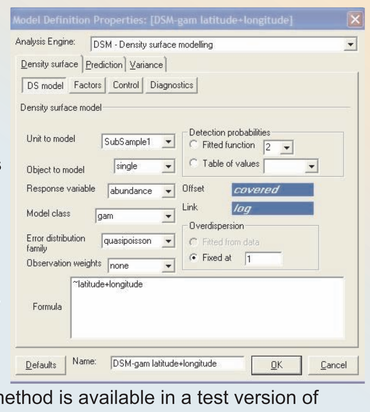
Spatial density surface modelling

Issue

Estimation of animal abundance or density from non-random survey designs; or estimation of abundance in small regions of the study area

Solution

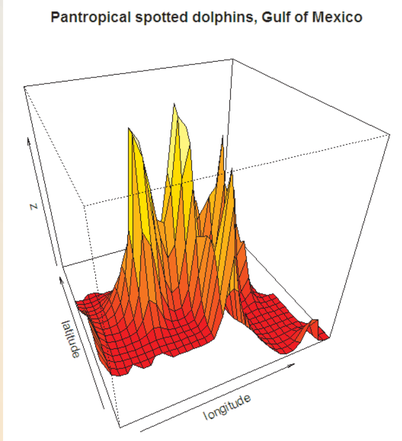
If survey effort is not randomly allocated, then there is no statistical guarantee that the area surveyed is likely to be representative of the study area in general. Hence, design-based estimators of abundance are not appropriate. Instead, model-based estimators can be employed. In this circumstance, the procedure is to model estimated abundance along survey transects using (possibly spatial) covariates. This modelling approach may also improve precision in estimates by explaining some variability attributable to modelled covariates. Predictions from the fitted model are available everywhere within the study region over a prediction grid. Abundance in small regions of the study area can be derived by summing the predicted abundance in each cell of the prediction grid falling within a region of interest.



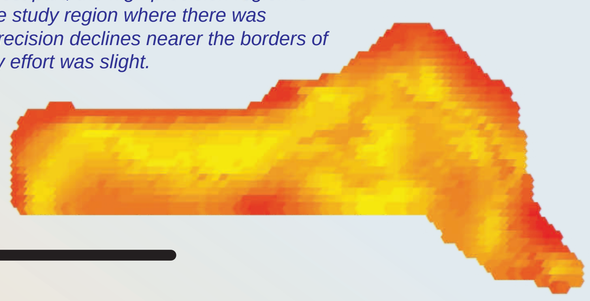
Hedley and Buckland (2004a, b) describe the estimation process using what they describe at the "count" method, to produce estimates of abundance using this model-based approach. This analysis method is available in a test version of Distance 6.0.

Example

Over the past 15 years, the U.S. National Marine Fisheries Service conducted surveys in the Gulf of Mexico. Data were pooled across all of those surveys to produce a map of pantropical spotted dolphin abundance in the Gulf of Mexico. The spatial density surface modelling produced not only point estimates of abundance for cells that comprise the study area (3D panel below), but also using a bootstrap procedure, produced measures of uncertainty for each of the prediction grid cells (below right).



Top right in this panel is the user interface of Distance for specifying details of fitting a density surface model. Visualization of density surface for pantropical spotted dolphins in the Gulf of Mexico (left). This view is looking northeastward across the Gulf of Mexico toward the Florida panhandle in the back corner of the image. Below right is a heat map of the uncertainty in the density surface model estimates. The uncertainty is measured as a coefficient of variation in the estimated density in each of the prediction grid cells. The colour coding grades from yellow (where uncertainty is small), to red where uncertainty is high. For the most part, the high precision region is associated with portions of the study region where there was considerable survey effort. Precision declines nearer the borders of the study region where survey effort was slight.



Future enhancements

Estimation under unequal coverage probability

Under some situations it may be impossible to produce equal coverage probability in a survey. If conventional estimation techniques are employed under these conditions, they will produce biased estimates of abundance or density. However, because the survey design engine computes the coverage probability for all points in the study region, this information can be used to produce unbiased estimates of abundance or density from surveys with unequal coverage probabilities. This feature will be added to a future version of Distance.

Simulation engine

Simulation not only of the survey designs, but an area populated with animals, has the potential to explore the expected precision of proposed designs. Robustness of estimation methods to various patterns in animal distribution within the study region can also be examined. We are currently seeking sponsors to fund the addition of this feature into Distance.

Photo: Peter Corkeron

Bibliography and further reading

Borchers, D., Laake, J., Southwell, C. and Paxton, C. 2006. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. Biometrics 62:372—378.

Hedley, S. L. and Buckland, S. T. 2004a. Spatial models for line transect sampling. Journal of Agricultural, Biological, and Environmental Statistics 9:181—199.

Hedley, S.L., Buckland, S.T., and Borchers, D.L. 2004b. Spatial distance sampling models. Pages 48—70 in Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., and Thomas, L. (eds.) Advanced distance sampling: estimating abundance of biological populations. Oxford University Press.

Marques, F.F.C., and Buckland, S.T. 2003. Incorporating covariates into standard line transect analyses. Biometrics 59:924—935.

Marques, F.F.C., and Buckland, S.T. 2004. Covariate models for the detection function. Pages 31—47 in Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., and Thomas, L. (eds.) Advanced distance sampling: estimating abundance of biological populations. Oxford University Press.

Marques, T. A., Thomas, L., Fancy, S. G. and Buckland, S. T. 2007. Improving estimates of bird density using multiple covariate distance sampling. The Auk 124.

Strindberg, S. and Buckland, S. T. 2004a. Zigzag survey designs in line transect sampling, Journal of Agricultural, Biological, and Environmental Statistics 9:443—461.

Strindberg, S., Buckland, S.T., and Thomas, L. 2004b. Design of distance sampling surveys and Geographic Information Systems. Pages 190—228 in Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., and Thomas, L. (eds.) Advanced distance sampling: estimating abundance of biological populations. Oxford University Press.

Thomas, L., Williams, R., and Sandilands, D. 2007. Designing line transect surveys for complex survey regions. Journal of Cetacean Research and Management.

Program Distance website http://www.ruwpa.st-and.ac.uk/distance/