An integrated data collection system for line transect surveys

Douglas Gillespie*, Russell Leaper+, Jonathan Gordon* and Kelly Macleod*

Contact e-mail: dg50@st-andrews.ac.uk

ABSTRACT

A computer based system for the collection of line transect survey data is described. The primary goals of the system were to measure (rather than estimate) distances and angles wherever possible, to provide accurate time-stamps for surfacing events as an aid to duplicate identification and to facilitate accurate data collection by using computers to automate data collection wherever possible. Distance and angle measurements were made using established photogrammetric techniques. Collection of photogrammetric data from video was automated and included a system of data buffering so that several seconds of data prior to each observer sighting could be captured. An additional goal of the system was to eliminate the need for post-cruise data entry and validation through the use of on-board data validation software. The system was successfully used during the 2005 SCANS-II and the 2007 CODA surveys.

KEY WORDS: SURVEY–VESSEL; PHOTOGRAMMETRY

INTRODUCTION

Visual surveys to estimate cetacean abundance rely on observers detecting cues from the target species during brief periods when animals are at the surface. A common analysis approach for line transect surveys is to use the observed locations at which animals are initially sighted to estimate the relative detection probability as a function of perpendicular distance from the track-line (Buckland *et al.*, 2001). Critical data for such analyses include species, group size and sighting location relative to the vessel (usually recorded as range and bearing). More complex analyses that attempt to estimate absolute detection probability frequently use more than one independent team of observers and require some method of assessing whether sightings reported by different observers are of the same animal or group and can thus be classified as duplicates. Detection probability can also be a function of several covariates including number of observers, sea state and weather conditions. These effort-related parameters need to be recorded along with the track of the survey vessel. On surveys, data are often collected in difficult conditions, for example in a cramped location exposed to strong winds and cold and observers are required to continuously scan the sea; all factors that make it difficult to make accurate written notes. A further complication is that for some methods (such as those relying on independent observers) it can be important that independent observers are unaware of the observations of others, while other personnel (e.g. data recorders or duplicate identifiers) need to be able to receive data from all observers.

The potential for computerised data entry systems has long been recognised. However there are considerable challenges to operating complex electronic systems at sea and an understandable reluctance to move away from simple reliable systems based on pencil and paper. The use of paper forms takes the attention of observers away from looking for whales and considerable subsequent effort is required to enter such data into a computer for analysis. An additional disadvantage is that it is not possible to make use of an automatic time-stamp. Simple dictaphones to record verbal commentaries have also often been employed. These have the advantage over paper forms that the observer does not have to take their attention away from the sighting and it is frequently possible to identify events with an automatic time-stamp. However, data still need to be transcribed offline, which can be a more onerous task with verbal recordings than with paper forms.

One of the first real-time computerised data entry systems was the Logger software developed by Lex Hiby and Phil Lovell of Conservation Research Ltd in the early 1990s. This was intended to facilitate data collection during commercial whale watch cruises where it was required that data could be entered in a standardised manner with a minimum of effort by people primarily involved in other activities (Leaper *et al.*, 1997). The original Logger software ran under the MS-DOS operating system. It automatically logged GPS and wind instrument data and also had a fixed number of forms for manual entry of effort, environmental and sightings data. Apart from the content of drop down lists, the forms could not be altered by the user. Following several years of successful use by a number of groups, a new version, Logger 2000[1], was developed with a much more flexible user interface, enabling users to create any number of data entry forms with user defined data entry fields and also interface to external hardware, such as sightings buttons, sound cards and video cameras. This has been used to collect survey and behavioural data on a number of studies in the past decade, (e.g. Gillespie *et al.*, 2005; Matthews *et al.*, 2001). This software forms the basis for the integrated data collection system developed for the SCANS-II survey in 2005 (SCANS-II, 2008) and subsequently used on the CODA survey in 2007 (CODA, 2009).

[1] The Logger 2000 software was developed by the International Fund for Animal Welfare to assist with benign research on cetaceans and is available to download from *http://www.ifaw.org/sotw*.

* *Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, KY16 8LB, UK.*
+ *International Fund for Animal Welfare, 87–90 Albert Embankment, London, SE1 7UD, UK.*

Even with the addition of real-time computer data entry, the majority of surveys still rely entirely on human observers to estimate and collect key data items, with limited scope for identifying or rectifying errors. This contrasts with most other fields of science and engineering where the use of calibrated instruments to take and record measurements is considered the norm. The SCANS-II system attempted to measure data wherever possible, to record data in ways that allowed errors to be identified, and allowed backwards comparability with previous surveys.

The problems of using subjective human judgement to estimate the radial distances and angles that distance based methods rely on have been identified in a number of studies (Leaper *et al.*, 1997; Schweder, 1997) but rarely accurately quantified. This is because attempts to quantify errors in judgement have relied on experiments that were not fully representative of the process involved (Williams *et al.*, 2007). In addition, survey data may be subject to recording errors, but during most surveys there is generally little scope for validating data beyond excluding values that are beyond the possible range. Systematic bias in distance and angle data will result in biased estimates of total abundance. In addition, random errors can affect both the accuracy and precision of estimates. Surveys using independent observers also rely on matching data on time, location and a description of what was seen in order to identify duplicate sightings. Failure to correctly identify duplicates can also cause bias in abundance estimates (Hammond *et al.*, 2002).

When an observer detects a sighting cue the key data are species, group size, time, distance and bearing together with ancillary data related to behaviour. For observers using binoculars mounted on a stand, a protractor on the stand is generally used to read out angles relative to the vessels heading and distances are measured through use of binocular reticules. Reticule readings allow the angle of dip between the horizon and the whale to be measured and this allows the distance from an observation platform of known height to be calculated. These measurements are difficult to make when the object (a surfacing cetacean for example) is only fleetingly visible and the boat itself is pitching and rolling in a seaway. For observers searching with the naked eye, angles are generally measured through use of an angle board (a protractor fixed to the vessel, with a movable pointer which the observer lines up on the animals or in the direction of their last observed location). Estimation by eye is the primary method of distance measurement, with observers generally being trained in distance estimation at the start of the survey using objects at known distances (e.g. a navigation buoy tracked with the vessels radar). Observers may also use sightings sticks – a marked stick held at arms length or attached round the observers neck with a horizon mark and marks indicating various distances as an additional aid to distance estimation.

For observers using binoculars, Leaper and Gordon (2001) describe methods for measuring distances and angles using video and stills cameras mounted on the same stands as the binoculars. The distance measuring system measures the angle of dip from the horizon in the same way that reticules or distance sticks do, but provide a record of the data which can be measured carefully offline and at much higher resolution than the other methods. The angle measuring system simply takes photos of fixed marks (typically parallel lines) on the deck beneath the binocular stand. For observers using only the naked eye it is much more difficult to develop systems which will capture an image that allows the location of the sighting to be measured. Trials with low resolution helmet mounted video cameras were not successful due to the movement of the human eye within the head requiring the camera to have such a wide field of view that few sightings could be detected on the images.

This paper describes both the Logger 2000 system in general terms, giving an overview of its functionality and how it may be used in a wide variety of applications as well as more specific information about the configuration employed during the SCANS-II and CODA surveys. A separate piece of software was developed for off-line data entry, validation, processing of verbal commentaries, and photo-grammetric measurements from images. The practical problems encountered and recommendations for development of systems for future surveys are also discussed. Results are given for the overall performance of the system, but comparison of data collected by different methods during the SCANS-II survey and the potential effects of measurement error on abundance estimation are the subject of a companion paper (Leaper *et al.*, 2010).

LOGGER 2000 SOFTWARE

The Logger 2000 software has been used since 1999 by a number of research groups for the collection of both visual and acoustic data. The Logger 2000 software provides the user with a flexible interface which can be configured in a variety of ways and a number of modifications were made specifically for the SCANS-II survey in order to deal with high volumes of data coming in from five different observers. These included the option to have multiple instances of the same form open at once, the opening of forms by remote 'action buttons' and automatic video and stills image capture.

The Logger 2000 software stores data collected automatically (e.g. from a GPS or other NMEA compatible wind and navigation instruments or temperature probes and other sensors linked to an analog to digital data acquisition device) or entered by the operator, in a Microsoft Access database. In its most basic configuration, the software does nothing but store GPS data on a timer (default every 10 seconds) and has a form for comments entered by the operator. Under normal use, however, a number of forms will have been configured for entry of other types of data, such as sightings and environmental data. The software can also be configured to make sound recordings, capture webcam images and display forms or buttons which can initiate a variety of actions.

User defined forms

User Defined Forms (UDFs) are used to specify the characteristics of all the non-standard forms in a particular Logger configuration. Their creation is realised by adding tables to the Access database with table names beginning with the characters 'UDF_' and having a pre-determined column format. On start-up, Logger reads the content of the UDF tables and for each UDF table, a form for data entry and an output database table are created. For example, the existence of the table UDF_Sightings would cause the creation of a data entry form on the Logger display called 'Sightings' and a corresponding table in the database also called 'Sightings'. Each row of data in the UDF table constitutes an instruction, which either sets a parameter governing the behaviour or appearance of the form, or creates a data entry field on the display panel and a corresponding column for data storage in the output data table. Any number of UDF tables may be created and the data entry forms are laid out on a tabbed panel, enabling the operator to navigate between them easily. If multiple forms of the same type are required to be open simultaneously (as is often the case with sightings forms) they are laid out on a secondary tab panel contained within the main Logger tab panel display. The Logger display also contains a map, showing the vessel's track, coastal outline and depth contours. Summary information from each data entry form (such as the locations of sightings) can be overlaid on the logger map and there is scope for customising line colours and symbols used for plotting contours, sightings and detections.

The main data entry types allowable within the Logger 2000 UDF system are detailed in Table 1. For a complete set of commands, readers are referred to the Logger 2000 Help (available once the software is installed). Some of the data types are either preset (such as counters) or can be collected automatically (e.g. time-stamps, NMEA data). Drop down lists of selectable items, such as species or observer names can also be created. Operators may create forms that mix both automatic data and manually entered data. For instance an environmental data form may contain an NMEA data field which collects wind speed directly from the ships instruments and have a separate field where the operator enters sea state. If a form is created which only contains data fields which are filled automatically, then the software can be configured to save those data automatically either on a timer or every time GPS data are read. All data recorded by Logger are automatically cross referenced in the database to the most recent GPS data record.

Sound recording

The software can be configured to contain one or more sound recorders which acquire sound data from PC sound cards or other data acquisition devices and store the data in wave files on a hard drive. While in standby mode, the recorders can be configured to write acoustic data continuously to a circular buffer so that when recording starts, several seconds of data prior to the actual recording start time are saved. The sound recorders were originally developed to record cetacean sounds from underwater hydrophones. Recording can be continuous (in which case new files are started at regular intervals to stop individual files becoming too large), can be made at user defined timed intervals or can be started and stopped manually or triggered automatically when Logger is used in conjunction with other acoustic detection program, or the buttons described below.

Video image capture

The software can be configured to capture still images from DV camcorders, USB webcams, PCI capture cards, TV cards, USB capture devices and IEEE 1394 (Firewire) cameras. The moving video images and captured stills are displayed in a window on a Logger form for quality checking. Images are grabbed and stored in either jpeg or bitmap format in response to a user pressing a button or to a timer that can operate at fixed or randomised time intervals. A sequence of images can be stored in a buffer (similar to the buffered sound recorder) so that images recorded prior to a trigger event (button press or timer) can be stored.

Action buttons

Logger 2000 can be also configured to contain one or more forms of action buttons. As well as existing on the screen as standard software buttons which can be clicked on using the mouse, the buttons can be linked to the keyboard function keys (F1 … F12 at the top of most keyboards) and can also be linked to external physical buttons via a digital interface card (e.g. Measurement Computing USB-1208LS[2] or similar).

Button configuration is realised in a similar way to the user defined forms with UDB_ tables in the database defining each form of buttons; each row in the UDB_ table specifies a button on the form. When a button is pressed, the time-stamp and button reference code are immediately stored

[2] Measurement Computing Corporation, 10 Commerce Way, Norton, MA 02766, USA. See: *http://www.mccdaq.com*.

Table 1

Logger data entry and command fields.

Data type	Description	Required configuration data
SHORT/INTEGER	16 or 32 bit signed integer data	
FLOAT/DOUBLE	32 or 64 bit double precision data	
CHAR	Character/text data	Maximum length
LOOKUP	Character data, selectable from a dropdown list such as a list of species or a list of observer names	A list of selectable items
COUNTER	Automatically incrementing integer number	
TIMESTAMP	Date and time data	
TIME	Time only	
NMEAINT/NMEACHAR/ NMEAFLOAT	Integer, character or floating point data from ships or instruments outputting data in National Marine Electronics Association (NMEA) format	NMEA sentence name and position within the sentence
ANALOG	An analogue voltage from either a Measurement Computing or a National Instruments data acquisition board	Channel number, channel gain, multiplicative and additive scaling factors to convert voltage (e.g. into a pressure measurement)
DIGITAL	A digital bit from either a Measurement Computing or a National Instruments data acquisition board	Channel number

in the associated output database table and a number of other actions can also be initiated:

(1) open single or multiple data entry forms;

(2) start a sound recording (the user sets which sound recorder to start and how long recording should continue for);

(3) start a series of video frame captures;

(4) send a command to a serial (RS-232) port;

(5) after a defined delay, automatically 'press' another button on the same or on a different button form. This allows sequences of commands to be created;

(6) broadcast software messages which can be picked up and acted on by other Windows programs.

Clock synchronisation

Data containing an accurate time-stamp are output by GPS systems at intervals varying between 0.5 and several seconds (more recent models tending to output data more frequently). Logger 2000 automatically updates the PC clock with a time-stamp from the GPS each time the software starts using one of the first data strings received from the GPS. All subsequent times used by the program are then read from the PC clock. This avoids either using a GPS time that is slightly out of date or having to wait for a new GPS time-stamp. During data collection, each GPS record is written to the database with both its own GPS time-stamp and the PC clock time so that any drift in the PC clock is detectable offline. In our experience, PC clocks are very accurate, it has never been necessary to correct PC clock times even when Logger has been running continuously and therefore not made further clock updates for many days. All times recorded in the database can therefore be considered accurate to better than 1s.

SCANS-II DATA COLLECTION

The SCANS-II survey was primarily designed to estimate the abundance of harbour porpoises and other small cetaceans in European shelf waters. The survey protocol closely followed that of the SCANS I survey of 1994 (Hammond *et al.*, 2002) using mark recapture distance sampling methods (Buckland and Turnock, 1992). Two teams of observers searched for animals. Two observers on the 'Tracker' platform used binoculars to search as far ahead of the survey vessel as possible, while two observers on the 'Primary' platform searched with the naked eye on either side of the track-line within 500m of the ship. The Tracker sightings were then used as trials for whether an animal first seen at a particular location was seen by the Primary observers. By maximising the detection range of the Trackers it was hoped that animals would be seen before they responded to the approaching vessel. In an attempt to maximise team efficiency over a range of different conditions, one Tracker used 7 × 50 and the other used 25 × 100 binoculars. Once an animal or group had been detected, Tracker observers attempted to record all subsequent surfacings in order to allow possible duplicates with the sightings from Primary observers to be identified. A fifth person acted as a Duplicate Identifier, receiving data in real time from all observers and making a judgement as to whether or not sightings were duplicates. A sixth person acted as a Data Recorder. Mark recapture distance sampling implemented in this way requires that the Primary observers have no knowledge of detections made by the Tracker platform, but the Trackers can be aware of Primary sightings. The Trackers, Duplicate Identifier and Data Recorder were therefore accommodated on the same platform on each vessel, whereas the Primary observers were stationed as far away as practically possible.

The primary goals of the SCANS-II data collection system were to:

Measure distances and angles wherever possible, rather than estimate them;

Automate data collection; and

Cross validate though multiple measures of critical data items.

Logger 2000 was used as the main data entry program. Data from the four observers and the Duplicate Identifier were audible to the Data Recorder sitting at the Logger computer. The Logger configuration contained the following data entry forms:

(1) primary Sightings (sightings from the primary platform);

(2) primary Resightings (resightings from the primary platform);

(3) tracker Sightings (sightings from the tracker platform);

(4) tracker Resightings (resightings from the tracker platform);

(5) effort (activity, observer and weather information);

(6) personnel Data (including eye heights for video range measurements);

(7) incidental Sightings (any other sightings made by non-observers).

In addition to the data entry forms, Logger was also configured to contain:

(1) two sound recorders (one for Primary and one for Tracker observers);

(2) a sightings button form;

(3) a webcam frame capture form for angle measurement (see below);

(4) a buttons form sending serial port commands to the video capture system used for distance measurement (see below).

The SCANS data collection system is shown schematically in Fig. 1. A data flow diagram, including actions and information from observers, the software and the data recorder is shown in Fig. 2.

Sightings buttons

In order to record sighting times accurately and to ensure that sightings were not missed by the Data Recorder, each observer was equipped with a microphone and two buttons, one for sightings and one for re-sightings. The buttons were mounted close to each observer, connected to the main data

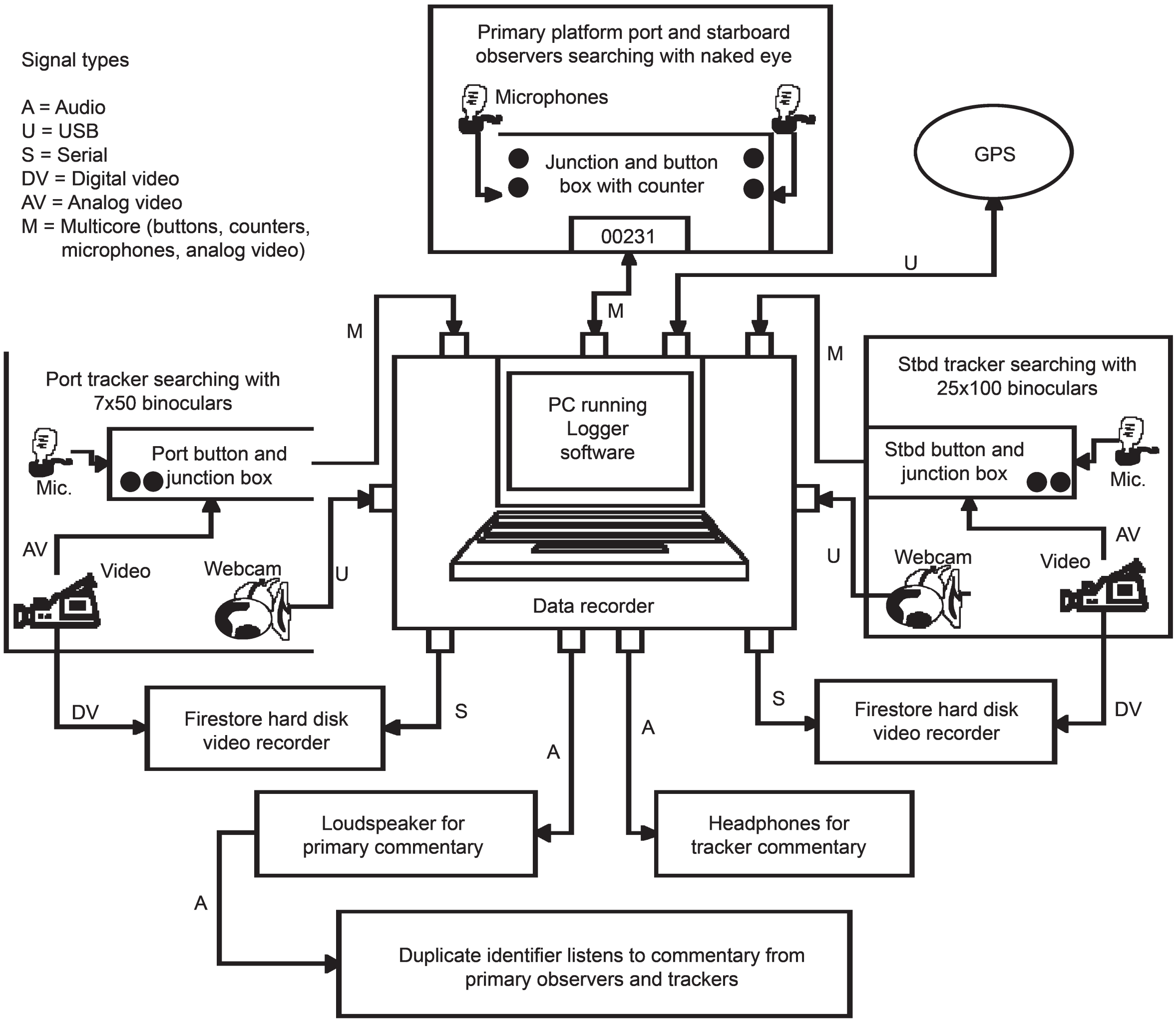

Fig. 1. Schematic diagram of the data collection system.

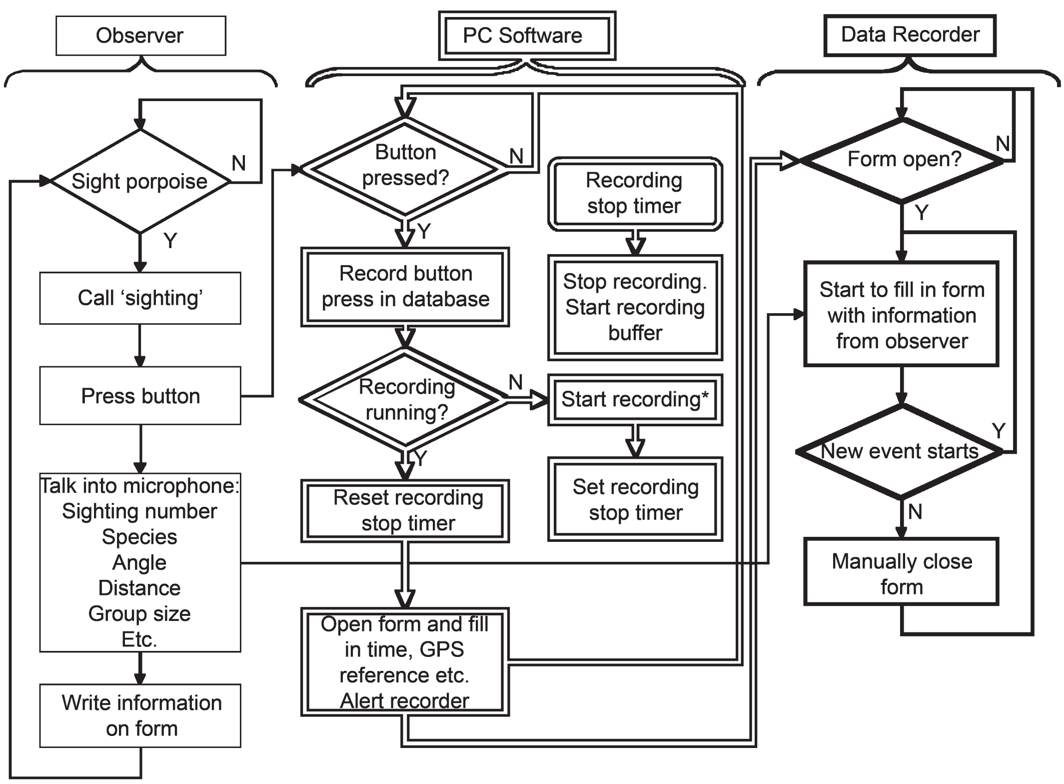

Fig. 2. Observer and data recorder actions and Software flow chart for the SCANS-II data collection system. *Tracker platform also takes a sequence of webcam photos for angle measurement and starts/stops video capture.

collection computer by cable and linked to Logger software buttons as described above. When a button was pressed, a number of actions took place within the Logger software:

(1) the date, time and button id was written to a table in the database;

(2) the appropriate sighting or resighting form opened on the Logger display, with fields filled in with the time, the button id and the incremental sighting/resighting number;

(3) the appropriate sound recorder started to record a voice track or, if it was already running, the timer controlling the recorder stop time would be reset.

If the button was a Tracker sighting or resighting button, the following actions also took place:

(1) video recording started (see below), or if it was already running, the stop timer reset so that recording would continue for a further 6s;

(2) a sequence of webcam images (see below) for angle measurement were stored.

Sighting numbers

The Primary platform button box also contained a counter which showed the same incremental sighting number as the Logger sighting form. The observers included this number in their commentary for cross-referencing during data validation. Tracker platform observers, who were stationed close to the data recorder, were given sighting numbers by the data recorder so that they did not have to take their eyes from the binoculars.

Sound recording

Two sound recorders were configured in Logger. Each had a 10 second data buffer allowing it to acquire voice commentary prior to each button press and would record for two minutes after each button press before stopping. Each sound recorder recorded data from a separate external USB sound card (Edirol UA20) and was configured to write stereo 16 bit wav files so that a separate channel could be used for each observer: the port and starboard Primary observers using the left and right channels of one recorder and the two Trackers the two channels of the other. Microphones were generally tucked inside the clothing of each observer to keep them out of the wind. The Data Recorder monitored the headphone output of the sound cards to listen to the Primary platform observers. The Data Recorder was generally close enough to the Tracker observers to hear them, but would use headphones or speakers on the output of the tracker sound card if required.

Observers were instructed to give information in the same order that data entry fields appeared in the Logger forms. The Data Recorder could talk back directly to the Tracker observers. A two-way radio was used to talk to the Primary observers.

Video capture and range measurement

When using the photo-grammetric video range methods described in Leaper and Gordon (2001) an animal on the captured video image may only be a few pixels in size. The video is therefore never used to detect an animal and identifying it on the video can usually only be reliably achieved by using information from the observer viewing the animal through binoculars either in the form of an audio commentary or some other accurate time-stamp to indicate an animal surfacing. Previous experiments had shown that it was not possible to rely on the video system to measure distances to 100% of sightings and so reticule readings were always recorded as well. The aim was to obtain sufficient measurements from the video to be able to measure any bias in the reticule readings made when video measurements were not available. In addition, comparison of measured distances and visual estimates from naked eye or reticules was informative in the context of other surveys.

The major factors in the choice of video camera for distance measurement are the quality of still images that can be achieved and the field of view of the lens. A narrow field of view maximises the size of distant cues in terms of pixels but at the risk of missing either the horizon or the cue. A wider field of view allows measurements from closer cues, but at lower resolution. Leaper and Gordon (2001) used a video camera with a field of view 2.7° vertically which was narrower than the 7° field of view of the 7 × 50 binoculars, but this was compensated for by a natural tendency for observers to place the object of interest in the centre of the binoculars field of view. For the SCANS-II survey, Canon XM1 cameras were used on the 25 × 100 binoculars and Sony HC90E camcorders on the 7 × 50 binoculars. These had a slightly wider field of view and were chosen in order to allow tracking closer to the vessel and to make the system easier for less experienced observers. The camera used on the CODA survey was the Canon HV20 high definition (HD) camera with a vertical field of view of 4.9°. On the 25 × 100 binoculars the same camera was used with a Canon TL-43 2× converter giving a vertical field of view of 2.9°. This HD video camera gave an effective image resolution of 1920 × 1080 pixels which was a considerable improvement over the 720 × 576 pixels of standard digital video used on SCANS-II. All cameras used had the progressive scan facility such that both interlaced fields in the video image could be captured simultaneously and shutter speeds were set to 1/1000s or faster. It was found that auto-focus systems were often not effective when scanning and so the camera was always set to manual focus at ∞ (a camera which has a control to set the focus to ∞ is much easier to use than one where this has to be done through manual adjustment).

In previous implementations, digital video tape was used to store data. The obvious disadvantages of using tape is that it must either be left running continuously, which generates vast quantities of data to review following the cruise, or recording must be started once animals are sighted. In this latter case the initial surfacing will be missed because the cameras generally enter a standby mode or turn off completely if not operated for a few minutes and it can take several seconds to re-load the tape and start recording when the record button is pressed.

Although the Logger software can capture sequences of still images from video and buffer sequences of images, it cannot perform video buffering in order to store video prior to a button press. Therefore a hard drive based video capture system was used (Firestore FS-4 from Focus Enhancements). These units are basically a computer hard drive configured to store data from the IEEE 1394 (Firewire) output of a digital video camera. Data are stored as video files which can be uploaded to a computer for editing or analysis. The great advantage of the Firestore units is that, like the sound recording system in Logger, they can be set to write data continuously to a 6s circular buffer so that when recording is started, data are recorded from 6s prior to the operator starting the system. This allows the first surfacing of any sighting to be captured on video which is particularly useful for harbour porpoises. The Firestore units have a serial (RS-232) interface connection which was used to start and stop recordings and set up the buffered operation. Sequences of Logger buttons were programmed to send the necessary sequences of commands to the Firestore units in order to make a 12s recording for each Tracker sighting, which would start 6s before the button press and end 6s after it. In the event of multiple resightings occurring within 6s of each other, recording would continue until 6s after the last button press.

Commentary from the camera microphone, or an external microphone mounted on the camera, was also recorded in the Firestore data. The commentary was the same as that recorded by Logger (but truncated to 6s either side of the button press). Generally, it was found that identifying surfacings within the short 12s video clips using the commentary, or the simple expectation of a surfacing about five seconds into each video clip, was considerably easier than searching for surfacing in the longer taped sequences captured in previous studies.

Accurate alignment of the video camera with the binoculars on which it is mounted is critical for successful video distance measurement as is the need for the horizon to be visible in every image. For the 7 × 50 binoculars a custom mount was designed that located the video camera above the binoculars and also held the webcam for bearing measurement (Fig. 3). This was supported on a monopod with a tripod ball head which allowed the observer to move freely to compensate for the motion of the vessel. The 25 × 100 binoculars were mounted on a fixed stand with the video camera mounted on top of the binoculars. There were no easy attachment points for mounting the camera and so a plate was fixed to the binoculars with steel bands which passed around the whole binocular body. On the SCANS-II survey, a custom mounting was built to allow adjustment of the alignment of the camera while on the CODA survey geared tripod heads (Manfrotto Junior Geared Head) were used. These proved easier to use, allowing small alignment corrections to be made when necessary.

It was often not practical to view the screens of the digital video cameras, so the analogue video output from each camera was fed back to the Data Recorder position and input to a small monitor. The Data Recorder could therefore make periodic checks of vertical and horizontal camera alignment. A video switch was used to monitor both cameras alternately with a single monitor.

Angle measurement (webcam capture)

Leaper and Gordon (2001) describe a method of angle measurement using downward pointing digital stills cameras mounted on Tracker binoculars which are used to photograph marks (generally parallel lines) on the deck below the observer. This basic methodology was updated for the SCANS-II survey by using the video capture utilities in Logger to acquire images from low cost webcams. When a Tracker sighting or resighting button was pressed, a series of eleven webcam images were stored, one per second from five seconds prior to the button press to five seconds after it. Images were also acquired at random time intervals in order to investigate observer scanning patterns.

Implementation

The data collection system described above is relatively complicated, requiring a computer with a number of external interfaces and cables to each observer position to carry button, audio, digital and analogue video signals and webcam images. Mains power was also required to run the computer, the Firestores and the video cameras.

The computer, sound cards, USB hub, USB to serial

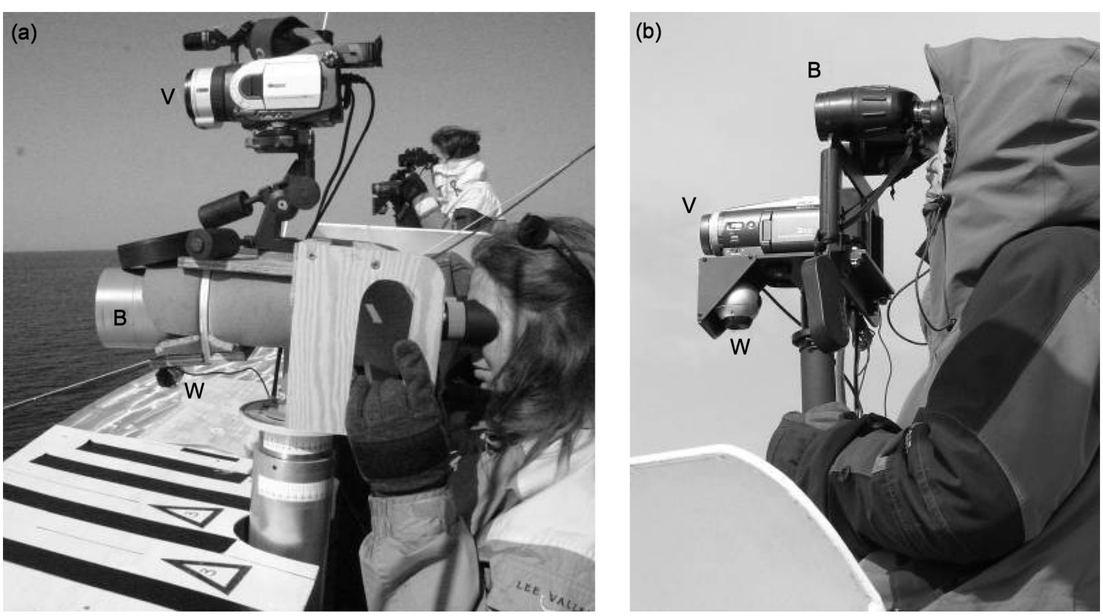

Fig. 3. (a) 25 × 100 and (b) 7 × 50 binocular stands on the tracker platform showing the webcams (W) the video cameras (V) and binoculars (B). Also visible are lines on the angle measurement board mounted below the webcam in front of the 25 × 100 stand.

adapters, button interface card and audio junction box were all built into an aluminium flight case which could be easily carried to the observation platform each day, provided a degree of environmental protection, and also provided a shield to aid viewing of the laptop screen in bright sunlight.

The system was built using standard components, requiring no specialist electronic expertise, the most complicated item being the junction boxes and cabling which combined audio and button signals into a single 10 or 15m long multi core cable to the two Primary observers and cables carrying button, audio and analog video to each of the Trackers. Digital video signals were carried down standard Firewire cables, which have a maximum length of 4.5m and the USB webcam signals were carried on 5m long USB extension cables which restricted the arrangement of the Trackers and data recorders on the platform.

In the event of inclement weather it was necessary to shut down the system quickly because the cameras and much of the other equipment used was not in any way water resistant. Most could be put into the aluminium case with the computer.

Two important aspects of the data collection system were: (a) that it would be backwards compatible with the data collection system employed in the first SCANS survey of 1994 (Hammond *et al.*, 2002), so that a direct comparison of the two surveys could be conducted; and (b) that it would contain sufficient redundancy that data collection could always continue. Thus, in the event of damage to the cable from the Primary platform, two way radios could be used for communication with the data recorder and an audio output from the radio at the data recorder station input to the sound card to record the Primary platform voices, with the software buttons in Logger being used in place of the wired buttons. The Trackers were within talking distance of the Data Recorder and the software buttons could again be used should the wired buttons fail. Video range tracking was very much an add-on to the survey protocol, and had no backup beyond the binocular reticule measurements. Compass roses and pointers on the binocular stands could be used to read angles manually. All vessels were also provided with paper forms for data recording. The laptop used was a standard, non specialised model, so in principle the software could have been reinstalled on a new machine, without specialist help, should the need have arisen.

Testing and training

The system was tested, reviewed and where necessary modified during a dedicated two week pilot survey in late April, 2005. This left a two month period prior to the main SCANS-II survey for construction of hardware interfaces and software development. Cruise leaders from each survey vessel were also trained in using the system during the pilot survey. Additional tests and training of one additional person per vessel were conducted during a one week passage on a platform of opportunity in early June. Cruise leaders were responsible for training other observers on each vessel.

DATA VALIDATION

At sea data validation

One of the potential benefits of entering as much data as possible into a computer in real-time is that it allows for data validation algorithms to flag potential problems that may be possible to correct while the details of the sighting are still fresh in the observers' minds. This type of processing needs to be balanced against the need for efficient and reliable data entry. For SCANS-II, with so many other new components to the data collection system, it was decided that data validation should be done entirely off-line in order not to compromise or further complicate the data collection task.

An additional requirement of the system was that at the end of the survey there should be a database with complete data on each sighting including measurements of distances and angles and that no further processing should be required (such as watching video or listening to commentaries) to extract the basic data. The data recording system was designed such that the Data Recorder would enter as much data for each sighting as possible. However, if there were several events happening at the same time then data items could not all be entered in real time and would need to be entered from recordings at the end of the day. In these circumstances the Data Recorder's main task was to monitor the commentaries and check that they were clear (e.g. all microphones were working and well positioned) and to remind observers if they had missed key data items. At the end of each observation session there were thus a full set of sightings and resightings forms, but often with gaps in some data fields. Hence an off-line data entry and validation system was developed which allowed observers to listen to commentaries, analyse video and measure bearings from the webcams at the end of the survey period each day. For each record the validation software listed possible problems as 'errors' or 'warnings'. 'Errors' were problems with the data that would preclude standard distance analysis (e.g. distance, angle, species) and 'warnings' were problems with ancillary data (e.g. swim direction, cue type).

As well as identifying blank fields, errors and warnings were triggered by the validation algorithms if identified values were outside of predetermined ranges for all parameters, inconsistencies between the observer reporting the sighting and the effort status, or inconsistencies in duplicate status.

For records that had errors or warnings, the observer listened to the verbal commentary to fill in blanks and try to resolve any discrepancies. The record was then saved with a code indicating whether there were no warnings or errors, whether further processing was planned, or if the data could not be resolved.

The validation software could be run simultaneously on multiple PC's, all networked to the central Logger laptop containing the database. Some vessels ran a network cable from the Logger laptop during data collection so that observers could process and validate their recent sightings during rest periods, immediately after coming off watch as other observers continued data collection.

The validation software also allowed cruise leaders to extract summary statistics of the type of errors that were occurring and plots of distances and bearings (diagnostics similar to those recommended by the International Whaling Commission Scientific Committee for cruise reports). These could be selected by species or observer and were particularly intended to identify problems such as rounding in estimated values that might be corrected during the survey.

Post-survey data validation

The database from each vessel contained records with two values for some parameters for many of the sightings (e.g. range from reticule, range from video, angle from angle board, angle from webcam). These included estimated and measured values of angles and distances. These were not directly compared by the data validation software at sea in order to allow for a more informative post-cruise analysis of errors. Comparisons of distances and bearings indicated a variance associated with estimation error but also occurrence of major errors which were assumed to be due to mistakes in data entry. Visual inspection of plots suggested that checking the 10% of sightings with the largest discrepancies should capture the majority of these gross errors. Although the photo-grammetric systems allowed for careful measurement, this was still done by an operator and so there was scope for error in these measurements too. All video and still images used for distance and angle measurement were linked to the database and so could be retrieved and re-measured. In cases where the image was clear and a discrepancy remained, it was assumed that the error was in the estimated distance or angle.

OVERALL PERFORMANCE OF THE SYSTEM

The full system was used on seven vessels on the SCANS-II survey in 2005, and five vessels on the CODA survey in 2007. In addition, systems were supplied to three vessels on the 2007 T-NASS survey of the North Atlantic. Components of the system have also been used on the IWC SOWER cruises.

Video distance measurement

The proportions of sightings that were successfully captured on video such that distance measurements could be made are given in Table 2. Success rates for the 7 × 50 and 25 × 100 binoculars were similar but varied considerably among vessels as a result of different conditions experienced and some technical problems. One vessel on the CODA survey experienced a total technical failure of the video equipment. If these data are discounted then the overall success rate for the CODA survey (66%) was higher than that for SCANS-II (37%). This was probably due to the use of high definition video cameras that resulted in much better image quality meaning that fewer surfacings were missed due to camera resolution and the fact that harbour porpoises, which made up the vast majority of sightings during SCANS-II but were absent on CODA, were particularly challenging subjects.

The most common problems encountered were with control of the Firestore hard-disc recording units; it was later found that these function more reliably on mains power if their internal rechargeable battery is removed. On one vessel on the CODA survey these failed completely which seemed to have been mainly a result of a failure of communication between the PC and the Firestore.

Bearing measurement

On the SCANS-II survey, the bearing cameras generally worked well, with an overall 94% success rate. On CODA there were more problems due to hardware conflicts related to the number of USB devices connected to the computer resulting in a lower success rate of 85%. Achieving a high success rate of bearing measurement using webcams should be possible, however recent developments in other angle measurement devices (e.g. magnetic sensors) may ultimately give better results.

The light levels during surveys varied from very dull conditions to bright sunlight. The video cameras used for distance measurement were designed for such a range of conditions and the automatic exposure compensation worked well. The webcams were generally designed for indoor use under artificial light and needed additional filters over the

Table 2

Number of Tracker sightings for all species and the percentage of sightings recorded for each measurement from SCANS-II and CODA surveys.

Survey	Observer	Vessel code	Number of sightings	Angle		Distance	
				Estimated %	Measured %	Estimated %	Measured %
SCANS-II	25 × 100	GO	128	98.4	86.7	98.4	17.2
		IN	91	95.6	95.6	89.0	51.6
		MC	20	45.0	100.0	100.0	0.0
		SK	77	97.4	90.9	100.0	36.4
		VH	108	99.1	100.0	100.0	73.1
		WF	57	93.0	93.0	86.0	22.8
		ZI	86	95.3	94.2	98.8	8.1
SCANS-II	7 × 50	GO	144	95.1	93.8	100.0	14.6
		IN	124	97.6	95.2	94.4	67.7
		MC	66	51.5	86.4	100.0	10.6
		SK	61	98.4	98.4	100.0	21.3
		VH	97	97.9	97.9	97.9	62.9
		WF	52	98.1	98.1	82.7	51.9
		ZI	100	100.0	92.0	99.0	41.0
	Total	**All**	**1,211**	**93.9**	**94.0**	**96.7**	**37.2**
CODA	25 × 100	IN	147	99.3	83.7	100.0	68.7
		MC	38	92.1	84.2	86.8	63.2
		RA/GE	100	100.0	88.0	100.0	0.0
CODA	7 × 50	IN	345	98.8	80.6	98.8	67.8
		MC	71	97.2	95.8	95.8	69.0
		RA/GE	142	97.9	90.1	98.6	0.0
	Total	**All**	**843**	**98.5**	**85.1**	**98.3**	**48.4**

lens to prevent overexposure in bright sunlight. Dark lenses from cheap sunglasses were found to be an effective form of filter.

An additional advantage of the bearing measurement system was that bearings were collected at random intervals (with a mean interval of 30s) in order to examine the scanning patterns of observers. Detailed analyses of scanning patterns and relative sighting rates are the subject of further analyses but simple plots made during surveys could be used to identify whether observers were scanning the appropriate angle sectors.

Data validation and workload for observers

Unlike the Logger software which had evolved over more than 10 years with considerable feedback from users, the validation software was written specifically for the SCANS-II survey. The software suffered from a lack of flexibility which meant that changes implemented during the pilot survey involved writing new code. Thus the first real test of the system was the survey itself. The system performed adequately in allowing the playback of audio recordings and measurements from images. However, this did prove a very time consuming process, especially for observers who had not seen the system before. Although several improvements that would speed up data entry were identified, the main problems that caused most lost time were small bugs and glitches. Vessels varied in their ratio of survey time to bad weather time and while some teams were able to keep up with data entry and validation, for others it proved a rather onerous task.

DISCUSSION

The Logger 2000 software has been used for several years on a variety of projects and has provided a reliable and flexible method for semi-automated data collection during line transect surveys, mitigation monitoring, photo-identification and behavioural studies of marine mammals. The system of user defined forms allows different users to configure the software in different ways without having to modify the program code itself. However, in order to deal better with high volumes of concurrent data coming in from five different people and to improve the way in which sightings data are recorded more generally, a number of modifications were made to the Logger code specifically for the SCANS-II survey. All of these modifications have been included in software releases since the SCANS-II survey and are freely available. Information on circuits for external interfaces are available from the authors on request.

Although accurate input data are clearly critical for line-transect surveys and serious biases can be caused by measurement error, surprisingly little attention has been given to data collection compared to that devoted to data analysis methodology (Williams *et al.*, 2007). In addition, experiments that have attempted to quantify range and bearing measurement error have used static targets which are unlikely to provide realistic error data. The data collection methods used on the SCANS-II and CODA surveys generated range and bearing measurements for a proportion of sightings allowing both better abundance estimation from these surveys and also comparison of estimated and measured distances and angles under real survey conditions. Such comparisons are of more general value in revealing the likely extent of measurement error on other surveys that have used conventional reticule binoculars and angle-boards. Full analyses of these results are described in Leaper *et al*. (2010).

The data collection system worked effectively on all seven vessels taking part in the SCANS-II survey, although the complexity of the system and the large number of interconnected components working in a harsh environment required a certain level of enthusiastic vigilance on the part of the operators to keep it running. The most commonly encountered problems were with the video capture system. This appears to be due to either the Firestore units overheating or a failure of communication between the video cameras and the Firestores or the Firestores and the computer. At the time of the survey, the Firestore units were a new and innovative product. It is likely that this technology will become more standard, and therefore cheaper and more reliable in the future. During the CODA survey, problems were encountered with the webcams used for angle measurement. This appears to be due to the use of newer webcams sending higher quality data and also an audio signal to the PC, which overloaded the USB system. We also note that the Edirol sound cards used in 2005 are no longer manufactured and we have yet to identify a replacement with the same specification. Recreating the system for future surveys would therefore require a certain amount of re-development in order to recreate a stable system.

As computer hardware capabilities develop it is likely that the optimum means of implementing a system like this may change more fundamentally. For example, some of the rather cumbersome cabled connections used here might be replaced by wireless links.

The Logger 2000 software is written in C++ and runs on a PC under Windows. The software itself is not open source and cannot therefore be modified by the operator. We hope, in the future, to incorporate Logger features into the PAMGUARD software (*http://www.pamguard.org*; Gillespie *et al.*, 2008) which is both free and open source and is also more likely to be supported in the future. The validation software is coded in Visual Basic and was written specifically for the SCANS-II survey. Unlike Logger, the database structure used by the validation software is fixed, so the validation system can currently only be used with the Logger SCANS-II configuration. For future systems, validation should have the same flexibility as Logger in terms of user defined data base structure based on the user's choice of forms.

ACKNOWLEDGEMENTS

Development of the Logger software was funded by the International fund for Animal Welfare to promote benign and non-invasive research. The SCANS-II survey was supported by the EU LIFE Nature programme, project LIFE04NAT/GB/000245 and the governments of all 12 range states: Belgium, Denmark, France, Germany, Ireland, The Netherlands, Norway, Poland, Portugal, Spain, Sweden and the UK. We are particularly grateful to all those who have provided feedback and suggestions for the Logger software over the years, to cruise leaders and other members of the SCANS Data collection team, to the observers who kept the system running through sometimes difficult conditions and to Phil Hammond for comments on earlier drafts of this paper.

REFERENCES

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford, UK. vi+xv+432pp.

Buckland, S.T. and Turnock, B.J. 1992. A robust line transect method. *Biometrics* 48: 901–09.

CODA. 2009. Cetacean Offshore Distribution and Abundance in the European Atlantic. SMRU, Gatty Marine Labratory, University of St. Andrews, St. Andrews, Fife, KY16 8LB, UK. (Available at: *http://biology.st-andrews.ac.uk/coda/*).

Gillespie, D., Berggren, P., Brown, S., Kuklik, I., Lacey, C., Lewis, T., Matthews, J., McLanaghan, R., Moscrop, A. and Tregenza, N. 2005. Relative abundance of harbour porpoises (*Phocoena phocoena*) from acoustic and visual surveys of the Baltic Sea and adjacent waters during 2001 and 2002. *J. Cetacean Res. Manage.* 7(1): 51–57.

Gillespie, D., Gordon, J., McHugh, R., McLaren, D., Mellinger, D., Redmond, P., Thode, A., Trinder, P. and Deng, X.Y. 2008. PAMGUARD: Semi-automated, open source software for real-time acoustioc detection and localisation of cetaceans. *Proceedings of the Institute of Acoustics* 30.

Hammond, P., Benke, H., Berggren, P., Borchers, D.L., Buckland, S.T., Collet, A., Heide-Jørgensen, M.P., Heimlich-Boran, S., Hiby, A.R., Leopold, M. and Øien, N. 2002. Abundance of harbour porpoises and other cetaceans in the North Sea and adjacent waters. *J. Appl. Ecol.* 39: 361–76.

Leaper, R., Burt, L., Gillespie, D. and MacLeod, K. 2010. Comparisons of measured and estimated distances and angles from sighting surveys. *J. Cetacean Res. Manage.* 11(3): 229–237.

Leaper, R., Fairbairns, R., Gordon, J., Hiby, A., Lovell, P. and Papastavrou, V. 1997. Analysis of data collected from a whalewatching operation to assess relative abundance and distribution of the minke whale (*Balaenoptera acutorostrata*) around the Isle of Mull, Scotland. *Rep. int. Whal. Commn* 47: 505–11.

Leaper, R. and Gordon, J. 2001. Application of photogrammetric methods for locating and tracking cetacean movements at sea. *J. Cetacean Res. Manage.* 3(2): 131–41.

Matthews, J.N., Brown, S., Gillespie, D., Johnson, M., McMclanaghan, R., Moscrop, A., Nowacek, D., Leaper, R., Lewis, T. and Tyack, P. 2001. Vocalisation rates of the North Atlantic right whale (*Eubalaena glacialis*). *J. Cetacean Res. Manage.* 3(3): 271–81.

SCANS-II. 2008. Small Cetaceans in the European Atlantic and North Sea. Final report to the European Commission under project LIFE04NAT/GB/000245. Available from SMRU, Gatty Marine Laboratory, University of St. Andrews, St. Andrews, Fife, KY16 8LB, UK. (Available at: *http://biology.st-andrews.ac.uk/scans2/*).

Schweder, T. 1997. Measurement error models for the Norwegian minke whale survey in 1995. *Rep. int. Whal. Commn* 47: 485–88.

Williams, R., Leaper, R., Zerbini, A. and Hammond, P.S. 2007. Methods for investigating measurement error in cetacean line-transect surveys. *J. Mar. Biol. Assoc. U.K.* 87: 313–20.

Date received: June 2009
Date accepted: May 2010