Methods in Statistical Ecology

S.T. Buckland
E.A. Rexstad
T.A. Marques
C.S. Oedekoven

Distance Sampling: Methods and Applications

Springer

Methods in Statistical Ecology

Series Editors

Andrew P. Robinson, Melbourne, Victoria, Australia
Stephen T. Buckland, St. Andrews, United Kingdom
Peter Reich, St. Paul, USA
Michael McCarthy, Parkville, Australia

More information about this series at http://www.springer.com/series/10235

S.T. Buckland • E.A. Rexstad • T.A. Marques
C.S. Oedekoven

Distance Sampling: Methods and Applications

Springer

S.T. Buckland
CREEM The Observatory
University of St. Andrews
St. Andrews, UK

E.A. Rexstad
CREEM The Observatory
University of St. Andrews
St. Andrews, UK

T.A. Marques
CREEM The Observatory
University of St. Andrews
St. Andrews, UK

C.S. Oedekoven
CREEM The Observatory
University of St. Andrews
St. Andrews, UK

ISSN 2199-319X ISSN 2199-3203 (electronic)
Methods in Statistical Ecology
ISBN 978-3-319-19218-5 ISBN 978-3-319-19219-2 (eBook)
DOI 10.1007/978-3-319-19219-2

Library of Congress Control Number: 2015941311

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

This is the fourth book dedicated to distance sampling. The first (Buckland et al. 1993) appeared at around the same time as the first version of the `Distance` software (Laake et al. 1993). This ground-breaking book was largely replaced by *Introduction to Distance Sampling* (Buckland et al. 2001). A further book, *Advanced Distance Sampling* (Buckland et al. 2004), mostly covered advances since 1993. In parallel, many advances were introduced into the `Distance` software, summarized in Thomas et al. (2010). Since 2004, there have been many new developments, and it is becoming clearer which of these advances are of most use. In this book, we seek to cover the basic methods, and the various advances that we believe are of greatest value to practitioners. There is an increasing trend towards conducting distance sampling analyses using the statistical software `R`, rather than using the `Distance` software. Several of our examples reflect this.

In Chap. 1, we build on conventional plot sampling ideas to outline the basic methods of conventional distance sampling and the respective assumptions. Prior to Buckland et al. (1993), there was very little work on relaxing the four key distance sampling assumptions. Subsequently, methods for when animals on the line or at the point are not certain to be detected were developed, and Buckland et al. (2004) include a comprehensive treatment of double-observer methods. Since 2004, there has been work on incorporating measurement error, and allowing non-uniform animal distribution. Further, the importance of modelling animal availability in the context of estimating probability of detection on the line or at the point has been recognized in recent work. Incorporating animal movement models into distance sampling is a focus of current research.

In a marked departure from previous books on distance sampling, we bring design issues and field methods forward. In terms of the practical application of these methods, it is fundamental to get the design right and the field methods optimized. Data arising from poor survey design and/or field methods might or might not be salvageable using sophisticated analysis methods, but data from a good and carefully thought-out design with optimized field methods are generally simpler to analyse, and give more reliable estimates of abundance.

In Chap. 2, we address survey design issues for points and lines, covering aspects of stratification, survey effort guidelines and assessing designs by simulation. We also include a chapter on the design of distance sampling experiments (Chap. 3). Increasingly, distance sampling is being used not to estimate abundance in a study area, but to assess whether animal density varies by type of plot. For example, some plots may be subject to a treatment, while others act as a control, and we wish to assess whether the treatment has an effect on density. The treatment might relate to management of habitat, use of a pesticide, construction of a structure, *etc*.

Field methods and data checking are covered in Chap. 4. Trying to address data quality issues at the analysis stage is no substitute for gathering good quality data, with procedures for identifying and correcting recording errors.

At the heart of distance sampling methods is the concept of a detection function. The detection function represents the probability $g(y)$ of detecting an animal, given that it is at a distance y from a line or point. Intuitively, this should be a non-increasing function of distance. In Chap. 5, we consider models for the detection function $g(y)$. For conventional distance sampling (CDS), we assume that this is a function of y alone, with $g(0) = 1$ (*i.e.* an animal on the line or point is certain to be detected). For multiple-covariate distance sampling (MCDS), we allow the function to vary with other covariates in addition to distance. For mark-recapture distance sampling (MRDS), we also allow $g(0) < 1$. Modelling $g(y)$ is crucial, because it allows us to estimate the proportion that we detect of animals that are located on the sampled strips or circles (the sampling units, which together comprise the covered region).

For CDS, we adopt a hybrid approach to modelling, so that model-based methods are used to estimate the detection function (that is, we fit a model to distances, to estimate for each animal detected its probability of detection), but design-based methods are used to give robust extrapolation from the sampling units to the wider survey region. In other words, we estimate abundance on the plots, and extrapolate to the whole study area, based on the random design, which ensures that the plots are representative of the area from which they were drawn. In Chap. 6, we address issues associated with design-based estimation of animal density and abundance.

An alternative to CDS methods is to consider distance sampling data as counts of detected animals within a set of sampling units. Viewed in this way, then it is natural to use count models such as Poisson or negative binomial to analyse the data. Chapters 7 and 8 present such methods. Distance sampling spatial models fit into this framework. Seminal work in this area stemmed from Hedley et al. (1999) and Hedley (2000). As the name implies, extrapolation beyond the sampling units to a wider study area is based on a model indexed in space.

In Chap. 7, we adopt such a model-based approach for counts. We consider an 'offset' within a count regression model to account for detectability. Thus, the approach has two stages: the detection function is estimated in the first, which provides an estimated offset to include in the count model in stage two. In designed

distance sampling experiments, this allows us to test whether treatment effects in the count model are significant, while for standard distance sampling studies, we can extrapolate beyond the samplers by modelling how counts vary through the survey region. We can also estimate an animal density surface, which allows density to be modelled as a function of spatial covariates such as habitat, and abundance estimates to be obtained for any sub-area by integrating the corresponding section of the estimated density surface.

While the two-stage approach is easier to implement, it is not conceptually elegant. Perhaps worse is that it raises issues relating to uncertainty propagation across the two stages. Therefore, in Chap. 8 we take these models further, exploring maximum likelihood and Bayesian methods to give flexible options for analysing distance sampling data adopting a single, integrated likelihood.

In Chap. 9, we describe variations on the theme of distance sampling, concentrating on extensions of line or point transect sampling that either are already in wide use or that have the potential to see wide use: cue counting, in which cues (for example, whale blows or calls, song bursts or calls for bird surveys, or primate calls) are recorded, from which animal density is estimated; trapping and lure point transects, in which a trap or lure is placed at each survey point; indirect surveys in which signs produced by the animal of interest (usually dung or nests) are surveyed; and three-dimensional surveys (such as sonar surveys of fish or radar surveys of migrating birds).

Distance sampling is used for a wide range of taxa and habitats. While most of the advice is general, some issues are taxon-specific. Chapter 10 discusses special issues that arise in applying distance sampling methods to the following taxa: songbirds, seabirds, cetaceans, primates, ungulates, butterflies and plants.

While we recommend to design surveys and collect field data to ensure that key assumptions are met, this is not always possible, and options are needed for when assumption failure cannot be ignored. Therefore, in Chap. 11, we return to assumptions, and consider ways to allow assumptions to be relaxed or removed. This will be done at the expense of requiring additional data and/or additional assumptions.

In Chap. 12, we finish by summarizing key issues, and providing a checklist for designing, conducting and analysing a distance sampling survey. We also discuss some of the new technologies that are being used to change the way surveys are implemented. Although this section is likely to become outdated very quickly, we believe that it is useful to practitioners to be aware of the technological advances that we feel will have most impact on the future implementation of distance sampling methods.

For key case studies, we provide data and files to allow the reader to implement the methods on the book website:

https://synergy.st-andrews.ac.uk/ds-manda/.

Readers' Guide

We provide suggestions on how to use the book for several classes of readers. We hope that the book will be read in its entirety, but on first reading, we provide some suggested paths through the book.

- Those new to distance sampling should read Chaps. 1 and 2 for basic acquaintance with distance sampling principles. In addition, these readers should read Chap. 4 for thoughts on data collection procedures, Chaps. 5 and 6 for fitting detection functions and the estimation of density and abundance, and the checklist of Chap. 12.
- Those familiar with distance sampling should review Chaps. 1–6, noting in particular the discussion of experiments to assess effects of treatments upon abundance in Chap. 3. In addition, Chaps. 9–11 provide additional tools for dealing with special circumstances where traditional distance sampling methods may not perform satisfactorily.
- Readers interested in model-based methods to exploit the data from surveyed plots to draw inference on the study area should read Chaps. 7 and 8 in lieu of Chap. 6. The mathematical treatment of model-based methods in the unified likelihood framework of Chap. 8 is fairly sophisticated.
- We see distance sampling integrated with the assessment of experimental treatment effects (Chap. 3) as well as increased use of model based-inference (Chaps. 7 and 8) as areas of future methodological expansion. These chapters, along with the final chapter (Chap. 12), in which we speculate on the effects of technological advances upon distance sampling, should be read by those making contributions to methodological development of distance sampling.

Acknowledgements

This book includes many illustrative examples based on real case studies. It is a pleasure to thank everyone who kindly provided data for our case studies.

The conservation buffer case studies with indigo buntings and northern bobwhite quail coveys (Sects. 3.2.2.1, 8.5.1 and 8.5.2): The national CP33 monitoring program was coordinated and delivered by the Department of Wildlife, Fisheries, and Aquaculture and the Forest and Wildlife Research Center, Mississippi State University. The national CP33 monitoring program was funded by the Multistate Conservation Grant Program (Grants MS M-1-T, MS M-2-R), a program supported with funds from the Wildlife and Sport Fish Restoration Program and jointly managed by the Association of Fish and Wildlife Agencies, U.S. Fish and Wildlife Service, USDA-Farm Service Agency, and USDA-Natural Resources Conservation Service-Conservation Effects Assessment Project.

The cork oak woodland data (Sect. 5.2.2.4): Filipe S. Dias, Miguel N. Bugalho, Joana Marcelino (Centro de Ecologia Aplicada, Prof. Baeta Neves/Instituto Superior

de Agronomia, Universidade de Lisboa) and Jorge Orestes Cerdeira (Departamento de Matemática, Faculdade de Ciências da Universidade Nova de Lisboa).

The black-capped vireo data used in Sect. 5.4.2.3 were provided by Dr Bret Collier and collected under The U.S. Army Integrated Training Area Management Program, Office of the Secretary of Defense, Department of Defense, who provided funding support.

The long-tailed duck data used in Sect. 7.4.3 were collected by the National Environmental Research Institute of Denmark.

The crossbill lure point transect survey (Sect. 9.2.1): the Royal Society for the Protection of Birds and Scottish Natural Heritage co-commissioned this national survey.

The red deer dung data (Sect. 9.3.3): the Royal Society for the Protection of Birds conducted the surveys and funded the development of the survey design.

The chimpanzee nest data (Sect. 9.3.4): Hjalmar Kühl at the Max Planck Institute for Evolutionary Anthropology.

The DTAG data from Sect. 10.3.3: Peter Tyack for facilitating access to DTAG data. Tagging was performed under US National Marine Fisheries Service Research Permit Nos. 981-1578-02 and 981-1707-00 to PLT and with the approval of the Woods Hole Oceanographic Institution Animal Care and Use Committee. Jessica Ward and NUWC colleagues for processing data to obtain localizations of the animals.

The hare survey data (Sect. 11.1.2): David Tosh and Robbie Macdonald for access to the data; the Northern Ireland Environment Agency (formerly known as the Environment and Heritage Service) through their Natural Heritage Research Partnership with Queen's University Belfast.

The eastern grey kangaroo data survey (Sect. 11.1.3): Brett Howland for facilitating access to the data; ACT Government Ecologists Dr Don Fletcher and Claire Wimpenny for providing GPS collar data.

We thank Eric Howe, who reviewed a draft of the book, and Laura Marshall, who put together the DSsim case study. Suresh Kumar from the SPi TeX Support team was very helpful in sorting bibtex formatting issues.

STB thanks his wife Marlene for unwavering support, and his colleagues at CREEM, who provide a great environment to work on projects such as this book. He is especially grateful to those colleagues who have relieved him of various responsibilities, leaving him (relatively) free to do what he enjoys. He also thanks various collaborators from around the world, who have generated and/or helped solve many interesting problems in distance sampling.

ER acknowledges the support of his colleagues at CREEM, workshop participants and contributors to the distance sampling email list for presenting to him many facets of distance sampling; making him think more deeply about issues faced by practising field biologists. He is grateful to his wife Charisse for her support in this and other adventures.

TAM thanks his family and close friends for the continued support, in particular his grandmother Gabriela Silva for funding his first trip to St Andrews back in 2001. It has been a while since then, but nothing would have happened without it. Dinis

Pestana for pulling him to the dark side, Russell Alpizar-Jara for the first contact with distance sampling and the decisive push towards a PhD in St Andrews. Steve Buckland during the PhD and Len Thomas ever since, for believing in him and providing more support than he could have wished for. Anything I ever do in life will be partly yours. A big thank you to CREEMinals in general: it has been both a pleasure and a privilege to work in such a vibrant yet friendly environment. To Regina, with love. To Filipe, with a message: now both mom and dad have written a book, so it's your turn!

CSO thanks the CREEM team, friends and family for the continuous support during her PhD studies and as a postdoctoral research fellow. Furthermore, she is grateful to scientists outside CREEM: Kristine Evans and Wes Burger, both from Mississippi State University, Jeff Laake from the National Oceanic and Atmospheric Administration and Hans Skaug from the University of Bergen for collaborations over the past few years which are mentioned in this book. CSO was part-funded through the National Centre for Statistical Ecology by EPSRC/NERC grant EP/1000917/1.

St. Andrews, UK

Stephen T. Buckland
Eric A. Rexstad
Tiago A. Marques
Cornelia S. Oedekoven

Contents

Part I Introduction, Survey Design and Field Methods

1 The Basic Methods 3
- 1.1 Introduction 3
- 1.2 Line and Point Transect Sampling 5
- 1.3 The Montrave Case Study 6
- 1.4 Plot Sampling 8
  - 1.4.1 Estimating Density and Abundance 8
  - 1.4.2 Limitations of Plot Sampling 9
- 1.5 Line Transect Sampling 10
- 1.6 Point Transect Sampling 11
- 1.7 Assumptions 12

2 Designing Surveys 15
- 2.1 Point Transect Survey Designs 15
- 2.2 Line Transect Survey Designs 17
- 2.3 Stratification 19
  - 2.3.1 Geographic Strata 19
  - 2.3.2 Post-stratification 22
- 2.4 Survey Effort 23
  - 2.4.1 Minimum Sample Size 23
  - 2.4.2 Estimating Total Effort Required 23
- 2.5 Comparing Designs 25
  - 2.5.1 Assessing Designs by Simulation 25
  - 2.5.2 Example: Comparing the Efficiency of Parallel-Line and Zig-Zag Designs 26

3 Designing Distance Sampling Experiments 29
- 3.1 Introduction 29
- 3.2 Designed Experiments 30
  - 3.2.1 Completely Randomized Designs 30
  - 3.2.2 Randomized Block Designs 30

3.2.3 Factorial Designs 31
3.2.4 Repeated Measures Designs 31
3.2.5 BACI Designs 32
3.3 Impact Assessment Studies 33

4 Field Methods and Data Issues 35
4.1 Field Methods 35
4.1.1 Point Transect Sampling 36
4.1.2 Line Transect Sampling 37
4.2 Data Issues 42
4.2.1 Data Recording 42
4.2.2 Data Checking 43
4.2.3 Field Testing 48

Part II Design-Based and Model-Based Methods for Distance Sampling

5 Modelling Detection Functions 53
5.1 Model Selection and Goodness-of-Fit 54
5.1.1 Pooling Robustness 54
5.1.2 Information Criteria 56
5.1.3 Likelihood Ratio Test 58
5.1.4 Goodness-of-Fit Tests 58
5.2 Conventional Distance Sampling 61
5.2.1 Models for the Detection Function 61
5.2.2 Line Transect Sampling 64
5.2.3 Point Transect Sampling 74
5.2.4 Summary 82
5.3 Multiple-Covariate Distance Sampling 82
5.3.1 Adding Covariates to the Detection Function Model 82
5.3.2 MCDS Case Studies 84
5.4 Mark-Recapture Distance Sampling 93
5.4.1 Modelling the Detection Function 93
5.4.2 MRDS Case Studies 97

6 Design-Based Estimation of Animal Density and Abundance 105
6.1 Introduction 105
6.2 Plot Sampling 106
6.2.1 Quantifying Precision 106
6.2.2 The Montrave Case Study: Plot Sampling 107
6.3 Conventional Distance Sampling 109
6.3.1 Line Transect Sampling 109
6.3.2 Point Transect Sampling 113
6.3.3 Generalizing the Estimation Equations 115
6.4 Horvitz–Thompson-Like Estimators 116
6.4.1 The Horvitz–Thompson Estimator 116

6.4.2 The Horvitz–Thompson-Like Estimator for Conventional Distance Sampling 117
6.4.3 The Horvitz–Thompson-Like Estimator for Multiple-Covariate Distance Sampling 118
6.4.4 The Horvitz–Thompson-Like Estimator for Mark-Recapture Distance Sampling 123

7 Model-Based Distance Sampling: Two-Stage Models 127
7.1 Introduction 127
7.2 Plot Count Models 128
7.2.1 Stage One: Estimating Probability of Detection 128
7.2.2 Stage Two: Modelling the Counts 129
7.3 Plot Abundance Methods 132
7.4 Examples 132
7.4.1 The BACI Case Study 132
7.4.2 Density Surface Modelling: Spotted Dolphins in the Gulf of Mexico 135
7.4.3 The Nysted Windfarm Example 138

8 Model-Based Distance Sampling: Full Likelihood Methods 141
8.1 Introduction 141
8.2 Formulating the Likelihood 142
8.2.1 Conventional Distance Sampling 142
8.2.2 Multiple-Covariate Distance Sampling 144
8.2.3 Mark-Recapture Distance Sampling 146
8.2.4 Plot Count Models 146
8.2.5 Model Extensions 150
8.3 Model Fitting: Maximum Likelihood Methods 153
8.3.1 Maximum Likelihood Estimation 153
8.3.2 Estimating Precision 154
8.4 Model Fitting: Bayesian Methods 155
8.4.1 MCMC Algorithm 155
8.4.2 Model Averaging: Reversible-Jump MCMC 156
8.4.3 Bayesian Inference 157
8.5 Conservation Buffer Case Study 158
8.5.1 Indigo Buntings: A Full Likelihood Approach Using Maximum Likelihood Methods 158
8.5.2 Northern Bobwhite Surveys: A Bayesian Approach 160

Part III Distance Sampling Variations, Special Issues and Assumptions

9 Variations on a Theme 167
9.1 Trapping Webs and Trapping Line Transects 169
9.2 Trapping and Lure Point Transects 170
9.2.1 Lure Point Transect Case Study: Scottish Crossbills 171

9.2.2 Trapping Point Transect Case Study: Key Largo Woodrats 174
9.3 Indirect Surveys 179
9.3.1 Clearance Plot Method 180
9.3.2 The Standing Crop Method 181
9.3.3 Red Deer Dung Decay Rate Experiment 182
9.3.4 Chimpanzee Nest Survey 184
9.4 Cue Counting 187
9.4.1 Cue Counting for Whales 188
9.4.2 Cue Counting for Birds 189
9.5 Surveys in Three Dimensions 194
9.6 Radar, Sonar and Acoustic Surveys 195
9.6.1 Radar Surveys 195
9.6.2 Active Sonar Surveys 198
9.6.3 Passive Acoustic Surveys 199

10 Taxon-Specific Issues 201
10.1 Songbirds 201
10.1.1 Line Transect Versus Point Transect Surveys 201
10.1.2 Minimizing Bias from Bird Movement 202
10.1.3 Minimizing Bias from Failure to Detect All Birds on the Line or Point 205
10.1.4 Minimizing Bias from Inaccurate Distance Measurement 206
10.2 Seabirds 207
10.2.1 Shipboard Surveys 207
10.2.2 Aerial Surveys 208
10.3 Cetaceans 211
10.3.1 Shipboard Surveys 211
10.3.2 Aerial Surveys 215
10.3.3 Acoustic Surveys 216
10.4 Primates 220
10.4.1 Dealing with Primate Groups 221
10.4.2 Other Approaches 222
10.5 Ungulates 223
10.5.1 Ground-Based Surveys 223
10.5.2 Aerial Surveys 224
10.6 Butterflies 225
10.7 Plants 226

11 Exchanging Assumptions for Data 231
11.1 Non-random Transect Placement 232
11.1.1 Bias 232
11.1.2 Correcting for Bias 233
11.1.3 Example: Eastern Grey Kangaroos 235
11.1.4 Estimating Bias by Simulation 237
11.1.5 Other Approaches 239

11.2 Uncertain Detection on the Transect 240
11.3 Measurement Error Models 243
11.3.1 Bias Arising from Measurement Error 243
11.3.2 Accounting for Measurement Error 244
11.3.3 A Practical Example 246
11.3.4 Take-Home Messages 248
11.4 Addressing Bias from Animal Movement 249
11.4.1 Movement Independent of the Observer 249
11.4.2 Responsive Animal Movement 251

12 Summary 253
12.1 Check-List for a Good Survey 254
12.2 New Technology 257
12.2.1 Using Technology to Detect Animals 257
12.2.2 Data Recording 259
12.2.3 Distance Estimation 259
12.2.4 Aerial Surveys 260
12.2.5 Satellites and GPS 260

Erratum E1

Erratum 1 ..... E3

References 263

Index 275

The original version of the Preface was revised. An erratum can be found at https://doi.org/10.1007/978-3-319-19219-2_14

Part I
Introduction, Survey Design and Field Methods

Chapter 1
The Basic Methods

1.1 Introduction

One of the most common and pervasive questions in applied ecology relates to the size of a given population. How many animals are there? The question is intrinsically interesting, but perhaps even more important from an applied perspective, the actual answer has implications for most ecological processes affecting that population. The effective management of a population is not possible without knowing at least approximately how many individuals it includes. As an example, for a small population, a given mortality rate due to a newly introduced human disturbance might be important and a matter of concern, quickly leading the population to local extinction, but essentially not ecologically relevant for an abundant population. Therefore, knowledge about abundance is required to adequately interpret a wide variety of ecological processes affecting a given population.

Under all but the simplest of scenarios, total enumeration, known as a *census*, is not an option. Therefore, researchers depend on some kind of sampling method to answer this question, and will have to draw inferences on a population based on a sample. Conventional sampling ideas like randomness, replication and simple random sampling are important in this context. However, due to the special characteristics of wild populations, it is hard to implement conventional survey methods. One cannot simply divide the area of interest into a large number of sampling units and count all animals within a random selection of these. The difficulty will stem from three special characteristics of estimating abundance of wild populations. To begin with, we are interested in estimating the size of the population. Contrast that with most traditional sampling scenarios, where population size is assumed known, and we want to draw a random sample of individuals from the population to estimate a characteristic of the members of the

The original version of this chapter was revised. An erratum to this chapter can be found at https://doi.org/10.1007/978-3-319-19219-2_14

© Springer International Publishing Switzerland 2015

S.T. Buckland et al., *Distance Sampling: Methods and Applications*,
Methods in Statistical Ecology, DOI 10.1007/978-3-319-19219-2_1

population. For example, we might wish to estimate the mean height of students at a school based on selecting a small number of students at random. Secondly, even after selection of spatial sampling units, we cannot guarantee that all individuals present are detected. Thirdly, wild populations exist in time and space but most are not static, and hence the traditional concept of sampling and sampling units gets blurred. Animals can move around and avoid observers, approach observers, *etc*. The relative speed (observer to animal speed) at which we observe the areas of interest will have an influence on the outcome of the sampling.

Because of these features, it is unsurprising that various methods have been developed explicitly to deal with them. A review of available methods is beyond the scope of this book, but the reader is referred to the excellent article by Schwarz and Seber (1999) for an overview. We note in passing that George A.F. Seber is one of the authors who contributed most to synthesizing the knowledge in this area with two earlier reviews (Seber 1986, 1992) and a book (Seber 1982) on the theme. The most commonly-used methods in this context are capture–recapture and distance sampling. While for the former we can think of sampling individuals, with information about detectability contained in the capture histories of different individuals, for the latter, we can think of sampling areas, with information on detectability being contained in the detection distances. The underlying idea to both approaches is that, if we can estimate the probability of detection for those objects of interest that we detect, we can also estimate how many were undetected. From that estimate of detectability, we can correct the observed counts for individuals missed on the sampled plots.

The term 'distance sampling' was introduced by Buckland et al. (1993) to include a suite of methods, including line and point transect sampling, in which animal density or abundance is estimated from a sample of distances to detected individuals. The first version of the `Distance` software (Laake et al. 1993) was released to allow users to implement the methods of the book.

Introduction to Distance Sampling (Buckland et al. 2001) updated the earlier book, and remains the standard reference for 'conventional distance sampling', corresponding with the `cds` engine of `Distance`. *Advanced Distance Sampling* (Buckland et al. 2004) addressed a range of advances, including multiple-covariate distance sampling (the `mcds` engine of `Distance`), density surface modelling (the `dsm` engine of `Distance`), mark-recapture distance sampling for double-observer data (the `mrds` engine of `Distance`), and automated survey design (implemented in `Distance` with the aid of various Geographic Information System algorithms). The stand-alone version of `Distance` is summarized by Thomas et al. (2010).

As more advanced methods were developed, it became easier to implement them using the statistical software `R`. Initially, developments were implemented in `Distance`, with `R` called by `Distance`, so that the user was largely unaware of the role of `R`. However, to give users the full power of the `mrds` analysis engine with a simplified syntax and data structures, a version of `Distance` was released as an `R` package (Miller 2013).

Datasets and the means to implement methods (computer code or `Distance` projects) are provided for the book case studies at the book website:
https://synergy.st-andrews.ac.uk/ds-manda/.

Some of the case studies (Montrave, Sect. 1.3, and Hawaiian song birds, Sect. 5.3.2.1) are based upon the `Distance` software. The northern bobwhite quail (Sect. 8.5.2) and Scottish crossbill (Sect. 9.2.1) case studies use specialized software and cannot be accomplished using `Distance`. Finally, the spotted dolphin (Sect. 7.4.2) and simulation (Sect. 2.5.2) case studies can be done either using `Distance` or using only `R` packages `dsm` and `DSsim`.

1.2 Line and Point Transect Sampling

There are two basic methods of distance sampling: line transect sampling and point transect sampling. All other methods may be considered as extensions of one or the other. In line transect sampling, a series of lines is distributed according to some design (usually a systematic grid of parallel lines, Chap. 2), and an observer travels along each line, searching for animals or animal *clusters*. We define a cluster to be a group of animals with a well-defined location for the group centre. For each animal or cluster detected, the observer measures or estimates the (perpendicular) distance x of the animal or cluster centre from the nearest part of the line. In many surveys, it is easier to measure or estimate the observer-to-animal (radial) distance r at the time the detection is made. If the sighting angle θ is also measured, then the perpendicular distance may be found by simple trigonometry: $x = r \sin \theta$ (Fig. 1.1).

For point transect sampling, the design comprises a set of points, often spaced on a systematic grid (Chap. 2). The observer travels to each point, and records any animal (or cluster) detected from the point, together with its distance r from the point.

Conceptually, distance sampling is a form of plot sampling in which not all animals on the plot are detected. In line transect sampling, the plots are rectangles of size $2wl$ where l is the length of a given line (which may vary by line) and w is

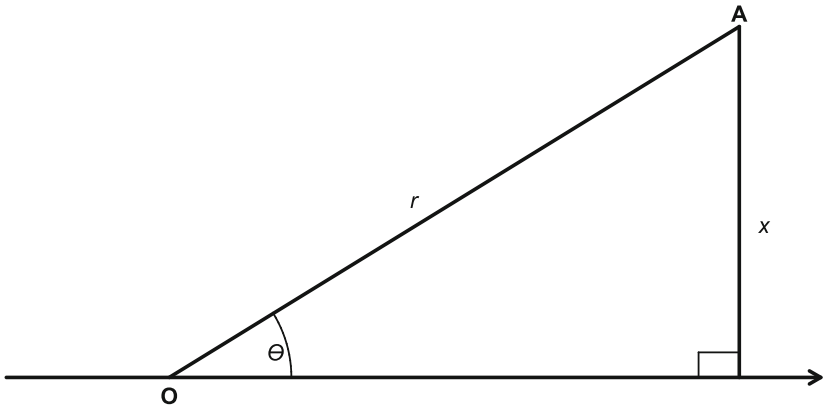

Fig. 1.1 An observer travels along the transect from left to right. When the observer is at point O, an animal is detected at A. The observer can record either the perpendicular distance x directly, or both the radial distance r and the sighting angle θ, from which $x = r \sin \theta$

the half-width of the strip; it is also the truncation distance for distances from the line, which might be pre-determined, so that detections further than w from the line are not recorded, or may be set by the analyst, so that distances greater than w are excluded from analysis. In point transect sampling, the plots are circles of size πw^2, where w, the plot radius, is again the truncation distance.

1.3 The Montrave Case Study

We will use the surveys of Buckland (2006) as a case study to illustrate standard distance sampling methods. We will refer to this study as the Montrave case study; it was conducted in an area of woodland and parkland at Montrave in Fife, Scotland, in the spring of 2004. Four species were recorded: common chaffinch (*Fringilla coelebs*), winter wren (*Troglodytes troglodytes*), great tit (*Parus major*) and European robin (*Erithacus rubecula*). Only territory-holding male birds were recorded, as some females may be undetectable even at close range. In addition to standard line and point transect sampling methods, a snapshot point transect method and a cue count survey were conducted (Buckland 2006). We use line transect and snapshot point transect data on the robin (Fig. 1.2) as a case study. We also use the robin data to illustrate both cue counting and plot sampling. For plot sampling, we truncate detections at 35 m, and assume that all male birds within that distance were detected.

Fig. 1.2 The European robin is a strongly territorial species, occurring in woodland habitats throughout Europe. Photo: Steve Buckland

Survey design for this case study will be shown in Chap. 2. Estimation of the probability of detection will be addressed in Chap. 5, and estimating bird density will be covered in Chap. 6. For plot sampling only, density will be estimated in Sect. 6.2.2. A cue counting analysis is presented in Sect. 9.4.2.

We provide the data used to illustrate line transect sampling in Table 1.1. Note that no robins were detected from three of the lines. These lines must be included in the analysis; if they are not, mean density and hence abundance will be overestimated. The distances are plotted in a histogram in Fig. 1.3.

The data for point transect sampling are shown in Table 1.2 and Fig. 1.4. Note the different shape of the histogram relative to Fig. 1.3. This arises because there is

Table 1.1 The line transect data for European robins from the Montrave study

Line	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Length (m)	208	401	401	299	350	401	393	405	385	204	39	47	204	271	236	189	177	200	20
Distances	12	5	0	15	15	15	15	23	15	55		45	5	5	5		15	10	
	26	5	5	35	21	15	30	27	45			50	25	10	20		25	25	
	70	20	5	72	22	20	50	38	80				38	27	25		33	70	
		24	15	75	28	35	60	55					45	40	45		35		
		30	20	100	33	50	85	58					45	45					
		45	22		35	65		75						45					
		50	27		45	70		100											
		53	28		60														
			32																
			40																
			60																

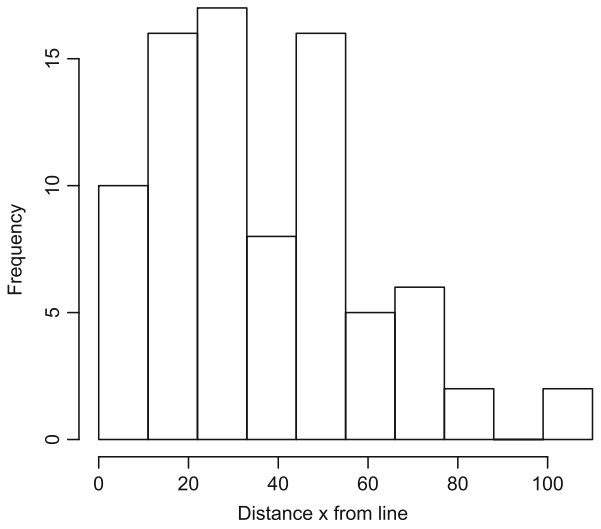

Fig. 1.3 Distribution of robin detections by distance (line transect sampling)

Table 1.2 The snapshot point transect data for European robins from the Montrave study

Point	1	2	3	6	7	8	9	11	13	14	15	16	19	20	21	22	24	25	26	27	28	30	31	32
Distances	100	35	40	40	115	60	40	65	42	45	35	45	20	55	27	30	25	45	25	35	55	100	25	25
			85		120	80			70	45	45	65	30	100	35	40	40	80	35	45				36
									70	60	55		65		45	70	50		40	60				
											55						60		50					

No robins were detected from points 4, 5, 10, 12, 17, 18, 23 and 29

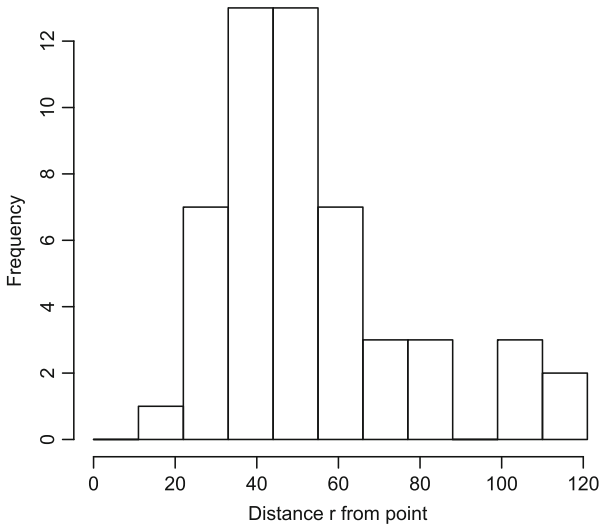

Fig. 1.4 Distribution of robin detections by distance (snapshot point transect sampling)

very little area within a small distance of a point, and so few birds are detected close to a point. Thus numbers of detections increase with distance, until the distance is sufficiently large that we fail to detect many of the birds.

1.4 Plot Sampling

1.4.1 Estimating Density and Abundance

Consider first the case that all animals on the sampled plots are detected. For line transects, we term this special case *strip transect sampling*, and for point transects, *circular plot sampling*. To estimate animal density D and hence abundance $N = DA$, where A is the size of the study region, we simply take the total number n of animals detected across the plots, and divide by the total size a of the plots: $\hat{D} = n/a$ where the hat ($\hat{}$) indicates that this is an estimate based on our sample data.

The corresponding abundance estimate is $\hat{N} = \frac{nA}{a}$. For strip transect sampling with K lines, we have $a = 2wL$ where $L = \sum_{k=1}^{K} l_k$ is the total length of the K lines. Thus we have

$$\hat{D} = \frac{n}{2wL} \tag{1.1}$$

and

$$\hat{N} = \frac{nA}{2wL} \, . \tag{1.2}$$

For circular plot sampling with K points, $a = K\pi w^2$ so that

$$\hat{D} = \frac{n}{K\pi w^2} \tag{1.3}$$

and

$$\hat{N} = \frac{nA}{K\pi w^2} \, . \tag{1.4}$$

1.4.2 Limitations of Plot Sampling

Because we assume that all animals on each plot are detected, either we must make the plots small, so they can be fully surveyed from the centre point or line, or effort must be expended searching the plot. The more time that is spent on the plot, the greater the bias arising from new animals entering the plot. Active searching of the plot may also disturb animals, which may move off the plot undetected; and it can be difficult to keep track of those animals already counted. Whether plots are small, and searched quickly, or large, and searched more slowly, many detected animals are likely to be beyond the plot. These detections are ignored, reducing precision. Further, it can be difficult to judge whether an animal close to a boundary, especially an animal that is moving around, is inside or outside the plot; observers have a tendency to include such animals in their counts, causing upward bias in abundance estimates. For standard line or point transect sampling, for which probability of detection at the plot boundary is typically low, such bias causes some upward bias in estimated probability of detection at the boundary, but there is a corresponding upward bias in the estimate of the effective size of the plot surveyed, and the two tend to cancel, so that there is little if any bias in resulting abundance estimates.

A further limitation of plot sampling is that it can be very difficult to judge whether the observer has successfully detected all animals on each plot. If distances of detected animals from the line or point are recorded, then we can explore whether numbers at larger distances are fewer than we would expect, if all animals are being detected. However, if we are recording these distances anyway, then we can make full use of the data by using line and point transect methods.

In some studies, no attempt is made to detect all animals on the plots, yet the plot counts are not corrected for detectability, on the grounds that only a measure of relative abundance is needed, given the study objectives. This rationale is valid provided detectability does not vary in any systematic way across plots or over time. Given that detectability is not estimated, this assumption typically cannot be tested. Further, studies show that detectability varies by many factors, including habitat, weather conditions, animal behaviour, time of day, time of year, observer experience, observer age, *etc*. It seems preferable therefore to adopt methods that allow detectability to be estimated and controlled for, if it is not practical to carry out complete counts on plots.

1.5 Line Transect Sampling

In conventional line transect sampling, we assume that all animals on the line are detected, but that probability of detection falls off smoothly with distance from the line. We use the recorded distances x of detected animals from the line to model the *detection function*, $g(x)$, which is defined to be the probability of detecting an animal that is at distance x ($0 \leq x \leq w$) from the line. For the standard method, we assume that animals on the line are certain to be detected: $g(0) = 1$.

We may regard line transect sampling as plot sampling, in which the plots are long, narrow strips of half-width w, centred on the transect line, and for which not all animals on a plot are counted. For those animals on a plot, we denote the expected proportion counted by P_a. Thus for strip transect sampling, $P_a = 1$, but for line transect sampling, $0 < P_a < 1$. We show in Fig. 1.5 that $P_a = \mu/w$ where $\mu = \int_0^w g(x)\,dx$. Hence, given an estimate $\hat{g}(x)$ of the detection function $g(x)$, we can obtain an estimate $\hat{P}_a$ of P_a. Equation (1.1) of strip transect sampling is replaced by:

$$\hat{D} = \frac{n}{2wL\hat{P}_a}\,. \tag{1.5}$$

Thus if we estimate that one half of the animals on the plot of half-width w are counted, we double the density estimate that would be obtained if we were to assume that all animals on the plot were counted. This is equivalent to estimating density under the assumption that our count of n animals was a complete count of animals on a strip of half-width μ, which has led to μ being termed the effective strip half-width.

We need a method for estimating $g(x)$. As we will see in Sect. 5.2.2.1, we can derive the *probability density function* corresponding to the distances x simply by rescaling $g(x)$ so that it integrates to one — a property that all probability density functions share. Thus $f(x) = g(x)/\mu$ for $0 \leq x \leq w$, and in particular, $f(0) = 1/\mu$ because we assume that $g(0) = 1$. The advantage of this is that a probability density function represents the relative likelihood of different observations, and we have general methods for fitting such functions (Sect. 5.2.2.1). Having obtained the

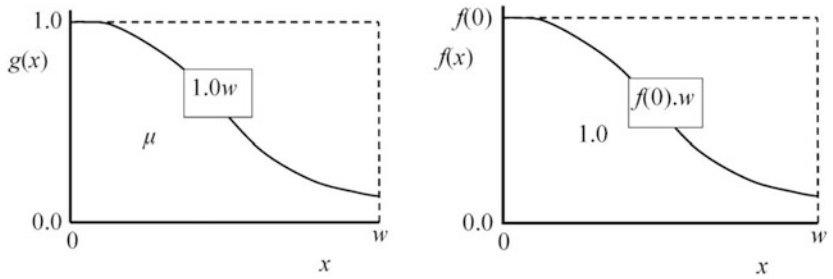

Fig. 1.5 The probability P_a that an animal within distance w of a line is detected may be represented by the proportion of the rectangle that is under the curve in these plots. If the curve is the detection function $g(x)$ (*left plot*), this gives $P_a = \mu/w$ where $\mu = \int_0^w g(x)\,dx$ is the area under the curve, and the area of the rectangle is $g(0).w = w$, as $g(0) = 1$. For the probability density function (*right plot*), the area under the curve is 1.0 (a property of all probability density functions), and so $P_a = 1/[w.f(0)]$. Hence the alternative forms for $\hat{D}$ in Eq. (1.6)

estimated function $\hat{f}(x)$, and evaluated it at $x = 0$ to give $\hat{f}(0)$, we can write

$$\hat{D} = \frac{n}{2wL\hat{P}_a} = \frac{n}{2\hat{\mu}L} = \frac{n\hat{f}(0)}{2L}\,. \tag{1.6}$$

The conceptual relationship between P_a, μ and $f(0)$ is shown in Fig. 1.5.

We address modelling of the detection function in greater detail in Sect. 5.2.2.1, and the estimation of density and abundance in Sect. 6.3.1.

1.6 Point Transect Sampling

As for line transect sampling, we start by estimating the detection function, which we now represent by $g(r)$, where r is distance from the point. Thus $g(r)$ is the probability that an animal at distance r from a point is detected ($0 \leq r \leq w$). We again assume $g(0) = 1$. Having fitted a model for the detection function, we can estimate the proportion P_a of animals available for detection (*i.e.* within distance w of a point) that are detected. Then Eq. (1.3) is extended as:

$$\hat{D} = \frac{n}{K\pi w^2 \hat{P}_a}\,. \tag{1.7}$$

As with line transect sampling, for the purpose of estimating P_a, it is convenient to work with the probability density function. However, unlike line transect sampling, this function does not now have the same shape as the detection function. We will see in Sect. 5.2.3 that the probability density function $f(r)$ of detection distances r from the point is proportional to the detection function $g(r)$ multiplied by distance

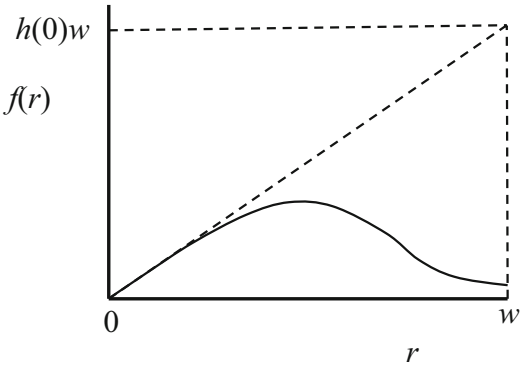

Fig. 1.6 The probability P_a that an animal within distance w of a point is detected may be represented by the proportion of the triangle that is under the curve in this plot. The area under the curve is one, as the curve is a probability density function. The triangle has base w and height $h(0)w$ where $h(0)$ is the slope of the hypotenuse, so that its area is $h(0)w^2/2$, giving $P_a = 2/[h(0)w^2]$. Note that, if all animals out to distance w were detected, then the probability density function would be represented by the triangle, and we would have $P_a = 1$

r. As illustrated in Fig. 1.6, we have $P_a = 2/[h(0)w^2]$ where $h(0)$ is the slope of the probability density function evaluated at zero distance. Thus we can write Eq. (1.7) as:

$$\hat{D} = \frac{n}{K\pi w^2 \hat{P}_a} = \frac{n\hat{h}(0)}{2K\pi} \,. \tag{1.8}$$

We give details of estimating the detection function in Sect. 5.2.3, and we address density and abundance estimation in Sect. 6.3.2.

1.7 Assumptions

There is one key design assumption of distance sampling and three key model assumptions. These four assumptions are:

1. Animals are distributed independently of the lines or points.
2. Objects on the line or at the point are detected with certainty.
3. Distance measurements are exact.
4. Objects are detected at their initial location.

Most extensions of the above theory are to address failure of one or more of these assumptions. The first assumption is the key design assumption. Stated more fully, we assume that animals are distributed uniformly with respect to distance from the line (so that $\pi(x) = 1/w$) for line transect sampling or that animals are

distributed according to a triangular distribution with respect to distance from the point (so that $\pi(r) = 2r/w^2$) for point transect sampling. We call this a design assumption because a suitably randomized design will ensure that the assumption holds. (So-called plus-sampling into a bufferzone around the study region ensures that edge effects do not cause violation of this assumption (Strindberg et al. 2004), but for large study regions, edge effects tend to be negligible if this issue is ignored.) Thus lines or points should be positioned according to a random design. Usually, a systematic grid of lines or points is placed randomly over the study region.

Model assumptions (assumptions 2–4 above) are usually of greater concern, because we cannot ensure that they hold by adopting a suitable design. Instead we need to consider whether field methods can be adopted that will ensure low bias when they fail. Assumption 4 is related to animal movement, which may be in response to the observer or independent of the observer. If animals show responsive movement, then field methods should if possible ensure that animals are detected (and their locations recorded) before they respond. Movement that is independent of the observer should be slow relative to observer speed. If this is not possible (for example for point transect sampling, in which the observer does not move), then the standard methods may require modification. In Chap. 11, we will consider methods to address failures of each of these three model assumptions.

Other assumptions are usually of little consequence. We assume that detections are independent, but adopt estimation methods that are very robust to failures of this assumption, although assessing model fit is more problematic if there is strong violation. For clustered populations, we assume that the size of detected clusters is recorded without error. Using the default method in `Distance` for estimating mean cluster size, we can relax this assumption, so that for clusters on or close to the line or point, cluster size estimates are assumed to be unbiased. Another minor assumption is that, having detected an animal, the observer is able to avoid counting it a second time from the same line or point. Counting the same animal across multiple lines because it has moved between surveying one line and the next should not be a problem, provided that the animal movement is not in response to the observer. In the absence of directed movement such as migration, on average for every animal moving from line 1 to line 2, there is another animal that moves the other way, so that any effect tends to cancel out.

Chapter 2
Designing Surveys

Survey design is covered in greater detail by Buckland et al. (2001, pp. 230–248) and by Strindberg et al. (2004). We concentrate here on basic designs, and on issues that commonly need to be addressed. Some of those issues relate to analysis; in such cases, we provide forward references to where they are addressed. Although through most of the book, we describe line transect methods before point transect methods (for which analysis has greater complexity), for survey design, we consider point transect sampling first. This is because point transect surveys are simpler to design than are line transect surveys.

2.1 Point Transect Survey Designs

A standard point transect survey comprises points spanning the entire study area, and systematically spaced, so that they form a grid. The Montrave study is an example of this design: points are at the intersections of a grid comprising squares of side 100 m (Fig. 2.1).

In some studies, a simple random sample of units is preferred to a systematic sample. In this case, it is again convenient to define a grid of squares, so that the length of the side of each square is at least $2w$, where w is the distance from the point beyond which either detected animals are not recorded, or such detections are truncated at the analysis stage. (In the latter case, this value might be assessed from similar past studies.) A simple random sample of grid squares is then selected, and a point placed in the middle of each selected square. This ensures that the circular plots around selected points cannot overlap. See Fig. 7.2 of Strindberg et al. (2004).

For large study areas with wide separation between points, the above systematic design is inefficient, as the observer must travel a large distance from one point to the next, and then record just animals detectable from that point. In this case, cluster sampling is a better option: at each selected point, locate a small grid of points, with

© Springer International Publishing Switzerland 2015

S.T. Buckland et al., *Distance Sampling: Methods and Applications*,
Methods in Statistical Ecology, DOI 10.1007/978-3-319-19219-2_2

Fig. 2.1 The survey design for the Montrave study. The surveyed woodland and parkland totalled about 33 ha in two blocks, surrounded by grazing and crops. For the point transect surveys, a systematic grid of 32 points (indicated by *circles*) was placed over the study area, with 100 m separation between neighbouring points. Each point was visited twice. For the line transect survey, there were 19 lines (11 in one block and eight in the other), with about 70 m separating successive lines. Each line was surveyed twice

separation distance between neighbouring points in the vicinity of $2w$. For analysis, each cluster (*i.e.* grid of points) will be treated as a single sampling unit.

Often, points are positioned along lines. If the separation of neighbouring points on the same line is less than the separation of neighbouring points on different lines, then this should be treated as cluster sampling, where each line represents one cluster of points. If the separation is the same, then the surveyed points form a regular grid through the study area, and each point may be treated as a sampling unit. See Fig. 2.2.

When points are placed at random within a study area, some will fall within distance w of the study area boundary. In such cases, the intersection of the circular plot with the study area is no longer a complete circle, which creates an 'edge effect' on animal availability. Such edge effects cause bias in abundance estimates.