SC/48/O 15

A Strategy for Adaptive Sampling in Shipboard Line Transect Surveys

J.H. Pollard and S.T. Buckland

School of Mathematical and Computational Sciences, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK

ABSTRACT

This paper considers an adaptive approach to shipboard line transect surveys, in which the total effort available is fixed. This strategy is aimed at increasing sampling efficiency for sparse but highly clustered populations, whilst recognising the practical and financial restrictions of a shipboard survey. Effort is increased by initiating a zigzag trackline when the number of observations exceeds some pre-set limit and a basic formula is provided to calculate the appropriate effort increase, known as the *effort factor*. The formulae derived in the paper are conditional on the effort factors, and are therefore not *design-unbiased* (Thompson, 1992, p.17). A simulation program is used to test the theory and to estimate the efficiency compared with conventional line transect methods. Simulation results, for clustered populations, demonstrated reduced density estimate variance for adaptive surveys, compared with conventional line transects, and that the efficiency increased as the clustering increased. Mean efficiency increases of 1.07 were recorded for highly clustered populations and, as expected, a loss in efficiency (0.96) was detected for populations which exhibited complete spatial randomness.

KEYWORDS: SURVEY-VESSEL; METHODOLOGY; SIMULATION

INTRODUCTION

Adaptive sampling allows survey effort to be increased in areas of high animal density. This leads to larger numbers of detections, and potentially higher precision, than for conventional designs. A disadvantage is that it is difficult to predict in advance the total survey effort required for an adaptive design. Shipboard line transect surveys are expensive, and the number of ship-days is usually predetermined, so a design in which total effort is a function of the number of animals detected is unlikely to be acceptable. Further, a proportion of the planned survey effort is typically lost, for instance through bad weather, and a mechanism to reallocate the remaining effort is desirable.

We develop an adaptive design for shipboard line transect surveys in which the number of ship-days is fixed. A minimum amount of survey effort, termed the nominal effort, is predetermined, and the degree to which survey effort increases in areas of high density is a function of the difference between the total effort still available and the nominal effort remaining.

The survey effort is adapted by increasing the effort, above the nominal straight line effort, when the number of observations exceeds some limit. The increased effort is achieved by zigzagging for a period, after which the ship returns to the nominal (straight line) cruise track (Fig. 1). The increase in effort is measured by the *effort factor*, which is the ratio of the length of transect line travelled to the nominal (straight line) length. Thus a transect may be divided into a number of sub-transects, or *legs*, each with a different effort factor. The formulae derived in this paper are conditional on the effort factors, and are therefore not *design-unbiased* (Thompson, 1992, p.17). Simulations show that little bias is introduced by conditioning on the effort factors, and this relaxation of the requirements of adaptive sampling allows a much wider class of designs than those considered by Thompson (1992, p.263-318).

The formulae are derived for passing-mode surveys only, where the ship does not detour to investigate observations. It may be possible to modify the formulae to accommodate closing-mode surveys, but this has not been investigated.

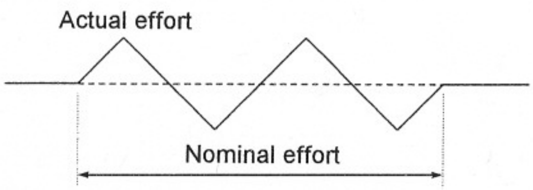

Fig. 1. Actual effort is increased by zigzagging, nominal effort refers to the equivalent straight line track.

METHODS

Adapting the nominal effort

The mechanism used to increase effort is by zigzagging when the number of observations exceeds some limit. After a number of zigzags the survey then reverts to a straight line again. The ratio of the zigzag effort to the respective straight line effort is defined as the effort factor, and denoted by λ.

The change in effort can be varied by altering the zigzags, either in length, angle, number or a combination of all these. Thus the adaptive component can be modified as the survey progresses to allow it to be completed, using fixed effort.

The adaptive scheme ensures that greater effort is expended in areas of higher animal density. If analysis is carried out ignoring this, abundance is overestimated. This is avoided by downweighting data from zigzag sections in the analysis. The weight is inversely proportional to the effort factor, so that each section of transect is given weight in proportion to the length of *nominal* effort through that section.

We define the nominal number of observations in a section to be the number of detections had the nominal search effort been carried out. This is estimated in zigzag sections by dividing the actual number by λ.

Notation

Nominal values refer to the values expected if a conventional straight line transect is followed. The nominal effort is signified by a dash, such as L', whereas corresponding actual effort is denoted by L. A detection may consist of one or more animals; throughout this paper 'school' is used to indicate the target detected and 'school size' to indicate the number of animals within a target.

Each transect is divided into a number of sub-transects or legs, where the start and finish of each leg occurs at a change in direction (Fig. 2). Typically the straight sections within a zigzag will all have the same angle, and hence effort factor, so the complete zigzag section can be considered as a single leg (Fig. 3). This simplifies the labelling of transect components.

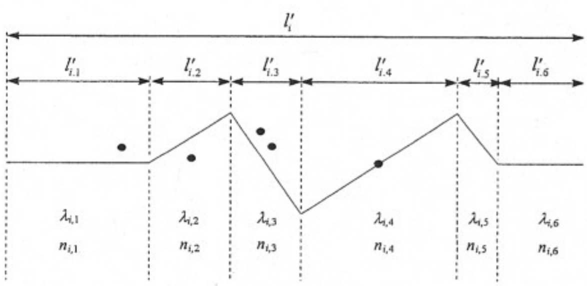

Fig. 2. Notation used for an adaptive transect, where the angle is varied at each change of direction within a zigzag section.

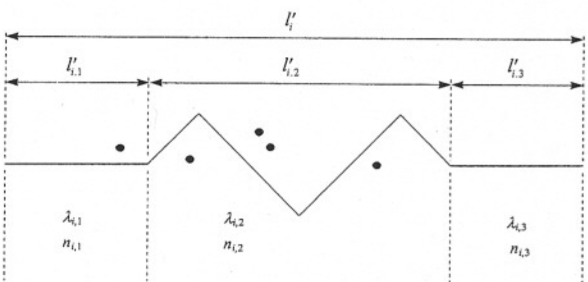

Fig. 3. Simplified notation when the angle of the zigzags remains constant for the length of the zigzag section.

Where present, subscript i represents the i^{th} transect, $i = 1...k$, and subscript j represents the j^{th} leg within a transect, $j = 1...m_i$.

- L is the total actual effort;
- L' is the total nominal effort (i.e. the straight line effort in the absence of zigzags);
- l_i is the actual effort for the i^{th} transect;
- l'_i is the nominal effort for the i^{th} transect;
- l_{ij} is the actual effort for the j^{th} leg of the i^{th} transect;
- l'_{ij} is the nominal effort for the j^{th} leg of the i^{th} transect;
- k is the total number of transects.

Similarly: n is the total number of schools detected; s the observed school size; λ the effort factor; e the encounter rate; and D the population density. So:

- $E(n)$ is the expected number of schools detected in the surveyed area;
- $E(n_i)$ is the expected number of schools detected for the i^{th} transect;
- n_{ij} is the number of schools detected for the j^{th} leg of the i^{th} transect;
- $E(s)$ is the expected school size for the population;
- $E(s_i)$ is the expected school size for the i^{th} transect;
- $E(s_{ij})$ is the expected school size for the j^{th} leg of the i^{th} transect;
- s_{ij} is the mean observed school size for the j^{th} leg of the i^{th} transect;
- s_{ijX} is the school size for the X^{th} ($x = 1...n_{ij}$) observation for the j^{th} leg of the i^{th} transect;
- λ_{ij} is the effort factor for the j^{th} leg of the i^{th} transect;
- e_{ij} is the encounter rate for the j^{th} leg of the i^{th} transect;
- D is the population density (animals per unit area);
- D_i is the population density for the i^{th} transect;
- $f(0)$ is the value of the probability density function of perpendicular distances to detections, evaluated at zero distance.

The effort factor as a function of survey schedule

The change in effort is represented by the effort factor, λ, where

$$\lambda = \frac{\text{Actual Effort}}{\text{Nominal Effort}}$$

Thus the effort factor for the j^{th} leg of the i^{th} transect is given by

$$\lambda_{ij} = l_{ij} / l'_{ij}$$

The effort factor is calculated based on the excess effort available (i.e. any effort over and above the nominal effort required to complete the survey) and the expected number of times the effort will be increased (i.e. the expected number of times the observer will enter zigzag mode).

The excess effort available, at a given point in time, is any effort remaining after subtracting both the effort used and the nominal effort remaining from the total effort available for the survey, L. Let:

- L_U be the total actual effort used at any point in time;
- L_E be the amount of excess effort available at any point in time;
- L'_R be the nominal effort remaining at any point in time;
- ξ be the expected number of times the actual effort will increase above its nominal level at any point in time.

Then

$$L_E = L - L_U - L'_R$$

Let the increase in effort, following an observation, be the excess effort available, L_E, divided by the expected number of times effort will increase above its nominal level, ξ, plus the current increase. So the increase in effort for a leg is given by

$$l_{ij} - l'_{ij} = L_E / (1 + \xi)$$

By definition, $l_{ij} = l'_{ij} \cdot \lambda_{ij}$, so

$$l'_{ij} \cdot (\lambda_{ij} - 1) = L_E / (1 + \xi)$$

thus the effort factor is given by

$$\lambda_{ij} = 1 + \frac{L_E}{l'_{ij} \cdot (1 + \xi)} = 1 + \frac{(L - L_U - L'_R)}{l'_{ij} \cdot (1 + \xi)}$$

When each effort increase is applied for a fixed distance along the nominal trackline, then ξ can easily be calculated from an estimate of the encounter rate. Let l'_Z be the nominal effort over which the effort increase is applied and γ be the encounter rate estimate, which might be obtained from previous surveys or a best guess provided by the user. Then

$$\xi = \gamma \cdot (L'_R - \xi \cdot 1'_Z)$$

so

$$\xi = \frac{\gamma \cdot L'_R}{1+\gamma \cdot l'_Z}$$

Thus when effort is increased over a fixed distance along the nominal trackline l'_Z (i.e. $l'_{ij} = l'_Z$ for all i,j), then the effort factor is calculated as

$$\lambda_{ij} = 1 + \frac{\left(L - L_U - L'_R\right)}{l'_Z\left(1 + \frac{\gamma \cdot L'_R}{(1+\gamma \cdot l'_Z)}\right)}$$

Assumptions

In deriving these estimation equations, the following standard line transect assumptions are made:

(a) probability of detection on the line, $g(0)$, is 1;
(b) there is no size bias (the probability of detection is independent of the school size);
(c) there is no responsive movement of schools in advance of detection, and any non-responsive movement is slow relative to the speed of the ship.

These assumptions could be weakened or removed using similar strategies as for conventional line transect sampling. Additional assumptions are:

(d) the expected encounter rate for a zigzag track is the same as the expected encounter rate for the corresponding nominal (straight line) track;
(e) the expected school size for an observation on a leg following a zigzag track is the same as the expected school size for an observation when following the corresponding nominal track;
(f) conditional on the location of the actual (as distinct from the nominal) trackline, detections are independent events. That is, the probability of detecting a school is only a function of its perpendicular distance from the actual line (and is only a function of whether another school is detected through its potential influence on the location of the line).

The estimating equations

Conventional line transect density estimate

From Buckland *et al.* (1993, p.56) for a conventional line transect survey, and setting the sampling fraction c and the probability $g(0)$ to unity, the density is given by

$$D = \frac{E(n)\cdot f(0)\cdot E(s)}{2L'}$$

If we assume $f(0)$ is constant, then for the i^{th} transect, the density D_i, is given by

$$D_i = \frac{E(n_i)\cdot f(0)\cdot E(s_i)}{2l'_i}, \qquad i=1..k$$

Let $\hat{f}(0)$ be a single pooled estimate of $f(0)$ for the survey. Then, replacing the parameters by their estimators, an estimate of the density for the i^{th} transect is

$$D_i = \frac{\hat{E}(n_i)\cdot \hat{f}(0)\cdot \hat{E}(s_i)}{2l'_i} \qquad (1)$$

So, from Buckland *et al.* (1993, p.92),

$$\hat{D} = \frac{1}{L'}\sum_{i=1}^{k} l'_i \hat{D}_i \qquad (2)$$

To estimate the variance of the density for a conventional line transect survey, if the density components are estimated on a per transect basis, then from Buckland *et al.* (1993, p.92),

$$\hat{V}(\hat{D}) = \frac{1}{L'\cdot(k-1)}\sum_{i=1}^{k}\left(l'_i\cdot\left(\hat{D}_i - \hat{D}\right)^2\right)$$

However, the estimate $f(0)$ is made by pooling data across transects. Dividing out this common estimate, we have

$$\hat{V}\left(\frac{\hat{D}}{\hat{f}(0)}\right) = \frac{1}{L'\cdot(k-1)}\sum_{i=1}^{k}\left(l'_i\cdot\left(\frac{\hat{D}_i}{\hat{f}(0)} - \frac{\hat{D}}{\hat{f}(0)}\right)^2\right) \qquad (3)$$

So an estimate of the variance of the density estimate has two components, $\hat{V}\left(\frac{\hat{D}}{\hat{f}(0)}\right)$ and $\hat{V}\left(\hat{f}(0)\right)$.

Using the delta method (Seber, 1982, p.5-7), an estimate of the variance of the density estimate is given by

$$\hat{V}\left(\hat{D}\right) = \hat{D}^2\cdot\left[\frac{\hat{V}\left(\frac{\hat{D}}{\hat{f}(0)}\right)}{\left(\frac{\hat{D}}{\hat{f}(0)}\right)^2} + \frac{\hat{V}\left(\hat{f}(0)\right)}{\left(\hat{f}(0)\right)^2}\right] = \hat{D}^2\cdot\left[\frac{\hat{V}\left(\hat{H}\right)}{\left(\hat{H}\right)^2} + \frac{\hat{V}\left(\hat{f}(0)\right)}{\left(\hat{f}(0)\right)^2}\right], \qquad (4)$$

where $\hat{H} = \frac{\hat{D}}{\hat{f}(0)}$

Adaptive line transect density estimate

In an adaptive line transect survey, greater effort is systematically placed in areas of higher density. Thus the overall encounter rate is a biased estimate of the expected encounter rate for a conventional survey design. The effort factors are used to downweight observations made while zigzagging as follows. Let

$\hat{f}(0)$ be a single pooled estimate of $f(0)$ for the complete survey;

$\hat{E}\left[n_i|l'_i\right]$ an estimate of the expected number of observations for the i^{th} transect following a straight line track;

$\hat{E}\left[s_i|l'_i\right]$ an estimate of the expected school size for the i^{th} transect following a straight line track.

Evaluation of $\hat{f}(0)$, $\hat{E}\left[n_i|l'_i\right]$ and $\hat{E}\left[s_i|l'_i\right]$ is discussed in the sections that follow.

Replacing the conventional survey parameters $E(n_i)$ and $E(s_i)$ by their estimators from an adaptive survey, an estimate of the density for the i^{th} transect is given by

$$\hat{D}_i = \frac{\hat{E}\left[n_i|l'_i\right]\cdot \hat{f}(0)\cdot \hat{E}\left[s_i|l'_i\right]}{2l'_i} \qquad (5)$$

This estimate of the adaptive line transect density is used in the same manner as the conventional transect estimator to get estimates $\hat{D}$ and $\hat{V}(\hat{D})$ using equations (2) - (4).

f(0)
Data are pooled across all transects to estimate $f(0)$, using conventional techniques. Thus we assume that there is no correlation between $f(0)$ and density.

Effort, L
By definition the nominal effort for the j^{th} leg of the i^{th} transect is

$$l'_{ij} = l_{ij} / \lambda_{ij}$$

with nominal transect length and total survey effort of

$$l'_i = \sum_{j=1}^{m_i} l'_{ij} \qquad \text{and} \qquad L' = \sum_{i=1}^{k} l'_i$$

Sample size, n, and encounter rate, e
Let $\hat{E}[n|L']$ be an estimate of the expected total sample size, and $\hat{E}[e|L']$ the corresponding estimate of the expected encounter rate, if only the nominal effort had been carried out. Similarly let, $\hat{E}[n_i|l'_i]$, $\hat{E}[n_{ij}|l'_{ij}]$, $\hat{E}[e_i|l'_i]$ and $\hat{E}[n_{ij}|l'_{ij}]$ represent the expected sample size and encounter rate if only the nominal effort is used for i^{th} transect and ij^{th} leg.

An estimate of the expected sample size if only the nominal effort had been used for the j^{th} leg of the i^{th} transect is given by

$$\hat{E}[n_{ij}|l'_{ij}] = \frac{n_{ij}}{\lambda_{ij}}$$

and the corresponding transect and survey estimates are

$$\hat{E}[n_i|l'_i] = \sum_{j=1}^{m_i} \hat{E}[n_{ij}|l'_{ij}] \qquad \text{and} \qquad \hat{E}[n|L'] = \sum_{i=1}^{k} \hat{E}[n_i|l'_i]$$

The encounter rate for the j^{th} leg of the i^{th} transect is given by

$$e_{ij} = n_{ij} / l_{ij}$$

and hence by assumption D, an estimate of the expected encounter rate for the j^{th} leg of the i^{th} transect is given by

$$\hat{E}[e_{ij}|l'_{ij}] = \frac{n_{ij}}{l_{ij}} = \frac{\hat{E}[n_{ij}|l'_{ij}]}{l'_{ij}}$$

Thus using weighted averages, an estimate of the expected nominal encounter rate for the i^{th} transect is

$$\hat{E}[e_i|l'_i] = \frac{\sum_{j=1}^{m_i}\left(l'_{ij} \cdot \hat{E}[e_{ij}|l'_{ij}]\right)}{\sum_{j=1}^{m_i} l'_{ij}} = \frac{\sum_{j=1}^{m_i} \hat{E}[n_{ij}|l'_{ij}]}{\sum_{j=1}^{m_i} l'_{ij}} = \frac{\hat{E}(n_i|l'_i)}{l'_i}$$

and an estimate of the expected nominal encounter rate for the survey is

$$\hat{E}[e|L'] = \frac{\sum_{i=1}^{k}\left(l'_i \cdot \hat{E}[e_i|l'_i]\right)}{\sum_{i=1}^{k} l'_i} = \frac{\sum_{i=1}^{k} \hat{E}[n_i|l'_i]}{\sum_{i=1}^{k} l'_i} = \frac{\hat{E}[n|L']}{L'}$$

An estimate of the variance of $\hat{E}[n/L']$ is

$$\hat{V}\left(\hat{E}[n|L']\right) = \frac{L'}{k-1} \sum_{i=1}^{k}\left(l'_i \cdot \left(\hat{E}[e_i|l'_i] - \hat{E}[e|L']\right)^2\right)$$

and an estimate of the variance of $\hat{E}[e|L']$ is

$$\hat{V}\left(\hat{E}[e|L']\right) = \hat{V}\left(\frac{\hat{E}[n|L']}{L'}\right) = \frac{\hat{V}\left(\hat{E}[n|L']\right)}{(L')^2}$$

School size, s
Let $\hat{E}[s|L']$ be an estimate of the expected school size if only the nominal effort had been used for the survey. Similarly let $\hat{E}[s_i|l'_i]$ and $\hat{E}[s_{ij}|l'_{ij}]$ represent the expected school size for the i^{th} transect and the ij^{th} leg.

The mean observed school size for the j^{th} leg of the i^{th} transect is

$$\bar{s}_{ij} = \frac{\sum_{X=1}^{n_{ij}} s_{ijX}}{n_{ij}}$$

Assuming there is no size biased detection, and that the expected school size for a leg following a zigzag track is the same as the expected school size when following the corresponding nominal track, i.e.

$$E[s_{ij}] = E[s'_{ij}]$$

then an estimate of the expected school size for the j^{th} leg of the i^{th} transect is

$$\hat{E}[s_{ij}] = \bar{s}_{ij}$$

so an estimate of the expected school size for the j^{th} leg of the i^{th} transect using nominal effort is

$$\hat{E}[s_{ij}|l'_{ij}] = \bar{s}_{ij}$$

and the expected total number of animals observed for the j^{th} leg of the i^{th} transect following a nominal trackline is

$$\hat{E}[n_{ij}|l'_{ij}] \cdot \hat{E}[s_{ij}|l'_{ij}] = \frac{n_{ij}}{\lambda_{ij}} \cdot \frac{\sum_{X=1}^{n_{ij}} s_{ijX}}{n_{ij}} = \frac{\sum_{X=1}^{n_{ij}} s_{ijX}}{\lambda_{ij}}$$

Using weighted averages, an estimate of the mean school size for the i^{th} transect is given by

$$\hat{E}[s_i|l'_i] = \frac{\sum_{j=1}^{m_i}\left(\hat{E}[n_{ij}|l'_{ij}] \cdot \hat{E}[s_{ij}|l'_{ij}]\right)}{\sum_{j=1}^{m_i}\left(\hat{E}[n_{ij}|l'_{ij}]\right)} = \frac{\sum_{j=1}^{m_i}\left(\left(\sum_{X=1}^{n_{ij}} s_{ijX}\right) \Big/ \lambda_{ij}\right)}{\hat{E}[n_i|l'_i]}$$

and an estimate of the mean school size for the survey is

$$\hat{E}[s|L'] = \frac{\sum_{i=1}^{k}\left(\hat{E}[n_i|l_i'] \cdot \hat{E}[s_i|l_i']\right)}{\sum_{i=1}^{k}\left(\hat{E}[n_i|l_i']\right)} = \frac{\sum_{i=1}^{k}\left(\hat{E}[n_i|l_i'] \cdot \hat{E}[s_i|l_i']\right)}{\hat{E}[n|L']}$$

An estimate of the variance of $\hat{E}[s|L']$ is given by

$$\hat{V}\left(\hat{E}[s|L']\right) = \frac{1}{\hat{E}[n|L'] \cdot (k-1)} \sum_{i=1}^{k} \left(\hat{E}[n_i|l'_i] \cdot \left(\hat{E}[s_i|l'_i] - \hat{E}[s|L']\right)^2\right)$$

SIMULATION STUDY

RATS

Simulation was conducted using the computer program RATS (Restricted Adaptive Transect Sampling), written by the first author. This allows comparison of results using conventional and adaptive line transect methods.

RATS simulates a population in a square frame with side 100 units. Conventional and/or adaptive line transect surveys can then be run on the simulated population, with nominal transects running from left to right across the population frame. The results from a simulated survey are analysed using the formulae developed in this paper.

Automated runs are able to simulate a number of populations in sequence and, for each population, an adaptive and a conventional line transect survey are run. For both surveys, the same transect start points are used but, given the transect locations, the detection process is simulated independently. Some sample simulations are shown in Appendix 2.

Population parameters

Populations are simulated using randomly located parent clusters, each of which consists of a number of schools distributed around the parent cluster centre. Each 'school' has a fixed size of 1. That is, the simulated populations comprise individual animals, each of which belongs to a loose cluster of animals. The default set-up provides a Poisson cluster process (Diggle, 1983, p.55).

The population is created as follows.

(i) The number of parent clusters is simulated using an appropriate distribution selected by the user from the options available.
(ii) For each parent cluster the following is then performed.

(1) The number of animals within the parent cluster is simulated, using a distribution selected by the user from the list available.
(2) The centre of the parent cluster is simulated using continuous uniform variates between 0 and 100 to simulate its horizontal and vertical position within the main population frame. Thus there is no gradient in the simulated population densities in these simulations.
(3) For each animal in a parent cluster, its position, relative to the parent cluster centre, is simulated using a radial angle and distance. The angle is simulated using a continuous uniform variate between 0 and 2π. The radial distance from the centre to the animal is simulated using a distribution selected by the user. This relative position is then converted to the animal's actual position relative to the population frame. If this position lies within the frame then the animal is included in the population. If the animal lies outside the frame, then the distance to the animal is wrapped around to the opposite edge. This is performed both horizontally and vertically as necessary, and is repeated until the animal lies inside the population frame. The components of a simulated population are summarised in Fig. 4.

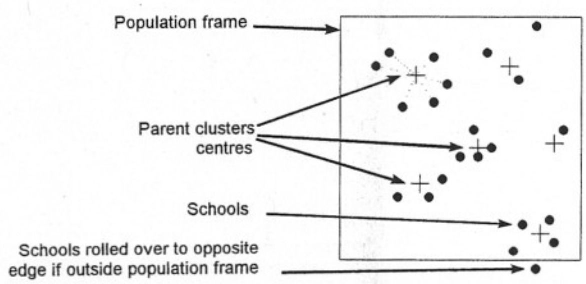

Fig. 4. Populations are simulated in clusters within a population frame. Schools lying outside the frame are repeatedly 'wrapped around' until they lie within the frame.

The transect start positions are simulated to generate randomly positioned lines parallel to the top and bottom edges of the population frame. The total number of transects can be set by the user.

Sampling parameters

The transect is traversed in horizontal steps of size 1 unit. For a conventional line transect survey, at each step an area is sampled using a rectangle centred on the transect. With an adaptive survey, the rectangle becomes a parallelogram for the zigzag sections. The perpendicular offset to the edge of the rectangle or parallelogram, on each side of the transect, is the value w (truncation width) specified by the user (Fig. 5).

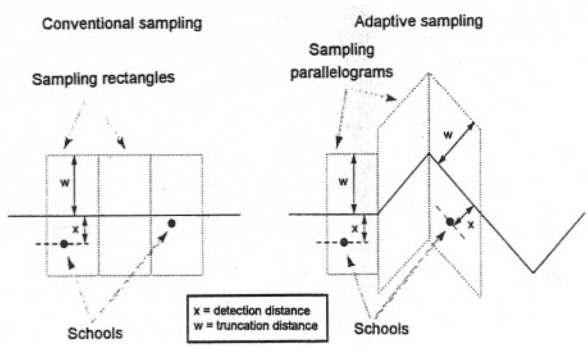

Fig. 5. Population is sampled using parallelograms.

For any animal within the parallelogram, detection is simulated using a half-normal detection function.

The trigger to start zigzagging, in these simulations, is a single detection within a parallelogram on the nominal transect. Zigzagging occurs for a fixed nominal track length of 8 units, with the angle of the zigzags adjusted to increase effort to the level indicated by the effort factor. As implemented, if an object is encountered on the last leg of a zigzag, then there will be another zigzag for 8 nominal units. Edge effects, due to the simulation still following a zigzag track when the transect intersects the population frame, are ignored. In this case the transect stops at the boundary of the population frame, and the next transect, if there is one, starts in straight line mode.

Transect length
The nominal length of each transect can also be specified as one of two options.

(1) The conventional and adaptive surveys use transects of equivalent nominal length. In this case, the transects run the full width of the survey area for both survey types. This means there are fewer transects in the adaptive survey than the conventional survey.
(2) The adaptive survey uses a nominal length which is scaled such that there is an equivalent number of transects for the two types of survey. When the adaptive transect length is scaled down, the transects all start from the extreme left hand side of the area. In our simulations, this is not biased as there is no gradient in the simulated populations. In a real survey, such an option would not be used, but it was implemented here to aid comparison between adaptive and conventional sampling.

Output
The data sets were simulated using a half-normal detection function. RATS has an option for analysing the data, assuming the half-normal model and untruncated data. Provided the truncation distance w is sufficiently large, the performance of this model should therefore be very good. In reality, the true detection function would not be known. RATS can use the computer package DISTANCE (Laake *et al*., 1994) to fit a selection of other models.

Summary analyses
Three types of population were considered:

(1) a population with Complete Spatial Randomness (CSR);
(2) a population with medium clustering (Clustered);
(3) a population with high clustering (Highly Clustered).

The parameters used to simulate the populations are given in Table 1. For each population type, 1,000 simulations were run, with an adaptive and a conventional line transect survey performed each time.

The survey parameters are given in Table 2. The truncation distance of 1 is well into the tail of the half-normal detection function with parameter $\sigma = 0.3$. Thus the RATS estimation of $f(0)$, which assumes no truncation, should work well, given that the probability of detection of an object beyond the truncation width is 0.000858.

A brief summary of the results follows; more detailed results are provided in Appendix 1.

Table 2

Details of parameters used to simulate the surveys.

Parameter	Value
Total effort	1,500
Nominal effort	1,300
Equal nominal length transects	TRUE
Truncation distance (w)	1
Detection function	Half-normal (σ=0.3)
Expected encounter rate	0.045

The comparative efficiency of the two methods was calculated by dividing the mean conventional variance estimate by the mean adaptive variance estimate. The efficiencies are summarised in Table 3 and Fig. 6. The significant improvement in efficiency for the adaptive estimate of $V[n|L']$ is misleading. Sightings made when zigzagging are downweighted using the effort factor, so for an adaptive survey $\hat{E}[n|L']$ is generally smaller than for a conventional survey, and consequently the variance of the estimate is also smaller. It is the efficiency of the encounter rate estimate, and not the number of observations estimate, that is important for the density estimate.

DISCUSSION

The results from the computer simulations indicate that it is acceptable to condition on the effort factors, λ_{ij}. In addition, the results demonstrate that, for populations displaying a reasonable degree of clustering, adaptive line transect sampling offers potential for improvement in the precision of density estimators, and generates a larger number of detections.

The 95% confidence intervals for the mean percent relative bias of the expected encounter rate estimates are given in Table 4. As expected, for all three population types there was no significant bias for the conventional expected encounter rate estimates, as all three intervals included zero. There was also no significant bias for the adaptive estimates for the CSR populations, but there was some evidence of a tendency to underestimate for both the clustered and highly clustered populations, although the bias was small.

The RATS estimate of $f(0)$ assumes there is no truncation. As the simulations had a truncation width of 1, this was expected to lead to an overestimation of both $f(0)$ and its variance. This was supported by the confidence intervals for the mean bias of the $f(0)$ estimates, which had ranges above zero for all three population types for both conventional and adaptive estimates.

Table 1

Details of the population simulation parameters. These parameters give an expected population size of 600 for all three population types.

	Population		
Component	CSR	Clustered	Highly clustered
Number of parent clusters	Constant (600)	Poisson (40)	Poisson (15)
X position of parent cluster centres	Uniform [0, 100]	Uniform [0, 100]	Uniform [0, 100]
Y position of parent cluster centres	Uniform [0, 100]	Uniform [0, 100]	Uniform [0, 100]
Number of objects in each parent cluster	Constant (1)	Poisson (15)	Poisson (40)
Object angle	Uniform [0, 2π]	Uniform [0, 2π]	Uniform [0, 2π]
Object radial distance	Constant (0)	Normal (0, 4)	Normal (0, 4)
School size	Constant (1)	Constant (1)	Constant (1)

Table 3

Efficiency of adaptive estimates in comparison to conventional estimates.

Parameter	Population CSR	Clustered	Highly clustered
Adaptive efficiency of $\hat{V}\left(\hat{E}\left[n\|L'\right]\right)$	1.237	1.341	1.368
Adaptive efficiency of $\hat{V}\left(\hat{E}\left[e\|L'\right]\right)$	0.929	1.007	1.027
Adaptive efficiency of $\hat{V}\left(\hat{f}(0)\right)$	1.006	1.156	1.423
Adaptive efficiency of $\hat{V}(D)$	0.958	1.027	1.072

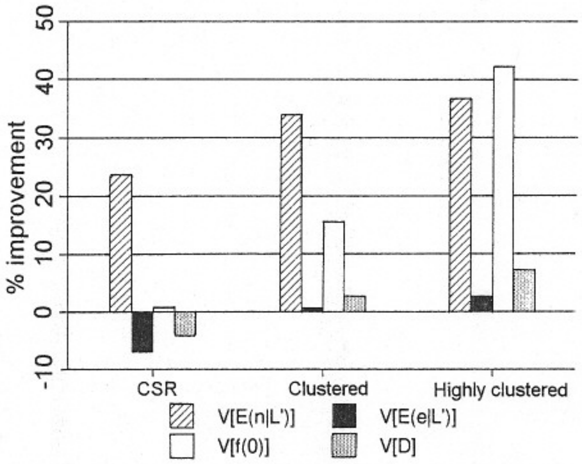

Fig. 6. Summary of mean efficiency of adaptive estimates for three types of population simulated. The bars represent the percentage improvement in efficiency for the adaptive estimate over the conventional estimate. The efficiency is calculated by dividing the conventional variance estimate by the adaptive variance estimate.

There was also some evidence of overestimation for the conventional density estimates, and a likely cause for this is the overestimation of $f(0)$. For the adaptive density estimates the positive bias of the $f(0)$ estimates appears to be partially compensated for by the negative bias of the adaptive encounter rate estimates, particularly for the clustered and highly clustered populations.

For the CSR populations, adaptive sampling was less efficient than conventional sampling. This is not surprising, as a zigzag search path following a detection will only enhance the expected number of detections if there is an increased probability of detecting another animal, having detected one. If animals are distributed entirely at random, this is not the case. The adaptive efficiency for the variance of the density estimate increased from around 0.96 for the CSR populations to 1.07 for the highly clustered populations. The increase in efficiency is correlated with the degree of clustering and work is necessary to identify at what level adaptive sampling provides a worthwhile benefit.

Heterogeneity in $f(0)$

Sightings made while zigzagging are not downweighted when estimating $f(0)$, which may lead to bias in the presence of heterogeneity. For example good sighting conditions may lead to an increase in the number of adaptive triggers, which in turn may lead to increased observations on the zigzag track and so negatively bias the $f(0)$ estimate. (Effective strip width is wider in good sighting conditions, so that $f(0)$ is smaller.) It is recommended that survey results are carefully examined to check for such bias. A potential approach is to pool the data for the observations while following a nominal track and separately, to pool the observations from the zigzag tracks. The resultant two estimates of $f(0)$, $\hat{f}_N(0)$ for the nominal observations and $\hat{f}_A(0)$ for the adaptive observations, can then be tested for differences. Three potential tests are:

(1) a basic z test of whether the expectations of $\hat{f}_{A-}(0)$ and $\hat{f}_N(0)$ are the same;
(2) a χ^2 test of whether the perpendicular sighting distance distribution for observations made on the nominal track is the same as that for observations made when adapting;
(3) Akaike's Information Criterion (AIC).

The AIC approach could be applied by comparing the sum of the AICs for modelling $\hat{f}_A(0)$ and $\hat{f}_N(0)$ separately with the AIC for modelling $f(0)$ using sightings pooled across the two survey modes. If the AIC value for the pooled model is less than the sum of the other two AICs then this suggests that a single model approximates the data better than two separate models. As a rough guide, if the AIC value for the pooled model is greater than the sum of the other two AICs, then this could be taken as a sign of heterogeneity.

Initial simulation trials suggest these tests have low power, though further investigation is required.

Table 4

Estimated 95% confidence intervals, assuming a normal distribution, for the mean percent relative bias of the estimates of expected encounter rate, f(0) and density over all 1,000 simulations. In each cell the top confidence interval is for the adaptive simulations and the bottom one relates to the corresponding conventional simulations.

Estimate	Population CSR	Clustered	Highly clustered
$\hat{E}\left[e\|L'\right]$	[-1.00 %, 0.49 %]	[-2.36 %, -0.05 %]	[-4.25 %, -1.14 %]
	[-1.10 %, 0.35 %]	[-1.40 %, 0.89 %]	[-2.07 %, 1.01 %]
$\hat{f}(0)$	[1.40 %, 2.51 %]	[1.30 %, 2.42 %]	[1.21 %, 2.18%]
	[1.77 %, 2.88 %]	[1.23 %, 2.37 %]	[1.70 %, 2.93 %]
$\hat{D}$	[0.78 %, 2.69 %]	[-0.61 %, 2.03 %]	[-2.77 %, 0.49 %]
	[0.99 %, 2.81 %]	[0.27 %, 2.88 %]	[-0.10 %, 3.28 %]

Ideally the observation data used to estimate $f(0)$ should be weighted by the effort factor, to give pooling robustness. A basic method of weighting the $f(0)$ estimate can be achieved using the computer package DISTANCE (Laake *et al.*, 1994), and entering the perpendicular sighting distances as data grouped into intervals. Normally each sighting would contribute 1 to the count of sightings in its appropriate distance interval. However in this case, to weight the data, the contribution from each sighting is taken as $1/\lambda_{ij}$, where λ is the effort factor for the leg on which the sighting was made. Fortunately DISTANCE will allow the count of sightings in each interval to be a non-integer value. This approach requires validation through simulation.

Extensions/future work

There are several areas where further work is required as discussed below.

Extending theory

Development of the theory to avoid the need to condition on the effort factors would be of interest. The strategy of zigzagging to increase effort conflicts with the requirements of adaptive sampling as defined by Thompson (1992). Thompson (pers. comm.) has suggested an approach in which parallel transects are systematically spaced. If the number of observations on a transect exceeds some limit, additional transects are added parallel to the transect. Although this method allows the adoption of Thompson's adaptive sampling methods, it has the disadvantage of additional costs in travelling off-effort to the start points of the extra transects.

As described above, heterogeneity in $f(0)$ may present an issue, so tests to detect this need to be evaluated. Ideally, to make the $f(0)$ estimate pooling robust, methods of weighting the observation data by the effort factor should be developed.

Effort factor and adapting effort

The effort factor calculation is a key feature of our adaptive sampling strategy. We have so far used a simple approach. The following areas need further research.

(1) The trigger function is very simple: effort is increased if the number of observations within a section exceeds some value (zero in our simulations). This does not cater for surveys of multiple species, where different trigger functions may be required. The issue is further complicated by the appropriate behaviour of the trigger function during a period of increased effort. Currently, primarily in the interests of acceptable field methods, the effort is not increased further when a detection occurs on a zigzag section. If observations are detected on the last leg of a zigzag, then the effort factor is re-calculated, and a new series of zigzags begins.

(2) The survey returns to nominal effort, following an adaptive trigger, after a fixed number of zigzags. There is potential to develop more sophisticated stopping functions.

(3) No facilities are included to detect or adjust for a density gradient in the population. The method will not therefore be fully efficient if a gradient is present. Provided the nominal tracklines are perpendicular to density contours, loss in efficiency should be slight.

(4) The expected encounter rate is fixed at the beginning of the survey, which requires that either an initial estimate (or guess) is available or a pilot survey is carried out. Adjustment of the expected encounter rate using the data that accumulate as the survey progresses may prove useful, particularly when a reliable initial estimate is not available.

(5) The design of the zigzag sections (angle and number of zigzags, and length of section) requires investigation. When each leg in a zigzag is not large relative to the truncation distance w, end and edge effects could be problematic, and field procedures will need to be carefully defined to minimise bias.

Field methods

Due to the complexities of identifying whether observations lie within the searching parallelograms, a computer program may be required to record the observations; calculate the effort factor to use; identify the track to take; and analyse results. The authors are currently exploring the possibility of developing appropriate software.

Simulation and analysis

The bootstrap is a widely used method for quantifying variance in line transect sampling when it is thought that analytic variance estimators do not incorporate all sources of variance. It is normal to resample transects, but this option is not available for adaptive sampling. Instead, both the population (with appropriate clustering) and the adaptive strategy must be simulated in each bootstrap resample. Research is needed on how best to achieve this.

ACKNOWLEDGEMENTS

We would like to thank P. Hammond (Sea Mammal Research Unit, Gatty Marine Laboratory, University of St Andrews, St Andrews, Fife KY16 8LB, UK) and D. Borchers (School of Mathematical and Computational Sciences, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS, UK) for their comments and input to this manuscript. We would also like to thank two anonymous reviewers for their helpful remarks and suggestions.

REFERENCES

Buckland, S.T., Anderson, D.R., Burnham, K.P. and Laake, J.L. 1993. *Distance Sampling: Estimating Abundance of Biological Populations*. Chapman and Hall, New York and London. xii+446pp.

Diggle, P.J. 1983. *Statistical Analysis of Spatial Point Patterns*. Academic Press, London. ix+148pp.

Laake, J.L., Buckland, S.T., Anderson, D.R. and Burnham, K.P. 1994. *DISTANCE User's Guide (Version 2.1)*. Colorado Cooperative Fish & Wildlife Research Unit, Colorado State University, Fort Collins, CO. 84pp.

Seber, G.A.F. 1982. *The Estimation of Animal Abundance and Related Parameters*. 2nd Edn. Edward Arnold, London. xvii+654pp.

Thompson, S.K. 1992. *Sampling*. Wiley, New York. xv+343pp.

[Appendix 1 next page]

Appendix 1

SIMULATION RESULTS

Results are summarised for the population types in Tables 5, 6 and 7 below. The tables contain mean values/estimates for he 1,000 simulated adaptive and conventional surveys. Where appropriate the respective standard deviation of the 1,000 values/estimates is also included.

Table 5

CSR population results.

	Adaptive		Conventional	
	Mean	SD	Mean	SD
Total area	10,000		10,000	
Total effort per survey	1496.2	3.69	1500	-
Nominal effort per survey	1300	-		
Population size	600	0	600	0
Population Index of Dispersion	1.02	0.40	1.02	0.40
Number of observations	67.23	8.07	67.37	7.92
Nominal number of observations	58.46	7.06		
True encounter rate	**0.0451**	**0**	**0.0451**	**0**
Estimated nominal encounter rate	0.0450	0.00543	0.0449	0.00528
Nominal encounter rate bias	-0.00011	0.00543	-0.00017	0.00528
Estimated variance for nominal encounter rate estimate	0.000032	0.000013	0.000029	0.000012
Number of nominal encounter rate estimates within 95% CI	971		977	
True f(0)	**2.6619**		**2.6619**	
RATS estimate of f(0)	2.7140	0.2390	2.7238	0.2398
Bias of RATS estimate of f(0)	0.0521	0.2390	0.0619	0.2398
RATS estimate of V $[\hat{f}(0)]$	0.0560	0.0121	0.0563	0.0126
Number of f(0) estimates within 95% CI	971		973	
True density	**0.06000**	**0**	**0.06000**	**0**
RATS density estimate	0.06104	0.00922	0.06114	0.00879
Bias of RATS density estimate	0.00104	0.00922	0.00114	0.00879
Number of density estimates within 95% CI	985		983	
Efficiency using RATS for detection function estimates	**0.958**			

Table 6

Clustered population results.

	Adaptive		Conventional	
	Mean	SD	Mean	SD
Total area	10,000		10,000	
Total effort per survey	1485.6	14.08	1500	-
Nominal effort per survey	1300	-		
Population size	603.25	97.75	603.25	97.75
Population Index of Dispersion	12.27	4.35	12.27	4.35
Number of observations	77.87	18.21	67.81	16.82
Nominal number of observations	58.22	14.47		
True encounter rate	**0.0453**	**0.00734**	**0.0453**	**0.00734**
Estimated nominal encounter rate	0.0448	0.01113	0.0452	0.01121
Nominal encounter rate bias	-0.00054	0.00842	-0.00012	0.00833
Estimated variance for nominal encounter rate estimate	0.000068	0.000038	0.000069	0.000037
Number of nominal encounter rate estimates within 95% CI	959		964	
True f(0)	**2.6619**		**2.6619**	
RATS estimate of f(0)	2.7115	0.2408	2.7099	0.2452
Bias of RATS estimate of f(0)	0.0496	0.2408	0.0480	0.2452
RATS estimate of V $[\hat{f}(0)]$	0.0504	0.0162	0.0583	0.0200
Number of f(0) estimates within 95% CI	971		976	
True density	**0.06033**	**0.00978**	**0.06033**	**0.00978**
RATS density estimate	0.06075	0.01613	0.06127	0.01620
Bias of RATS density estimate	0.00042	0.01279	0.00094	0.01263
Number of density estimates within 95% CI	963		970	
Efficiency using RATS for detection function estimates	**1.027**			

Table 7

Highly clustered population results.

	Adaptive		Conventional	
	Mean	SD	Mean	SD
Total area	10,000		10,000	
Total effort per survey	1466.1	28.34	1500	-
Nominal effort per survey	1300	-		
Population size	599.18	151.68	599.18	151.68
Population Index of Dispersion	31.36	11.12	31.36	11.12
Number of observations	89.83	28.81	67.17	23.81
Nominal number of observations	56.95	20.31		
True encounter rate	**0.0450**	**0.01140**	**0.0450**	**0.01140**
Estimated nominal encounter rate	0.0438	0.01563	0.0448	0.01587
Nominal encounter rate bias	-0.00121	0.01132	-0.00024	0.01122
Estimated variance for nominal encounter rate estimate	0.000126	0.000084	0.000129	0.000083
Number of nominal encounter rate estimates within 95% CI	960		949	
True f(0)	**2.6619**		**2.6619**	
RATS estimate of f(0)	2.7069	0.2082	2.7235	0.2626
Bias of RATS estimate of f(0)	0.0450	0.2082	0.0616	0.2626
RATS estimate of $V[\hat{f}(0)]$	0.0463	0.0212	0.0659	0.0588
Number of f(0) estimates within 95% CI	987		975	
True density	**0.05992**	**0.01517**	**0.05992**	**0.01517**
RATS density estimate	0.05923	0.02133	0.06087	0.02209
Bias of RATS density estimate	-0.00068	0.01581	0.00095	0.01634
Number of density estimates within 95% CI	954		960	
Efficiency using RATS for detection function estimates	**1.072**			

Appendix 2

EXAMPLE SIMULATIONS

CSR population

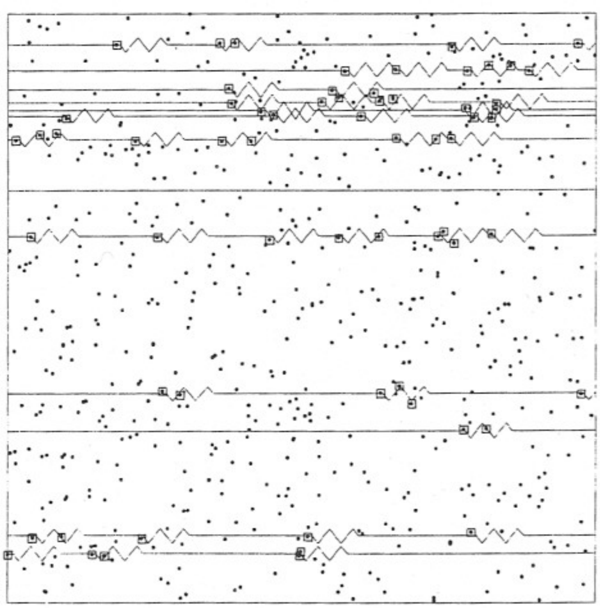

Fig. 7. Simulation of an adaptive line transect survey of a CSR population. Schools are represented by dots and each observed school is bounded by a square. In this case the total population is 600 and the number of observations is 67.

Clustered population

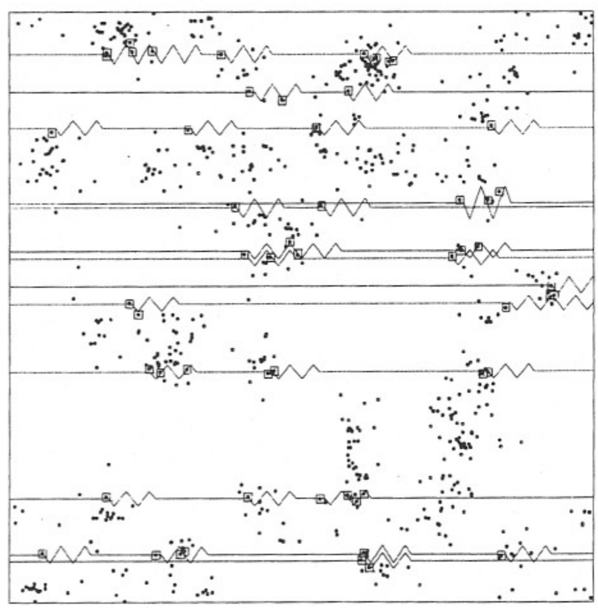

Fig. 8. Simulation of an adaptive line transect survey of a Clustered population. Schools are represented by dots and each observed school is bounded by a square. In this case the total population is 706 and the number of observations is 59.

Highly Clustered population

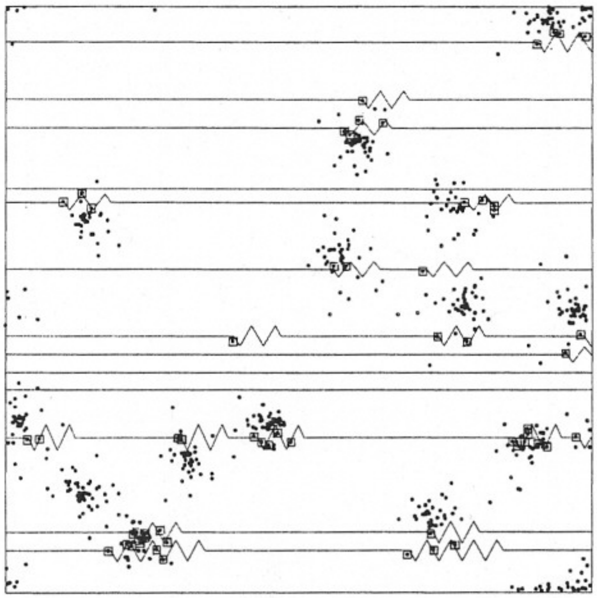

Fig. 9. Simulation of an adaptive line transect survey of a Highly Clustered population. Schools are represented by dots and each observed school is bounded by a square. In this case the total population is 625 and the number of observations is 66.

Forty-Seventh Report
of the
International Whaling Commission

Covering the forty-seventh financial year 1995-1996
Approved by the Commission at its forty-eighth meeting in Aberdeen, Scotland, May/June 1996
Authorised to be printed together with the Chairman's Report of the forty-eighth meeting.

Cambridge
1997