Contents

1 Introduction to advanced distance sampling 1
   S. T. Buckland and D. R. Anderson

2 General formulation for distance sampling 6
   D. L. Borchers and K. P. Burnham
   2.1 Introduction 6
   2.2 CDS revisited 7
      2.2.1 Conventional line transect estimator 8
      2.2.2 Conventional point transect estimator 9
   2.3 Horvitz–Thompson: a versatile estimator 9
      2.3.1 Animals that occur as individuals 10
      2.3.2 Animals that occur in clusters 11
      2.3.3 CDS estimators 11
   2.4 Maximum likelihood estimation 11
      2.4.1 ‘Covered’ animals 12
      2.4.2 Random detection with known probability 13
      2.4.3 CDS likelihoods 16
   2.5 Summary so far and preview of advances 17
      2.5.1 Summary 17
      2.5.2 Preview of advances 18
   2.6 Advanced methods for detection function estimation 19
      2.6.1 Multiple covariate distance sampling 19
      2.6.2 Mark-recapture distance sampling 21
      2.6.3 Estimation when \( \pi(y) \) is unknown 23
   2.7 Estimating animal density surfaces 24
      2.7.1 The count method 25
      2.7.2 The waiting distance method 25
      2.7.3 Cluster size surface estimation 26
   2.8 Survey design 26
      2.8.1 Likelihood-based inference 26
      2.8.2 Design-based inference 27
      2.8.3 Adaptive distance sampling 27
2.9 Model selection 28
2.10 Summary 29

3 Covariate models for the detection function 31
F. F. C. Marques and S. T. Buckland

3.1 Introduction 31
3.2 A conditional likelihood framework for distance sampling 32
3.3 Line transect sampling 33
  3.3.1 The conditional likelihood 33
  3.3.2 Incorporating covariates into semiparametric models for the detection function 35
  3.3.3 Abundance estimation 38
3.4 Point transect sampling 43
3.5 Example 45
3.6 Discussion 47

4 Spatial distance sampling models 48
S. L. Hedley, S. T. Buckland, and D. L. Borchers

4.1 Introduction 48
4.2 Spatial line transect models 49
  4.2.1 Deriving a likelihood 50
  4.2.2 A likelihood based on inter-detection distances 52
  4.2.3 Clustered populations 54
4.3 Practical implementations of spatial line transect models 55
  4.3.1 A waiting distance model 55
  4.3.2 A count model 57
4.4 Spatial distribution of Antarctic minke whales 60
4.5 Spatial point transect models 63
  4.5.1 Deriving a likelihood 63
  4.5.2 A point transect count model 65
4.6 Discussion 66

5 Temporal inferences from distance sampling surveys 71
L. Thomas, K. P. Burnham, and S. T. Buckland

5.1 Introduction 71
5.2 Concepts 73
  5.2.1 Sampling and population variation 73
  5.2.2 Sampling covariance 75
  5.2.3 Empirical and process models 76
  5.2.4 Trend 76
  5.2.5 Abundance as a fixed or random quantity 77
CONTENTS

5.3 Trend estimation from global abundance estimates 79
  5.3.1 Graphical exploration 79
  5.3.2 Linear trend models 79
  5.3.3 Smoothing 84
  5.3.4 Trend estimation when samples covary 88
5.4 Spatio-temporal analysis 91
  5.4.1 Transect-level models of trend 91
  5.4.2 Spatio-temporal modelling 93
5.5 Process models 93
  5.5.1 State-space models 94
  5.5.2 Generalizing state-space models 96
5.6 Other analysis methods 98
  5.6.1 Time series methods 98
  5.6.2 Quality control methods 98
5.7 Survey design 99
  5.7.1 Repeating transects 99
  5.7.2 Sample size 101
  5.7.3 Planning long-term studies 105

6 Methods for incomplete detection at distance zero 108
  J. L. Laake and D. L. Borchers

  6.1 Introduction 108
  6.2 Likelihood and Horvitz–Thompson 111
    6.2.1 Constant detection probability 112
    6.2.2 Detection probability changing with distance 114
    6.2.3 Independence issues 117
    6.2.4 Multiple covariates 121
    6.2.5 Unobserved heterogeneity 122
  6.3 State and observation models 123
    6.3.1 State models 124
    6.3.2 Observation models 125
    6.3.3 Observation configurations 137
  6.4 Example data 140
  6.5 Estimation for independent configuration 142
    6.5.1 Distance only 142
    6.5.2 Distance and covariates 154
  6.6 Estimation for trial configuration 159
    6.6.1 Distance and covariates 159
    6.6.2 Distance, covariates, and responsive movement 162
  6.7 Estimation for removal configuration 166
  6.8 Dealing with availability bias 168
    6.8.1 Static availability 168
    6.8.2 Hazard-rate models for dynamic availability 170
CONTENTS

6.8.3 Discrete availability: animal-based 171
6.8.4 Discrete availability: cue-based 174
6.8.5 Intermittent availability 175
6.8.6 Design-based availability estimation 178

6.9 Special topics 180
6.9.1 Uncertain duplicate identification 180
6.9.2 When should double-observer methods be used? 182

6.10 Field methods 185
6.10.1 Marked animals 185
6.10.2 Observation configuration 186
6.10.3 Data collection and recording 188

7 Design of distance sampling surveys and Geographic Information Systems 190
S. Strindberg, S. T. Buckland, and L. Thomas

7.1 The potential role of GIS in survey design 190
7.2 Automated survey design 191
7.2.1 Point transect design 192
7.2.2 Line transect design using lines of fixed length 199
7.2.3 Line transect design using lines that span the full width of the survey region 206
7.2.4 Zigzag samplers 211
7.3 Estimation for uneven coverage probability designs 224
7.3.1 Objects that occur singly 225
7.3.2 Objects that occur in clusters 226
7.3.3 Variance estimation 226
7.4 Choosing between survey designs by simulation 226

8 Adaptive distance sampling surveys 229
J. H. Pollard and S. T. Buckland

8.1 Introduction 229
8.2 Design-unbiased adaptive point transect surveys 230
8.2.1 Survey design 230
8.2.2 Estimation 233
8.2.3 Simulated example 237
8.2.4 Discussion 239
8.3 Design-unbiased adaptive line transect surveys 240
8.3.1 Survey design 240
8.3.2 Estimation 241
8.3.3 Discussion 246
8.4 Fixed-effort adaptive line transect surveys 247
8.4.1 Survey design 247
8.4.2 Estimation 247
8.4.3 Simulation 254
8.4.4 Discussion 259

9 Passive approaches to detection in distance sampling 260
P. M. Lukacs, A. B. Franklin, and D. R. Anderson
9.1 Introduction 260
9.2 Trapping webs 262
9.2.1 Density estimation 262
9.2.2 Including data from recaptures 263
9.2.3 Design of trapping webs 266
9.2.4 An example 266
9.2.5 A critique of the trapping web 269
9.3 Trapping line transects 270
9.3.1 Density estimation 271
9.3.2 Including data from recaptures 271
9.3.3 Design of trapping line transects 272
9.3.4 An example 272
9.3.5 A critique of the trapping line transect 278
9.4 Discussion and summary 278

10 Assessment of distance sampling estimators 281
R. M. Fewster and S. T. Buckland
10.1 Introduction 281
10.2 Estimation framework 282
10.3 Model and design 283
10.3.1 Model-based inference 283
10.3.2 Design-based inference 284
10.3.3 Distance sampling: a composite approach 286
10.4 Simulation framework 286
10.4.1 Testing the design 287
10.4.2 Testing the model 288
10.4.3 Testing the full line transect estimation procedure 291
10.5 Example: testing the design 292
10.5.1 Testing equal coverage designs for \(\text{var}(\hat{N})\) 293
10.5.2 A design without equal coverage probability 297
10.6 Example: non-uniformity within the strip 298
  10.6.1 Estimation of $N_c$ 298
  10.6.2 Asymptotic result when $\theta$ is estimated 301
10.7 Example: full estimation procedure 302
10.8 Trial by simulation: a completely model-based approach 302
10.9 Summary 306

11 Further topics in distance sampling 307
  K. P. Burnham, S. T. Buckland, J. L. Laake, D. L. Borchers,
  T. A. Marques, J. R. B. Bishop, and L. Thomas

11.1 Distance sampling in three dimensions 307
  11.1.1 Three-dimensional line transect sampling 307
  11.1.2 Three-dimensional point transect sampling 309
11.2 Conventional distance sampling: full likelihood examples 312
  11.2.1 Line transects: simple examples 312
  11.2.2 Point transects: simple examples 318
  11.2.3 Some numerical confidence interval comparisons 321
11.3 Line transect surveys with random line length 327
  11.3.1 Introduction 327
  11.3.2 Line transect sampling with fixed $n$ and random $\widetilde{L}$, under Poisson object distribution 328
  11.3.3 Technical comments 331
  11.3.4 Discussion 333
11.4 Models for the search process 335
  11.4.1 Continuous hazard-rate models 335
  11.4.2 Discrete hazard-rate models 340
  11.4.3 Further modelling of the detection process 343
11.5 Combining mark-recapture and distance sampling surveys 350
11.6 Combining removal methods and distance sampling 352
  11.6.1 Introduction 352
  11.6.2 Combining removal methods with distance sampling 354
11.7 Point transect sampling of cues 356
  11.7.1 Introduction 356
  11.7.2 Estimation 357
11.8 Migration counts 359
  11.8.1 Background 359
  11.8.2 Modelling migration rate 360
CONTENTS

11.8.3 Modelling detection probabilities 360
11.8.4 An example: gray whales 364
11.9 Estimation with distance measurement errors 371
  11.9.1 Conventional distance sampling: \( g(0) = 1 \) 371
  11.9.2 Independent multiplicative measurement errors 373
  11.9.3 Mark-recapture distance sampling: \( p(0) < 1 \) 375
  11.9.4 Maximum likelihood vs pdf correction approach 377
11.10 Relating object abundance to population abundance for indirect sampling 377
  11.10.1 Introduction 377
  11.10.2 Discrete-time modelling 378
  11.10.3 Continuous-time modelling 378
  11.10.4 Conclusions 384
11.11 Goodness of fit tests and q–q plots 385
  11.11.1 Quantile–quantile plots 385
  11.11.2 Kolmogorov–Smirnov test 387
  11.11.3 Cramér-von Mises test 388
  11.11.4 The Cramér-von Mises family of tests 388
11.12 Pooling robustness 389

References 393

Index 413
1

Introduction to advanced distance sampling

S. T. Buckland and D. R. Anderson

Distance sampling, primarily line transect and point transect sampling, has had a relatively short history. The earliest attempts to use distances to detected animals to estimate abundance date back to the 1930s, and the first line transect estimator with a rigorous mathematical basis was due to Hayne (1949). Nearly 20 years later, Gates et al. (1968) and Eberhardt (1968) made important contributions to the development of line transect sampling methodology. Neither the radial distance model of Hayne (1949) nor the negative exponential model of Gates et al. (1968) is based on plausible assumptions about the detection process. Eberhardt’s (1968) work was more conceptual, and attempted to provide a class of models that were robust to differing detection processes. None of these early methods are now recommended. Three papers in the early 1970s prompted Burnham and Anderson (1976) to develop the general theory needed for reliable estimation. The first of these papers was Anderson and Pospahala (1970), who used polynomials to fit the distance data, but who did not provide underlying theory. The field experiments of Robinette et al. (1974) were important in providing data sets with known abundance, on which estimation methods could be tested. The third paper was by Sen et al. (1974), which gave an erroneous formulation. Burnham and Anderson (1976) corrected this formulation, and provided a general framework for both parametric and nonparametric methods, applied to data that were either grouped or ungrouped, and truncated or untruncated. The first comprehensive treatment of the topic was by Burnham et al. (1980). Point transect sampling (or variable circular plots) was conceptualized in the early 1970s for songbird surveys, although the initial work was not published until 1980 (Reynolds et al. 1980), by which time several papers using the technique had already been published. The method is still largely restricted to avian studies (Rosenstock et al. 2002), although other applications are now starting to appear in the literature. Reviews of these historical developments are given by Buckland et al. (2000, 2001).
INTRODUCTION

Given its short history, it is perhaps surprising that distance sampling is the most widely used technique for estimating abundance of wild animal populations. The use of mark-recapture in fisheries dates as far back as Walton (1653), although work by Petersen (1896) seems to have led to its use for estimating abundance (see Buckland et al. 2000). The use of harvest models based on the ‘catch equations’ to estimate abundance was first documented by Baranov (1918); catch-per-unit-effort by Hjort and Ottestad (1933); change-in-ratio by Kelker (1940); and removal methods by Moran (1951). The success of distance sampling perhaps stems from the fact that it provides more robust estimation of abundance more cheaply than methods based on catching animals, at least for populations for which the key assumptions hold to a good approximation. Further, the effects on abundance estimates when key assumptions fail tend to be more readily understood than for most competitive methods. Distance sampling represents a suite of methods which are extensions of complete counts of sample plots (called sample counts here). Their advantage over sample counts is that not all animals within the sampled plots (strips in the case of line transect sampling and circles for point transect sampling) need be counted, so that the approach can usually achieve a given level of precision at lower cost than a comparable method based on sample counts. Relative to mark-recapture, the modelling element in distance sampling is straightforward. The detection function model seldom requires more than four parameters, and one is often sufficient. Only a handful of contending models need be considered. By contrast, mark-recapture models commonly require more than 20 parameters—possibly substantially more—and there are many possible models for a given data set.

The methods that are currently considered standard are largely as set out by Buckland et al. (1993a). For an updated introduction to standard methodology, the reader is referred to the companion volume to this book (Buckland et al. 2001). Here, we concentrate on more advanced methodologies, some of which are reviewed by Buckland et al. (2002). A few of the more advanced sections from Buckland et al. (1993a), especially from chapter 6 of that book, are updated and reproduced here.

For a general methodological framework for distance sampling and other methods of estimating abundance of closed populations, the reader is referred to Borchers et al. (2002). That book is intended as an advanced student text, and provides a basis for many of the developments outlined here. Williams et al. (2002) cover a range of methods for estimating animal abundance, and show how these methods are used in the management of animal populations.

In Chapter 2, a general likelihood framework is presented. The three components of estimation for standard methodology, encounter rate, effective area surveyed, and mean cluster size (when objects occur in clusters), are integrated into a single Horvitz–Thompson-like estimator, in which the
inclusion probabilities are estimated

\[
\hat{N} = \sum_{i=1}^{n} \frac{s_i}{P_i},
\]  

(1.1)

where \( \hat{N} \) is estimated population size, \( n \) is the number of animal clusters detected in the covered area, \( s_i \) is the size of the \( i \)th detected cluster, and \( \hat{P}_i \) is the estimated inclusion probability for that cluster. This probability has two components: a coverage probability \( P_c \), which is determined by design, and given by \( P_c = a/A \), where \( A \) is the size of the study region and \( a \) is the covered area (the total area of surveyed strips or circles); and an estimated detection probability \( \hat{P}_a \), given that a cluster is in the covered area. If we choose to estimate a single \( \hat{P}_a \) for all clusters, then \( \hat{P}_i = P_c \times \hat{P}_a \) and the effective area surveyed is \( a \times \hat{P}_a \), estimated as \( a \times \hat{P}_a \). Alternatively, if we obtain cluster-specific estimates, then \( \hat{P}_i = P_c \times \hat{P}_{ai} \) for cluster \( i \).

If objects occur singly, then \( s_i = 1 \) for every detection, and

\[
\hat{N} = \sum_{i=1}^{n} \frac{1}{P_i}.
\]  

(1.2)

In the case of clustered populations, this equation gives the estimated number of clusters in the study area.

Full likelihood and conditional likelihood methods are described in Chapter 2. Because the distribution of covariates in the study population or area is generally unknown and not easily estimated, inference is generally based on conditional likelihood methods, for which we condition on the values of covariates observed.

Covariate models for the detection function (Chapter 3) potentially yield more efficient estimates of abundance, and eliminate the bias that may arise, for example, when abundance estimates by stratum are required, but data are pooled across strata for estimating the detection function. They also offer the potential for more reliable estimates of trend in abundance when surveys are conducted from platforms of opportunity, such as ferries or fishing vessels at sea, although they do not address the problems that arise because the region is not randomly sampled in such surveys.

Spatial line transect models (Chapter 4) allow a surface to be fitted, representing animal density throughout the study region. This in turn allows estimation of abundance for any subset of the area, by integrating over the relevant section of the surface. It also allows abundance to be related to spatial covariates, so that managers can assess the importance of habitat and environment to the population of interest. Spatial models also potentially reduce bias in abundance estimates from platforms of opportunity survey data, in which survey effort is non-random.
Spatial models estimate variation in density or abundance through the region. Wildlife managers are often more interested in modelling temporal trends, to identify whether management action is required, or to assess the effects of such action. Methods for trend estimation addressed in Chapter 5 fall into two categories: empirical estimation of trend from a series of abundance estimates; and fitting of a population dynamics model to the time series of estimates. The second approach has the advantages that estimated trends are consistent with biological reality, and the effects of management actions that affect survival or productivity can be modelled and predicted.

For some populations, such as whales or porpoise, or burrowing animals such as rabbits or tortoise, one of the key assumptions that any animal on the line or at the point will be detected may be violated. Double-platform methods (Chapter 6) allow distance sampling methodology to be combined with mark-recapture methods, so that this assumption is no longer required. The second ‘platform’ may be a standard sightings platform on a ship or aircraft, or it may comprise an independent method of locating a subset of the surveyed animals, such as a radio-tagging experiment. In the latter case, individual animals are marked, or are identified from natural markings, whereas in the former case, no marking takes place. Instead, it is necessary to develop field methods so that duplicate detections (animals detected by both platforms) can be identified.

Automated design algorithms, linked with Geographic Information Systems (GIS) functionality, are covered in Chapter 7. They allow quick and easy generation of survey designs, and enable different designs to be compared for efficiency and accuracy of the subsequent abundance estimates, using simulation. For complex surveys in which coverage probability is not uniform, they also allow estimation of coverage probability by location. This in turn allows valid abundance estimation.

For populations that typically have an aggregated spatial distribution, adaptive distance sampling surveys (Chapter 8) potentially yield more precise estimates of abundance. They also give more detections than can conventional surveys with the same overall effort, which can be an important advantage for scarce species, for which sample size may otherwise be inadequate for modelling the detection function.

In most distance sampling surveys, one or more observers actively try to detect objects, usually animals. In passive distance sampling methods (Chapter 9), animals are not actively searched for. Instead, detections are made, for example, by using traps, or devices to secure a sample of hair or feathers, or remote systems such as cameras. A detection occurs when an animal enters a trap or a sensed area at a known distance from a central point (trapping web) or line (trapping line transect). The density of traps or sensed areas is greater near the centre point or line. Trapping webs (Anderson et al. 1983) have their roots in point transect sampling theory. Trapping transects (Chapter 9) make similar use of line transect sampling
theory. These methods can be particularly useful for species that do not

generally meet the assumptions of standard distance sampling, perhaps

because they hide and are undetectable for much of the time (e.g. rep-

tiles) or because they are too small to be reliably detected by line transect

observers (e.g. beetles).

Standard distance sampling methods blend model-based statistical

methods (to model detectability within the surveyed strips or circles) with

design-based statistical methods (to estimate the number of animals outside

the surveyed strips). (Spatial line transect models, discussed in Chapter 4,

replace the design-based element by a spatial model of animal density.)

In Chapter 10, we give a rigorous basis for the composite approach, show

why it leads to robust estimation of animal abundance, and explore the

limitations of this robustness.

Other advanced topics are covered in Chapter 11: three-dimensional dis-

tance sampling methods; full likelihood methods for conventional distance

sampling; line transect surveys with random line length; models for the

search process in sightings surveys; combined mark-recapture and distance

sampling surveys; combined removal methods and distance sampling sur-

veys; point transect sampling of cues; migration counts; measurement error

models; theory for indirect surveys of animal signs (usually dung or nests);

quantile–quantile plots, the Kolmogorov–Smirnov test and Cramér–von

Mises tests; and pooling robustness.

Most of the above advances have been implemented, or will shortly be

implemented, in version 4 of the software distance (Thomas et al. 2003).

Version 4 also incorporates the standard methods of Version 3.5 (Thomas

et al. 1998). Version 3.5 is the companion software for Buckland et al.

(2001), and Version 4 is the companion software for this book.