IWC Database-Estimation and Software System

M.L. Burt[1] and S. Strindberg[2]

Contact email: lb9@st-andrews.ac.uk

ABSTRACT

The International Whaling Commission's (IWC) Database-Estimation and Software System (DESS) was designed to standardise IWC line-transect survey data storage and streamline abundance estimation, including abundance estimation by small management area to inform decision making. Not only was DESS designed to accommodate all International Decade of Cetacean Research (IDCR) and Southern Ocean Whale and Ecosystem Research (SOWER) surveys, but also national surveys, such as the Japanese surveys in the Antarctic and North Pacific, and multi-national surveys conducted in the North Atlantic. The standard software for estimating abundance was program Distance, which could estimate abundance in designed survey blocks, however, management procedures required estimation not just in survey blocks but also in smaller regions that could be defined by the user. This added a spatial component to data selection which led to the inclusion of a geographical information system (GIS). This paper describes the data stored in DESS and the features available for data retrieval and analysis.

INTRODUCTION

One of the primary reasons for conducting the SOWER surveys was to estimate the density and abundance of cetacean species within IWC Antarctic Management Areas I–VI. While Antarctic minke whales (*Balaenoptera bonaerensis*) were prioritised, sightings data were collected for all species. Surveys were conducted using line-transect distance sampling methods (Buckland *et al*., 2001; 2015), and Laake *et al*. (1994) developed a program called Distance, which could estimate density from distance sampling data. Therefore, given the large quantities of data collected on the SOWER (and other) surveys, combined with management procedures which required abundance estimates for small regions (which may not coincide with the designed survey blocks and could include data from several surveys), a system was needed to store data and automate the analysis through links to the Distance software. Thus, the sightings database project originated in 1992 and later a Working Group was established to review the project (IWC, 1993; IWC, 1996).

The IWC Database-Estimation and Software System (DESS) was commissioned in 1993 and implemented by the Research Unit for Wildlife Population Assessment (RUWPA), now part of the Centre for Research into Ecological and Environmental Modelling at the University of St. Andrews. A relational database was a natural choice for data storage due to the hierarchical structure of these data. Estimating abundance in smaller regions (known as small management areas; SMAs), which could be defined by the user, added a spatial component to data selection, which led to the inclusion of a geographical information system (GIS). Datasets and facilities were further added as data became available and analysis methods were developed. Hedley *et al*. (1997) provides a review of these methods. Here we describe the components of DESS – the data and software features which permit data extraction and analysis – and begin with a brief introduction to abundance estimation using line-transect sampling methods (see Buckland *et al*., 2001 for more details).

[1] CREEM, University of St. Andrews, The Observatory, Buchanan Gardens, St. Andrews, UK
[2] Wildlife Conservation Society, 2300 Southern Boulevard, Bronx, NY10460, USA

Estimating abundance using line-transect sampling

In general, cetacean populations for which an abundance estimate is required occur in regions of interest that are too large to undertake a complete census. In addition, due to the cryptic behavior of these species, some proportion of animals in the area covered during a survey is missed. Line-transect distance sampling (Buckland *et al*., 2001; 2015) can account for this imperfect detectability in the estimation process. Line-transect sampling protocol requires that lines are laid down at random (or with some random component) within the region and observers travel along these lines collecting information about the animals that are detected. Cetacean groups (one or more animals) are the unit of detection. For boat-based surveys, it is also crucial to collect information about the species, sighting angle (from the trackline), radial distance to the centre of the group, and the number of animals in the group. Other information, such as whether the group size was confirmed, may also be collected. In conventional line-transect sampling, it is assumed that, if an animal is on the transect line (or trackline), the observer is certain to detect it (frequently denoted by $g(0) = 1$). From the sighting angle and radial distance, the perpendicular distance from the transect to the centre of the animal group can be calculated (y) and the shape of the distribution of these perpendicular distances contains information about the probability of detection.

The density (D) and abundance (N) of a cetacean species in a region can be estimated by:

$$\hat{D} = \frac{n}{2wL\hat{P}_a} . E[s]; \hat{N} = A.\hat{D} \qquad \text{Equation 1}$$

where n is the number of groups of animals detected in the area covered during the survey of size $2wL$ (i.e., assuming a strip-half width w and L is the total length of transects covered), $E[s]$ is the estimated expected group size and A is the size of the study region. The term $\hat{P}_a$ is the probability of detecting a group within w of the transect line. Line-transect distance sampling methods use the perpendicular distances, y, to estimate the average probability of detection as follows:

$$\hat{P}_a = \frac{\int_0^w \hat{g}(y)dy}{w} = \frac{\hat{\mu}}{w} \qquad \text{Equation 2}$$

where $g(y)$ is the detection function and the parameter μ is known as the effective strip half-width (esw). To summarise, there are three components to the estimator given in the Equation 1: encounter rate (n/L), $E[s]$ and μ. These are mentioned specifically because DESS contains some purpose-written software for combining data to estimate these components by user-defined areas.

SURVEY DATA

One of the original motivations for DESS was to store SOWER survey data, but over the years, several other datasets have been added. DESS now contains the following datasets:

- IDCR/SOWER: Southern Ocean annual data spanning 32 survey seasons (1978/79 to 2009/10) with substantial variation in the type and format of data collected through the years (e.g., Branch & Butterworth, 2001b), including species codes and types.

- North Atlantic Sighting Surveys (NASS): aerial and shipboard survey data collected over a number of years by four countries: Faroes, Iceland, Norway, and Spain (Lockyer & Pike, 2009).

- Japanese Scouting Vessel (JSV): data spanning 23 years (1965 to 1987) collected from scouting boats accompanying the Japanese whaling fleets in the Antarctic. These data do not permit abundance estimation as data are only available as daily summaries (Ohsumi & Yamamura, 1982; Miyashita *et al*., 1994).

- Japanese Whale Research Programme under Special Permit in Antarctica (JARPA): data collected in the Antarctic spanning six survey seasons (1987/88 to 1992/93) (e.g., Kato *et al*., 1988; 1989).

- North Pacific (NP) surveys: data from fourteen annual surveys (1982 to 1995) undertaken by Japan (e.g., Fujise *et al.*, 1995; 1996).
- IWC & Commission for the Conservation of Antarctic Marine Living Resources (IWC-CCAMLR): data from the krill synoptic survey conducted in 2000 (Reilly *et al.*, 2000).
- Trans NASS (TNASS): the 2007 Icelandic component of the NASS data (Desportes & Pike, 2006; Desportes *et al.*, 2007).

Not all the datasets contained in DESS are sufficient to estimate abundance using distance sampling methods (e.g., JSV), but in general, the data collected during each survey could be classed as information about survey blocks, transects or detections, and there is a natural hierarchical ordering. For example, each year a region is chosen to be surveyed and divided into survey blocks. Within each block, observers search along predefined transect lines. Within each transect line, observers record information about each cetacean group detection. In practice, substantially more information was collected than the minimum suggested here. For example, search effort was conducted under different observer configurations and protocols. Branch & Butterworth (2001b) provide a clear description of SOWER data; Strindberg & Burt (2003) list all data fields collected for all the datasets.

SOFTWARE

DESS is mainly an interface that combines three separate software packages: a database, GIS, and Distance v3.5. It also contains customised Fortran code for implementing estimation by SMA and 'smearing' to accommodate distance-measurement errors. The data storage and manipulation are based on Paradox for Windows, a relational database management system. Data are stored in tables which are linked together allowing two or more tables to be queried simultaneously. This type of system naturally captures the hierarchical structure of the survey data. Density and abundance estimation is based on the program Distance (Laake *et al.*, 1994) and Fortran 77 modules were written for any estimation components specific to these surveys which Distance was not able to perform (i.e., SMALLMAN, SMEARPAR). The GIS MapInfo is used to disaggregate survey data spatially for estimation by SMA and for the geographically explicit visual display of data. DESS was developed using Paradox for Windows (version 5) and MapInfo 3 and then updated to run under later releases of both programs. Paradox 9 and MapInfo Professional 7.5 are now in use.

DATA MANIPULATION AND EXTRACTION

DESS contains user-friendly software for importing, restructuring and exporting survey data, including provision for changing data formats over time. Due to the complexity of SOWER data, this component contains specific procedures for the importation of sightings, effort, weather and survey block boundary data, and specific menu-driven procedures for viewing, filtering, and exporting the data. Other datasets are handled with more generic procedures. With the flexibility that DESS offers to enable the import and export of data in a variety of formats, the user is not tied to the facilities provided by the interface but can create their own queries to extract data.

The visual display and manipulation of data within DESS facilitates stratified abundance estimation and is a central feature of estimation by SMA. DESS enables users to obtain plots of cruise tracks, sighting positions and block boundaries, while providing access to many other GIS features.

FACILITIES

Abundance estimation by survey block

A frequent use of DESS was the estimation of density and abundance from SOWER surveys (e.g., Branch & Butterworth, 2001a; 2001b) using distance sampling methods. These studies focused on estimating density by survey and for each survey block. Branch (2001) provides a user-friendly guide to doing this in DESS. As mentioned previously, Equation 1 has three components to be estimated (encounter rate n/L, effective strip half-width μ

and expected group size $E[s]$). Ideally, μ and $E[s]$ are estimated separately for each survey block to account for regional differences (e.g., differences in group size in strata near the ice edge compared to further away). If the number of detections is insufficient to do this, data from survey blocks can be combined to estimate a single μ and $E[s]$, so that each component can be stratified differently during analysis. A Fortran module called SMALLMAN was written specifically to deal with multi-level stratification and produce estimates for each survey block and for all blocks combined.

Abundance estimation by SMA

The standard units for abundance estimation are the survey blocks that were used to design the survey. However, given the interest in allowing the user to define a region of interest that was generally smaller than the standard survey block, such as a small management area (SMA), DESS provided features for the definition of the region of interest and the extraction and inclusion of the user-selected survey data in the analysis (Borchers & Chalis, 1996). The SMA can be specified in two different ways: by drawing a closed polygon on a map using the GIS features or by defining a regular grid. The system lists all survey strata that coincide with the defined SMA, and the user is asked to mark all surveys to be included in the analysis. All survey data (i.e., transects and sightings) selected and falling within the SMA are extracted. Then, using SMALLMAN, the user can specify the level of data pooling to estimate μ and $E[s]$ to be used in an abundance estimator as before.

Extraction of data to perform spatial modelling analysis

The aim of DESS was to estimate abundance in a region of interest. Hedley *et al*. (2004) developed methods which related animal density to spatial variables, such as topography, habitat and other environmental factors. With the 'count' method of Hedley *et al.* (2004), transects are divided into small segments and the number of animals estimated to be within the segment provides the response variable in a generalised additive model (GAM) for non-uniform spatial density estimation. Spatial and environmental variables are potential explanatory variables in the model. Specialised statistical software is required to fit these models, which falls outside the scope of DESS. The data format required for this method is fundamentally different to that required for conventional distance sampling methods. Restructuring the data to enable this type of model-based analyses is a considerable task. However, these methods were sufficiently appealing that code was written to extract the SOWER survey data in the required format so that it could be exported and analysed using statistical software, such as R (R Core Team, 2024).

DISCUSSION

DESS was originally designed to store SOWER survey data and estimate abundance for both the original survey areas and user-defined SMAs. A large number of additional datasets were subsequently incorporated into DESS while associated data-manipulation features were added to the system. As analysis methods and information requirements changed over time, many refinements were made to DESS to enable these changes, such as extracting data in a format for spatial modelling. However, analysis methodologies and software have moved on since DESS was developed, which means DESS could now be further refined. For example, the command-line version of the Distance software used in DESS could be replaced with the current Windows-based version (Thomas *et al*., 2010) and incorporate more analysis options and methodologies, such as multiple covariate distance sampling (Marques & Buckland, 2004) and mark-recapture distance sampling (Laake & Borchers, 2004). In this case, given the extraction utilities available in DESS, it is also possible to extract the requisite data for a separate analysis using the most recent version of Distance.

Many data from costly cetacean and other wildlife surveys disappear in the mists of time, either because the data are lost or because the data are not either formatted or documented to make them accessible at a later date, but due to the IWC's vision and investment, key cetacean survey datasets are validated, stored and retrievable from a single software system with detailed metadata. Thus, even if the specific automated analyses for which DESS was originally designed become obsolete, these data will remain accessible to cetacean scientists and researchers in the future.

ACKNOWLEDGEMENTS

Thanks to D. Borchers, M. Hughes, and P. Best for their assistance to improve this manuscript.

REFERENCES

Borchers, D.L., & Chalis, S. (1996). An overview of the IWC Database-estimation software system (DESS). SC/48/O36 presented to the IWC Scientific Committee, Aberdeen, UK, 1996. [Available from the IWC Publications Team]

Branch, T.A. (2001). To insanity and beyond: A simpler guide to DESS 3.0. [Available from the author]

Branch, T.A., & Butterworth, D.S. (2001a). Southern Hemisphere minke whales: Standardised abundance estimates from the 1978/79 to 1997/98 IDCR-SOWER surveys. J. Cetacean Res. Manage. 3(2): 143–174. [Available at: *https://doi.org/10.47536/jcrm.v3i2.886*]

Branch, T.A., & Butterworth, D.S. (2001b). Estimates of abundance south of 60°S for cetacean species sighted frequently on the 1978/79 to 1997/98 IDCR-SOWER sighting surveys. J. Cetacean Res. Manage. 3(3): 251–270. [Available at: *https://doi.org/10.47536/jcrm.v3i3.877*]

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., & Thomas, L. (2001). *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press.

Buckland, S.T., Rexstad, E.R., Marques, T.A., & Oedekoven, C.S. (2015). *Distance Sampling: Methods and Applications*. Springer, Cham, Switzerland.

Desportes, G., & Pike, D.G. (2006). Trans North Atlantic Sightings Survey (TNASS). SC/58/O21 presented to the IWC Scientific Committee, St. Kitts, 2006. [Available from the IWC Publications Team]

Desportes, G., Pike, D.G., Gunnlaugsson, T., Mikkelsen, B., Lawson, J., Simon, M., Vikingsson, G., Witting, L., Øien, N., Zabavnikov, V., & Acquarone, M. (2007). Trans North Atlantic Sightings Survey – TNASS 2007. SC/59/O19 presented to the IWC Scientific Committee, Anchorage, Alaska. [Available from the IWC Publications Team]

Fujise, Y., Kishiro, T., Zenitani, R., Matsuoka, K. Kawasaki, M., & Shimamoto, K. (1995). Cruise report of the Japanese whale research program under a special permit for North Pacific minke whales in 1994. SC/47/NP3 presented to the IWC Scientific Committee, Dublin, Ireland, 1995. [Available from the IWC Publications Team]

Fujise, Y., Iwasaki, T., Zenitani, R., Araki, J., Matsuoka, K., Tamura, T., Aono, S., Yoshida, T., Hidaka, H., Nibe, T., & Tohyama, D. (1996). Cruise report of the Japanese whale research program under a special permit for North Pacific minke whales in 1995 with the result of a preliminary analysis of data collected. SC/48/NP13 presented to the IWC Scientific Committee, Aberdeen, UK, 1996. [Available from the IWC Publications Team]

Hedley, S.L., Borchers, D.L., & Buckland, S.T. (1997). A review of analysis options for implementation in the IWC Database and Estimation Software System. *Rep. Int. Whal. Comm*. 47: 885–895.

Hedley, S.L., Buckland, S.T. & Borchers, D.L. (2004). Spatial distance sampling models. In: S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers & L. Thomas) *Advanced Distance Sampling* (pp.48–70). Oxford University Press.

International Whaling Commission (1993). Report of the Scientific Committee. *Rep. Int. Whal. Comm.* 43: 55–86.

International Whaling Commission (1996). Report of the Scientific Committee: Annex L. *Rep. Int. Whal Comm.* 46: 212–215.

Kato, H., Hiroyama, H., Fujise, Y., & Ono, K. (1988) Preliminary report of the 1987/88 Japanese feasibility study of the special permit proposal for Southern Hemisphere minke whales. SC/40/MI18 presented to the IWC Scientific Committee, Auckland, New Zealand, 1988.

Kato, H., Fujise, Y., Yoshida, H., Nakagawa, S., Ishida, M., & Tanifuji, S. (1989) Cruise report and preliminary analysis of the 1988/89 Japanese feasibility study of the special permit proposal for Southern Hemisphere minke whales. *Rep. Int. Whal. Comm*. 40: 289–300.

Laake, J.L., & Borchers, D.L. (2004). Methods for incomplete detection at distance zero. In: S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers & L. Thomas) *Advanced Distance Sampling* (pp.108–189). Oxford University Press.

Laake, J.L., Buckland, S.T., Anderson, D.R., & Burnham, K.P. (1994). *DISTANCE User's Guide V2.1.* Colorado Cooperative Fish & Wildlife Research Unit, Colorado State University, CO, USA.

Lockyer, C., & Pike, D. (2009). North Atlantic Sightings Surveys: Counting whales in the North Atlantic 1987–2001. *NAMMCO Sci. Pub*. 7.

Marques, F.F.C. and Buckland, S.T. (2004). Covariate models for the detection function. In: S.T. Buckland, D.R. Anderson, K.P. Burnham, J.L. Laake, D.L. Borchers & L. Thomas) *Advanced Distance Sampling* (pp.31–47). Oxford University Press.

Miyashita, T., Shigemune, H., & Kato, H. (1994). Outline of sighting strategy of scouting vessels attached to Japanese whaling fleets. *Rep. Int. Whal. Comm*. 44: 273–276.

Ohsumi, S., & Yamamura, K. (1982). A review of the Japanese whale sightings system. *Rep. Int. Whal. Comm*. 32: 581–586.

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reilly, S., Hedley, S., Hewitt, R., Leaper, R., Thiele, D., Pitman, R.L., Naganobou, M., Watkins, J., & Holland, R. (2000). SOWER 2000: Initial results from the IWC-CCAMLR programme to study whales and krill in the Southern Ocean. SC/52/E21 presented to the IWC Scientific Committee, Adelaide, Australia, 2000. [Available from the IWC Publications Team]

Strindberg, S., & Burt, M.L. (2003). *IWC Database-Estimation Software System (DESS) User Manual*. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK.

Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Bishop, J.R.B., Marques, T.A., & Burnham, K.P. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. J. Appl. Ecol. 47: 5–14. [Available at: *https://doi.org/10.1111/j.1365-2664.2009.01737.x*]