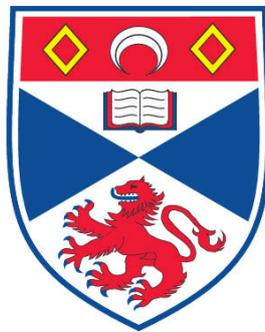


**STATISTICAL MODELS FOR THE LONG-TERM MONITORING OF  
SONGBIRD POPULATIONS:  
A BAYESIAN ANALYSIS OF CONSTANT EFFORT SITES AND  
RING-RECOVERY DATA**

**Vanessa M. Cave**

**A Thesis Submitted for the Degree of PhD  
at the  
University of St. Andrews**



**2010**

**Full metadata for this item is available in the St Andrews  
Digital Research Repository  
at:**

**<https://research-repository.st-andrews.ac.uk/>**

**Please use this identifier to cite or link to this item:  
<http://hdl.handle.net/10023/885>**

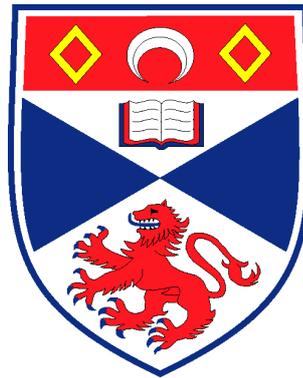
**This item is protected by original copyright**

**This item is licensed under a  
Creative Commons License**

# Statistical Models for the Long-Term Monitoring of Songbird Populations:

*A Bayesian Analysis of Constant Effort Sites  
and Ring-Recovery Data*

Vanessa M. Cave



Thesis submitted for the degree of  
DOCTOR OF PHILOSOPHY  
in the School of Mathematics and Statistics  
UNIVERSITY OF ST ANDREWS

March 2010

Copyright ©2010 Vanessa M. Cave



# Declaration

I, Vanessa Marion Cave, hereby certify that this thesis, which is approximately 70,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in July 2007 (credited with 15 months of study at the University of Cambridge) and as a candidate for the degree of Doctor of Philosophy in Statistics in July 2007; the higher study for which this is a record was carried out in the University of St Andrews between 2007 and 2009.

date: 8 March, 2010

signature of candidate:

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in Statistics in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date: 8 March, 2010

signature of supervisor:

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to Printed copy and electronic publication of thesis through the University of St Andrews.

date: 8 March, 2010

signature of candidate:

date: 8 March, 2010

signature of supervisor:



# Abstract

To underpin and improve advice given to government and other interested parties on the state of Britain’s common songbird populations, new models for analysing ecological data are developed in this thesis. These models use data from the British Trust for Ornithology’s Constant Effort Sites (CES) scheme, an annual bird-ringing programme in which catch effort is standardised. Data from the CES scheme are routinely used to index abundance and productivity, and to a lesser extent estimate adult survival rates. However, two features of the CES data that complicate analysis were previously inadequately addressed, namely the presence in the catch of “transient” birds not associated with the local population, and the sporadic failure in the constancy of effort assumption arising from the absence of within-year catch data. The current methodology is extended, with efficient Bayesian models developed for each of these demographic parameters that account for both of these data nuances, and from which reliable and usefully precise estimates are obtained.

Of increasing interest is the relationship between abundance and the underlying vital rates, an understanding of which facilitates effective conservation. CES data are particularly amenable to an integrated approach to population modelling, providing a combination of demographic information from a single source. Such an integrated approach is developed here, employing Bayesian methodology and a simple population model to unite abundance, productivity and survival within a consistent framework. Independent data from ring-recoveries provide additional information on adult and juvenile survival rates. Specific advantages of this new integrated approach are identified, among which is the ability to determine juvenile survival accurately, disentangle the probabilities of survival and permanent emigration, and to obtain estimates of total seasonal productivity.

The methodologies developed in this thesis are applied to CES data from Sedge Warbler, *Acrocephalus schoenobaenus*, and Reed Warbler, *A. scirpaceus*.



# Acknowledgements

First, and foremost, I wish to express my sincere gratitude to Dr Ruth King, Lecturer in Statistics at the University of St Andrews, who so kindly and so proficiently adopted the role of primary supervisor at the start of my second post-graduate year. Ruth's guidance, academic insight, and constant encouragement were invaluable and are hugely appreciated.

I would also like to convey my wholehearted thanks to Professor Steve Brooks, Director of Research at ATASS Ltd, United Kingdom, for his wonderful supervision throughout the first year of my post-graduate studies. In particular, special thanks are due to Steve for the significant part he played in developing the proposal for, and organising the funding of, this research.

I also wish to express my appreciation to Dr Stephen Freeman, Statistician at the Centre for Ecology & Hydrology, United Kingdom, for the supervision, numerous ideas, biological insight and the useful discussion he has provided throughout my post-graduate research. Stephen was also very instrumental in the establishment of this project, in particular obtaining access to the Constant Effort Sites (CES) and National Ring-Recovery (NRR) data sets.

The research in this thesis would not have been possible without the immense effort, and large time commitment, the volunteer ringers spend collecting CES data annually, for which I thank them. The UK ringing scheme is jointly funded by the British Trust for Ornithology, the Joint Nature Conservation Committee (on behalf of Natural England, the Countryside Council for Wales, Scottish Natural Heritage, and the Environment and Heritage Service of Northern Ireland) and the ringers themselves.

My special thanks is extended to the British Trust for Ornithology for granting me access to their data, and showing a keen interest in my research. In particular I would like to acknowledge Dr Stephen Baillie, Dawn Balmer, Jacquie Clark, Mark Grantham, Professor Jeremy Greenwood and Dr Rob Robinson.

This research was made possible by funding from the National Centre for Statistical Ecology (NCSE) through an Engineering and Physical Sciences Research Council grant, for which I am most appreciative.

Special thanks are also due to AgResearch Ltd, New Zealand, who so generously provided additional funding. I thank my colleagues in the Bioinformatics,

Mathematics and Statistics section for the encouragement, support, and inspiration they have given me prior to, and during, my post-graduate career.

I am greatly indebted to all those countless people who have supported, inspired, advised and taught me over the years. Collectively I thank: my secondary school teachers at Menzies College (New Zealand), my undergraduate Lecturers and Tutors at the University of Otago (New Zealand) who nurtured and developed my interest in, and love of, statistical analysis, my friends and past colleagues at the Statistical Laboratory, University of Cambridge, that made my time there such a pleasure, everyone at the Centre for Research into Ecological and Environmental Modelling, University of St Andrews, who are such a lovely group of people to work with, and the members and Directors (Professor Steve Buckland and Professor Byron Morgan) of the NCSE for the wonderful academic opportunity this centre has afforded me.

My heartiest thanks goes out to all my wonderful friends for the endless kindness and support they have furnished me with. To name but a few, Katie and John Enlow for their constant encouragement, Florence Cash who reminded me to laugh, Simon Tan for always being there for me, James Cruise who kept me realistic, Jane Hunter, Stephanie Tse and Willy Flockton for the fantastic holidays, Angelika Studeny and Christiana Spyrou for being such wonderful office mates and friends, Toby Reynolds for all the help with “Uni stuff”, and Debbie Russell who helped me stay sane.

Finally, I wish to express my heartfelt thanks to my marvelous family for all the support, love and laughter they endlessly, and so patiently, provided. My terrific parents, Mervyn and Tricia, my fantastic brothers, Hadley and Rowan, and my wonderful partner, Graeme. Thank you.

# Contents

<b>Abstract</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vii</b>
<b>List of Figures</b>	<b>xvi</b>
<b>List of Tables</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Monitoring Bird Populations: Why, How and Who . . . . .	1
1.2 Integrated Population Modelling . . . . .	4
1.3 Constant Effort Sites Scheme . . . . .	6
1.3.1 BTO Bird Monitoring Schemes . . . . .	6
1.3.2 CES Field Protocol . . . . .	8
1.3.3 CES Data . . . . .	11
1.4 Sedge and Reed Warblers . . . . .	14
1.5 Bayesian Methodology . . . . .	18
1.5.1 The Bayesian Approach . . . . .	19
1.5.2 Bayesian Inference . . . . .	20
1.5.3 The Metropolis-Hastings Algorithm . . . . .	21
1.6 Thesis Outline . . . . .	25
	<b>ix</b>

<b>2</b>	<b>Independent Models for CES Data with Complete Coverage</b>	<b>27</b>
2.1	Data . . . . .	28
2.2	Estimating Abundance . . . . .	29
2.2.1	Model Fitting and Results . . . . .	30
2.2.2	Validity of Model Assumptions . . . . .	33
2.3	Estimating Productivity . . . . .	36
2.3.1	Traditional Approach . . . . .	36
2.3.2	Model Fitting and Results . . . . .	38
2.3.3	Advantages of using CES Data to Monitor Productivity .	41
2.4	Estimating Adult Survival . . . . .	43
2.4.1	Models . . . . .	44
2.4.2	Model Fitting . . . . .	57
2.4.3	Comparing the Alternative Survival Models . . . . .	59
2.4.4	The Best Model for CES Live-Recapture Data . . . . .	67
2.4.5	Analysis of Sedge and Reed Warbler Data . . . . .	69
2.4.6	Validity of Model Assumptions . . . . .	73
2.4.7	Comment on Robust Design Models . . . . .	77
2.4.8	Concluding Remarks . . . . .	78
2.5	Chapter Summary . . . . .	80
<b>3</b>	<b>Integrated Population Model for Complete Coverage Data</b>	<b>82</b>
3.1	Integrated Population Modelling . . . . .	82
3.2	An Integrated Model for CES Data . . . . .	84
3.2.1	CES Live-Recapture Data . . . . .	86
3.2.2	Ring-Recovery Data . . . . .	86
3.2.3	Joint Analysis of Ring-Recovery and Live-Recapture Data	91
3.2.4	Bayesian Fitting of the Integrated Population Model . .	93
3.2.5	“Baseline” Models . . . . .	94

---

3.3	Analysis of Sedge Warbler Data . . . . .	94
3.3.1	Integrated Model with Ring-Recovery Data . . . . .	94
3.3.2	Including CES Live-Recapture Data . . . . .	97
3.4	Analysis of Reed Warbler Data . . . . .	104
3.5	Investigating the Relationship between the Demographic Parameters . . . . .	108
3.6	Concluding Remarks . . . . .	111
3.6.1	Integrated Modelling of UK Bird Populations . . . . .	111
3.6.2	An Integrated Model for CES Data - advantages . . . . .	112
3.6.3	Non-Independence of Likelihoods . . . . .	114
3.6.4	Goodness-of-Fit . . . . .	114
3.6.5	Potential Developments . . . . .	115
3.7	Chapter Summary . . . . .	117
<b>4</b>	<b>Independent Models for CES Data with Incomplete Coverage</b>	<b>119</b>
4.1	Data . . . . .	121
4.2	Estimating Abundance from Censored Data . . . . .	123
4.2.1	Classical Model of Peach et al. (1998) . . . . .	123
4.2.2	Alternative Bayesian Model for Censored Counts . . . . .	126
4.2.3	Model Assumptions . . . . .	131
4.2.4	Testing the Bayesian Model - a Simulation Study . . . . .	132
4.2.5	Analysis of Sedge and Reed Warbler Data . . . . .	138
4.2.6	Modelling CES Count Data at the Visit Level . . . . .	144
4.2.7	Concluding Remarks . . . . .	145
4.3	Estimating Productivity from Censored Data . . . . .	148
4.3.1	Simulation Study . . . . .	148
4.3.2	Analysis of Sedge and Reed Warbler Data . . . . .	150
4.3.3	Traditional Approach . . . . .	151
4.4	Estimating Adult Survival from CES Data with Missing Visits . . . . .	156

4.4.1	Effect of Missed Visits and Transient Birds . . . . .	157
4.4.2	Pattern of Within-Year Captures . . . . .	158
4.4.3	Survival Model for CES Data with Missed Visits . . . . .	161
4.4.4	Survival Model for CES data with Missed Visits and Transient Birds . . . . .	173
4.4.5	Analysis of Sedge and Reed Warbler Data . . . . .	185
4.4.6	Concluding Remarks . . . . .	187
4.5	Comparing Analyses which Exclude or Include Incomplete Coverage Data . . . . .	191
4.6	Chapter Summary . . . . .	194
<b>5</b>	<b>Extensions to the Integrated Population Model</b>	<b>195</b>
5.1	Including Years with Incomplete Coverage . . . . .	196
5.1.1	Bayesian Model Fitting . . . . .	198
5.1.2	“Baseline” Models . . . . .	200
5.1.3	Analysis of Sedge and Reed Warbler Data . . . . .	200
5.1.4	Advantage of Including Years with Incomplete Coverage	204
5.1.5	Relationship between the Demographic Parameters . . .	207
5.2	Allowing for Permanent Emigration . . . . .	209
5.2.1	An Integrated Population Model for CES Data in the Presence of Permanent Emigration . . . . .	209
5.2.2	Analysis of Sedge and Reed Warbler Data . . . . .	211
5.2.3	Assessing Adult Site-Fidelity . . . . .	218
5.2.4	True Seasonal Productivity . . . . .	223
5.3	Concluding Remarks . . . . .	226
5.4	Chapter Summary . . . . .	228
<b>6</b>	<b>Discussion</b>	<b>229</b>
	<b>Appendix A - Notation</b>	<b>243</b>
	<b>References</b>	<b>249</b>

# List of Figures

1.1	Images of CES mist-netting . . . . .	10
1.2	Photographs of a Sedge Warbler and a Reed Warbler . . . . .	15
2.1	Estimated trends in adult and juvenile abundances from Sedge and Reed Warbler complete coverage data . . . . .	32
2.2	Annual trends in standardised $\ln(n_{gt}^a)$ and $\ln(n_{gt}^j)$ counts plotted for separately each site . . . . .	35
2.3	Estimated indices of productivity for Sedge and Reed Warbler . . . . .	39
2.4	Adult survival estimates under the alternative models . . . . .	60
2.5	Expected and observed values of $\varepsilon$ plotted against $c$ . . . . .	64
2.6	Expected values of $\varepsilon$ plotted against $c$ allowing capture heterogeneity between three groups . . . . .	65
2.7	Estimates of adult survival from Sedge and Reed Warbler complete coverage data . . . . .	70
2.8	Expected and observed values of $\varepsilon$ plotted against $c$ from Sedge and Reed Warbler complete coverage data . . . . .	73
3.1	Survival estimates for Sedge and Reed Warblers from ring-recovery data . . . . .	90
3.2	Survival estimates for Sedge and Reed Warblers from the joint analysis of CES and NRR data . . . . .	92
3.3	Directed Acyclic Graph corresponding to the fully integrated model . . . . .	93
3.4	Estimates of the Sedge Warbler demographic parameters from the integrated model which uses ring-recovery data only in the estimation of adult survival . . . . .	95
3.5	Integrated analysis with a large, simulated, ring-recovery data set . . . . .	97

3.6	Estimates of the Sedge Warbler demographic parameters from the integrated model which uses ring-recovery data and CES data in the estimation of adult survival . . . . .	99
3.7	Estimates of the Sedge Warbler demographic parameters from the fully integrated model in which a split data analysis is used to deal with the lack of independence between $L_a$ and $L_{live}$ . . .	101
3.8	Comparison of two further split data integrated analyses of the Sedge Warbler data . . . . .	103
3.9	Estimates of $k\phi_{j,t}$ from the integrated analysis with, and without, ring-recovery data . . . . .	104
3.10	Estimates of the Reed Warbler demographic parameters from the integrated model which uses ring-recovery data and CES data in the estimation of adult survival . . . . .	105
3.11	Estimates of the Reed Warbler demographic parameters from the fully integrated model in which a split data analysis is used to deal with the lack of independence between $L_a$ and $L_{live}$ . . . .	107
3.12	Graphical checks of goodness-of-fit of the integrated model . . .	115
4.1	Estimated trend in abundance from simulated Poisson data . . .	134
4.2	Posterior distributions of imputed censored counts . . . . .	135
4.3	Estimated trends in adult and juvenile abundances from Sedge Warbler data artificially censored . . . . .	137
4.4	Estimated trends in adult and juvenile abundances from Sedge and Reed Warbler data that include years with incomplete coverage	139
4.5	Estimated trend in abundance when presence of missed within-year visits is ignored . . . . .	142
4.6	Estimated trend in abundance when CES count data are modelled at the visit level . . . . .	146
4.7	Estimated productivity trend from Sedge Warbler CES data artificially censored . . . . .	149
4.8	Estimated indices of productivity from Sedge and Reed Warbler CES data that include years with incomplete coverage . . . . .	151
4.9	Estimated trend in productivity from Sedge Warbler data artificially censored . . . . .	154
4.10	Trend in productivity calculated using the BTO's classical approach, and our novel Bayesian approach, from Sedge and Reed Warbler data with incomplete coverage . . . . .	155

---

4.11	Hypothetical pattern in the availability for capture over the 12 visit CES mist-netting season . . . . .	158
4.12	Annual within-year catch sizes from Sedge and Reed Warbler CES data . . . . .	159
4.13	Estimated within-year recapture probabilities . . . . .	161
4.14	The within-year recapture probabilities used in the simulation of a “simple” live-recapture data set . . . . .	166
4.15	Annual survival probabilities estimated from artificially simulated “simple” data . . . . .	167
4.16	Recapture regression parameters estimated from artificially simulated “simple” data . . . . .	168
4.17	Estimated survival probabilities from artificially simulated “realistic” data . . . . .	171
4.18	Estimated trend in the within-year recapture probabilities from the simulated “realistic” data . . . . .	172
4.19	Comparing the adult survival estimates under the different models which account for, or fail to account for, missed visits and/or transient birds . . . . .	182
4.20	Estimates of adult survival from CES “best” sites data with incomplete coverage . . . . .	186
4.21	Estimates of the demographic parameters from the Bayesian models which exclude or include years with incomplete coverage . . . . .	193
5.1	Directed Acyclic Graph corresponding to the fully integrated model that accommodates years with incomplete coverage . . . . .	199
5.2	Estimates of the Sedge Warbler demographic parameters from the integrated model that accommodates years with incomplete coverage . . . . .	201
5.3	Estimates of the Reed Warbler demographic parameters from the integrated model that accommodates years with incomplete coverage . . . . .	202
5.4	Comparing the estimates of the demographic parameters for Sedge Warblers when years with incomplete coverage are excluded, or included, in the fully integrated model . . . . .	205
5.5	Comparing the estimates of the demographic parameters for Reed Warblers when years with incomplete coverage are excluded, or included, in the fully integrated model . . . . .	206

5.6	Directed Acyclic Graph, for complete coverage CES data, corresponding to the fully integrated model that allows for permanent emigration . . . . .	211
5.7	Estimates of the Sedge Warbler demographic parameters from the integrated model that allows for permanent emigration . . .	213
5.8	Comparing the estimates of true adult survival if, and if not, permanent emigration is allowed . . . . .	214

# List of Tables

1.1	Location and habitat type of CES sites (1987-2005) . . . . .	11
1.2	Summary of CES data on Sedge and Reed Warbler (1987-2005)	14
1.3	Sedge and Reed Warbler CES recaptures (1987-2005) . . . . .	14
1.4	Breeding statistics for Sedge and Reed Warbler . . . . .	17
2.1	Sedge and Reed Warbler complete coverage data sets . . . . .	29
2.2	Correlations between Sedge and Reed Warbler indices of abundance and productivity . . . . .	40
2.3	Percentage of adult Sedge and Reed Warblers ringed at each visit	56
2.4	Adult Sedge Warbler “30 best” sites CES live-recapture data . .	59
2.5	Proportion of the adult Sedge Warbler “30 best” sites data utilised under the alternative survival models . . . . .	61
2.6	Estimates of $\tau$ , under the alternative models, from the adult Sedge Warbler “30 best” sites data . . . . .	62
2.7	Percentage of single capture and known resident adult Sedge Warblers caught at each visit . . . . .	66
2.8	Attributes of the alternative adult survival models . . . . .	67
2.9	Published estimates of Sedge and Reed Warbler adult survival rates . . . . .	71
2.10	Posterior means and 95% credible intervals for $\tau$ from the Sedge and Reed Warbler complete coverage data . . . . .	72
2.11	Assessing time and site dependence of $c$ in Sedge and Reed Warbler data . . . . .	74
2.12	Assessing time and site dependence of $\tau$ in Sedge and Reed Warbler data . . . . .	75

3.1	Ring-recovery data for Sedge and Reed Warblers . . . . .	89
3.2	Correlations between changes in adult abundance and the demographic parameters . . . . .	109
3.3	Correlations between the demographic parameters . . . . .	110
4.1	Sedge and Reed Warbler data including years with incomplete coverage . . . . .	122
4.2	Percentage of times individual within-year visits are missed . . .	122
4.3	Observed and expected percentages of missed within-year visits	122
4.4	Level of censoring in simulated Poisson data . . . . .	133
4.5	Level of censoring imposed upon Sedge Warbler data . . . . .	136
4.6	Sedge and Reed Warbler yearly-site counts corrected using the Peach et al. (1998) approach . . . . .	138
4.7	Correlations between the corrected censored counts under the classical Peach et al. (1998) approach and the Bayesian approach	140
4.8	Comparing the corrected counts, under the Peach et al. (1998) and Bayesian approaches, to their observed lower bounds . . . .	140
4.9	Annual numbers of missed within-year visits and the proportions of censored counts in Sedge and Reed Warbler data . . . . .	143
4.10	Precision of parameter estimates . . . . .	143
4.11	Estimates of annual recapture probability from the simulated “simple” data set . . . . .	168
4.12	Estimates of recapture regression parameters from the simulated “realistic” data set . . . . .	172
4.13	Summary of the parameter estimates under the different adult survival models using artificially simulated data . . . . .	179
4.14	Comparison of the parameter estimates under the different adult survival models for CES data . . . . .	183
4.15	Comparing Sedge and Reed Warbler “best” sites data sets which either exclude or include years with incomplete coverage . . . .	192
5.1	Estimates of $\tau$ and $\xi$ from the integrated model which include years with incomplete coverage for Sedge and Reed Warblers . .	203

5.2	Correlations between changes in adult abundance and the demographic parameters when data with incomplete coverage are included in the integrated analysis . . . . .	208
5.3	Correlations between the demographic parameters when data with incomplete coverage are included in the integrated analysis . . . . .	208
5.4	Posterior estimates of $\gamma$ , the probability of not permanently emigrating . . . . .	216
5.5	Change in demographic parameters when the fully integrated population model allows for permanent emigration . . . . .	217
5.6	Assessing the importance of allowing for permanent emigration .	223
6.1	Exploratory environmental covariate analysis . . . . .	234



# Chapter 1

## Introduction

### 1.1 Monitoring Bird Populations: Why, How and Who

Monitoring of wildlife populations is essential if they, and the habitats on which they depend, are to be managed and conserved effectively (Atkinson et al., 2006). The United Kingdom (UK) government is committed to a range of international agreements concerning the conservation of biodiversity (see UK Biodiversity Partnership Standing Committee, 2006). These include the Convention on Biological Diversity, an international treaty to sustain the rich diversity of life on earth, and the 2010 Biodiversity Target, a commitment to achieving a “significant reduction” in the rate of biodiversity loss by 2010. The monitoring and protection of wild bird populations constitutes an important part of these obligations. In addition, the UK is bound by the European Union (EU) Wild Birds Directive which provides a comprehensive framework for the conservation and management of wild birds in Europe.

Human-driven environmental factors such as urbanisation, global warming, agricultural intensification, deforestation, and the over-exploitation of natural resources all pose a threat to the well-being of wildlife. For example, widespread agricultural intensification has been associated with the decline of many granivorous farmland birds in the UK (Evans and Smith, 1994; Fuller et al., 1995; Peach et al., 1999; Chamberlain et al., 2000; Donald et al., 2001). The effect such environmental change has on wildlife needs to be carefully monitored and the results made readily available to politicians, their advisors, and the general

public to ensure appropriate and timely conservation measures are undertaken when required (Van Strien et al., 2001).

For assessing the state of wildlife populations in general, monitoring wild birds is particularly useful as they act as indicators to the health of the wider environment (Eaton et al., 2008; King et al., 2008b; Saracco et al., 2008; RSPB Website, 2009). Birds' usefulness as indicators arises because they occupy a wide range of habitats, tend to be at, or near, the top of the food chain (Furness and Greenwood, 1993), and are susceptible to environmental change (e.g. Peach et al., 1991; Julliard et al., 2004; Robinson et al., 2004). Their importance as indicators is highlighted by the use of bird population "healthiness" measures during policy making processes by government, conservation organisations etc. For example, the Farmland Bird Indicator, based on the abundance trends of 19 widespread bird species associated with agricultural land, is used by the UK government to assess the success of its policies for conserving wildlife in the countryside (Eaton et al., 2008). Furthermore, the Wild Bird Index, developed specifically to give a broad summary of the general health of wildlife in the British countryside, is one of the 15 headline indicators used by the UK government in its "Quality of Life" measure (Gregory et al., 2004).

Monitoring of bird populations is not limited to surveillance (Greenwood, 2003). For the effective management and conservation of wild birds it is essential to monitor populations through time and from the resulting data produce accurate and precise measurements of changes in abundance (Fewster et al., 2000; Brooks et al., 2008). Furthermore, it is necessary to distinguish important population changes that warrant conservation concern, perhaps resulting from human actions, from the "noise" created by short-term fluctuations in the natural environment (Baillie, 1990; Underhill and Prÿs-Jones, 1994; Siriwardena et al., 1998a; Greenwood, 2003). For instance, in a study of 28 common songbirds Peach et al. (1998) found that many of the annual fluctuations in abundance could be attributed to individual weather events (i.e. unfavourable weather conditions in either the UK, or for migratory species, in their African wintering quarters). However of greater conservation concern were the long-term changes in abundance for several species. In order to make this distinction an understanding of the normal patterns of population variability is required (Baillie, 1990).

Bird monitoring schemes must be designed appropriately so that species for which conservation action is needed are identified (Atkinson et al., 2006). However, in order to adequately protect these species monitoring schemes which

collect data on the underlying demographic rates (productivity, survival, movement), in addition to abundance, are particularly valuable since it is the demographic rates which are directly affected by environmental factors (Freeman et al., 2007b; Saracco et al., 2008). Typically abundances (or indices of abundance) are derived from annual count data (Peach et al., 1998; Fewster et al., 2000; Atkinson et al., 2006), estimates of productivity from records of breeding success (Bibby, 1978; Crick et al., 2003; Freeman et al., 2007b), and estimates of survival from ring-recovery data (Freeman and Morgan, 1992; Brooks et al., 2000b; Mazzetta, 2009) or live-recapture data (Peach et al., 1990; Hines et al., 2003; Clavel et al., 2008).

Monitoring the status of wild bird populations at a national (or even higher) level is becoming increasingly common. For example, the North American Breeding Bird Survey (Sauer et al., 1997) run by the Patuxent Wildlife Research Centre and the Pan European Common Bird Monitoring Scheme (Van Strien et al., 2001) established by the European Bird Census Council. Information from such monitoring schemes forms the knowledge base that underpins management and conservation action.

In the UK the British Trust for Ornithology (BTO) is as an independent, scientific research trust that investigates the populations, movements, and ecology of wild birds in Britain (BTO Website, 2009). Founded in 1933, the BTO gathers long-term data on bird abundance and demography through several monitoring schemes (see Baillie et al., 2007). Unlike the Royal Society for the Protection of Birds (RSPB), the BTO is not a campaigning organisation but an impartial advisor on the state of the UK's wild bird populations to government and other interested parties. Research by the BTO not only enables governmental agencies to set priorities and monitor the effectiveness of their conservation strategies, but also informs campaigners, enabling the making of well-researched and detailed arguments for policy change (BTO Website, 2009).

Birds are popular with the public; the RSPB has over one million members (RSPB Website, 2009). Furthermore, bird watching is a major recreational activity in the UK, with the BTO's Garden BirdWatch project attracting some 16500 participants (BTO Website, 2009). Volunteer-based monitoring schemes thus can provide a practical and cost-effective means of collecting demographic information on wild bird populations. The BTO runs a number of national monitoring schemes through its volunteer network. In this thesis we focus on one such scheme, the BTO's Constant Effort Sites (CES) scheme, which provides

possibly the best source of long-term data available to monitor populations of Britain's common songbirds in scrub and wetland habitats. We elaborate further on this in Section 1.3.

### 1.2 Integrated Population Modelling

Monitoring of wildlife populations often involves the collection of several different types of demographic data; data used to estimate abundance, productivity and survival rates (Besbeas and Freeman, 2006). Despite the biological relationship between these demographic variables, their temporal trends are often estimated separately. However, of increasing interest is the relationship between abundance and the underlying demographic rates that drive it, an understanding of which facilitates effective conservation (Freeman and Crick, 2003; Eaton et al., 2008). This relationship is usually expressed mathematically in the form of a matrix-based population model (Caswell, 2001), which serves to relate the abundance in a given year as a function of the abundance in the previous year(s) and the intermediate survival and reproductive rates. The state-space modelling framework (Buckland et al., 2004b) has proved very useful in this regard (e.g. Besbeas et al., 2002).

To investigate the causes of population decline for several species of wild songbirds, productivity and survival estimates have been incorporated into the models for abundance to see whether they explain the observed annual variations in population size (e.g. Peach et al., 1999; Siriwardena et al., 2001; Freeman and Crick, 2003). However these approaches are piecemeal, and do not use the data to its full potential (Besbeas et al., 2002). The simultaneous estimation of abundance and the underlying demographic parameters, integrating all available data within a single framework to produce parameter estimates that are consistent with both the totality of the data, and according to the underlying population dynamics, with one another, has been the focus of recent research (e.g. Besbeas et al., 2002; Brooks et al., 2004; King et al., 2008b; Besbeas et al., 2009; Borysiewicz et al., 2009; Reynolds et al., 2009).

An analysis which combines several sources of data is typically referred to as an "Integrated Population Model". Previous integrated analyses of wildlife populations have combined various different types of abundance data (e.g. census or survey counts) with information on survival arising from ring-recoveries (Bes-

beas et al., 2002; Brooks et al., 2004), live-recaptures (Schaub et al., 2007; Véran and Lebreton, 2008), multi-site live-recapture-recoveries (Borysiewicz, 2008), or independent data from both ring-recoveries and live-recaptures (Reynolds et al., 2009). Independent productivity data have also been incorporated (Freeman et al., 2007b; Gauthier et al., 2007; Reynolds et al., 2009). These integrated analyses have been conducted using both classical (Besbeas et al., 2002; Besbeas and Freeman, 2006; Véran and Lebreton, 2008; Borysiewicz et al., 2009) and Bayesian (Brooks et al., 2004; Schaub et al., 2007; King et al., 2008b; Reynolds et al., 2009) approaches.

An integrated approach to population modelling affords several important advantages over the separate analyses of the individual data sets. By combining information from various studies in a coherent manner, parameter estimates are more robust, fully reflecting all the information available (Brooks et al., 2004). In addition, the sharing of information between the component analyses can lead to improved precision in estimation (Besbeas et al., 2002; Borysiewicz et al., 2009). This may be particularly important when only sparse data on a parameter of interest are available (Schaub et al., 2007; Cave et al., 2009b). An integrated analysis may also be capable of estimating parameters not obtainable, or separately identifiable, when using data from only a single source (Besbeas et al., 2005; Cave et al., 2009b; Reynolds et al., 2009). Furthermore, integrated approaches to population monitoring have proved useful in identifying the likely demographic mechanism driving observed changes in population size (Besbeas et al., 2002; Freeman et al., 2007b).

The use of integrated population modelling with an aim to monitoring UK's wild bird populations is not new. Indeed the importance of an integrated approach is highlighted by the establishment of the BTO's Integrated Population Monitoring Programme, a research programme focused on developing integrated population models for data from the BTO's various long-term bird monitoring schemes (BTO Website, 2009). Previous research has combined data from several of the BTO's monitoring schemes (e.g. Peach et al., 1999; Besbeas et al., 2002; Brooks et al., 2004; Besbeas and Freeman, 2006; Freeman et al., 2007b; Baillie et al., 2009a). However, none of these previous approaches have utilised data collected under the CES scheme. The CES scheme is particularly amenable to an integrated approach to population modelling as this single survey provides a cheap, practical, and high quality source of information on several of the key demographic parameters (adult abundance, juvenile abundance, productivity and adult survival) for certain common songbird species (see Section 1.3). De-

veloping an integrated model appropriate for CES data is timely, and of benefit to those governmental agencies responsible for the monitoring and conservation of UK's common songbird species.

### 1.3 Constant Effort Sites Scheme

#### 1.3.1 BTO Bird Monitoring Schemes

The effective monitoring of songbird populations requires the measurement of three main demographic variables; abundance, productivity and survival (Baillie et al., 1986). The BTO's CES scheme, an annual programme of standardised bird-ringing across a large number of sites, was developed in recognition that previous BTO monitoring schemes were each only able to provide part of the required information, and were often inadequate for monitoring some songbird populations (Baillie et al., 1986; Peach et al., 1998).

The Common Bird Census (CBC), which ran from 1962 to 2000, mapped the breeding territories of common farmland and woodland birds (Fuller et al., 1985; Baillie et al., 2007), and provided a proven index of adult population size for these birds (Baillie et al., 1986). Weaknesses as a monitoring scheme were largely related to the time-consuming nature of both fieldwork and analysis (Greenwood, 2003; Baillie et al., 2007). With relatively few volunteers participating in the scheme, the CBC concentrated on farmland and woodland habitats, thus bird populations in urban areas and the uplands were poorly represented (Fuller et al., 1985; Baillie et al., 2007). Moreover, as the sampling sites were chosen by the volunteers there may have been some bias towards bird-rich habitats (Baillie et al., 2007).

Censusing, by mapping territories, can be difficult as birds may nest outside the habitat under study, change territories, or be polygamous (Simms, 1985). The Breeding Bird Survey (BBS), a line-transect based survey, which began in 1994, superseded the CBC as the major scheme for monitoring the abundance of UK's common and widespread bird species (Baillie et al., 2007). As it is less labour-intensive than the CBC, more sampling sites are included enabling more precise monitoring (Greenwood, 2003). Furthermore, on account of its rigorous, stratified random sampling design, and its simplicity in the field, the BBS produces data that better cover the previously under-represented urban

and upland bird populations (Baillie et al., 2007).

Data from the CBC and BBS schemes are commonly used to derive annual indices of adult (breeding) bird abundance (Peach et al., 1999; Fewster et al., 2000; Freeman and Crick, 2003; Robinson et al., 2004), and are now regularly analysed jointly (Freeman et al., 2007a). However, bird populations in wetland and scrub habitats have not been adequately monitored by the CBC, and the BBS does not provide sufficient information on species confined to habitats difficult to survey, for example reedbed birds such as the Reed Warbler (Peach et al., 1998). Furthermore, neither scheme provides direct information on the underlying demographic rates (survival and productivity) that drive changes in population size. Studies that aim to relate changes in abundance detected in CBC or BBS data to changes in the underlying demographic rates require separate sources of information on these variables (e.g. Siriwardena et al., 2001; Besbeas et al., 2002; Freeman and Crick, 2003; Brooks et al., 2004; King et al., 2008b), raising the issue of whether these separate data sets relate to the same population.

The Nest Record Scheme (NRS), which began in 1939, collates standardised information on the components of productivity, for example clutch size, brood size, egg failure rates, and fledging success rates (Crick and Baillie, 1996; Crick et al., 2003; Baillie et al., 2007). NRS data enables productivity per nest attempt (fledged young per breeding attempt) to be indexed but may not give very useful estimates of total seasonal productivity (fledged young per breeding pair per year) since repeat and replacement nesting attempts are poorly sampled (Bibby, 1978; Baillie et al., 1986; Crick et al., 2003).

The National Ring-Recovery (NRR) database gathers information on changes in survival, however the analysis is not habitat specific (Coiffait et al., 2008). Importantly, for most small songbird species, especially long-distance migrants, recoveries are too few to provide precise estimates of survival (Baillie et al., 1986; Peach et al., 1990; Cave et al., 2009b; Freeman, 2008).

The CES scheme provides direct information on abundance, as well as the key demographic parameters, productivity and survival, for many British common songbirds enabling their populations to be successfully monitored through time (Peach et al., 1996, 1998; Clark et al., 2007; Cave et al., 2009b; Freeman, 2008). Unlike the CBC and BBS, demographic data on bird populations in wetland, scrub and reedbed habitats are obtained (Baillie et al., 1986). In addition, CES monitoring is not limited to that of the adult population, with the scheme also

providing information to monitor juvenile abundance (population of recently fledged birds). Previously very little was known about this life-cycle stage despite its importance to population dynamics (Peach et al., 1998). CES data can be used to estimate total seasonal productivity (Cave et al., 2009b), as opposed to per nesting attempt as NRS data do, and are particularly valuable for estimating survival rates, although only rates for adult birds tend to be estimable due to juvenile dispersal (Peach, 1993; Cave et al., 2009b).

### 1.3.2 CES Field Protocol

The BTO's CES scheme is a long-running and extensive bird ringing programme conducted annually at a large number of sites across the UK. Data are collected by volunteers following the standard CES mist-netting and ringing protocol (see Baillie et al., 1986; Peach et al., 1996, 1998; Freeman et al., 2001; Miles et al., 2007) as described below. Potential CES sites are proposed by individual volunteer ringers interested in operating that particular site, and are generally accepted into the scheme provided they are located in suitable habitats not undergoing major successional changes. The CES scheme is funded by a partnership between the BTO, the Joint Nature Conservation Committee, and the National Parks and Wildlife Service (Ireland), as well as the volunteers themselves.

Each year, the breeding season (May-August) is divided into 12 consecutive visit periods of 10-11 days by the BTO. Volunteers aim to visit their particular site on one day selected (by the volunteers) within each of these 12 visit periods, and use a standardised mist-netting procedure, within and between years, to capture birds. The precise timing and duration of mist-netting, and the position, number, length and type of mist-nets are not specified by the BTO but decided by the individual volunteers and thus varies between sites according to the local conditions and with personal preference but, crucially, remains constant at every visit to a particular site both within and between years. A typical CES regime is to begin mist-netting around dawn and to continue until a fixed time during mid to late morning, when catch numbers have generally declined.

Mist-nets are a cheap and practical method for catching a large number of songbirds. These fine nets, designed to catch birds in flight, are usually made of a dark coloured nylon rendering them almost invisible against the background foliage. In structure, a mist-net consists of a series of horizontal "shelves" which

essentially form long “hammocks”. Any bird striking the net is likely to fall into the hammock and become sufficiently entangled that it cannot escape (Bub, 1991). To ensure the well-being of a captured bird it must not be left in the mist-net for too long, thus regularly throughout the CES visit captured birds are carefully disentangled from the mist-net (Figure 1.1a) and processed. All birds caught are fitted with individually identifying, light-weight, aluminium, leg rings (Figure 1.1b). Each ring also bears the BTO postal address, and in recent years the web address, so that any subsequent live resightings or dead recoveries can be recorded. For mist-netting to be conducted the weather must be suitable, that is no rain and little wind. Mist-nets are approximately 2.5m high, and are of variable length ranging from 15 -300m (average length = 115m). Figures 1.1c and 1.1d show mist-nets erected in typical CES reedbed and dry scrub sites respectively.

Mist-net captures include both juvenile birds (i.e. fledged young born within the current CES mist-netting season that are a few weeks old at most), and adult birds (i.e. those born in a previous CES mist-netting season). All CES volunteers are highly trained and experienced; they must hold a current BTO ringing permit, a legal requirement for anyone ringing birds in the UK, that has to be renewed annually (Clark et al., 2007). According to aging criteria, based on plumage differences such as Svensson (1992), CES volunteers are able to classify most birds as either adults or juveniles in the hand. For each bird caught, either for the first time or at another visit in the same or in a subsequent year, the volunteers record its unique ring ID, species, age (juvenile or adult), and the site, year and visit of capture. Note, recaptures of a bird already caught within the same visit are ignored.

Successive CES visits, within a year, are never less than 3 days apart, usually more, to minimise the problem of birds learning to avoid the mist-nets. For the same reason, mist-netting is not permitted at any CES site 3 days prior to a CES visit. Visit dates, and information pertaining to the next CES mist-netting season are issued to all volunteers prior to the start of the season by the BTO.

At the end of the CES mist-netting season (September onwards) volunteers submit their data to the BTO (see the BTO Website, 2009, for details). A print-out for each site, produced at the BTO, is later returned to the volunteers for verification. Additional information on habitat type is collected every 2 to 3 years from the volunteers about their site.



**Figure 1.1** Images of CES mist-netting<sup>†</sup>, a) disentangling a Bullfinch from a mist-net, b) fitting an individually identifying aluminium leg ring on a Greenfinch, c) mist-net on a reedbed habitat, d) mist-net on a dry scrub habitat.

<sup>†</sup>Photos provided by Mark Grantham (a, b) and Dawn Balmer (c, d), British Trust for Ornithology.

The scheme began in 1983, after a brief pilot period, with 46 sites, and quickly proved popular with volunteers. In 1990 the number of sites reached 102, and since then has fluctuated between 98 (in 1991) and 143 (in 2000). From these the scheme routinely monitors 25 common songbirds species (Grantham and Robinson, 2008). Note that as well as joining the scheme sites may leave, thus any two sites might neither start nor finish contributing data (i.e. are operational) in the same year, and within any year a different number and set

of sites may be operational. We consider data from 1987-2005, having omitted the first few years, when there were relatively few sites and the scheme was "settling down".

Although the sites cover the entire UK they are concentrated in the south east of England where the human population is highest, and where weather is more suitable for mist-netting. Sites are mostly wet scrub, dry scrub, woodland or reedbed; sites which are amenable to capturing an appreciable number of birds and in habitats at which succession can be controlled by appropriate management. Table 1.1 summarises the location (country) and habitat type of the 354 sites contributing data to the CES scheme between 1987 and 2005. The range and proportion of habitats and locations covered are comparable between years.

**Table 1.1** The location and habitat type of the 354 sites contributing information to the CES scheme between 1987 and 2005.

<b>Location</b>	England	Ireland	Scotland	Wales
	282	19	31	22
<b>Habitat</b>	Dry Scrub	Wet Scrub	Reedbed	Woodland
	134	106	74	40

### 1.3.3 CES Data

The CES database contains the capture histories for every individual bird encountered during CES mist-netting. For example, a typical segment of CES data for a bird would be:

```
A334831 SEDWA AD 301 1994 0 0 0 1 0 1 0 0 0 1 0 0
A334831 SEDWA AD 301 1995 0 0 1 0 0 0 0 0 0 0 0 0
A334831 SEDWA AD 301 1996 0 0 1 1 1 1 0 0 1 0 0 0
A334831 SEDWA AD 301 1998 0 0 0 1 0 1 0 0 0 0 0 0
```

From left to right the columns represent the ring ID, species, age, site of capture, year of capture and the 12-visit encounter history in which "1" denotes capture and "0" that the bird was unobserved. Therefore, the Sedge Warbler with ring ID A334831 was caught as an adult at site 301 in 1994, 1995, 1996 and 1998, and not seen thereafter. In 1994, for example, A334831 was caught at visits 4, 6 and 10, and not encountered at any other visit that year.

From this subsection of data it is impossible to ascertain whether mist-netting had been carried out at site 301 during any other year in which bird A334831 was not encountered. For example, CES mist-netting took place in 1997 at site 301, but as A334831 evaded capture that year there is no record for 1997. To construct the full capture history the above data must be augmented by information giving the years each site was operated. In the above example the full capture history for bird A334831, from 1994 to 1998, is therefore:

```
A334831 SEDWA AD 301 1994 0 0 0 1 0 1 0 0 0 1 0 0
A334831 SEDWA AD 301 1995 0 0 1 0 0 0 0 0 0 0 0 0
A334831 SEDWA AD 301 1996 0 0 1 1 1 1 0 0 1 0 0 0
A334831 SEDWA AD 301 1997 0 0 0 0 0 0 0 0 0 0 0 0
A334831 SEDWA AD 301 1998 0 0 0 1 0 1 0 0 0 0 0 0
```

The fourth row, corresponding to 1997, now indicates that A334831 was not caught during that year, although mist-netting was conducted and A334831 was clearly alive.

Under the assumption of constant effort any variability between years in the numbers of birds caught can be attributed to changes in the population level and to stochastic variation, and not to varying intensity of capture effort. Thus, for each species monitored under the CES scheme, the total annual catches of adults and juveniles can be used to estimate between-year and long-term changes in species' abundance and productivity (Peach et al., 1998; Miles et al., 2007; Robinson et al., 2007; Cave et al., 2009a,b). In addition, from annual recaptures of adult birds, adult survival is estimable (Peach et al., 1990; Cave et al., 2009b; Freeman, 2008). Such a monitoring scheme, based on ringing across a number of sites, is perhaps unique in providing this combination of demographic information.

Note that in general juvenile songbirds, in contrast to adults, are not site faithful. Due to the high dispersal of juvenile birds from their natal sites, a juvenile ringed under the CES scheme is unlikely to be recaptured again and hence CES data alone will not enable accurate estimation of juvenile survival. Furthermore, any estimates of juvenile survival will be confounded with natal dispersal.

### *Missing CES Visits*

In practice there is considerable deviation from the "constant effort" ideal underpinning the CES scheme with some years receiving fewer than the full 12

visits. These visits may be missed entirely due to extremely bad weather, volunteer ringers' personal circumstances etc, or else not achieve the minimum required catch effort. That is, effort spent mist-netting may be unavoidably reduced, for example, by bad weather or net damage necessitating the early cessation of mist-netting at the site. If the standard mist-netting effort, in terms of the total net length and the time the net is open, falls below 80% of the norm at the site (as indicated by the volunteer) the visit is considered to be missed. Any visit where catch effort is at least 80% is deemed to be made.

In the CES database a missed visit is denoted by "9". In the above example say mist-netting had not been carried out at site 301, in 1998, at visits 1 and 12. Further, say the additional information giving the years, and visits within those years, site 301 was operated indicated that CES mist-netting had only taken place in 1997 at several visits (1 to 5). The full capture history for bird A334831, from 1994 to 1998, would then be:

```
A334831 SEDWA AD 301 1994 0 0 0 1 0 1 0 0 0 1 0 0
A334831 SEDWA AD 301 1995 0 0 1 0 0 0 0 0 0 0 0 0
A334831 SEDWA AD 301 1996 0 0 1 1 1 1 0 0 1 0 0 0
A334831 SEDWA AD 301 1997 0 0 0 0 0 9 9 9 9 9 9 9
A334831 SEDWA AD 301 1998 9 0 0 1 0 1 0 0 0 0 0 9
```

The data now indicate that A334831 evaded capture at all visits made in 1997 (visits 1 to 5), and for those visits made in 1998 (visits 2 to 11), A334831 was caught only at visits 4 and 6.

The presence of missed visits in the data, and the resulting violation of the constant effort assumption, require particular attention when building models to monitor abundance, productivity and adult survival from CES data. As the standard practices for analysing CES data do not adequately address the problem of missed visits, this is one of the key areas of research considered in this thesis.

Note, no attempt is made to correct data for any reduction in catch effort above the 80% threshold (i.e. for those visits deemed made but in which catch effort was reduced by up to 20%), as this information is difficult to obtain. However, as most frequently it is the end of a visit period that is prematurely curtailed, the time in which numbers of birds captured have typically declined, such reduction in the catch effort is believed to incur only a small proportional cost to catch and re-sighting numbers (Miles et al., 2007).

## 1.4 Sedge and Reed Warblers

We consider in detail CES data for Sedge Warbler, *Acrocephalus schoenobaenus*, and Reed Warbler, *Acrocephalus scirpaceus*. These are wetland birds and two of the most frequently encountered species under the CES scheme. Other *Acrocephalus* species are very rare in the UK. For species like Sedge and Reed Warbler, which have specialised habitat requirements, the CES scheme provides the best information available with regards to their demography.

Tables 1.2 and 1.3 provide summaries of the CES data on Sedge and Reed Warbler. For both species a large number of yearly-site records exist, i.e. the total number of times annual mist-netting was conducted at CES sites between 1987 and 2005. The large number of juvenile and adult captures provides high quality data on abundance and productivity, whilst the repeated adult captures enable precise estimation of adult survival. Note, sparse juvenile recaptures, particularly for Sedge Warbler, prevents accurate estimation of juvenile survival.

**Table 1.2** Summary of CES data on Sedge and Reed Warbler (1987 to 2005) giving: the number of sites at which each species was caught, the total number of yearly-site records, the number (percentage) of missed within-year visits, the total adult and juvenile catch (excluding multiple captures within years), and the number of recaptures, at the same site but in a subsequent year, for birds ringed as adults and as juveniles.

	Sedge Warbler	Reed Warbler
Sites	283	241
Yearly-site records	1848	1564
Missed within-year visits	1775 (8.0%)	1469 (7.8%)
Adult catch	22085	34792
Juvenile catch	28605	33347
Adult recaptures	2415	5938
Juvenile recaptures †	687	1819

† Birds ringed as juveniles and later recaptured as adults.

**Table 1.3** Sedge and Reed Warbler CES adult and juvenile recaptures (1987 to 2005): the number of years an individual bird is observed, the total number of birds ringed, and the total number (percentage) of birds caught in more than 1 year.

	<i>Years Caught</i>							ringed	recaptured
	1	2	3	4	5	6	≥7		
<b>Sedge Warbler</b>									
Adult	17750	1533	302	69	10	5	1	19670	1920 (9.8%)
Juvenile	28061	437	81	19	6		1	28605	544 (1.9%)
<b>Reed Warbler</b>									
Adult	24771	2889	722	258	110	37	17	28804	4083 (14.2%)
Juvenile	32036	994	210	55	33	14	6	33348	1312 (3.9%)

A Sedge Warbler and a Reed Warbler are pictured in Figure 1.2a and Figure 1.2b respectively. The white stripe above the eye makes Sedge Warblers readily distinguishable from Reed Warblers. Both are comparable in size, weighing about 13g, with a wingspan of approximately 19cm, and length approximately 13cm (Robinson, 2005). Both are short-lived, with a typical lifespan of up to 5 years (Hume, 2003).



a



b

**Figure 1.2** a) Sedge Warbler *Acrocephalus schoenobaenus* †, b) Reed Warbler *Acrocephalus scirpaceus*. ‡

†Photo courtesy of Dawn Balmer, British Trust for Ornithology.

‡Photo courtesy of Kevin Carlson, British Trust for Ornithology.

Studies of Sedge Warblers are particularly interesting as they are known to be susceptible to environmental change; Peach et al. (1991) identified a relationship between the survival of adult Sedge Warbler and the rainfall in their west African wintering quarters. Direct human related impacts on Sedge and Reed Warbler populations have also been established, for example a study by Graveland (1999) linked the harvesting of reed *Phragmites australis* (for thatching) to a lowered seasonal productivity brought on by an increased nest predation risk and delayed nesting times in cut reedbeds.

Sedge and Reed Warblers are long-distance trans-Saharan migrants, breeding in Europe and over-wintering in west Africa (Bibby and Green, 1981; Simms, 1985; Peach, 2002; Redfern and Alker, 2002). Although they are both small, insectivorous, wetland dwelling passerines (i.e. perching birds) that inhabit common summer and winter quarters, they occupy separate ecological niches and undertake quite different breeding (Bibby, 1978) and migration (Bibby and Green, 1981) strategies. These dissimilarities render them more, or less, vulnerable to different environmental changes, and make comparison between these species especially interesting.

Sedge and Reed Warbler breed after their first winter, aged around one, and every year thereafter (Cramp, 1992; Robinson, 2005). Adult birds are markedly site-faithful, returning to the same breeding grounds annually, whereas juveniles tend to disperse from their natal areas (Catchpole, 1974; Peach, 2002).

Sedge Warblers breed extensively throughout the UK during the summer (Simms, 1985; Peach, 2002). The first birds begin to arrive at their breeding sites in early April, with arrivals reaching a peak in mid May (Simms, 1985; Peach, 2002). Waterside habitats, such as lowland marshes, waterways, and flooded gravel pits, are preferred although Sedge Warblers can be found breeding in drier areas (Simms, 1985; Peach, 2002). Reed Warblers are not as widely distributed as Sedge Warblers, but in recent decades their breeding range has spread north into south-west Scotland and west into eastern Ireland (Simms, 1985; Redfern and Alker, 2002). As their spring migration is later and more protracted than Sedge Warbler, with arrivals first appearing in mid April, peaking in May, and continuing through to mid June (Simms, 1985; Redfern and Alker, 2002), breeding begins slightly later (Catchpole, 1974; Bibby, 1978; Robinson, 2005). Reed Warblers, being more strictly reedbed birds than Sedge Warblers, have a strong preference for mature beds of common *Phragmites* reed but can be found breeding in nearby vegetation (Catchpole, 1974; Simms, 1985; Couzens, 2006).

Within a breeding season, Sedge and Reed Warbler are considered to be predominantly monogamous (Simms, 1985). Upon arrival at the breeding grounds male birds secure a territory and attract a mate. Breeding statistics for both species are summarised in Table 1.4 (from Catchpole, 1974; Bibby, 1978; Simms, 1985; Cramp, 1992; Alker and Redfern, 1996; Robinson, 2005). Sedge Warbler clutches are typically 5 eggs, larger than those of Reed Warblers which are typically 4 eggs, although the mean clutch size for both species decreases uniformly as the breeding season progresses (Bibby, 1978). Approximately 17-32% of Reed Warblers will attempt a second brood after an early successful fledging of the first, or if the first nest completely failed (Cramp, 1992). It is believed that a much smaller proportion of Sedge Warblers attempt a second brood (Bibby, 1978; Simms, 1985), although a 3-year study by Alker and Redfern (1996) found that the majority of Sedge Warblers at a small UK scrubland site were double brooded.

**Table 1.4** Typical clutch size, number of broods, egg incubation period, fledging period and timing of first clutch for Sedge and Reed Warbler. Statistics were obtained from Catchpole (1974); Bibby (1978); Simms (1985); Cramp (1992); Alker and Redfern (1996); and Robinson (2005).

	Sedge Warbler	Reed Warbler
Clutch size	5	4
Broods	1	1-2
Incubation	13-15 days	9-12 days
Fledging	13-14 days	10-12 days
First clutch	late May	early June

To varying levels, depending on the species and on the individual breeding pair, both the male and the female will incubate the eggs and feed the chicks (Simms, 1985). Sedge and Reed Warblers feed on insects and their larvae, and adult birds will regularly fly out of their breeding territories to collect food for their young (Simms, 1985). Whereas Reed Warblers are able to catch airborne insects (e.g. Diptera), Sedge Warblers are restricted to less mobile prey, mainly feeding by picking slow moving insects off a surface (Bibby and Green, 1981; Simms, 1985). Morning and late evening are the busiest feeding times, when insects are slower (Simms, 1985).

After the Sedge and Reed Warbler young leave the nest, for two to three weeks, they are still dependent on their parents for food (Bibby, 1978; Alker and Redfern, 1996). During this time the family group may stay in the vicinity of the breeding territory or move into a different area (Catchpole, 1974).

In late July, Sedge and Reed Warblers begin leaving their UK breeding sites for pre-migratory feeding grounds in southern England and France (Peach, 2002; Redfern and Alker, 2002). Departure times can extend into late September, particularly if a second brood is attempted (Simms, 1985). From late August, extending through to October, Sedge and Reed Warblers start their autumn migration to west Africa. Although both species are morphologically similar, very closely related, and heading to a similar location, they have quite different migration strategies (Bibby and Green, 1981; Peach, 2002; Redfern and Alker, 2002; Couzens, 2006). Sedge Warblers, as far as it is known, take the entire journey to west Africa in a single stretch, flying continuously for approximately 72 hours (Couzens, 2006). Reed Warblers' migration is much slower, proceeding in small legs punctuated by refuelling stops in Iberia (Bibby and Green, 1981; Simms, 1985; Bensch and Nielsen, 1999). In addition, Reed Warblers also leave the UK later (Bibby and Green, 1981). For both species there is a tendency for adults to leave earlier than juveniles, albeit more marked for Reed Warblers (Bibby and Green, 1981). Note, most of the knowledge regarding migration also comes from ringing data.

## 1.5 Bayesian Methodology

Within this thesis Bayesian techniques are adopted for obtaining parameter estimates. In situations where classical techniques are prohibitively complex to apply (Jamieson and Brooks, 2004; Reynolds et al., 2009), or require potentially inappropriate distributional and/or asymptotic assumptions to be satisfied (King et al., 2009), the Bayesian paradigm provides a very flexible approach to data analysis, permitting more general modelling frameworks (Brooks et al., 2004; Schaub et al., 2007). For example, in the integrated analysis of demographic data, based on state-space modelling, the classical Kalman filter approach for parameter estimation relies on normality and linearity assumptions (see Besbeas et al., 2002, 2005; Borysiewicz et al., 2009). The alternative Bayesian approach enables these assumptions to be relaxed (Brooks et al., 2004; King et al., 2008b). As Bayesian ideas and methods are now well known and commonly applied, we present only a brief description here, focusing on the techniques employed in this thesis. More detailed discussion can be found in Gilks et al. (1996); Brooks (1998); Carlin and Louis (2000); Gelman et al. (2004); McCarthy (2007); King et al. (2009) for example.

### 1.5.1 The Bayesian Approach

Let  $\boldsymbol{\theta}$  denote the model parameters, and  $\boldsymbol{x}$  the observed data. Central to the Bayesian paradigm is that the model parameters are described as having distributions: a *prior* distribution before data are observed, denoted by  $p(\boldsymbol{\theta})$ , and a *posterior* distribution after data are observed, denoted by  $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ . The posterior distribution, which represents our updated beliefs regarding the model parameters upon observing data, forms the basis of Bayesian inference.

Formally, the posterior distribution is obtained using Bayes' theorem,

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

which combines the prior information (in  $p(\boldsymbol{\theta})$ ) with the information provided by the data via  $f(\boldsymbol{x}|\boldsymbol{\theta})$ . Note,  $f(\boldsymbol{x}|\boldsymbol{\theta})$  is given by the probability density function (or in the case of discrete data, the probability mass function) associated with the observed data under the chosen model, and is commonly referred to as the “likelihood”.

#### *Choice of Prior Distribution*

The prior distribution,  $p(\boldsymbol{\theta})$ , describes our initial beliefs regarding the model parameters before the data are observed, and is specified entirely independently of the data. In the absence of any *a priori* knowledge a *vague* (uninformative) prior, so that the inference is not affected by any information external to the data, is required. Commonly this is achieved by choosing a prior distribution with a large variance. Note, when a parameter reparametrisation is of interest care must be taken when specifying a vague prior as the prior induced on the transformed parameter may no longer be uninformative (for example, see King and Brooks, 2008). If *a priori* information is available, for example expert opinion, this is summarised in an *informative* prior. Forming an informative prior that is consistent with the *a priori* beliefs can be difficult, often entailing choosing a suitable probability distribution and then attempting to find parameters for that distribution to accurately represent the available *a priori* information (O’ Hagan, 1998).

When the data are sufficiently informative the prior distribution should have little influence on the resulting inference, however its effect can be assessed using a *prior sensitivity* analysis. A simple sensitivity analysis, commonly used,

is to increase or decrease the prior variance to see what consequences this has on the resulting posterior distribution. Different probability distributions can also be tried. When an uninformative prior is required, or if there is some controversy regarding an informative prior, a prior sensitivity analysis is potentially very interesting. Note, the prior distribution should be considered part of the “model” and as such prior sensitivity itself is not an issue, however its presence may highlight problems such as overly restrictive prior assumptions or parameter redundancy (King et al., 2009).

### *Marginal Posterior Distributions*

When the model is multivariate the marginal posterior distributions for the individual parameters are frequently of interest. By integrating out over the rest of the parameter space, the marginal posterior distribution provides information on a single parameter. For example if  $\boldsymbol{\theta} = (\phi, \boldsymbol{\psi})$  the posterior marginal distribution of  $\phi$  is given by:

$$\pi(\phi|\mathbf{x}) = \int \pi(\phi, \boldsymbol{\psi}|\mathbf{x}) d\boldsymbol{\psi}.$$

Posterior marginal distributions are often used to summarise the full joint posterior distribution (e.g.  $\pi(\phi, \boldsymbol{\psi}|\mathbf{x})$ ); however the marginals do not retain information regarding the correlation structure between the individual parameters.

## 1.5.2 Bayesian Inference

The posterior distribution (or the corresponding posterior marginal distributions) provide a description of the parameters, however as they tend to be complex simply presenting them is typically of little interpretable use. Consequently, Bayesian inference is generally based upon posterior summary statistics, for example the posterior mean, median and standard deviation. Bayesian *credible intervals* are also commonly used to describe the posterior (marginal) distribution of a parameter. The interval  $(a, b)$  is a  $100 \times (1 - \alpha)\%$  credible interval for  $\phi$  if:

$$\int_a^b \pi(\phi|\mathbf{x}) d\phi = 1 - \alpha, \quad 0 \leq \alpha \leq 1.$$

In general this interval is not unique as many values of  $a$  and  $b$  could satisfy the above criteria. In this thesis we present the *symmetric credible interval*, a

unique solution where  $a$  and  $b$  correspond to the  $\alpha/2$  and  $1 - \alpha/2$  quantiles of  $\pi(\phi|\mathbf{x})$  respectively, i.e.

$$\int_{-\infty}^a \pi(\phi|\mathbf{x}) d\phi = \int_b^{\infty} \pi(\phi|\mathbf{x}) d\phi = \frac{\alpha}{2}.$$

To obtain the quantities required for inference, for example the marginal posterior distributions or the posterior mean, requires integration of the posterior distribution. However, explicit evaluation of such integrals is frequently extremely difficult, if not impossible, due to the complex nature typical of posterior distributions. *Markov chain Monte Carlo* (MCMC) provides a means of performing the integration implicitly. Here samples are drawn from the posterior distribution by constructing a *Markov chain* (a set of simulated values generated by sampling a new value dependent only upon the previous value) whose stationary distribution is the required posterior. After the chain has run long enough to reach the stationary distribution, these samples are then used to obtain sample estimates of the quantities of interest, a simulation technique known as *Monte Carlo integration*. For example, the sample mean is used to estimate the posterior mean. A potential downside of such an approach is that obtaining enough samples for reliable inference can be time consuming and computer intensive.

The *Metropolis-Hastings algorithm* provides a general framework for constructing a MCMC sampler (Chib and Greenberg, 1995). This algorithm is used throughout this thesis to obtain a sample from the posterior distribution from which estimates of the posterior quantities of interest (the posterior mean, median and standard deviation of the model parameters and their corresponding 95% symmetric credible intervals) are calculated.

### 1.5.3 The Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm provides a method for drawing samples from a given posterior distribution. The algorithm constructs a Markov chain whose stationary distribution is the required posterior. Values are drawn from a sensibly chosen proposal distribution and “corrected” so that, asymptotically, they behave as random observations from the posterior distribution. The algorithm itself is a form of rejection sampling that uses an acceptance/rejection rule to converge to the posterior. Once the chain has converged the simulated values can be treated as a sample from the posterior distribution and used to estimate

the marginal posterior distributions of the model parameters, and posterior summary statistics of interest.

To initialise the Markov chain, the Metropolis-Hastings algorithm begins by setting starting values for the model parameters, denoted by  $\boldsymbol{\theta}^1$ , chosen arbitrarily. At iteration  $h$  of the algorithm, where the current state of the chain is denoted by  $\boldsymbol{\theta}^h$ , the model parameters are updated using the following steps:

1. Sample a proposed value for  $\boldsymbol{\theta}$ , denoted by  $\boldsymbol{\theta}^*$ , from the proposal distribution  $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^h)$ .
2. Calculate the acceptance probability  $\alpha(\boldsymbol{\theta}^h, \boldsymbol{\theta}^*)$ , where

$$\alpha(\boldsymbol{\theta}^h, \boldsymbol{\theta}^*) = \min \left( 1, \frac{\pi(\boldsymbol{\theta}^*|\mathbf{x})q(\boldsymbol{\theta}^h|\boldsymbol{\theta}^*)}{\pi(\boldsymbol{\theta}^h|\mathbf{x})q(\boldsymbol{\theta}^*|\boldsymbol{\theta}^h)} \right).$$

3. With probability  $\alpha(\boldsymbol{\theta}^h, \boldsymbol{\theta}^*)$  accept  $\boldsymbol{\theta}^*$  and set  $\boldsymbol{\theta}^{h+1} = \boldsymbol{\theta}^*$ , else reject and set  $\boldsymbol{\theta}^{h+1} = \boldsymbol{\theta}^h$ .

In practice, to conduct this step, we draw a random variable  $u$  from the Uniform[0,1] distribution, and if  $u \leq \alpha(\boldsymbol{\theta}^h, \boldsymbol{\theta}^*)$  accept the move.

Importantly, the Metropolis-Hastings algorithm only requires the posterior distribution to be known up to proportionality, as the constants of proportionality cancel out in the expression for the acceptance probability,  $\alpha(\boldsymbol{\theta}^h, \boldsymbol{\theta}^*)$ .

### *Single Update Metropolis-Hastings Algorithm*

Not all model parameters need be updated simultaneously in the Metropolis-Hastings algorithm; rather parameters can be updated singly, one-at-a-time, using the *single update Metropolis-Hastings algorithm*. Here each iteration of the Metropolis-Hastings algorithm consists of an updating step for every individual parameter, which is analogous to the updating procedure described above. For example, say  $\boldsymbol{\theta}$  is  $d$ -dimensional. At iteration  $h$  suppose the current values are  $\boldsymbol{\theta}^h = \{\theta_1^h, \dots, \theta_d^h\}$ . To update the first parameter, denoted by  $\theta_1$ , the algorithm proceeds as follows:

1. Propose a value for  $\theta_1$  from the proposal distribution  $q_1(\theta_1^*|\theta_1^h)$ , where  $\theta_1^*$  and  $\theta_1^h$  denote the proposed value and the current value respectively.
2. Let  $\boldsymbol{\theta}_1^* = \{\theta_1^*, \theta_2^h, \dots, \theta_d^h\}$  and  $\boldsymbol{\theta}_1^h = \{\theta_1^h, \dots, \theta_d^h\}$ .
3. Calculate the acceptance probability  $\alpha_1(\theta_1^h, \theta_1^*)$ , where

$$\alpha_1(\theta_1^h, \theta_1^*) = \min \left( 1, \frac{\pi(\boldsymbol{\theta}_1^*|\mathbf{x})q_1(\theta_1^h|\theta_1^*)}{\pi(\boldsymbol{\theta}_1^h|\mathbf{x})q_1(\theta_1^*|\theta_1^h)} \right).$$

4. With probability  $\alpha_1(\theta_1^h, \theta_1^*)$  set  $\theta_1^{h+1} = \theta_1^*$ , else set  $\theta_1^{h+1} = \theta_1^h$ .

The remaining  $(d - 1)$  parameters are updated in analogous steps, noting that for the  $j$ th parameter the proposal distribution is  $q_j(\theta_j^* | \theta_j^h)$ , and the parameter vectors  $\theta_j^*$  and  $\theta_j^h$  contain the previously updated values where appropriate, i.e.  $\theta_j^* = \{\theta_1^{h+1}, \dots, \theta_{j-1}^{h+1}, \theta_j^*, \theta_{j+1}^h, \dots, \theta_d^h\}$  and  $\theta_j^h = \{\theta_1^{h+1}, \dots, \theta_{j-1}^{h+1}, \theta_j^h, \theta_{j+1}^h, \dots, \theta_d^h\}$ . Iteration  $h$  is completed when all  $d$  parameters have been updated.

Block updating, where sets of parameters are simultaneously updated, is also possible (Chib and Greenberg, 1995). This is often useful for highly correlated parameters.

### *The Proposal Distribution*

There are many choices for sensible proposal distributions. A common form, one that is adopted throughout this thesis, is to base the proposal distribution around the current value resulting in a *random walk Metropolis-Hastings algorithm*. For example, random walk proposal distributions for  $\theta_1^*$  include the Uniform $[\theta_1^h - \delta_1, \theta_1^h + \delta_1]$  distribution and the Normal $(\theta_1^h, \sigma_1^2)$  distribution. Note,  $\delta_1$  and  $\sigma_1^2$  are referred to as the proposal step length and proposal variance respectively.

As the Uniform and Normal proposal distributions above are symmetric the acceptance probability reduces to the ratio of the posterior distributions evaluated at the proposed and current values, i.e.

$$\alpha_1(\theta_1^h, \theta_1^*) = \min \left( 1, \frac{\pi(\theta_1^* | \mathbf{x})}{\pi(\theta_1^h | \mathbf{x})} \right).$$

Although the choice of proposal distribution is essentially arbitrary, if poorly chosen the Metropolis-Hastings algorithm will be inefficient. If small moves are proposed, the updates will generally be accepted, but it will take a long time to explore the posterior parameter space. Conversely, if large moves are proposed, the updates will generally be rejected, and the algorithm will be inefficient. So that the chain efficiently traverses the posterior parameter space, acceptance rates (the proportion of times the proposed move is accepted) between 0.2 and 0.4 are optimal (Gelman et al., 1996). Once the proposal distribution is chosen, by increasing or decreasing the proposal variance/step length, the proposal distribution can be *tuned* to achieve an acceptance rate within this desirable range. Throughout this thesis *a priori* proposal tuning is achieved by

running the algorithm for a “small” number of iterations, typically in the order of 10,000, calculating the resulting acceptance rates, and making appropriate adjustments to the proposal distributions. This is generally an iterative process.

### *The Gibbs Sampler*

The *Gibbs sampler* is a special case of the single update Metropolis-Hastings algorithm. Here the proposal distribution for a parameter,  $q$ , is its conditional posterior distribution given the current values of the other parameters, i.e.  $\theta_j^*$  is sampled from  $\pi(\theta_j | \theta_1^{h+1}, \dots, \theta_{j-1}^{h+1}, \theta_{j+1}^h, \dots, \theta_d^h, \mathbf{x})$ . The Gibbs sampler is highly efficient as the acceptance probability,  $\alpha(\theta_j^* | \theta_j^h)$ , is always one, but does require the ability to sample from the posterior conditional distributions. Ideally, these conditional distributions will be of a standard distributional form. This can be engineered by using a *conjugate* prior: a prior distribution where the resulting posterior distribution is of the same distributional form. For examples of conjugate priors see Gelman et al. (2004). Metropolis-Hastings updating steps can also be introduced into a Gibbs sampler, known as *Metropolis-Hastings within-Gibbs*. For parameters with standard conditional distributions Gibbs updates are used, else Metropolis-Hastings single updates are used (see King et al., 2009).

### *Number of Iterations*

The Metropolis-Hastings algorithm (or Gibbs sampler) must be run for a suitably large number of iterations to ensure that a) the chain has converged and b) enough samples are drawn from the posterior distribution (post-convergence) so that Monte Carlo (sampling) errors are small.

The initial samples drawn prior to convergence are discarded as *burn-in*. It is only the remaining samples that are used for inference. It is necessary to ensure that the length of the burn-in period is long enough so that the remaining samples can be assumed to arise from the posterior distribution of interest. Examining the trace-plots of the individual parameters, to check if the samples have “settled down” to values based around a constant mean, provides a simple means for detecting a lack of convergence. Another simple approach is to repeat the simulations, running the chain from multiple different overdispersed starting points. If essentially identical posterior estimates are obtained, this suggests convergence. The Brooks-Gelman-Rubin diagnostic, based upon the idea of

using an “Analysis of Variance” to assess whether or not each of the multiple chains has the same distribution, formalises this approach (Brooks and Gelman, 1998). Other, more formal, techniques exist (Cowles and Carlin, 1996), however it is important to note that convergence diagnostics can only provide evidence of a lack of convergence, no technique can prove that the Markov chain has converged to its stationary distribution. Standard convergence diagnostics are adopted throughout this thesis, although details are generally omitted. Further, long burn-in periods are used that are overly conservative.

After convergence, further iterations are needed to obtain samples for posterior inference. The more iterations, the more accurate posterior estimates will be. Accuracy of the posterior estimates can be assessed by the Monte Carlo error, the uncertainty arising from the finite number of samples (see Gilks et al., 1996).

## 1.6 Thesis Outline

The aim of this thesis is to develop Bayesian statistical methodology to monitor bird populations reliably using CES data. The research focuses on three main aspects, to i) extend the current methodology for estimating abundance, productivity and adult survival from CES data using models cast in a Bayesian framework, ii) address the problem of missed within-year visits in the modelling of CES data, the presence of which violates the assumption of constant effort, and iii) develop an integrated population model appropriate for CES data. Throughout this thesis the methodologies developed are applied to CES data from Sedge and Reed Warbler, and the results discussed in detail.

In Chapter 2 the thesis begins by developing Bayesian models for CES data from which independent estimates of the key demographic variables (adult abundance, juvenile abundance, productivity and adult survival) are obtained. We consider here only the subset of CES data from years where the full complement of 12 visits is made, for which the constant effort assumption is met. We produce a Bayesian integrated population model for this subset of CES data in Chapter 3, based on the independent models developed in Chapter 2 and augmented with additional information on adult and juvenile survival from National Ring-Recovery (NRR) data.

In Chapter 4 we address the problem of missing within-year visits, the presence of which violates the key assumption of constant effort. We extend the indepen-

dent Bayesian models of Chapter 2 so that data from years with missed visits can be included into the analyses without introducing bias into the estimates of the demographic variables. An integrated population model for such data is presented in Chapter 5. Furthermore, we modify the integrated population models, for both CES data excluding and including years with missed visits, to allow for a lack of breeding-site fidelity in adult birds.

The thesis concludes in Chapter 6 with a short discussion, with a particular focus on areas of potential research. A comprehensive list of the notation used during this thesis can be found in Appendix A.

Materials presented in this thesis have been published in two scientific papers, Cave et al. (2009a,b), both of which focus on CES data from Sedge Warblers. Cave et al. (2009a) describes the independent, Bayesian model for indexing abundance from CES data in which missed visits cause a violation of the constant effort assumption. This work forms the basis of thesis Section 4.2. Cave et al. (2009b) presents an integrated population model for the subset of CES data with no missed visits, with an adjustment for permanent emigration. This paper is a culmination of aspects of the research presented in Chapters 2, 3, and 5 of this thesis. Results pertaining to Reed Warblers, and the majority of the research on handling missed visits in the independent and integrated modelling of CES data, are as yet unpublished, but will be the focus of future scientific papers.

## Chapter 2

# Independent Statistical Models for CES Data with Complete Coverage

Data from the BTO's CES scheme are routinely used to index abundance and productivity, and to a much more limited extent estimate adult survival rates (Clark et al., 2007). Independent models have been developed for each of these demographic parameters and from which the BTO currently employs classical methodology for parameter estimation (see Peach et al., 1998; Robinson et al., 2007; Freeman, 2008).

In this chapter we cast the existing models into a Bayesian framework, the added flexibility of which will be advantageous when more complex models are developed in later chapters, for example the integrated population model of Chapter 3 and for the incorporation of missing visit data in Chapter 4. Estimates for the annual indices of adult abundance ( $A_t$ ), juvenile abundance ( $J_t$ ) and productivity ( $P_t$ ) are derived from models which adopt CES “count data”, the number of unique juvenile and adult birds caught per year. Annual adult survival rates ( $S_{a,t}$ ) are estimated from models which adopt CES “live-recapture data” that arise from the multiple encounters of individual adult birds across years.

We propose an alternative productivity model, that produces analogous indices of productivity to the model currently employed by the BTO, but will be shown to be more amenable to an integrated approach to population modelling. Further, we build upon the model of Pradel et al. (1997) from which unbiased

estimates of adult survival are obtainable from live-recapture data known to contain records from transient birds (i.e. non-resident birds captured once as they pass through the study site). We efficiently extend the approach developed by Pradel et al. (1997) to account for transients, making use of specific advantages in this respect provided by the within-year CES encounters.

Models are fitted to Sedge Warbler and Reed Warbler CES data, and the Bayesian and classical approaches compared. Further, adult survival estimates under the Pradel et al. (1997) model and the extended version are compared.

## 2.1 Data

As described in Section 1.3, CES protocol dictates that within each year a site is operated, it is visited on 12 separate occasions that span the British summer months (May-August). However, occasionally isolated individual visits within a site's period of operation are missed, some years receiving fewer than the full 12 visits. Initially we only consider data in years for which all 12 visits are made, i.e. "complete coverage data". For example, in the segment of data with missed visits presented in Section 1.3.3, records from site 301 in 1997 and 1998 would be excluded. When one or more visits are missed an incomplete capture record results, and the assumption of constant annual effort for that year is violated. This lack of constant effort requires special treatment, and methods for dealing with missed visits in the separate modelling of abundance, productivity and adult survival are discussed in Chapter 4. For both Sedge and Reed Warbler data, 56% of the times CES mist-netting was conducted, at a given site in a given year, the full set of 12 visits were made and complete coverage thus achieved.

The Sedge Warbler complete coverage data set contains 192 sites at which Sedge Warblers have been ringed. The Reed Warbler complete coverage data set contains 156 sites. However, many of these sites have been operated for only a few years, and/or caught only a few Sedge or Reed Warblers. Sites with such sparse data contribute little information to the analyses at the cost of extra site-specific parameters. Therefore we work with slightly reduced Sedge and Reed Warbler data sets. Included are only those sites with complete coverage for 4 years or more that caught appreciable numbers of Sedge/Reed Warblers (at least 10 adults, and at least 10 juveniles over the duration of their operation).

These reduced, “best” sites, Sedge and Reed Warbler data sets contain 71 and 55 sites respectively which, between them, contribute 78% and 82% of the total number of captures at all ages respectively (Table 2.1).

**Table 2.1** Sedge and Reed Warbler complete coverage data sets, 1987 - 2005: the number of adult and juvenile captures (excluding multiple within-year encounters), sites included, and yearly-site records in the reduced CES data set of “best” sites and the full CES data set. The right-hand column gives the percentage of the full set of complete coverage data represented in the reduced set from the “best” sites.

	Best sites	Full data	Percentage
<b>Sedge Warbler</b>			
Adult captures	10220	13055	78%
Juvenile captures	14627	18650	78%
Sites	71	192	37%
Yearly-site records	522	767	68%
<b>Reed Warbler</b>			
Adult captures	18664	22352	84%
Juvenile captures	18107	22581	80%
Sites	55	156	35%
Yearly-site records	418	661	63%

## 2.2 Estimating Abundance

For data collected under the CES scheme, by design, capture effort at any given site is assumed to be constant. Thus, any variability between years in the numbers of unique birds (adult or juvenile) caught is attributable only to annual changes in the population level and to stochastic variation, and not to varying intensity of capture effort. If all sites contributing data are operated over a series of years in accordance with the CES protocol, then the total number of unique birds caught provides an index of abundance.

Let  $G$  denote the number of sites, and  $T$  the number of study years. Given that all 12 visits are carried out at site  $g \in [1, G]$  in year  $t \in [1, T]$  the numbers of unique adult and juvenile birds caught once or more are denoted by  $n_{gt}^a$  and  $n_{gt}^j$  respectively, and referred to as “yearly-site counts”. Note that these yearly-site counts are not equal to the total number of adult/juvenile captures since individual birds can be caught at multiple visits within any given year. A natural estimate of the trend in adult abundance follows by assuming that the  $n_{gt}^a$  are independent, random variables from a Poisson distribution with parameter  $\lambda_{gt}^a$ . We express the natural logarithm of  $\lambda_{gt}^a$  as a linear combination of an intercept

term,  $\beta^a$ , plus site- and year-specific effects,  $s_g^a$  and  $y_t^a$  respectively (Peach et al., 1998). Mathematically, for  $g = 1, \dots, G$  and  $t = 1, \dots, T$ ,

$$\ln(\lambda_{gt}^a) = \beta^a + s_g^a + y_t^a. \quad (2.1)$$

This formulation naturally accommodates the absence of observations from some years, which are assumed to follow the observed trend (ter Braak et al., 1994). The site effects,  $s_g^a$ , control for the inherent differences in population level, and also for differences in capture effort, between individual sites. An estimated index of adult abundance in year  $t$  is given by  $A_t = \exp(y_t^a)$ . This index enables annual changes in the adult (breeding) population to be monitored.

Similarly, for juveniles, we assume that the  $n_{gt}^j$  are independent, random variables from a  $\text{Poisson}(\lambda_{gt}^j)$  distribution and form the analogous model for  $\lambda_{gt}^j$ :

$$\ln(\lambda_{gt}^j) = \beta^j + s_g^j + y_t^j, \quad (2.2)$$

where  $\beta^j$ ,  $s_g^j$ , and  $y_t^j$  denote the intercept term, and site- and year-specific effects respectively. An estimated index of juvenile abundance in year  $t$  is given by  $J_t = \exp(y_t^j)$ . As  $J_t$  is derived from both nest success rates and immediate post-fledging survival (Du Feu and McMeeking, 1991) it enables annual variations in the number of successfully fledged young to be monitored.

For identifiability, the first year effect (1987) and an arbitrary reference site effect in Equations (2.1) and (2.2) are constrained to zero.

### 2.2.1 Model Fitting and Results

Likelihoods  $L_a$  and  $L_j$  under the adult and juvenile abundance models, described by Equations (2.1) and (2.2), are easily formed and fitted, respectively, to the observations  $n_{gt}^a$  and  $n_{gt}^j$ . Here:

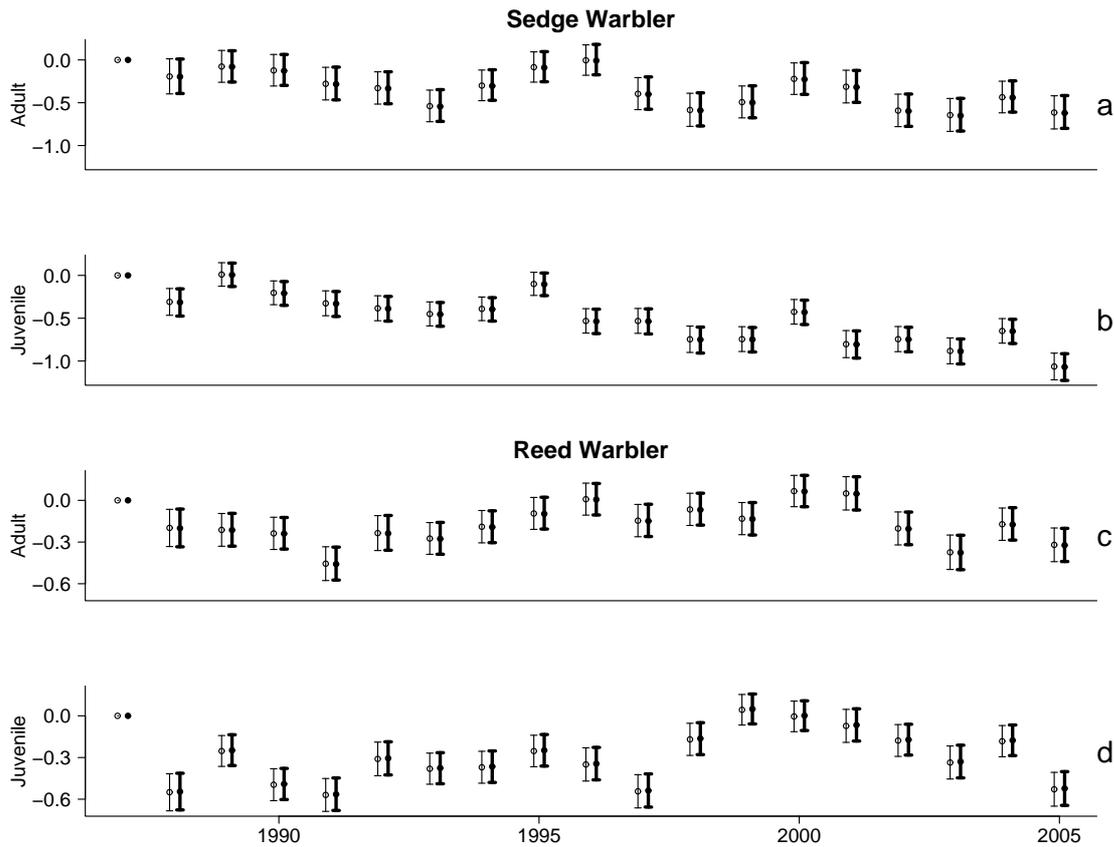
$$\begin{aligned} L_a &= f(\mathbf{n}_{gt}^a | \beta^a, \mathbf{s}^a, \mathbf{y}^a) \\ &= \prod_{g,t} \frac{\exp(-\exp(\beta^a + s_g^a + y_t^a)) \times \exp(\beta^a + s_g^a + y_t^a)^{n_{gt}^a}}{n_{gt}^a!}, \\ L_j &= f(\mathbf{n}_{gt}^j | \beta^j, \mathbf{s}^j, \mathbf{y}^j) \\ &= \prod_{g,t} \frac{\exp(-\exp(\beta^j + s_g^j + y_t^j)) \times \exp(\beta^j + s_g^j + y_t^j)^{n_{gt}^j}}{n_{gt}^j!}. \end{aligned} \quad (2.3)$$

Estimates of the unknown parameters are obtainable either classically, by maximum likelihood (as conventionally), or using Bayesian MCMC techniques. Using a classical approach, the models for adult and juvenile abundance are readily fitted in any Generalized Linear Modelling package, or as in the current research using the `glm` function in R (R Development Core Team, 2007). Likewise, Bayesian parameter estimation is straightforward. We specify vague, independent, Normal priors with mean 0 and variance 10,000 for all parameters, and use a random walk, single update, Metropolis-Hastings algorithm with Normal proposal distributions (see Section 1.5). For each parameter the variance of the corresponding Normal proposal distribution is tuned *a priori*. The MCMC simulations are run for 200,000 iterations with the first 100,000 iterations discarded as burn-in. Essentially identical posterior estimates are obtained from independent replications with different overdispersed starting points, so that we assume the chain has converged. The Bayesian analysis is readily conducted in WinBUGS (Spiegelhalter et al., 2003), although bespoke code was also written in C for this purpose.

Figure 2.1 provides estimates of  $y_t^a$  and  $y_t^j$ , from the adult and juvenile abundance models given in Equations (2.1) and (2.2) respectively, derived using both classical and Bayesian techniques, along with measures of their uncertainty (95% confidence intervals and 95% symmetric credible intervals). Note that for the Bayesian model we have specified vague priors, and with the resulting posterior distributions approximately symmetrical, the Bayesian posterior means are nearly identical to the classical maximum likelihood estimates. Further, the measures of uncertainty are also similar.

The year effects are relative to a base year (1987 in this case), hence it is changes in these effects that are meaningful, not their absolute value. Under the constant effort protocol, changes in the year effects are assumed to reflect true changes in abundance. Therefore, Figure 2.1 enables comparisons between the estimated trends in abundance for adult and juvenile birds of the same species, and between Sedge and Reed Warblers, to be made.

The abundances of both Sedge and Reed Warbler exhibit large inter-annual fluctuations, although, for both, the long-term trend from 1987 to 2005 appears fairly stable (Figure 2.1). The true trends in adult and juvenile abundance are unlikely to be smooth as short-lived species with potentially environmentally dependent survival and productivity rates might well fluctuate in number.



**Figure 2.1** Classical maximum likelihood estimates and the 95% confidence intervals, denoted by thin lines (left), and Bayesian posterior means and the 95% symmetric credible intervals, denoted thick lines (right), of  $y_t^a$  and  $y_t^j$  from Sedge Warbler (SW) and Reed Warbler (RW) CES data. a) SW:  $y_t^a$ , b) SW:  $y_t^j$ , c) RW:  $y_t^a$ , d) RW:  $y_t^j$ .

Within a species, there is some similarity in the trends of adult and juvenile abundances in that the major peaks and troughs are preserved. Between species there is considerable correspondence in the annual trend of adult abundance. For example the steady increase in adult abundance between 1993 and 1996, followed by the sharp decline in abundance in 1997, is evident in both Figures 2.1a and 2.1c. Conversely, the overall trend in juvenile abundance is more inconsistent between the two species. Nevertheless, several of the major peaks and troughs are evident in both Figures 2.1b and 2.1d, for example the increased abundance in 1989, 1995 and 2004, and the sudden declines in 1988 and 2005.

For adult and juvenile birds of both species, but in particular for adult Sedge Warbler, an apparent cyclic pattern in abundance is noticeable. Such a phenomenon is consistent with density-dependence, although there is no evidence that Sedge or Reed Warblers have filled all their suitable habitats to capacity. Moreover, the pattern is synchronised in timing between species, and despite

some inter-specific competition at their British breeding sites (Catchpole, 1974), Sedge and Reed Warblers occupy different ecological niches (Bibby and Green, 1981; Simms, 1985). Therefore the apparent cyclic trend is more likely to be related to common environmental factors, in their British breeding grounds or African wintering quarters, affecting both species. For example in the analysis of CES count data for 28 songbird species, Peach et al. (1998) ascribed many of the pronounced drops in abundance to unfavourable weather conditions in either Britain, or for migratory species, west Africa.

### 2.2.2 Validity of Model Assumptions

Pivotal to the use of CES count data to monitor abundance is the assumption that changes in the yearly-site counts ( $n_{gt}^a$  or  $n_{gt}^j$ ) reflect true changes in the population size. Several studies have been undertaken to test this assumption such as Baillie et al. (1986); Du Feu and McMeeking (1991); Peach et al. (1998); and Silkey et al. (1999). In their review of the CES scheme Baillie et al. (1986) emphasised that between year comparisons in the numbers of birds caught are likely to provide an accurate representation of changes in abundance if:

1. The within-year seasonal pattern of capture is consistent across years, as substantial differences, resulting from behavioural change, could mask changes in abundance.
2. A high proportion of birds at each site are caught.

For the five species Baillie et al. (1986) considered (Blackcap *Sylvia atricapilla*, Willow Warbler *Phylloscopus trochilus*, Wren *Troglodytes troglodytes*, Blackbird *Turdus merula*, and Bullfinch *Pyrrhula pyrrhula*) they concluded that obtaining meaningful indices of adult and juvenile abundance from CES data is possible.

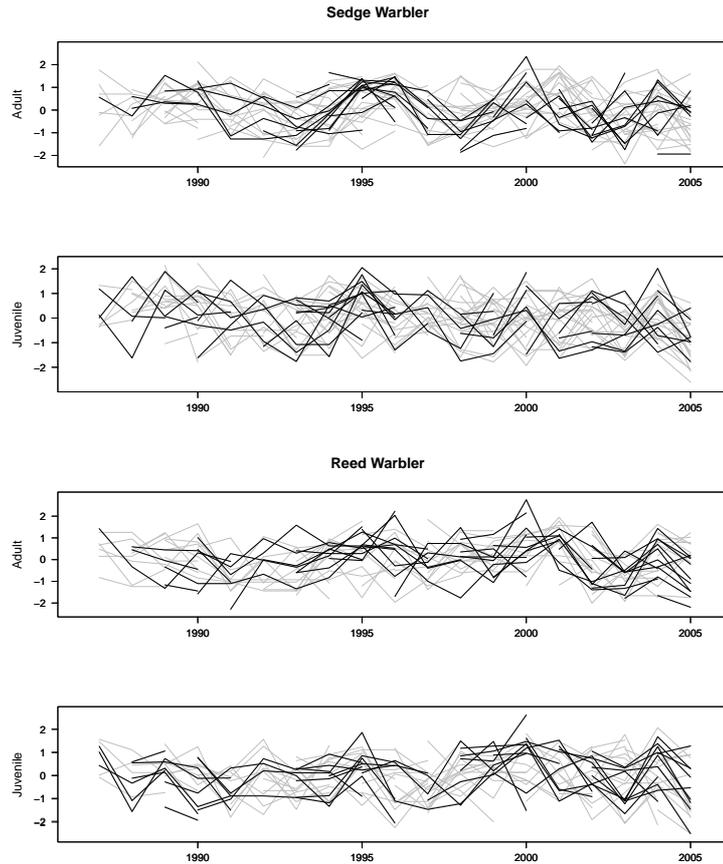
To validate the use of adult CES count data ( $n_{gt}^a$ ) to measure changes in adult abundance Peach et al. (1998) compared the average annual rate of population change calculated from adult CES counts, using a linear model analogous to that given by Equation (2.1) in which the year effects,  $y_t^a$ , were constrained to be log-linear across time, to the change in counts of breeding pairs from Common Bird Census (CBC) data. The analysis was conducted for 22 songbird species including Sedge Warbler, but not Reed Warbler as the CBC data for this species were deemed too sparse. They found good agreement between the independent measures of population change from the CBC data and the trends derived from the CES count data. This led them to conclude that trends in adult CES count

data are broadly representative of changes in adult songbird abundance in the wider countryside. In an analogous study of coastal bird populations at Point Reyes (California, U.S.A.), Silkey et al. (1999) found that relative changes in breeding density between years could be inferred by changes in mist-net catch sizes for 3 of the 4 species studied.

For adults it is possible to compare indices of population change derived from CES count data with those from an independent source, like the CBC (e.g. Peach et al., 1998), however for juvenile CES count data ( $n_{gt}^j$ ) there is no comparable independent data source readily available (Du Feu and McMeeking, 1991). Nonetheless, in a small-scale validation study, for a single CES site only, Du Feu and McMeeking (1991) found evidence that CES data provide a useful index of juvenile abundance by comparing the juvenile CES counts to independent data on the number of nestlings ringed.

Local habitat change can bias estimates of annual change in abundance (Harrison et al., 2000). However, the CES scheme's selection of sites not undergoing rapid successional change in habitat, the control of habitat succession at selected sites, the large number of sites contributing data, and the turn-over of sites, all mitigate against such bias. Further, on account of the similarity between estimates based on CES and CBC data, Peach et al. (1998) conjecture that habitat change is not a serious problem for CES monitoring.

Indices of adult and juvenile abundance produced via the models given in Equations (2.1) and (2.2) respectively assume that the annual trend in abundance is homogeneous across sites (Peach et al., 1998). Exploratory analyses of the current data sets (for example Figure 2.2 which shows the annual trend in  $\ln(n_{gt}^a)$  and  $\ln(n_{gt}^j)$  observed at each site), and previous studies (Baillie et al., 1986; Peach et al., 1996, 1998), indicate that this assumption is largely met, although to varying degrees for adult and juvenile birds. Peach et al. (1996) noted that annual changes in juvenile catch size were less consistent across sites than those of adults due to the aggregation of juvenile birds within a site. Juvenile birds tend to assemble, and thus be caught, in family groups or larger mixed-species flocks, whereas adults are more evenly distributed throughout the site as a consequence of territorial behaviour (Peach et al., 1996). The abundance model for juvenile CES data (and indeed adult data if need be) is readily adapted to incorporate such overdispersion of the count data relative to the Poisson distribution, either following the approach of Link and Sauer (2002) or by specifying a Negative Binomial distribution in place of the Poisson (King et al., 2008b).



**Figure 2.2** Standardised  $\ln(n_{gt}^a)$  and  $\ln(n_{gt}^j)$  counts for each site plotted against year. Consecutive annual counts for a particular site are connected with lines. The 20 sites with the highest mean annual count are plotted in black, the remainder are plotted in grey. Note, to standardise the observed site-specific log transformed counts the mean and standard deviation of the log transformed counts, at a particular site, over all years, are used.

As birds are mobile animals, the CES count data likely contain records from transient birds; birds that are not summer residents at the site of their capture but rather were caught as they migrated through it. Provided that the proportion of transient birds in the catch remains constant over time, their presence will not bias the indices of abundance, these being relative measures of the true population size. Such an assumption seems reasonable for CES data. We return to the issue of transients in Section 2.4.

Fundamental to the formulation of the adult and juvenile abundance models in this current section is the assumption of constant effort. Such an assumption is clearly violated when individual within-year visits to a particular site, during a year of operation, are missed. Here we only consider “complete coverage” data, so the assumption is not violated as a result of this. However, in Chapter 4 the models for adult and juvenile abundance are revisited, and adapted to accommodate data from years with missing visits.

## 2.3 Estimating Productivity

We define true seasonal productivity in year  $t$ , denoted by  $P_t^s$ , as the ratio of juvenile to adult birds in the population of birds breeding in the UK. Recall that  $A_t$  and  $J_t$  are indices of adult and juvenile abundance derived from Equations (2.1) and (2.2) respectively. The capture probabilities of adult and juvenile birds are unlikely to be equal due to behavioural differences (DeSante et al., 1999; Robinson et al., 2007), however, under the assumption of constant effort, these differences are expected to remain constant over time (Peach et al., 1996; Freeman et al., 2001; Miles et al., 2007; Robinson et al., 2007). Therefore, we let  $k$  denote an unknown positive scaling factor that accounts for the difference, assumed to be constant, between the indices and the true abundances, and write

$$\begin{aligned} P_t^s &= k \frac{J_t}{A_t} \\ &= kP_t, \end{aligned} \tag{2.4}$$

where  $P_t$  is the index of productivity in year  $t$ . Note that  $P_t = \exp(y_t^j - y_t^a)$ .

Biologically, juvenile abundance is influenced by both productivity and adult abundance. As such juvenile abundance could be high in a given year despite low productivity if there were sufficiently abundant adults to produce them, for example. Consequently  $P_t$  is a more useful parameter biologically than  $J_t$  itself as it adjusts for the numbers of adult birds producing young.

Annual variations in laying date may bias the productivity index; juveniles that leave the nest earlier are more likely to be captured than juveniles who leave later (Julliard et al., 2004). A trend for earlier laying dates, between 1987 and 2005, has been detected in Nest Record Scheme (NRS) data for both Sedge and Reed Warbler (Baillie et al., 2009b), however as this shift is small (less than 10 days) any bias in the current analysis is expected to be slight. In future analyses, should this trend continue, the potential for bias may need addressing.

### 2.3.1 Traditional Approach

Recall that the numbers of unique adult and juvenile birds caught at site  $g$  in year  $t$ , given that all 12 visits are made, are denoted by  $n_{gt}^a$  and  $n_{gt}^j$  respectively. Indices of productivity from CES catch data are typically calculated from a

simple Generalized Linear Model in which the number of unique juvenile birds in the annual unique catch are assumed to have a Binomial distribution (e.g. Freeman et al., 2001; Julliard et al., 2004; Robinson et al., 2007; Miles et al., 2007). That is, it is typically assumed

$$n_{gt}^j \mid (n_{gt}^j + n_{gt}^a) \sim \text{Binomial}(n_{gt}^j + n_{gt}^a, \theta_{gt}),$$

where the parameter  $\theta_{gt}$  is given by

$$\text{logit}(\theta_{gt}) = \beta^p + s_g^p + y_t^p, \quad (2.5)$$

in which  $\beta^p$ ,  $s_g^p$ , and  $y_t^p$  denote the intercept term and site- and year-specific effects respectively. For identifiability the first year effect (1987) and an arbitrary site effect in Equation (2.5) are constrained to zero. An index of productivity in year  $t$  is given by  $\exp(y_t^p)$  (Freeman et al., 2001; Robinson et al., 2007). It is this Binomial model that is currently employed by the BTO in their routine monitoring of UK songbird productivity from CES data. Following, we show that the Poisson abundance models for adult and juvenile counts, described by Equations (2.1) and (2.2) respectively, induce this traditional model, and hence our approach for indexing productivity is consistent with that currently used.

Recall that we assume  $n_{gt}^a \sim \text{Poisson}(\lambda_{gt}^a)$  and  $n_{gt}^j \sim \text{Poisson}(\lambda_{gt}^j)$ . Juveniles are dependent upon their parents for only a short period after fledging, three weeks at most (Bibby, 1978; Alker and Redfern, 1996). Then, the adults depart first for pre-migratory feeding grounds in July (Cramp, 1992). Assuming therefore that the Poisson distributions for  $n_{gt}^a$  and  $n_{gt}^j$  are independent, it can be readily shown that:

$$n_{gt}^j \mid (n_{gt}^j + n_{gt}^a) \sim \text{Binomial}(n_{gt}^j + n_{gt}^a, \theta_{gt}) \text{ where } \theta_{gt} = \frac{\lambda_{gt}^j}{(\lambda_{gt}^j + \lambda_{gt}^a)}.$$

Note that

$$\begin{aligned} \text{logit}(\theta_{gt}) &= \ln \left( \frac{\theta_{gt}}{1 - \theta_{gt}} \right) \\ &= \ln \left( \frac{\lambda_{gt}^j / (\lambda_{gt}^j + \lambda_{gt}^a)}{1 - \lambda_{gt}^j / (\lambda_{gt}^j + \lambda_{gt}^a)} \right) \\ &= \ln(\lambda_{gt}^j) - \ln(\lambda_{gt}^a). \end{aligned}$$

Under the Poisson abundance models for  $n_{gt}^a$  and  $n_{gt}^j$ , given in Equations (2.1)

and (2.2), we can write

$$\begin{aligned}
 \text{logit}(\theta_{gt}) &= \ln(\exp(\beta^j + s_g^j + y_t^j)) - \ln(\exp(\beta^a + s_g^a + y_t^a)) \\
 &= (\beta^j - \beta^a) + (s_g^j - s_g^a) + (y_t^j - y_t^a) \\
 &= \beta^p + s_g^p + y_t^p,
 \end{aligned} \tag{2.6}$$

where  $\beta^p = \beta^j - \beta^a$ ,  $s_g^p = s_g^j - s_g^a$ , and  $y_t^p = y_t^j - y_t^a$ . Note that this expression is identical to that for the traditional Binomial model described by Equation (2.5), assuming of course that the effects ( $s$  and  $y$ ) constrained to zero in Equations (2.1), (2.2) and (2.5) correspond to the same site and to the same year.

The index of productivity derived from the Binomial model,  $\exp(y^p)$ , can therefore be expressed as

$$\begin{aligned}
 \exp(y^p) &= \exp(y_t^j - y_t^a) \\
 &= \frac{J_t}{A_t} \\
 &= P_t,
 \end{aligned}$$

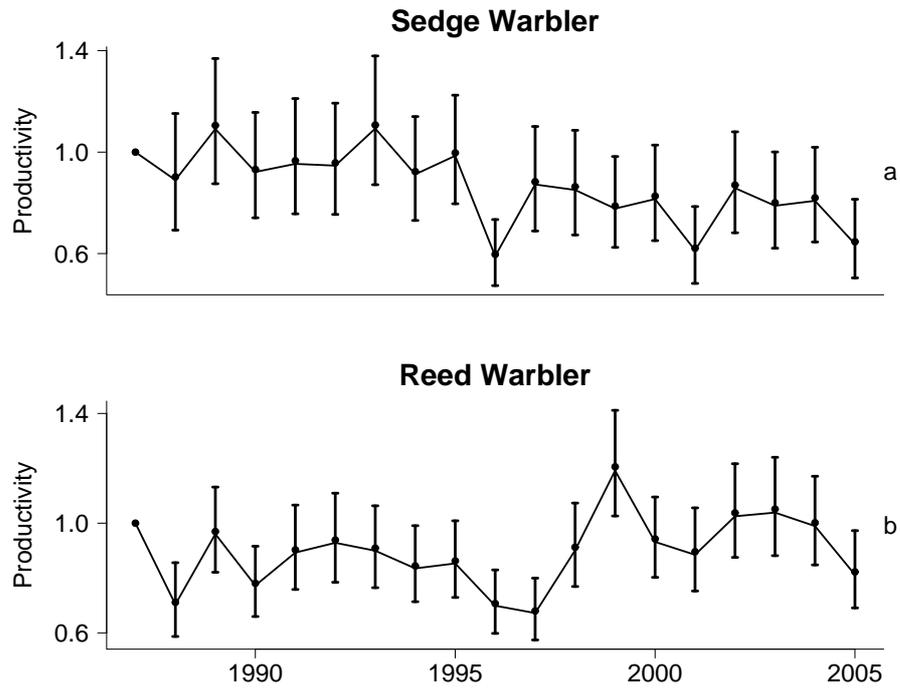
showing that our Poisson formulation for indexing productivity, given by Equation (2.4) via Equations (2.1) and (2.2), is analogous to the Binomial method often used.

We retain the Poisson formulation for indexing productivity, as this formulation is particularly useful in the integrated context since the joint likelihood for adult and juvenile counts, assuming independence, is simply the product of their individual likelihoods.

### 2.3.2 Model Fitting and Results

Bayesian estimates for the productivity indices,  $P_t$ , are readily derived by calculating the ratio of  $J_t$  to  $A_t$  using samples drawn from their posterior distributions. Classically, using the invariance property, the maximum likelihood estimator of  $P_t$  is given by  $\exp(\hat{y}_t^j - \hat{y}_t^a)$ . Here,  $\hat{y}_t^a$  and  $\hat{y}_t^j$  are the maximum likelihood estimators for  $y_t^a$  and  $y_t^j$  under the adult (Equation (2.1)) and juvenile (Equation (2.2)) abundance models respectively. For both Sedge and Reed Warbler, the classical maximum likelihood estimate of  $P_t$  is consistent with the Bayesian posterior mean (see Figure 2.3).

Due to the disparities in the magnitude of the annual fluctuations in adult and juvenile abundances over time (Figure 2.1), the productivity trend, for both species, is far from smooth (Figure 2.3). Between Sedge and Reed Warbler there is some correspondence between the estimated trend in productivity. For example, both species seem to experience a reduced productivity in 1988 followed by an increase. However, there are also some prominent differences, for example the rapid increase in productivity between 1997 and 1999 so evident for Reed Warbler is not reflected in the Sedge Warbler analysis.



**Figure 2.3** Indices of productivity for a) Sedge Warbler and b) Reed Warbler. The Bayesian posterior means are denoted by points, and the 95% symmetric credible intervals by vertical bars. The line represents the classical maximum likelihood estimates.

#### *Examining the Relationship Between the Indices of Abundance and Productivity*

The indices of abundance ( $A_t$  and  $J_t$ ) derived in Section 2.2 are clearly related to the indices of productivity ( $P_t$ ) derived in the current section; productivity is the ratio of juvenile to adult abundance (see Equation (2.4)). The Pearson's correlations ( $\rho$ , estimated classically) between the abundance and productivity indices for Sedge and Reed Warbler are given in Table 2.2. We note that the correlation between  $A_t$  and  $J_t$  is slightly stronger for Sedge Warbler than Reed Warbler (Table 2.2) implying a greater stability of productivity over time for Sedge Warbler.

For both Sedge and Reed Warbler,  $P_t$  is more strongly correlated with  $J_t$  than  $A_t$  (Table 2.2) implying that the non-smooth productivity trend is primarily due to the fluctuating nature of juvenile density over time. Despite, biologically, productivity directly affecting juvenile abundance, a causal relationship cannot be concluded. For example, whether productivity fluctuates because juvenile abundance is unstable, possibly due to variable chick and/or egg survival caused by poor weather during nesting (Freeman et al., 2001), or juvenile abundance fluctuates as fertility (brood number, clutch size etc) is unstable cannot be determined. Joint analysis with NRS data may shed light as to what underlying productivity factors (egg survival, chick survival, brood number, clutch size etc) are driving the observed changes (e.g. Freeman et al., 2001).

**Table 2.2** Correlations ( $\rho$ ) between Sedge and Reed Warbler indices of abundance ( $A_t$ ,  $J_t$ ) and productivity ( $P_t$ ). Each index is estimated by the mean of its posterior distribution.

		Sedge Warbler		Reed Warbler		
		$J_t$	$P_t$	$A_t$	$J_t$	$P_t$
<b>Sedge Warbler</b>	$A_t$	0.81	0.20	0.42	0.04	-0.35
	$J_t$		0.72	0.09	-0.02	-0.14
	$P_t$			-0.33	-0.13	0.12
<b>Reed Warbler</b>	$A_t$				0.67	-0.06
	$J_t$					0.70

The between species correlations are also given in Table 2.2. Comparing the abundance and productivity indices of Sedge and Reed Warblers is also of interest since these two species inhabit common summer and winter quarters (Peach, 2002; Redfern and Alker, 2002), albeit occupying different ecological niches (Bibby and Green, 1981; Simms, 1985), and thus similar environmental variables may act upon both populations.

As observed in Figure 2.1, the conformity in the annual trend of abundance between Sedge and Reed Warbler populations is much stronger for adult birds than juvenile birds, indeed for juveniles the correlation is nearly 0 (Table 2.2). Due to the inconsistencies in the annual trend of juvenile abundance between the two species, it is not surprising that Table 2.2 indicates that Sedge and Reed Warbler productivity are only weakly related at best. The influence of weather, and other environmental variables, on annual productivity most certainly operates in a complex manner throughout all stages of the breeding season (Freeman et al., 2001). The lack of association between Sedge and Reed Warbler with regards to the trends in productivity and juvenile abundance, despite both breed-

ing in wetland habitats and therefore being subjected to common environmental factors over the breeding season, may be attributable to their different breeding strategies. For instance, Reed Warbler nests, being commonly built suspended between reed stems, are more vulnerable to poor weather conditions and parasitism by Cuckoos, *Cuculus canorus*, whereas Sedge Warbler nests, being built concealed in dense tangles of vegetation, are more prone to mammalian predation (Bibby, 1978; Simms, 1985; Graveland, 1999). Further, Reed Warblers have a longer breeding season and double brooding is more common (Simms, 1985; Bibby, 1978; Robinson, 2005). Thus, in a particularly good breeding year a greater increase in productivity for Reed Warbler, than Sedge Warbler, might be observed due to a larger proportion of breeding pairs attempting a second brood. Conversely, in a poor breeding year, the shorter duration of the Sedge Warbler breeding season might restrict the number of breeding pairs attempting a second nest if the first one failed completely. We also note that the narrower food repertoire of Sedge Warbler, due to their inability to catch airborne insects (Bibby, 1978; Bibby and Green, 1981), may render them more susceptible to fluctuations in the availability of individual food sources.

### 2.3.3 Advantages of using CES Data to Monitor Productivity over NRS Data

For both Sedge and Reed Warbler, CES data and NRS data provide information on productivity, although the nature of this information is quite different. Whereas the indices of productivity,  $P_t$ , estimated from CES count data, provide a measure of breeding performance across the entire breeding season, integrating fecundity, multiple breeding attempts, egg and chick loss, and immediate post-fledging mortality (Freeman et al., 2001; Robinson et al., 2007), the NRS gathers information on the components of productivity: clutch size, brood size, and daily nest failure rates during the egg and nestling stages (Crick et al., 2003; Baillie et al., 2007). Measures of productivity across the entire breeding season are calculable from NRS data (see Peach et al., 1999; Siriwardena et al., 2001; Freeman and Crick, 2003), but, in general, constructing an informative measure is problematical. In part this is due to the unknown number of breeding attempts per pair (i.e. multiple broods or replacement nests) as a result of insufficient NRS sampling late in the breeding season (Bibby, 1978; Baillie et al., 1986), and as individual breeding birds are not monitored for the duration of the breeding season (Crick et al., 2003). This inability of the NRS to provide

information on the annual variation in the number of breeding attempts can prevent reliable estimation of true (i.e. total) seasonal productivity. For example, in a scenario described by Freeman et al. (2001), changes in agricultural practices, such as earlier harvesting, could cause a decline in farmland bird productivity via a reduction in the annual number of possible breeding attempts. This goes undetected in the NRS data. For species such as Reed Warbler which are typically double brooded (Catchpole, 1974), or species like Sedge Warbler for which the degree of double brooding is unknown (Alker and Redfern, 1996), CES count data provides a much better means of monitoring the true seasonal productivity over time.

Secondly, immediate post-fledging mortality is automatically incorporated into the CES index of productivity,  $P_t$ , for juveniles are defined to be fully fledged birds. As NRS data collection ends once the offspring have left the nest, it provides no information on post-fledging mortality. Consequently this important, and often appreciable (see Robinson et al., 2004), component of productivity is frequently over-looked in NRS reports on breeding performance (Freeman et al., 2001; Miles, 2005). Additional nestling ringing programmes may provide this information, but for many species, such as Reed Warbler, nests are not readily accessible rendering it impossible to ring the sufficiently large number of nestlings needed for accurate estimation (Freeman et al., 2001).

Although the CES scheme provides the only practical method of indexing annual productivity over the entire breeding season (Robinson et al., 2007), rather than on a per attempt basis as NRS data do, the worth of NRS data is not in doubt. They can yield information from habitats not amenable to CES monitoring, and have proved particularly valuable in highlighting the effects of climate change (Crick et al., 1997; Crick and Sparks, 1999). Along with joint interpretation with CES productivity indices, NRS can also, uniquely, provide an insight into which individual components of productivity are driving the observed changes in annual productivity.

## 2.4 Estimating Adult Survival

The CES live-recapture data consist of live resighting histories for individual birds over the study period and fall into the category of mark-recapture models (Lebreton et al., 1992; Williams et al., 2002). For example, the live-recapture data for an individual bird with ring K710666 are:

```
K710666 SEDWA AD 154 1997 0 0 1 0 1 0 1 1 0 0 0 0
K710666 SEDWA AD 154 1998 0 0 0 0 1 1 1 0 0 0 0 0
K710666 SEDWA AD 154 1999 0 1 0 0 0 0 1 0 0 0 0 0
K710666 SEDWA AD 154 2000 1 1 1 0 0 0 0 0 0 0 0 0
```

indicating that this bird, an adult (AD) Sedge Warbler (SEDWA), was initially caught and ringed at site 154 in 1997, and subsequently recaptured at this site in years 1998, 1999, and 2000. Furthermore, bird K710666 was caught at multiple visits within each of these years, e.g. in 1997 this bird was encountered at visits 3, 5, 7, and 8.

The annual recaptures of individual birds, under the CES scheme, provide valuable information on their survival. Adult Sedge and Reed Warblers are markedly site-faithful (Wernham et al., 2002), generally returning to the same breeding grounds annually, whereas juveniles disperse from their natal areas (Catchpole, 1974; Peach, 2002). On account of the high dispersal of juvenile birds, and consequently the low number of juveniles that are recaptured in years subsequent to ringing (see Table 1.3), we restrict this part of the analysis to the capture histories of adult birds.

An asset of the “constant effort” design is that probabilities of recapture (at a site) can be considered constant over years, simplifying the model structure (Peach et al., 1990; Pratt and Peach, 1991; Peach, 1993; DeSante et al., 1999). However, it is imperative that captured birds are representative of the population under interest, in this case resident breeders at the CES sites (Hines et al., 2003). The temporary presence of transient birds, i.e. birds not breeding in the vicinity of the site but migrating through it, needs to be accounted for by explicit probabilities within the analysis (Peach, 1993; Thaxter et al., 2006; Cave et al., 2009b; Freeman, 2008). As transient birds do not return to the site of their inaugural capture and ringing, if not adequately modelled, their presence in the CES live-recapture data produces negative bias in the estimates of survival (Peach et al., 1990; Pradel et al., 1997). Operationally, transient

birds can be thought of as having a zero probability of “survival” after their initial capture as they permanently leave the study area (Pradel et al., 1997).

The occurrence of transients is a common problem in the analysis of live-recapture data for birds and small mammals (Pradel et al., 1997). In this section we describe, compare, and contrast several alternative models appropriate for CES live-recapture data known to contain a mixture of resident and transient birds. We develop, and describe in detail, a new, modified approach to the standard Pradel et al. (1997) transient model that more fully utilises the information pertaining to residency provided by the CES scheme, and advocate its benefits over the alternative models.

We note that estimates of adult survival derived from CES data should be considered those of “apparent” adult survival (denoted by  $S_a$ ), which in addition to mortality, also incorporates a component due to the permanent emigration of birds away from their site of ringing (Pratt and Peach, 1991; Peach et al., 1995b; Thaxter et al., 2006). Assuming the rates of permanent emigration are constant over time, annual estimates of apparent adult survival will enable temporal changes in “true” adult survival (denoted by  $\phi_a$ ) to be detected (Peach, 1993). We address the conversion of apparent survival rates to true survival rates via the integrated population model in Chapter 5.

### 2.4.1 Models

Due to the high level of breeding site fidelity, and the geographic separation of CES sites, an individual adult Sedge or Reed Warbler is encountered at only a single site meaning that the capture histories of adult birds are site-specific. Under CES design protocol, catch effort is constant between years but may vary between sites, meaning that probability of recapture (denoted by  $c$ ) may do likewise. The same apparent survival rates ( $S_a$ ) are adopted across all sites, as the CES sites are assumed to sample with varying catch effort the same wider population of adult birds. Therefore, we define:

$c_g$  = Pr(a surviving resident bird, ringed at site  $g$ , is recaptured at this site in any given year),

$S_{a,t}$  = Pr(a resident bird alive and present at its site of ringing in year  $t$ , survives and is present in year  $t + 1$ ).

A bird regarded as a transient is assumed to permanently emigrate away from the study site immediately after ringing (Thaxter et al., 2006). That is, it does not breed at the site it was ringed but was merely caught in transit so that its chances of being available for future recapture can be assumed to be zero (Pradel et al., 1997; Thaxter et al., 2006). Following Pradel et al. (1997) we assume that at the time of ringing a bird has probability  $\tau$  of being a transient. We assume  $\tau$  is constant across both years and sites, an assumption which appears adequate (see Section 2.4.6), however this constraint can be relaxed. We now consider a number of different models for the CES live-recapture data.

### i Cormack-Jolly-Seber Model

If there were no transient birds in the data a Cormack-Jolly-Seber (CJS) model, with the appropriate parameter assumptions, would provide unbiased estimates of survival. Cormack (1964) and Lebreton et al. (1992) provide detailed descriptions of the general form of the CJS model. Following is a description of the particular CJS model suitable for CES live-recapture data, that is the model with time-dependent survival probabilities and site-specific recapture probabilities, given that the data contains only records from resident birds.

Initially, we formulate the model for a single site only, site  $g$ , in which mist-netting was carried out for  $T$  years. The live-recapture data from site  $g$  can be summarised in matrix form, commonly referred to as an “m-array” (Burnham et al., 1987), in which *for this site* (using our notation):

$R_{gi\bullet}$  = the number of birds caught and released in year  $i$ ,

$R_{git}$  = the number of birds recaptured in year  $t$  that were previously caught, and last released in year  $i$ ,

$Z_{gi}$  = the number of birds caught and released in year  $i$  never seen again

$$= R_{gi\bullet} - \sum_{t=i+1}^T R_{git},$$

for  $1 \leq i < T$  and  $i < t \leq T$ .

For example, when  $T = 4$  the m-array for site  $g$  would be:

$i/t$		2	3	4	<i>never seen again</i>
1	$R_{g1\bullet}$	$R_{g12}$	$R_{g13}$	$R_{g14}$	$Z_{g1}$
2	$R_{g2\bullet}$		$R_{g23}$	$R_{g24}$	$Z_{g2}$
3	$R_{g3\bullet}$			$R_{g34}$	$Z_{g3}$

The CJS model, which conditions on initial release, assumes that for site  $g$ :

1. each ringed bird, at time  $t$ , has the same probability of recapture,
2. each ringed bird has the same probability of surviving and remaining in the population of birds breeding at site  $g$  from time  $t$  to  $t + 1$ ,
3. the fate of each bird, with respect to recapture and survival, is independent of the fate of any other bird,
4. rings are not lost,
5. recaptures are accurately recorded.

Under these assumptions each row (or release cohort) of the m-array for site  $g$  has an independent Multinomial distribution. We denote the associated Multinomial cell probabilities (for site  $g$ ) by:

$$p_{git} = \Pr(\text{a bird caught and released in year } i \text{ is next recaptured in year } t),$$

$$\chi_{gi} = \Pr(\text{a bird caught and released in year } i \text{ is never seen again}).$$

Recall that for CES data from a single site, the probability of recapture ( $c_g$ ) is assumed to be time-invariant, whereas the probability of survival ( $S_a$ ) is allowed to vary annually. Therefore, we express the Multinomial cell probabilities as:

$$p_{git} = c_g(1 - c_g)^{t-i-1} \prod_{h=i}^{t-1} S_{a,h}$$

and

$$\chi_{gi} = 1 - \sum_{t=i+1}^T p_{git}. \quad (2.7)$$

The Multinomial based likelihood, based on data from a single site, is readily formed. A total likelihood for all  $G$  sites follows by multiplying the individual likelihoods for the independent sites. Recall that each site has its own time invariant recapture probability,  $c_g$ , but that the same temporally varying apparent survival rates,  $S_{a,t}$ , are adopted across all sites. The full likelihood, under the CJS model, denoted by  $L_{live(CJS)} = f(\mathbf{R}, \mathbf{Z} | \mathbf{S}_{a,1..T-1}, \mathbf{c}_{1..G})$ , is proportional to:

$$\prod_{g=1}^G \left[ \left( \prod_{i=1}^{T-1} \prod_{t=i+1}^T p_{git}^{R_{git}} \right) \left( \prod_{i=1}^{T-1} \chi_{gi}^{Z_{gi}} \right) \right].$$

Note that by conditioning on first release in the CJS model, the ringed birds constitute the population under study. We assume that the ringed birds are representative of the wider population. In particular, it is assumed that birds are selected at random for ringing, and that fitting and wearing of a leg ring has

no effect, adverse or beneficial, on subsequent survival (Lebreton et al., 1992). Many studies have shown that ringing, using BTO protocol, has no effect (BTO Website, 2009). Further, for Sedge and Reed Warbler, both being short-lived and neither having a strong association with salt water where ring loss due to corrosion and wear is often important, it is assumed that ring loss is negligible (Dobson, 1990), complying with assumption 4 above.

The presence of transient birds in the data violates assumptions 1 and 2 outlined above, for these birds do not return to their inaugural ringing site and therefore have a zero probability of being seen again after initial capture (Hines et al., 2003). Estimates of  $S_{a,t}$  under the CJS model, from data containing transients, will be negatively biased. In the following sections we describe various models, both *ad hoc* and theoretical, for dealing with the problem of transience.

## ii *Ad Hoc* Methods

There are a number of proposed *ad hoc* methods for separating the capture histories of known resident birds from those of transient birds. Typically such methods are easily used, but they can introduce different biases. One such approach is to model only the subset of data beyond the first year of recapture (Peach et al., 1990; Pradel et al., 1997; Cilimburg et al., 2002). Birds contributing data to this subset, by virtue of being seen in more than one year, are identified as residents, i.e. breeding at the CES site. Estimates of  $S_{a,t}$  from this reduced data set are readily obtained using the CJS model described above (Section 2.4.1.i). However as all captures from the first year of the study are omitted  $S_{a,1}$  cannot be estimated.

For short-lived species, like warblers, where the number of annual recaptures are low, see Table 1.3, this *ad hoc* approach necessitates discarding a large fraction of the data. A substantial reduction in the precision of estimation is to be expected (Peach et al., 1990). However, by design the CES scheme provides additional information with regards to residency status. Under the CES scheme individual birds can be encountered multiple times within the mist-netting season. This “within-year recapture” information is recorded and readily available. Given that successive CES visits are at least 3 days apart, usually more, we consider those birds caught in a single year, but repeatedly, also as residents. The *ad hoc* approach is easily modified, to utilise this extra information, by allowing the first recapture to occur within the year of ringing.

We refer to this approach as the “extended *ad hoc*” method. Using the within-year recaptures to identify residents, in addition to the between-year recaptures, potentially enables a larger fraction of the data to contribute to the analysis, increasing the precision of estimation. In addition,  $S_{a,1}$  can now be estimated.

As selected residents must have survived until recapture, the removal of birds caught only once is prone to selection bias (Freeman, 2008). We note that survival rates of individuals in an ecological population will differ according to their “fitness”. It is residents with higher survival rates that are more likely to be seen again and contribute data to these *ad hoc* analyses, whereas data from residents with poor survival are more liable for deletion. This non-random selection of residents with typically higher survival rates produces positive bias in the estimates of  $S_{a,t}$  (Peach et al., 1990). The use of within-year recaptures, in the extended *ad hoc* approach, will help to reduce this bias if mortality (of adults) over the mist-netting season is low, and if the majority of residents are identified as such by virtue of recapture within their inaugural ringing year.

### iii Pradel Model

Pradel et al. (1997) developed a theoretical approach to model the unknown mixture of transient and resident individuals in live-recapture data explicitly. This widely accepted model has been applied to a variety of data sets, for example those of Loery et al. (1997); DeSante et al. (1999); Sandercock and Jaramillo (2002); and Hilton and Miller (2003). Their approach is a generalisation of “model 2” of Brownie and Robson (1983) in which inaugural capture and ringing affects the survival probability of a bird in the year immediately after initial release. Except for newly ringed and released birds, all other birds are assumed to have the same probability of surviving. We describe the “Pradel model” for CES live-recapture data below.

The annual capture histories of each individual bird are conceptually thought to consist of (potentially) two parts:

1. “newly-marked” portion, i.e. the capture history from initial release to first year of recapture, or until the end of the study if the bird is not caught in any other year,

and if the bird is recaptured in a subsequent year,

2. “previously-marked” portion, i.e. the capture history from the first recapture year until the end of the study.

For example, take two birds, A and B, with the following five-year capture histories

<b>Bird</b>	<b>History</b>
A	1 0 1 1 0
B	1 0 0 0 0

where 1 denotes it was caught that year, and 0 that it was unobserved. The newly-marked and previously-marked parts of their capture histories are:

<b>Bird</b>	<b>Newly-Marked</b>	<b>Previously-Marked</b>
A	1 0 1	0 0 1 1 0
B	1 0 0 0 0	

Considering data from a single site  $g$  only, both these parts give rise to two site-specific m-arrays: one for the newly-marked birds and one for the previously-marked birds. For site  $g$  data, given  $T$  study years, for newly-marked birds we denote:

- $F_{gi\bullet}$  = the number of newly-marked birds caught and released in year  $i$ ,  
 $F_{git}$  = the number of newly-marked birds recaptured in year  $t$  that were caught, and last released in year in  $i$ ,  
 $X_{gi}$  = the number of newly-marked birds caught and released in year  $i$  never seen again

$$= F_{gi\bullet} - \sum_{t=i+1}^T F_{git},$$

where  $1 \leq i < T$  and  $i < t \leq T$ .

Similarly, for previously-marked birds, where  $2 \leq i < T$  and  $i < t \leq T$ , for site  $g$  data we denote:

- $R_{gi\bullet}$  = the number of previously-marked birds caught and released in year  $i$ ,  
 $R_{git}$  = the number of previously-marked birds recaptured in year  $t$  that were caught, and last released in year in  $i$ ,

- $Z_{gi}$  = the number of previously-marked birds caught and released in year  $i$  never seen again

$$= R_{gi\bullet} - \sum_{t=i+1}^T R_{git}.$$

In the example above, birds A and B would contribute the following data to

the site  $g$  m-arrays:

<i>Newly-Marked Birds</i>						
$F_{gi\bullet}$	$F_{git}$				$X_{gi}$	
$i/t$	$2$	$3$	$4$	$5$	<i>never seen again</i>	
$1$	$2$	$0$	$1$	$0$	$0$	$1$
$2$	$0$	$0$	$0$	$0$	$0$	$0$
$3$	$0$	$0$	$0$	$0$	$0$	$0$
$4$	$0$	$0$	$0$	$0$	$0$	$0$

<i>Previously-Marked Birds</i>					
$R_{gi\bullet}$	$R_{git}$			$Z_{gi}$	
$i/t$	$3$	$4$	$5$	<i>never seen again</i>	
$2$	$0$	$0$	$0$	$0$	
$3$	$1$	$0$	$1$	$0$	
$4$	$1$	$0$	$0$	$1$	

Each row of the newly-marked m-array, for site  $g$ , is assumed to have a Multinomial distribution, which given the number of newly-marked birds released in year  $i$ ,  $F_{gi\bullet}$ , are independent. Likewise it is assumed that each row of the previously-marked m-array, for site  $g$ , has an independent Multinomial distribution given the number of previously-marked birds released in year  $i$ ,  $R_{gi\bullet}$ .

Upon initial release the residency status of a newly-marked bird is unknown: it might be a transient (with probability  $\tau$ ) or a resident (with probability  $1 - \tau$ ). Conversely, by virtue of being seen in at least two years, the previously-marked birds are by definition residents, and the Multinomial cell probabilities, corresponding to previously-marked m-array, are readily derived. They are simply the straightforward Cormack-Jolly-Seber form given in Section 2.4.1.i.

To model the *newly-marked* segment of the data from site  $g$  we define:

$$\begin{aligned}
 f_{git} &= \Pr(\text{a bird ringed and released in year } i \text{ is next recaptured in year } t), \\
 x_{gi} &= \Pr(\text{a bird ringed and released in year } i \text{ is never recaptured in any} \\
 &\quad \text{subsequent years}).
 \end{aligned}$$

For newly-marked birds, ringed in year  $i$ , Pradel et al. (1997) gave the probability of surviving to the following year,  $i + 1$ , as

$$S_i^{new} = \tau_i S_i^{trans} + (1 - \tau_i) S_i^{res} \tag{2.8}$$

where  $S^{trans}$  and  $S^{res}$  are the annual survival probabilities of transients and residents respectively. That is, survival following initial capture and release is a mixture of survival rates for transients and residents, where  $\tau$  is the mixture parameter. Recall that in our notation the (apparent) survival probability for resident birds is denoted by  $S_a$ .

Estimation under this general model (given by Equation (2.8)) is not possible, but by operationally defining transients to be individuals that have a zero probability of survival after their initial capture Pradel et al. (1997) reduced Equation (2.8) to:

$$S_i^{new} = (1 - \tau_i)S_i^{res}. \quad (2.9)$$

By definition, a newly-marked bird “surviving” its year of ringing is a resident, and hence its survival probability in the following year,  $i + 1$ , is given by  $S_{i+1}^{res}$ , the resident survival rate.

Recall that for CES live-recapture data, the recapture probability for a resident bird,  $c$ , is assumed to be time-invariant due to the constant effort protocol, but is allowed to vary between sites.  $S^{res}$ , which we denote by  $S_a$ , is allowed to vary temporally but is constant across all sites, as these sites are assumed to sample the same wider population of birds. The probability a bird is a transient at the time of ringing,  $\tau$ , is assumed to be constant across both years and sites. Therefore, following the CJS model, described by Equation (2.7), but adjusting for transients via Equation (2.9), the Multinomial cell probabilities ( $f_{git}$ ,  $x_{gi}$ ) corresponding to the newly-marked m-array for CES data from site  $g$  are:

$$f_{git} = (1 - \tau)c_g(1 - c_g)^{t-i-1} \prod_{h=i}^{t-1} S_{a,h}$$

and  $x_{gi} = 1 - \sum_{t=i+1}^T f_{git}.$  (2.10)

Equation (2.10) explicitly acknowledges that newly-marked birds have a lower *apparent* survival in the first year immediately after their initial release owing to the presence of transient birds.

For live-recapture data from a single site the full likelihood is simply the product of the independent, component Multinomial likelihoods arising from the newly-marked and previously-marked m-arrays. For  $G$  independent sites, the total likelihood results by multiplying together the site-specific likelihoods, each of

which has common adult survival probabilities ( $S_{a,t}$ ) but their own recapture probability ( $c_g$ ), i.e.

$$L_{live(P)} = f(\mathbf{F}, \mathbf{X} | \tau, \mathbf{c}_{1...G}, \mathbf{S}_{a,1...T-1}) \times f(\mathbf{R}, \mathbf{Z} | \mathbf{c}_{1...G}, \mathbf{S}_{a,2...T-1}). \quad (2.11)$$

We now extend the Pradel model given the extra information contained in CES data relating to the numerous visits within a single year.

#### iv Extended Pradel Model

Pradel et al. (1997) assumed that all birds caught in more than one year were, by virtue of the fact, resident birds (for the duration of the breeding season), but those caught in only a single year were a mixture in unknown proportion of resident birds that evaded future recapture, and transients. The CES design protocol, however, means that additional residents can be identified: those encountered multiple times within the mist-netting season. Were this information exploited, improved precision in the estimates of  $S_{a,t}$  is expected as the uncertainty regarding residency status is reduced. We therefore propose an extension to the Pradel model that utilises the additional information in the CES live-recapture data pertaining to residency by considering birds caught in a single year, but repeatedly, also as residents (Cave et al., 2009b).

Following the approach of Pradel et al. (1997), described in Section 2.4.1.iii, we divide the capture histories into two segments. However, we allow the “first recapture” event to occur within the year of ringing. That is, we segment the data into:

1. From ringing to first recapture which may occur within the year of ringing or, if never seen again, from ringing to the end of the study.
2. From first recapture onwards, at which point the resident status is confirmed.

Initially we consider live-recapture data from a single site  $g$  only. To model the data in the first segment, for site  $g$ , recall that:

$$f_{git} - \text{Pr}(\text{bird at site } g, \text{ ringed in year } i, \text{ is first recaptured in year } t),$$

$$x_{gi} - \text{Pr}(\text{bird at site } g, \text{ ringed in year } i, \text{ is never seen again}),$$

but noting that “recaptures” may occur within the year of ringing. For example, formally:

$x_{gi} = \Pr(\text{bird at site } g, \text{ ringed in year } i, \text{ is never seen again either within year } i \text{ or in a subsequent year}).$

To derive expressions for the above probabilities we define:

$\varepsilon_g$  -  $\Pr(\text{resident bird, caught and ringed at site } g, \text{ is caught only once within its inaugural ringing year}),$

which is referred to as the ‘‘evasion’’ probability. Note that as a consequence of design  $\varepsilon_g$  is assumed to be time-invariant. Further, the assumption that a bird is caught independently at successive within-year visits is questionable, thus  $\varepsilon_g$  and  $c_g$  are estimated freely of one another.

The birds contributing data to the first segment are a mixture of transient and resident birds in unknown proportion. A resident bird only caught once has evaded future recapture, possibly due to death. Denoting the number of study years by  $T$ , for site  $g$  we derive

$$\begin{aligned} f_{git} &= (1 - \tau)(1 - \varepsilon_g) & i = t \leq T, \\ f_{git} &= (1 - \tau)\varepsilon_g c_g (1 - c_g)^{t-i-1} \prod_{h=i}^{t-1} S_{a,h} & i < t \leq T, \\ \text{and } x_{gi} &= 1 - \sum_{t=i}^T f_{git} & i \leq T. \end{aligned} \quad (2.12)$$

If, for site  $g$ ,  $F_{git}$  denotes the number of birds ringed in year  $i$  first recaptured in year  $t$ ,  $t \geq i$ , and  $X_{gi}$  the number that were never seen again, the histories up until the first recapture can be summarised in a matrix (an  $m$ -array) where the rows are assumed to have independent Multinomial distributions with the cell probabilities above. We denote the Multinomial based likelihood by  $f(\mathbf{F}, \mathbf{X} | \tau, \varepsilon_g, c_g, \mathbf{S}_{a,1..T-1})$ .

All birds ( $F_{gi\bullet} - X_{gi}$ ) contributing to the second segment, having been recaptured at least once, are by definition residents. The model for this segment of their capture histories is a product of Multinomial distributions, parametrised with recapture probabilities alongside adult survival rates in the standard fashion. The probability that one of these resident birds released in year  $i$  is next recaptured in year  $t$ , excluding multiple encounters within the same year,  $p_{git}$ , or never seen again,  $\chi_{gi}$ , is given by Equation (2.7), i.e. the CJS model of Section 2.4.1.i. The resulting Multinomial based likelihood is denoted by  $f(\mathbf{R}, \mathbf{Z} | c_g, \mathbf{S}_{a,1..T-1})$ .

The extended Pradel model is completed by recognising that we have further information on the evasion parameter,  $\varepsilon_g$ , arising from multiple within-year encounters from the data in segment 2. Under the assumption of “constant effort” the probability a resident is only caught once within its ringing year,  $\varepsilon_g$ , equals the probability a resident caught in any other year is only caught once within that year. Thus for  $t \in [2, T]$ , for site  $g$  data, we define:

$M'_{gt}$  - the number of ringed resident birds recaptured in year  $t$  exactly once,

$M''_{gt}$  - the number of ringed resident birds caught repeatedly in year  $t$ .

Therefore  $M'_{gt} | M'_{gt} + M''_{gt} \sim \text{Binomial}(M'_{gt} + M''_{gt}, \varepsilon_g)$ , and the Binomial based likelihood,  $f(\mathbf{M}'_g | \varepsilon_g, \mathbf{M}'_g + \mathbf{M}''_g)$ , results. Note that  $\mathbf{M}'_g + \mathbf{M}''_g = \sum_i R_{git}$ .

The likelihood for the capture histories of all birds (at site  $g$ ) is then given by the product of the above, independent, component likelihoods. A total likelihood for data from all  $G$  CES sites then follows by multiplying the likelihoods for the independent sites. This total likelihood is denoted by:

$$\begin{aligned} L_{live(EP)} = & f(\mathbf{F}, \mathbf{X} | \tau, \boldsymbol{\varepsilon}_{1\dots G}, \mathbf{c}_{1\dots G}, \mathbf{S}_{a,1\dots T-1}) \times f(\mathbf{R}, \mathbf{Z} | \mathbf{c}_{1\dots G}, \mathbf{S}_{a,1\dots T-1}) \\ & \times f(\mathbf{M}' | \boldsymbol{\varepsilon}_{1\dots G}, \mathbf{M}' + \mathbf{M}''). \end{aligned} \quad (2.13)$$

Note that if each site was only visited once per year, that is if  $\varepsilon_g = 1$  for all  $G$  sites, then our extended Pradel model reduces to the original Pradel et al. (1997) model.

The precise definition of  $\tau$  in the extended Pradel model requires care. Strictly speaking the parameter  $\tau$  also incorporates a component due to mortality within the mist-netting season, see Equation (2.12). Clearly this is inconsistent with the biological definition of transients being birds not normally resident at the CES site during the breeding season. For example, a resident bird breeding at a site may only be captured once due to its death shortly after ringing. Such a bird is operationally classified as a transient under the extended Pradel model, but biologically it is not. If significant mortality over the mist-netting season did occur  $S_{a,t}$  would be over-estimated (Pratt and Peach, 1991). However, despite the long CES sampling period (May-August) there is little evidence to suggest that small passerines experience significant mortality during this time (Peach, 1993), and thus the magnitude of such bias is expected to be small (Pratt and Peach, 1991).

## v c-dash Method

The approach currently under development by the BTO, which we refer to as the c-dash method, modifies each individual bird's annual capture history by adding a new term immediately after its inaugural ringing year. This new term indicates whether or not the bird had been encountered more than once within its year of ringing. The period of time after initial ringing, but in the same CES mist-netting season (year), is referred to as the "post-ringing period".

The c-dash method is described, and applied to CES, in an unpublished technical report by Freeman (2008), and provisional results for Reed Warbler presented in an article for the BTO News (Freeman et al., 2005), a magazine for BTO members. Like the extended Pradel method, the c-dash method is essentially an extension of the Pradel et al. (1997) approach that utilises the information in the multiple within-year encounters to identify residents. We briefly describe the method below.

Consider the capture history of K710666 given at the start of Section 2.4. This bird was initially ringed in 1997, and recaptured in 1998, 1999 and 2000. Its annual capture history from 1987 to 2005, represented by a vector of 1's (caught) and 0's (not caught), is

0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0

K710666 was captured at four visits, 3, 5, 7, and 8, within 1997, its inaugural ringing year. Its capture history is now modified, by the inclusion of an extra "1" after the entry corresponding to its initial capture in 1997, to indicate that this bird was caught multiple times within this year, i.e.

0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0

Had K710666 been captured only once in 1997, the modified capture history would be:

0 0 0 0 0 0 0 0 0 0 1 0 1 1 1 0 0 0 0 0

The modified histories are treated in the standard manner of Pradel et al. (1997) (described in Section 2.4.1.iii), i.e. the site-specific data are split into segments 1) from initial release to first recapture (which may now occur within the year of ringing), and 2) from first recapture till the end of the study. In the latter segment all birds are assumed to be residents and the standard Multinomial likelihood under the CJS model results. Likewise, a Multinomial likelihood for the data in the first segment results although formulating the appropriate cell

probabilities requires a little more attention, as the first “survival” probability now represents a combination of survival and residency in the post-ringing period.

In our notation, let:

$$c'_g = \text{Pr}(\text{a ringed resident bird at site } g \text{ is recaptured in the post-ringing period}),$$

$$S'_a = \text{Pr}(\text{a ringed resident bird survives the post-ringing period}).$$

Freeman (2008) assume both  $c'_g$  and  $S'_a$  are time-invariant, although as with the annual recapture probability  $c_g$ ,  $c'_g$  is allowed to vary between sites. Under the constant effort protocol assuming  $c'_g$  to be time-invariant is reasonable. The assumption of time-invariance is more tenuous for  $S'_a$  but it could be relaxed.

Consider a study over  $T = 3$  years, at site  $g$ . Letting  $t'$  denote the post-ringing period in year  $t$ , the probability matrix associated with the m-array for site  $g$  data in the first segment, initial release to first recapture, is:

$$\begin{array}{rccccccc}
 i/t & 1' & 2 & 2' & 3 & & 3' \\
 1 & (1-\tau)S'_a c'_g & (1-\tau)S'_a(1-c'_g)S_{a,1}c_g & & (1-\tau)S'_a(1-c'_g)S_{a,1}(1-c_g)S_{a,2}c_g & & \\
 2 & & & (1-\tau)S'_a c'_g & (1-\tau)S'_a(1-c'_g)S_{a,2}c_g & & \\
 3 & & & & & & (1-\tau)S'_a c'_g
 \end{array}$$

Note the difference in the period of time covered by  $S_{a,t}$  in ringing and non-ringing years is ignored. Effectively it runs from the end of one season to the start of the next, with zero mortality within the season assumed. Further, between individual birds the exact length of the post-ringing period will vary depending on what visit ringing occurred. This is also ignored. For example, a bird ringed at visit 1 has a post-ringing period that extends over the 11 remaining visits, whereas a bird caught in visit 12 has no post-ringing period. Potentially this could result in substantial individual heterogeneity in  $c'_g$  and  $S'_a$ , however as the majority of adult birds are ringed during the same visits (see Table 2.3) the impact of different post-ringing periods is probably slight.

**Table 2.3** Percentage of adult Sedge and Reed Warblers ringed at each of the 12 within-year CES visits.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>Sedge Warbler</b>	23.9	18.9	13.7	9.7	6.7	5.5	4.3	4.8	5.8	3.9	2.0	0.8
<b>Reed Warbler</b>	4.8	8.5	14.1	14.8	12.8	11.0	9.7	8.4	7.8	5.2	2.1	0.7

The parameters  $(1 - \tau)$  and  $S'_a$  are confounded, they always occur as a joint product. Their product,  $\varphi = (1 - \tau)S'_a$ , the probability a bird is a resident and once ringed survives long enough for recapture within its ringing year, is referred to as the “residence probability”.

If each site was only visited once per year then  $S'_a = 1$ , and  $c'_g = 0$ , reducing the c-dash method to the original Pradel et al. (1997) model.

#### *Correspondence Between the c-Dash Method and the Extended Pradel Model*

The key difference between the c-dash method of Freeman (2008) and our extended Pradel model is that we have also utilised the within-year recaptures in non-ringing years, including an extra Binomial term into the full likelihood. Otherwise, it can be readily shown that the c-dash method is structurally equivalent to the extended Pradel model (if the Binomial term is omitted). Clearly the models for the second segment of data, from first recapture onwards, are identical. The models for the first segment of data, initial release till first recapture, are essentially reparametrisations of one another. Here  $1 - \varepsilon_g = c'_g$ . However, as outlined below, the relationship between the transient/resident parameters of these two models is complicated by post-ringing period mortality.

For clarity, we denote  $\tau$  in the extended Pradel model and in the c-dash method by  $\tau_{EP}$  and  $\tau_{CD}$  respectively. With no mortality in the post-ringing period, as is expected to be the case for CES data (Peach, 1993), then  $\tau_{EP} = \tau_{CD}$ . However, when mortality occurs within the post-ringing period, due to its implicit incorporation into  $\tau_{EP}$  (see Section 2.4.1.iv), this equality no longer holds. Instead  $\tau_{EP} = 1 - (1 - \tau_{CD})S'_a = 1 - \varphi$ .

### 2.4.2 Model Fitting

For clarity we denote the models described above as:

- CJS - §i : Cormack-Jolly-Seber model
- AH - §ii : *Ad Hoc* method
- EAH - §ii : Extended *Ad Hoc* method
- P - §iii : Pradel model
- EP - §iv : Extended Pradel model
- CD - §v: c-Dash method

All six models are fitted using Bayesian techniques with bespoke code written in C. Vague, independent, Uniform[0,1] priors are specified for the survival ( $S_a$ ), transience ( $\tau$ ), residence ( $\varphi$ ), recapture ( $c$ ), and evasion ( $\varepsilon$ ) probabilities. We use a single update, random walk Metropolis-Hastings algorithm with Uniform proposals, appropriately truncated, for all model parameters. The step length for each proposal distribution is tuned *a priori*. For each model, the simulations are run for 200,000 iterations with the first 100,000 iterations discarded as burn-in. Further, as independent replications with different overdispersed starting points produced essentially identical posterior estimates, for each model considered, we assume convergence of the MCMC chain.

The various Cormack-Jolly-Seber models (CJS, AH, EAH) are readily fitted classically, at least in theory if not in practice, by Program MARK (White and Burnham, 1999). Model P can also be fitted in Program MARK using a two “age”-class model that separates the survival probability immediately after ringing ( $S^{new}$ ) from the survival probabilities in all subsequent years ( $S^{res} = S_a$ ). This two age-class model is functionally equivalent to model P (Sandercock and Jaramillo, 2002; Hilton and Miller, 2003). In practice, using Program MARK to fit these models for data from a large number of sites is cumbersome. Models EP and CD could be fitted classically using bespoke code but we consider the Bayesian approach. Model CD is currently fitted classically by the BTO using SAS software (SAS Institute Inc., 2008) although boundary issues occasionally prevent the estimation of standard errors, and convergence can be a problem (pers comm: Dr Rob Robinson, BTO).

Small scale simulation studies have shown that under ideal conditions (high numbers of releases, reasonable number of recaptures, and no individual heterogeneity in survival or recapture probabilities) models AH, EAH, P, EP, and CD all perform well. In Section 2.4.3 we compare the competing models, and the Bayesian and classical estimates where appropriate, adopting Sedge Warbler data from the “30 best” sites, i.e. the 30 sites, from the set of 71 sites in the complete coverage data set, that caught the most adult birds. This reduced data set was used to make it feasible to obtain classical estimates from Program MARK. Results from the Bayesian analysis of the full “best” sites complete coverage Sedge and Reed Warbler data sets, using our preferred model, EP (see Section 2.4.4), are presented in Section 2.4.5.

### 2.4.3 Comparing the Alternative Survival Models

Table 2.4 summarises the adult Sedge Warbler CES live-recapture data, all with complete coverage, from the “30 best” sites used in this section.

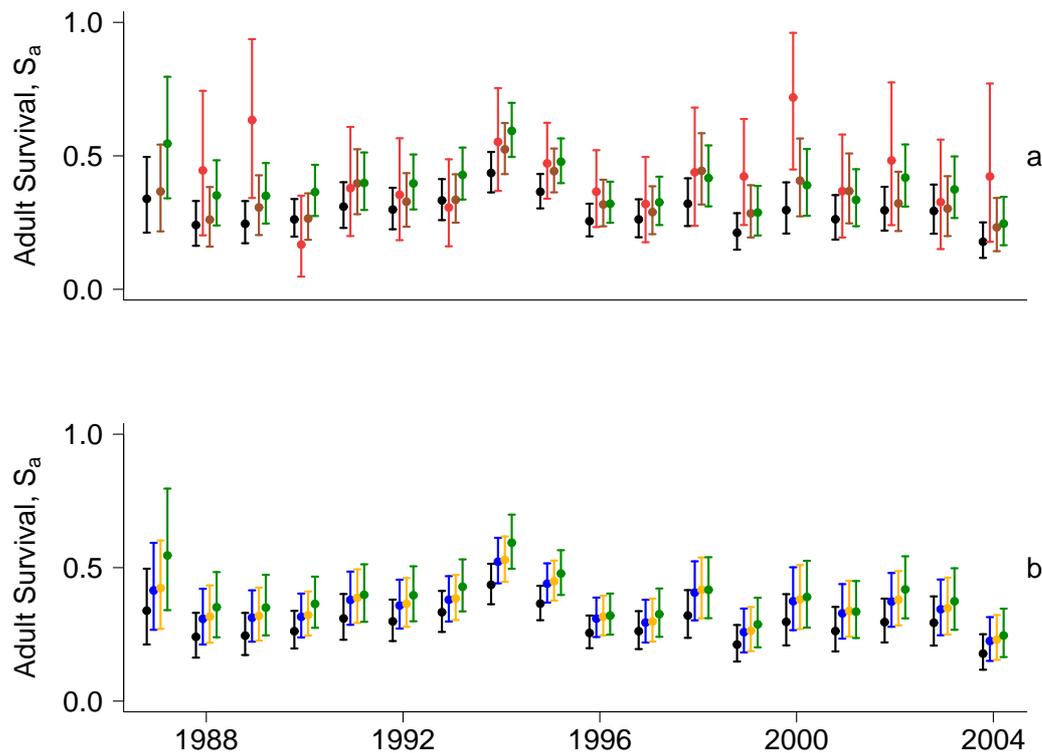
**Table 2.4** Adult Sedge Warbler “30 best” sites CES live-recapture data (1987-2005): number of unique adults caught, and the number and percentage of these adult birds that were recaptured in a subsequent year, caught more than once within their inaugural ringing year, and identified as residents by virtue of recapture either within or between years.

	Number	Percentage
Adults caught	6365	
Annual recaptures	638	10%
Within ringing year recaptures	2097	33%
Known residents	2358	37%

Models CJS, AH, EAH, and P were fitted using both classical and Bayesian techniques. Essentially identical results, in terms of estimation and precision, were obtained under both the classical and Bayesian approaches, so for clarity and brevity we present only the results from the Bayesian analysis.

Model P, being the standard approach to deal with transients in live-recapture data, is considered the “baseline” model from which survival estimates produced by the alternative models are compared. Further, comparisons with the estimates under model CJS, the “null” approach of simply ignoring the presence of transient birds in the data, enables the severity of the bias induced by transients to be assessed.

In Figure 2.4 the posterior means and 95% symmetric credible intervals for  $S_{a,t}$  under the alternative models are presented. As is to be expected, estimates of survival under the CJS model are lower than those under model P, for failing to account for transient birds in the data is known to induce negative bias into the estimates of  $S_{a,t}$  (Pradel et al., 1997). Nevertheless, there is a general overall agreement between the trend in survival estimated by the six models despite the inherent bias in some of the methods (i.e. CJS, AH, EAH). The baseline model P and the null model CJS are compared to the *ad hoc* methods (AH, EAH) in the top panel of Figure 2.4 (a), and to the theoretical approaches (EP, CD) in the bottom panel (b). We discuss these comparisons in turn.



**Figure 2.4** Posterior means and 95% symmetric credible intervals for  $S_{a,t}$  from the Sedge Warbler “30 best” sites CES data (1987-2005) under the alternative models. The baseline model P (right) and the null model CJS (left) are compared to a) the *ad hoc* methods (AH, EAH) and b) the theoretical approaches (EP, CD). black = CJS, red = AH, brown = EAH, green = P, blue = EP, and yellow = CD.

#### *Ad Hoc Methods, Figure 2.4a*

The *ad hoc* approaches, AH and EAH, both involve data reduction. Table 2.5 summarises the proportion of the full data utilised under these models. The improved precision in the estimation of  $S_{a,t}$  between the Cormack-Jolly-Seber based models, in ascending order, AH, EAH, and finally to CJS, is a by-product of the increasing amount of data employed by these models. In the standard *ad hoc* method (AH) the first annual capture is removed leaving only capture histories for 10% of the birds present in the full data set. On account of this vast data reduction estimates under model AH lack precision and deviate the most from the reasonably consistent survival trend estimated by the other models. In addition, we cannot estimate survival in the first study year, 1987. Estimates of survival obtained from model AH in several years (1988, 1989, 1999, 2000, 2002, 2004) seem to be positively biased. This maybe the result of discarding a disproportionate number of weaker resident birds that died shortly after ringing

in selecting the non-random sample of resident birds, caught in at least two years, contributing to the analysis.

The extended *ad hoc* approach (EAH) adopts a larger fraction of the data: 37% of the capture histories are retained. The use of within-year recaptures to identify residents is expected to reduce the positive bias in the estimates of  $S_{a,t}$  caused by the non-random selection of residents, as birds identified as such need not have survived an entire year. As the majority of birds (89%) included in the EAH data set were recaptured within their year of ringing, and as survival during the mist-netting season is believed to be high (Peach, 1993), the resulting bias is expected to be minimal. Indeed, model EAH seems to perform much better than model AH, with estimates of  $S_{a,t}$  consistent in scale and trend to those of the unbiased baseline model P. Further, there is not a substantial loss in precision and an estimate in 1987 is obtained.

**Table 2.5** Proportion of the adult Sedge Warbler “30 best” sites CES live-recapture data (1987-2005) utilised under the alternative survival models: the total number (and percentage) of birds caught, annual releases and annual recaptures.

Model	Birds Caught		Annual Releases		Annual Recaptures	
CJS, P, EP, CD	6365	(100%)	7180	(100%)	815	(100%)
EAH	2358	(37%)	2782	(39%)	554	(68%)
AH	638	(10%)	791	(11%)	177	(22%)

#### *Theoretical Approaches, Figure 2.4b*

The transient models, P, EP, and CD, so called as they explicitly account for transient birds in the data, provide a better, more rigorous, treatment than the *ad hoc* methods (AH and EAH) examined above. Reassuringly, estimates of survival obtained from models EP and CD, extensions of model P to utilise the within-year recapture information, are consistent with the baseline model P estimates, and with one another. As models EP and CD are essentially reparametrisations of each other, as is to be expected they provide nearly identical estimates of  $S_{a,t}$ . Indeed, when the extra Binomial term in the full likelihood for model EP was removed, estimates under models EP and CD were equivalent. Further, when the number of visits was constrained to 1, it was verified that both models EP and CD reduced to model P.

Model CJS is more parsimonious than any of the transient models (P, EP, CD) which require estimation of additional parameters, so although the estimates

of  $S_{a,t}$  under model CJS are negatively biased they are estimated with the greatest precision. Of the three transient models, model EP estimates  $S_{a,t}$  with the most precision, albeit only marginally more precisely than model CD - the average posterior standard deviation for  $S_{a,t}$ , over all years, for models P, EP, and CD are 0.056, 0.048, and 0.050 respectively. This improved precision is a consequence of model EP using the within-year recapture information in both ringing and non-ringing years. Model P is the least precise of the transient models as it does not utilise the within-year recapture information to identify residents, from the mixture of transients and residents in unknown proportion, in the sample of newly-marked birds (Hines et al., 2003).

Models P, EP, and CD provide estimates of the proportion of transients in the live-recapture data, a parameter of potential ecological interest. The posterior means, and 95% symmetric credible intervals for  $\tau$  (or  $1 - \varphi$  in the case of model CD) under these models are given in Table 2.6. As with  $S_{a,t}$ , and for the same reasons,  $\tau$  is estimated with the greatest precision under model EP, and the least precision under model P.

**Table 2.6** Posterior means and 95% symmetric credible intervals for  $\tau$  (under models P and EP) and  $1 - \varphi$  (under model CD) from the adult Sedge Warbler “30 best” sites data.

Model	Posterior mean	95% Credible Interval
P	0.496	(0.4187, 0.5629)
EP	0.386	(0.3551, 0.4140)
CD	0.414	(0.3806, 0.4446)

Models EP and CD also require the estimation of the site-specific nuisance parameters  $\varepsilon_g$  or  $c'_g$  respectively. These parameters are directly related via the expression  $1 - \varepsilon_g = c'_g$ , and a high correlation between the set of 30 posterior means for  $\varepsilon_g$  and  $c'_g$  was indeed observed ( $\rho = -0.99$ ). The extra Binomial term in the full likelihood for model EP translates into an increase in precision (the average posterior standard deviation, over all 30 sites, of  $\varepsilon_g$  and  $c'_g$  was 0.049 and 0.057, respectively). Further, we note that the estimated site-specific recapture probabilities,  $c_g$ , are in agreement between these two models ( $\rho = 1.00$ ).

#### *Between- and Within- Year Recapture Probabilities*

The between-year recapture probabilities,  $c_g$ , are clearly related to the measures of within-year recapture probabilities,  $\varepsilon_g$  and  $c'_g$ , in models EP and CD respectively. Intuitively, as the recapture probability increases, the probability a bird

is only encountered once in a year of capture decreases, or analogously, the more likely a bird is to be recaptured at a visit subsequent to ringing. Therefore, it is unsurprising that  $\varepsilon_g$  and  $c_g$  are negatively correlated ( $\rho = -0.46$ ), whereas  $c'_g$  and  $c_g$  are positively correlated ( $\rho = 0.49$ ).

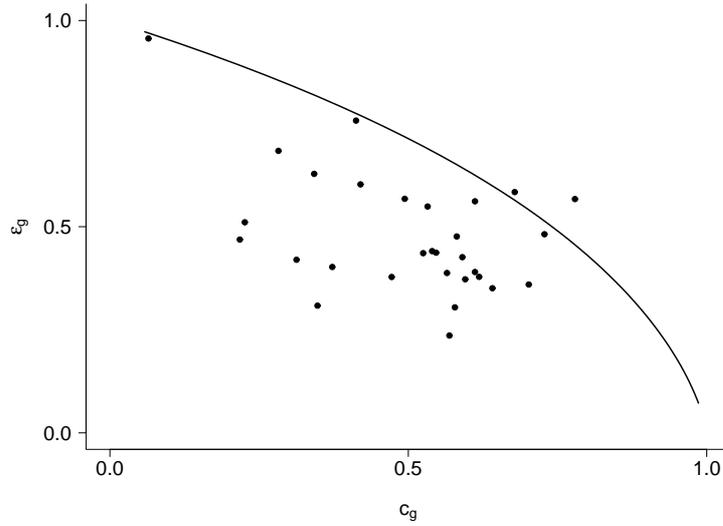
Recall that  $\varepsilon_g$  and  $c'_g$  are estimated independently of  $c_g$  in models EP and CD respectively. Freeman (2008) claimed that logic dictates that  $c'_g$  is less than  $c_g$  since the former quantity represents the probability of recapture in what remains of the mist-netting season after the initial capture, whilst the latter quantity corresponds to the recapture probability across the entire season. However, upon comparing the posterior means of  $c_g$  and  $c'_g$  for all 30 sites (the posterior distributions are symmetrical), only 11 times is  $c'_g < c_g$ . That  $c'_g$  is not strictly less than  $c_g$  implies that, for at least majority of sites, the probability of recapture within a year increases if the bird has already been caught that year, an apparent “trap-happy” response (Williams et al., 2002).

An apparent trap-happy response also occurs under model EP, as shown by comparing the estimated  $\varepsilon_g$  value to its expected value under the assumption of independent within-year recapture probabilities. Assuming, for simplicity, that within each year the recapture probability at visit  $k$ , for site  $g$  (denoted by  $v_{gk}$ ) is constant across all  $K = 12$  visits, then  $v_{gk} = 1 - (1 - c_g)^{1/K}$ . Under the assumption of independent within-year recaptures:

$$\begin{aligned} \mathbb{E}[\varepsilon] &= \Pr(\text{caught exactly once within a year} | \text{recaptured}) \\ &= \frac{\Pr(\text{caught exactly once within a year})}{\Pr(\text{recaptured})} \\ &= \frac{K v_{gk} (1 - v_{gk})^{K-1}}{c_g}. \end{aligned}$$

In Figure 2.5 the expected and estimated values of  $\varepsilon_g$  are plotted against  $c_g$  for all 30 sites. For the majority of sites the observed values of  $\varepsilon_g$ , estimated under model EP, are lower than their expected values. This indicates, as model CD did, that a larger number of birds are caught more than once within a year than is expected under the assumption of independent within-year recaptures.

Biologically, such a trap-happy response is unlikely as capture in a mist-net affords no benefits. This apparent positive dependence between the within-year recapture probabilities has an alternative, ecological explanation. Mist-nets are located at a fixed position within the site. Within any given year, those birds



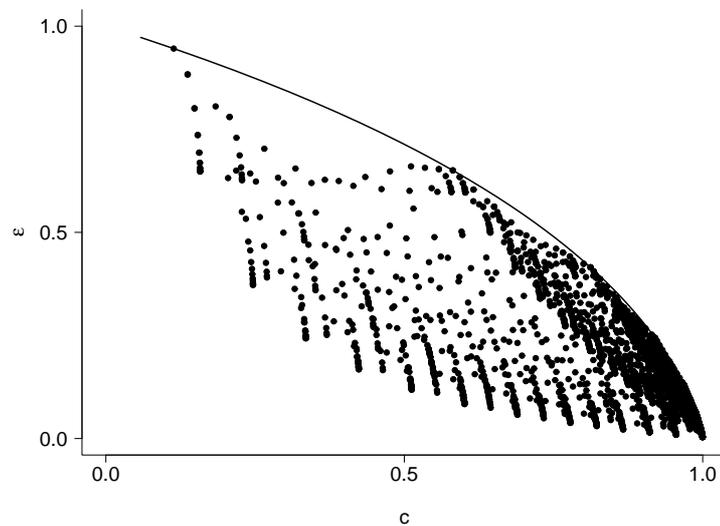
**Figure 2.5** Site-specific probability a bird is only caught at one visit during a year, given caught within that year ( $\varepsilon_g$ ), against the site-specific recapture probabilities ( $c_g$ ) under model EP. Expected values are represented by the solid line, estimated values, given by the posterior mean, are denoted by points.

with nesting sites and territories closer to the mist-net will have a higher capture rate than those further away (Baillie et al., 1986). This individual heterogeneity in capture rates may cause a “pseudo” trap-happy response - birds caught in a particular year have on average higher capture rates and thus are more likely to be seen multiple times within that year. In effect  $\varepsilon_g$ , or  $c'_g$ , adjusts for this individual heterogeneity by enabling birds closer to the mist-net to have a higher probability of being seen at repeated visits within the same year. Note that  $c_g$  is an average of the recapture probabilities of all birds at the CES site, and accommodates variations in an individual bird’s capture rate between years on account of a change in the precise location of its territory.

That heterogeneity in the recapture probabilities can cause a pseudo trap-happy response in  $\varepsilon_g$ , and by implication  $c'_g$ , is demonstrated in the following example. For a single site, let there be  $I$  groups with visit recapture probability  $v_i$  (constant across all  $K$  visits). Let  $p_i$  be the proportion of birds in group  $i$ . Under the assumption of independent within-year recaptures,  $c$  and  $\varepsilon$  are given by

$$c = \sum_{i=1}^I p_i (1 - (1 - v_i)^K) \quad \text{and} \quad \varepsilon = \frac{\sum_{i=1}^I p_i K v_i (1 - v_i)^{K-1}}{\sum_{i=1}^I p_i (1 - (1 - v_i)^K)}.$$

For a range of values of  $v_i$  and  $p_i$ , from  $I = 3$  groups,  $\varepsilon$  is plotted against  $c$  in Figure 2.6. Biologically this could represent a population of birds at a CES site in which one group of birds occupies territories that overlap the mist-net, a second with territories in close proximity to the mist-net, and a third with territories disjoint from the mist-net but whose feeding ranges encompass it. The solid line on Figure 2.6 corresponds to  $\varepsilon$  values when there is no heterogeneity in the recapture probabilities between the 3 groups. Figure 2.6 clearly demonstrates that heterogeneity in recapture rates can explain the discrepancy between the estimated and expected values of  $\varepsilon_g$ , for a fixed  $c_g$ , in Figure 2.5.



**Figure 2.6** Expected values of  $\varepsilon$ , under the assumption of independent within-year recapture probabilities, plotted against  $c$ . Points represent  $\varepsilon$  for a range of  $v_i$  and  $p_i$  values when there are three distinct groups with respect to probability of recapture. The solid line represents  $\varepsilon$  when there is no recapture heterogeneity between the three groups.

In reality the situation is more complicated than that presented in the simple example above. For instance, due to behavioural changes across the mist-netting season  $v$  is unlikely to be constant across all visits. The estimated values of  $\varepsilon_g$  (or  $c'_g$ ) reflect both the non-constant within-year recapture probabilities and heterogeneity between birds in recapture rates. Analogous exploratory studies have shown that allowing for temporal variation in  $v$  results in an increase in the expected value of  $\varepsilon_g$  relative to  $c_g$ .

Finally, we note that individual heterogeneity in recapture probabilities is a common feature of live-recapture data, but is not a detrimental issue in terms of our analyses as time-dependent survival estimates from Cormack-Jolly-Seber models are robust to such heterogeneity (Carothers, 1979).

*“Back of the Envelope” Calculations*

## PROPORTION OF MULTIPLE WITHIN-YEAR ENCOUNTERS

In the Sedge Warbler “30 best” sites data, 33% of the birds ringed were caught more than once within their ringing year (Table 2.4). Under model CD, the posterior mean of the probability of recapture within the ringing year,  $c'_g$ , across all 30 sites, is on average 0.551. The proportion of ringed birds in the data that are residents who survived the post-ringing period,  $\varphi$ , is estimated to be 0.587. Thus we would expect approximately  $(0.551 \times 0.587) \times 100\% = 32\%$  of birds to be caught more than once in their ringing year, in close agreement with the observed data. Similar verification calculations can be carried out for model EP. To be caught at only one visit within its ringing year the bird is either a transient or a resident that evaded future capture that year. Thus we would expect approximately  $(1 - \tau - (1 - \tau) \times \varepsilon) \times 100\%$  of birds to be caught more than once within their ringing year. Given that average of the posterior means of  $\varepsilon_g$ , across all 30 sites, is 0.481, and the posterior mean of  $\tau$  is 0.386, this equates to  $\approx 32\%$ , in agreement with the observed data.

## DISTRIBUTION OF WITHIN-YEAR TRANSIENT CAPTURES

Transient birds are caught as they pass through the CES site *en route* elsewhere. As Sedge Warblers migrate to their breeding grounds in early summer and depart for their pre-migratory fattening grounds at the end of summer, ecology suggests that transients should be caught by and large at the beginning and the end of the mist-netting season (Peach et al., 1990). In Table 2.7 the proportion of single capture birds seen at each visit is compared to the proportion of known residents captured at each visit. As is to be expected, a higher proportion of single capture birds (which are a mixture of residents and transients) are seen in the earlier and later visits than known residents. In particular, a far smaller proportion of known residents are caught in visit 1. This implies most transients are caught early in the summer, as ecology suggests.

**Table 2.7** Percentage of single capture and known resident adult Sedge Warblers caught at each of the 12 within-year CES visits.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>Single Capture Birds</b>	25.5	15.3	10.1	7.9	5.8	5.5	5.2	6.5	7.7	5.8	3.3	1.5
<b>Known Residents</b>	9.4	14.1	15.0	14.7	13.6	10.1	8.3	6.7	4.9	2.4	0.8	0.1

### 2.4.4 The Best Model for CES Live-Recapture Data

We have considered six different models for CES live-recapture data, five of which have been utilised to some extent previously, and one of which (model EP) is newly defined here. The attributes of these models are summarised in Table 2.8. With regards to providing precise and unbiased estimates of survival from CES data known to contain transients, our extended Pradel model (EP) is preferred. We now briefly explain why.

**Table 2.8** Attributes of the alternative adult survival models.

	CJS	AH	EAH	P	EP	CD
Readily implemented in Program MARK <sup>§</sup>	✓	✓	✓	✓	✗	✗
Unbiased estimates of survival	✗*	✗ <sup>†</sup>	✗ <sup>‡</sup>	✓	✓	✓
Estimates survival for the first study year	✓	✗	✓	✓	✓	✓
Estimates transience/residence	✗	✗	✗	✓	✓	✓
Adopts all annual capture histories	✓	✗	✗	✓	✓	✓
Uses within-year recaptures in ringing year	✗	✗	✓	✗	✓	✓
Uses within-year recaptures in non-ringing years	✗	✗	✗	✗	✓	✗

<sup>§</sup>At least for a small number of sites.

\*Negatively biased in the presence of transient birds.

<sup>†</sup>Positively biased if individual heterogeneity in survival.

<sup>‡</sup>Positively biased if individual heterogeneity in survival, but less so than AH.

Although models CJS, AH, and EAH provide quick and easy means of estimating  $S_{a,t}$ , these approaches are inherently biased, and do not provide an estimate of transience, which may be of interest in its own right. Further, the *ad hoc* approaches (AH and EAH) necessitate the removal of potentially a large fragment of data, increasing the uncertainty in the estimate of  $S_{a,t}$ , the ecologically interesting parameter.

Model P is a known and accepted method for modelling the captures of transients, but it does not fully utilise the ancillary information contained within the live-recapture data collected under the CES scheme to identify residents. In particular, a bird captured in a single year, but during multiple visits within that year, can be assumed to be a resident. Straight application of model P, as described by Pradel et al. (1997), does not accommodate this. We have extended this current methodology, formulating model EP, to use efficiently the additional information provided by the CES scheme with regards to residency. By utilising this extra information, improved precision in the estimate of  $S_{a,t}$  typically results. We note, however, the trade-off between the precision gained

from utilising the additional within-year recapture information, and the cost to precision from the estimation of extra site-specific nuisance parameters required to model it (i.e. parameters  $\varepsilon_g$  of model EP). Indeed, if there are fewer within-year recaptures, such that only a small number of resident birds are identified as such by virtue of recapture within their ringing year, model P may provide more precise inference.

Freeman (2008) provides an analogous extension to model P, model CD, which treats the previously-marked segment of the capture history in the same manner as model EP, but differs in the parametrisation used to model the newly-marked segment of the capture history. However, unlike model CD our extension (model EP) also utilises the information from within-year recaptures in non-ringing years, via the extra Binomial likelihood term, in the estimation of the parameters associated with the multiple within-year encounters ( $\varepsilon_g$  or  $c'_g$ ). The inclusion of the Binomial term in model EP leads to an improved precision in the estimation of these parameters, and consequently a more precise estimate of  $S_{a,t}$ . In the formulation of model CD, Freeman (2008) temporally orders the ringing and post-ringing period. As individual birds experience post-ringing periods of differing length, under the current formal definition of  $c'_g$  by Freeman (2008), it is unclear what the analogous term for model CD should be. If  $c'_g$  is assumed to be constant across all birds (which clearly it is not) then an identical Binomial likelihood term to that of model EP results.

Aside from the improved precision achieved by model EP compared to model CD, model EP is not subject to ambiguities in the definition of its parameters that the creation of the post-ringing period causes for model CD. To date the issue of differing post-ringing periods in the formal definition of  $c'_g$  (and  $S'_a$ ) has not been addressed. Obviously a bird ringed at an early visit should have a higher  $c'_g$  than a bird ringed near the end of the mist-netting season. The analogous parameter to  $c'_g$  in model EP,  $\varepsilon_g$ , is defined in a manner that bypasses the difficulty caused by the length of the post-ringing period, making it easier to justify convincingly, and therefore preferable. Moreover, model EP can be parsimoniously adjusted for missed visits, see Chapter 4.

Finally, we note that for Sedge Warbler CES data in particular, as the majority of known residents are identified by recapture within their ringing year, and as survival over the mist-netting season is expected to be high (Peach, 1993), model EAH appears to provide unbiased estimates of survival, albeit with typically less precision than model EP due to the necessary data reduction. Model EAH,

unlike model EP, can be fitted using the standard software, Program MARK. The ability of model EAH to provide quick and potentially unbiased estimates of survival for CES data is very appealing. However, this is true for only a limited number of sites; Program MARK does not readily handle the generally large number of sites in CES data very well.

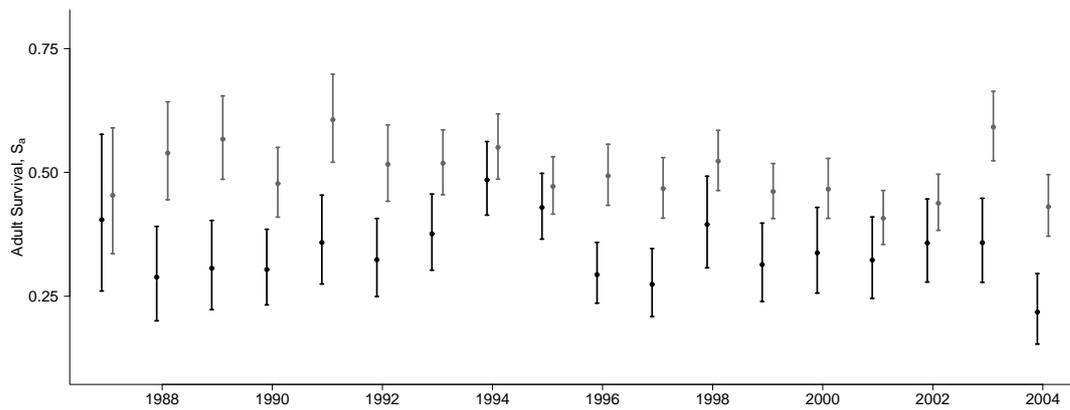
In conclusion, if one is most interested in estimating  $S_{a,t}$ , which is typically the case, then model EP is the best in terms of both accuracy and precision, given there are sufficient within-year recaptures. Estimates under this model are not prone to bias like models CJS, AH, and generally EAH, and furthermore as model EP utilises the full within- and between-year recapture information, its estimates of  $S_{a,t}$  are more precise than those of models P and CD.

#### 2.4.5 Analysis of Sedge and Reed Warbler Complete Coverage Data

Bayesian estimates of  $S_{a,t}$  are obtained from the Sedge and Reed Warbler complete coverage data, from the “best” sites, using the extended Pradel (EP) model, the preferred survival model for CES data, and presented in Figure 2.7. We observe that:

- a) Estimates of Reed Warbler survival are typically more precise than those of Sedge Warbler survival, as is to be expected on account of the greater amount of live-recapture data for Reed Warbler, see Table 2.1.
- b) There are some similarities in the estimated annual trend in survival for Sedge and Reed Warbler, for example the decrease in survival between 2003 and 2004.
- c) Survival rates for Reed Warbler are higher than those for Sedge Warbler.
- d) The estimates of survival presented here are consistent with those few from previously published research, see Table 2.9.

That there are some similarities in the annual trend of  $S_{a,t}$  between these two species suggests that they are subjected to some, but certainly not all, of the same environmental pressures. That the correlation in the posterior means is not stronger ( $\rho = 0.27$ ) is probably attributable to the different pre-migration and migration strategies employed by these two species. Migration is physically taxing and potentially hazardous (Bibby and Green, 1981). Death on migration is likely a major determinant of  $S_{a,t}$ . As Sedge and Reed Warbler undertake



**Figure 2.7** Estimates of adult survival from Sedge Warbler (left - black) and Reed Warbler (right - grey) “best” sites complete coverage data, 1987-2005. The points denote the posterior means and the vertical bars denote 95% symmetric credible intervals.

different migration strategies (Bibby and Green, 1981) different environmental factors will have different degrees of impact on their survival rates. For example, survival of adult Sedge Warblers is related to rainfall in the Sahel, their wintering quarters (Peach et al., 1991). No such association has been detected for Reed Warbler (Thaxter et al., 2006). Thaxter et al. (2006) suggest that Reed Warbler’s use of refuelling stops during migration renders them less susceptible to the environmental conditions in the Sahel. However, they note the speculative nature of this hypothesis since the wintering quarters of Reed Warbler are not precisely known.

The posterior means of  $S_{a,t}$ , for 1987 to 2004, range between 0.218 - 0.485 (mean = 0.342) and 0.408 - 0.606 (mean = 0.499) for Sedge and Reed Warbler respectively. These estimates are broadly consistent with the few published estimates that exist, see Table 2.9, despite these previous studies typically covering earlier time periods than CES data we consider, and generally being restricted to far smaller study areas within the UK.

That estimated survival rates of adult Reed Warbler are higher than those of adult Sedge Warbler is consistent with findings in other studies, for example Peach et al. (1990) and Peach (1993). It would seem likely that mortality during migration is higher for Sedge Warbler than Reed Warbler as they migrate in a single stage rather than in several shorter steps punctuated by refuelling stops as Reed Warbler do, however whether differential mortality on migration accounts for the lower Sedge Warbler survival rates is unknown (Bibby and Green, 1981).

**Table 2.9** Published estimates of Sedge and Reed adult Warbler survival rates.

	Sedge Warbler	Reed Warbler	Data
Peach et al. (1990)	0.477 (s.e. = 0.020)	0.534 (s.e. = 0.032)	Mark-recapture data collected at two southern England reedbed sites, one from 1969-1983, the other from 1967-1987.
Peach (1993)	0.182 - 0.440 (mean = 0.277)	0.465 - 0.530 (mean = 0.489)	CES live-recapture data from 4 sites, 1983 - 1991.
Siriwardena et al. (1998b)	0.284 (s.e. = 0.070)		Ring-recovery data, 1962 - 1995.
Freeman et al. (2005)		≈0.4 - 0.6 (mean ≈ 0.5)	CES live-recapture data from 80 sites, 1983 - 2003.
Thaxter et al. (2006)		female: 0.549 (s.e. = 0.103) male: 0.603 (s.e. = 0.060)	Mark-recapture data, collected at a northern England site from 1988 - 2004.
Thaxter et al. (2006)		female: 0.520 (s.e. = 0.224) male: 0.329 (s.e. = 0.160)	Mark-recapture data, collected at a southern England site from 1995 - 2004.

The proportion of transients in the data,  $\tau$ , also requires estimation. The posterior mean of  $\tau$  is slightly higher for the Reed Warbler data than the Sedge Warbler data, although the 95% credible intervals do, just, overlap (Table 2.10). For ecological reasons, the Reed Warbler live-recapture data may contain a larger proportion of transient birds. Reed Warblers return later from their wintering grounds and over a more protracted time period than Sedge Warblers (Bibby, 1978). Consequently, transient Reed Warblers are expected to be passing through CES sites for a greater proportion of the mist-netting season, and a larger number of transient captures may well result. We issue some caution regarding the interpretation of  $\tau$  as being the true proportion of transient birds in the data as  $\tau$  also contains a component due to survival within the mist-netting season. Conceivably one species could have a higher  $\tau$  than another simply due to a lower summer survival rate. However, we expect that the survival rates over the mist-netting season are high for adults of both species (Peach, 1993).

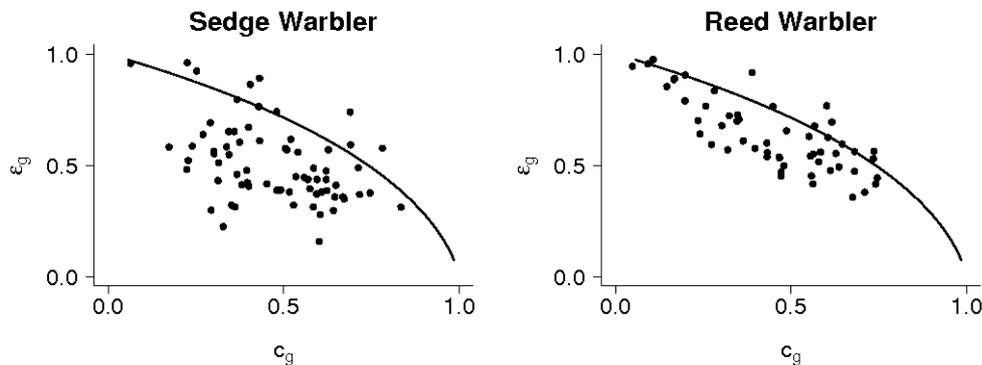
**Table 2.10** Posterior means and 95% symmetric credible intervals for  $\tau$ , under model EP, from the Sedge and Reed Warbler “best” sites complete coverage data.

	Posterior mean	95% Credible Interval
<b>Sedge Warbler</b>	0.370	(0.3427,0.3968)
<b>Reed Warbler</b>	0.411	(0.3924,0.4299)

Sedge and Reed Warbler “best” sites complete coverage data sets have 52 sites in common. A comparison of the site-specific parameters,  $c_g$  and  $\varepsilon_g$ , at these common sites does not reveal an association between the two species. The correlations in the posterior means are  $\rho = -0.10$  for  $c_g$ , and  $\rho = -0.04$  for  $\varepsilon_g$ . Therefore a site in which Reed Warblers are highly catchable does not mean that Sedge Warblers will also be highly catchable, for example. The analysis does however suggest that on average Sedge and Reed Warblers are similarly catchable (the average of the posterior means of  $c_g$  across these 52 common sites is 0.453 and 0.463 for Sedge and Reed Warblers respectively).

For both Sedge and Reed Warbler data,  $\varepsilon_g$  is, as expected, negatively correlated with  $c_g$  ( $\rho = -0.44$  for Sedge Warbler and  $\rho = -0.78$  for Reed Warbler). The analysis of the “30 best” site Sedge Warbler data in Section 2.4.3 revealed that  $\varepsilon_g$  could not be derived from  $c_g$  under the assumption of independent within-year recaptures. Rather,  $\varepsilon_g$  was lower than expected indicating that multiple within-year recaptures were more frequent than suggested by  $c_g$ . We hypothesised that this was a “pseudo” trap-happy response due to individual

heterogeneity in capture rates. Analysis of the entire “best” sites Sedge Warbler data set also gives rise to a “pseudo” trap-happy response, as do the Reed Warbler data (Figure 2.8). We note though that the estimated  $\varepsilon_g$  values for Reed Warbler are in closer agreement with the expected values than those of Sedge Warbler, and are much more strongly correlated with  $c_g$ . Under the individual heterogeneity hypothesis this would suggest that recapture heterogeneity may be less for Reed Warbler. Nevertheless, as time-dependent survival estimates from Cormack-Jolly-Seber models are robust to individual heterogeneity in recapture probability (Carothers, 1979), its presence in either the Sedge or Reed Warbler data is of little concern.



**Figure 2.8** Site-specific probability a bird is only caught at one visit during a year, given caught within that year ( $\varepsilon_g$ ), against the site-specific recapture probabilities ( $c_g$ ) under model EP. Expected values assuming homogeneity are represented by the solid line; the observed estimates, given by the posterior means, are denoted by points and calculated from the Sedge Warbler (left) and Reed Warbler (right) complete coverage data.

Note, if for another species one should find a close relationship between the expected and estimated values of  $\varepsilon_g$ ,  $\varepsilon_g$  could be constrained to be a function of  $c_g$  reducing the number of parameters in the model.

As is to be expected, the analysis of the reduced Sedge Warbler data from only the “30 best” sites (Section 2.4.3) gave very similar results, albeit less precise, to that where data from all 71 “best” sites are used.

## 2.4.6 Validity of Model Assumptions

### *Time-Invariant Recapture Probability, $c_g$*

Under the “constant effort” protocol, underpinning the CES scheme, recapture probability,  $c_g$ , at a given site is assumed to be constant across time. Such

a constraint on  $c$  is desirable - reducing the number of parameters improves the precision in survival estimates (Peach et al., 1990). Previous analyses of CES live-recapture data have provided support for this assumption, for example Pratt and Peach (1991); Peach (1993); Peach et al. (1995a,b). We provide further evidence.

Using Program MARK a series of Cormack-Jolly-Seber models with  $S_a$  time dependent and  $c$  either constant, time dependent, site dependent, or site and time dependent were fitted to Sedge and Reed Warbler “30 best” sites data. The AICs were used to distinguish between these competing models. For both species,  $c$  in the chosen model, the one with the lowest AIC, was time-invariant but site dependent ( $\Delta\text{AIC} > 50$ ).

Secondly, a Bayesian validation study was conducted using Sedge and Reed Warbler complete coverage data from all the “best” sites. Here the recapture probability at site  $g$  in year  $t$ ,  $c_{gt}$ , in the Cormack-Jolly-Seber model was expressed as a linear combination, on the logit scale, of an intercept term,  $\beta$ , plus site- and year-specific effects,  $s_g$  and  $y_t$  respectively, i.e.  $\text{logit}(c_{gt}) = \beta + s_g + y_t$ . For identifiability an arbitrary site and year effect are constrained to zero. If  $c$  is time-invariant then  $y_t = 0$  for all  $t$ ; if  $c$  is constant across all sites then  $s_g = 0$  for all  $g$ . Table 2.11 summarises the proportion of times the 95% symmetric credible intervals for  $s_g$  and  $y_t$  encompass 0. For both Sedge and Reed Warbler there is no evidence of time dependence in  $c$  but some evidence that  $c$  varies between sites.

**Table 2.11** Assessing time and site dependence of  $c$  in Sedge and Reed Warbler live-recapture data (1987-2005): the number (and percentage) of times the 95% symmetric credible interval for  $s_g$  and  $y_t$  includes 0. Note that  $S_{a,t}$  and  $y_t$  in the final year are confounded.

	$s_g$	$y_t$
<b>Sedge Warbler</b>	49/70 (70%)	15/16 (94%)
<b>Reed Warbler</b>	12/54 (22%)	16/16 (100%)

The use of credible intervals to detect time- or site-dependence is not statistically rigorous. Furthermore, we note that the conclusions drawn above are dependent on which site and year effects are constrained to zero. More formal alternatives to model selection in a Bayesian framework exist, for example the deviance information criterion (DIC), which is the Bayesian equivalent to the AIC (Gelman et al., 2004). In addition Reversible-jump MCMC (Green, 1995), a type of MCMC sampler that allows for dimensional changes in the probability

distribution being simulated, can also be used for model selection (see Chapter 5).

*Site-Invariant Survival Rates,  $S_{a,t}$*

A general feature of CES data is the sparsity of between-year recaptures at individual sites. Such sparsity results in a lack of precision in site-specific estimates of survival (Peach, 1993). Furthermore, as we are predominantly interested in the survival of the wider population of birds across the UK, rather than at individual sites, the models have adopted a common  $S_{a,t}$  for all sites. A study of CES live-recapture data, including those for Sedge and Reed Warbler, by Peach (1993) provides support for this assumption.

*Constant Transient Probability,  $\tau$*

In all analyses  $\tau$  was assumed to be site and time invariant, reducing the number of parameters requiring estimation. Violations of this assumption are conceivable, for example the location of sites relative to the migration path and/or the pre-migratory feeding grounds may cause site dependence in  $\tau$ , whereas a change in time of spring or autumn migration may cause time dependence in  $\tau$ . Model EP was extended to allow differences in  $\tau$  either between sites ( $\tau_g$ ) or between years ( $\tau_t$ ). The resulting parameter estimates, for both the Sedge and Reed Warbler complete coverage data, provided no strong evidence of such dependence (Table 2.12). The 95% symmetric credible intervals for  $\tau_g$  overlapped those of  $\tau$  (under the constant model), and encompassed its posterior mean, for nearly all the sites. Likewise, the 95% symmetric credible intervals for  $\tau_t$  overlapped and encompassed  $\tau$ , for most years. Importantly, estimates of  $S_{a,t}$  were virtually unchanged when allowing for site or time dependence in  $\tau$ .

**Table 2.12** Assessing time and site dependence of  $\tau$  in Sedge and Reed Warbler live-recapture data (1987-2005): the number (and percentage) of times the 95% symmetric credible interval (CI) for  $\tau_g$  and  $\tau_t$  overlap those of  $\tau$  (derived under the assumption of time and site invariance) and encompass the posterior mean (PM) of  $\tau$ .

		$\tau_g$	$\tau_t$
<b>Sedge Warbler</b>	<i>overlapping 95% CIs</i>	64 (90%)	17 (89%)
	<i>contains PM</i>	58 (82%)	15 (79%)
<b>Reed Warbler</b>	<i>overlapping 95% CIs</i>	44 (80%)	17 (89%)
	<i>contains PM</i>	44 (80%)	16 (84%)

An alternative approach for allowing temporal variability or site-dependence in  $\tau$ , not explored here, is the random effects model. For example, to allow for time invariance we assume that the annual estimates of  $\tau$  originate from a random process with a common mean and a temporal variance. That is, we set

$$\text{logit}(\tau_t) = \beta_\tau + \epsilon_t, \quad \epsilon_t \sim \text{Normal}(0, \sigma_\tau^2),$$

where  $\beta_\tau$  is the mean and  $\sigma_\tau^2$  is the temporal variance of the logit of  $\tau_t$ . Likewise, to allow site-dependence we assume

$$\text{logit}(\tau_g) = \beta_\tau + \epsilon_g, \quad \epsilon_g \sim \text{Normal}(0, \sigma_\tau^2),$$

where  $\beta_\tau$  is the mean and  $\sigma_\tau^2$  is the variance of the logit of  $\tau_g$ .

### *Sex Effects*

Typically in the analysis of CES data models with sex effects are not considered (for example Peach et al., 1990; Peach, 1993; Freeman, 2008), however Thaxter et al. (2006) did detect a difference in Reed Warbler survival between male and female birds at a single site (see Table 2.9). Approximately two-thirds of the Sedge and Reed Warblers caught have their sex recorded. Model EP was fitted to data from known female and known male birds separately. The resultant estimates of  $S_{a,t}$ ,  $\tau$ ,  $\epsilon_g$ , and  $c_g$  were in very close agreement between these models suggesting that sex is not important.

### *Use of Within-Year Recaptures to Identify Residents*

Crucial to modelling transience is the length-of-stay at a CES site of transient birds (Pradel et al., 1997). The assumption has been made that transients are only ever caught once, either within or between years. Were transients caught multiple times the estimates of  $S_{a,t}$  would be negatively biased, for such birds would appear as residents (Pradel et al., 1997; Hines et al., 2003). As CES visits are at least 3 days apart, a transient is likely to have departed the site before future mist-netting is conducted. Furthermore, even if a bird migrated to its breeding site via a particular CES site annually, it is unlikely to be caught in multiple years as 1)  $c$  and  $S_{a,t}$  are low, 2) its presence at the site is short, and 3) mist-netting is not continuous over the entire breeding season.

### 2.4.7 Comment on Robust Design Models

The CES scheme has a “Robust” style of design in the spirit of Pollock (1982): the “primary sampling periods”, years, contain 12 “secondary sampling sessions”, visits. Hines et al. (2003) propose a “likelihood approach” to analyse Robust design data, from the North American Monitoring Avian Productivity and Survivorship Program (MAPS), known to contain transient birds. Their likelihood approach explicitly models the capture histories both within and between years, necessitating the estimation of extra nuisance parameters corresponding to recapture and transient probabilities for each within-year visit. The model assumes the resident population is closed over all within-year visits, requiring annual sampling to begin after breeding commences and to end before birds began departing the breeding grounds (Hines et al., 2003). For CES data the closure assumption is clearly violated (although it may be a little better for non-migratory species): mist-netting spans both the beginning and end of the breeding season, and moreover individual birds neither arrive at, or leave from, the CES site at the same time.

Schwarz and Stobo (1997) and Kendall and Bjorkland (2001) have developed Robust design models with open secondary sessions that enable animals to arrive and depart the study site at different times during the secondary sampling session, as is typical for birds in the CES data. Hines et al. (2003) believe that these models can be adapted to deal with transients in a similar fashion to the Robust design model with closed secondary sampling sessions following their “likelihood approach”. They note that such a model would require the estimation of even more nuisance parameters - parameters associated with the probability of entry into the study site and apparent survival in the secondary period. Hines et al. (2003) present no details for this model as they doubt it can provide usefully precise estimates of survival on account of the vast number of additional nuisance parameters. This is certainly also true for CES data.

As an alternative, Hines et al. (2003) propose an “*ad hoc*” approach for incorporating the ancillary information about residence status resulting from the Robust design into the Pradel et al. (1997) transient model. Their approach, although using the within-year recaptures to identify residents, makes no attempt to model them. Here, birds caught multiple times within their inaugural ringing year contribute their entire (annual) capture history to the previously-marked m-array of the Pradel et al. (1997) model (see Section 2.4.1.iii). Such a model is readily fitted in Program MARK, albeit with the familiar implementation issues

stemming from the large number of sites. Clavel et al. (2008) have successfully used this model to analysis Reed Warbler CES style live-recapture data, albeit with an aim to estimate population size, not annual survival rates.

Model EP can be thought of as a compromise between Hines et al.'s (2003) *ad hoc* and likelihood-based approaches. Model EP explicitly models the within-year recaptures, at the cost of only  $G$  extra site-specific “evasion” parameters ( $\varepsilon_g$ ), but based on summary statistics - caught more than once within an encounter year or not. This negates the need to model the full within-year and between-year capture history as the likelihood approach does, substantially reducing the number of additional nuisance parameters requiring estimation.

### 2.4.8 Concluding Remarks

CES live-recapture data provide a valuable source of information on adult survival for songbird species, such as Sedge and Reed Warbler, in which this is otherwise problematical due to the few recoveries of ringed birds found dead (Baillie et al., 1986; Peach et al., 1990; Freeman, 2008). Adult Sedge and Reed Warblers are highly site faithful, although being migratory, birds passing through the site (transients) are occasionally caught. Failure to adequately account for the presence of transient birds in the CES data will result in negatively biased estimates of survival (Pradel et al., 1997).

Fitting Cormack-Jolly-Seber models, Peach (1993) made the first attempt to estimate adult survival from CES live-recapture data. Peach did not explicitly model transience, but assessed its effect by endeavouring to separate newly-marked birds into residents and transients on the basis of recapture at least 10 days after ringing. We have improved upon this initial approach, developing model EP, to estimate the proportion of transients in the data, making use of specific advantages in this respect provided by the CES scheme. Of the six alternative models described and compared in this chapter, model EP, an extension to the standard transient model of Pradel et al. (1997) which uses the ancillary within-year recapture information to identify residents, appears to be the best in terms of providing unbiased estimates of the biologically interesting parameter,  $S_{a,t}$ , that have the greatest precision.

The utility of model EP is not restricted to CES data alone. It can be applied to other “Robust design” (Pollock, 1982) style live-recapture data, providing an

efficient model. Indeed the Robust design is a feature common to many other bird monitoring schemes (Hines et al., 2003), and to mark-recapture studies in general (Kendall, 2001). By allowing temporal variation in  $c$  and  $\varepsilon$  model EP generalises to any two stage live-recapture study where transients are believed to be present, and inference on the primary sampling period is of interest

We have employed Bayesian methods to estimate the parameters of model EP. The use of Bayesian methodologies to estimate survival from mark-recapture data is not new (for example Brooks et al., 2000a; Poole, 2002; King and Brooks, 2004; King et al., 2008a). The advantage of such an approach, over the classical alternative, has been summarised by Brooks et al. (2000a). For this particular application the ease with which estimates of precision are obtained is a distinct advantage. Other advantages, although not exploited here, include the facility to readily incorporate additional information on  $S_{a,t}$  from other sources, such as information from the BTO's National Ring-Recovery (NRR) database or expert knowledge, by using an informative prior, and the ability to update the posterior distributions as more years of CES live-recapture data become available.

Currently the analysis has been restricted to the subset of data with complete coverage (i.e. data from years with the full complement of 12 visits). The ability to improve precision by including additional data from years with incomplete coverage is appealing. However, for these years the assumption of constant effort is violated, meaning that the recapture ( $c$ ) and evasion ( $\varepsilon$ ) probabilities (at a site) can no longer be considered time-invariant. In Chapter 4 model EP is parsimoniously extended to adjust  $c$  and  $\varepsilon$  for missing visits.

## 2.5 Chapter Summary

In this chapter independent Bayesian models for estimating indices of abundance and productivity from CES count data, and probabilities of adult survival from CES live-recapture data, have been described in detail. We have proposed a reparametrisation of the productivity model currently employed routinely by the BTO, that produces analogous indices of productivity but is more amenable to an integrated approach to population modelling. The standard Pradel et al. (1997) model, that accounts for transient birds in the live-recapture data, has been efficiently extended, making use of specific advantages in this respect provided by the design of the CES scheme, namely the observed within-year recaptures. These models have been successfully applied to Sedge Warbler and Reed Warbler complete coverage data. In Chapter 4 the models developed in this chapter are adapted to enable years with incomplete coverage to contribute to the analysis.

Biologically, abundance is related to the underlying demographic rates, productivity and survival. The simultaneous estimation of these parameters, via an integrated population model, to produce parameter estimates that are consistent with both the data, and according to the underlying population dynamics, with one another, is appealing. Further, as integrated modelling shares information between the component models, improved precision in estimation may result (Schaub et al., 2007; Besbeas et al., 2002). Building an integrated model, appropriate for CES data, is the focus of the following chapter.



# Chapter 3

## Integrated Population Model for Complete Coverage Data

### 3.1 Integrated Population Modelling

Projects for monitoring the status of wild bird populations at a national (or even higher) level are becoming increasingly common. Examples include the North American Breeding Bird Survey (Sauer et al., 1997) run by the Patuxent Wildlife Research Centre, the Pan European Common Bird Monitoring Scheme (Van Strien et al., 2001) established by the European Bird Census Council, and in the UK the BTO's Nest Record Scheme (NRS), Breeding Bird Survey (BBS), Waterways Bird Survey (WBS), and CES scheme (Baillie et al., 2009b). Such projects are generally orientated towards gathering data to estimate either indices of abundance or various demographic parameters (survival, or components of the production of young).

Typically, indices of abundance are derived from annual counts at a series of sites (Underhill and Prÿs-Jones, 1994; Peach et al., 1998; Fewster et al., 2000), survival estimates from ringing data (Peach et al., 1990; Siriwardena et al., 1999; Brooks et al., 2000a), and estimates of breeding success (at least on a per nesting attempt basis) from nest record data (Bibby, 1978; Crick et al., 2003; Freeman et al., 2007b). However, of increasing interest is the relationship between indices of abundance and the underlying demographic rates, an understanding of which facilitates effective conservation (Freeman and Crick, 2003). This relationship is usually expressed mathematically in the form of a matrix-based population model (Caswell, 2001), which serves to relate the abundance in a given year as a

function of the abundance in the previous year(s) and the intermediate survival and reproductive rates.

Many papers have now appeared in which abundance indices and demographic parameters are simultaneously estimated using data from two or more different sources (for example, Besbeas et al., 2002, 2003; Brooks et al., 2004; Besbeas and Freeman, 2006; Schaub et al., 2007; King et al., 2008b; Borysiewicz et al., 2009; Reynolds et al., 2009). Such analyses, which combine different sources of data, are typically referred to as “Integrated Population Models”. By combining information from various studies in a coherent manner, parameter estimates from integrated population models are more robust (Brooks et al., 2004) and typically more precise (Besbeas et al., 2002). An integrated analysis may also be capable of estimating parameters not obtainable from the independent analyses of the component data sets alone (Besbeas et al., 2005). Furthermore, integrated approaches to population monitoring have proved useful in identifying the likely demographic mechanism driving observed changes in population size (Siriwardena et al., 2001; Besbeas et al., 2002; Freeman et al., 2007b).

In this chapter we develop the first integrated population model to adopt information from the BTO’s CES scheme. Data collected under the CES scheme are particularly amenable to an integrated approach to population modelling as they provide information on three key demographic parameters: abundance, productivity and adult survival. Despite the CES scheme being especially important for the routine monitoring of abundance and productivity of songbird populations in the UK, the analyses of these two related demographic components are currently conducted independently (see Baillie et al., 2009b).

In Chapter 2, efficient, but independent, Bayesian models that adopt CES data with complete coverage were presented and/or developed for each of the three inter-related demographic components: abundance, productivity and adult survival. Building upon these methods, we formulate an integrated population model appropriate for CES data by specifying a meaningful functional relationship between these demographic parameters in the form of a deterministic population model uniting abundance, productivity and survival. This integrated approach has the advantage, over the independent analyses, of producing “robust” parameter estimates that are consistent with both the data, and according to the relationship imposed by the underlying population dynamics, with one another. Further, through the sharing of information across models, more precise inference is achieved.

Our integrated model is applied initially to CES complete coverage data for Sedge Warblers and Reed Warblers. Throughout, we assume that the population monitored is closed, with all transient movements between breeding sites occurring within the UK. Further, as adult songbirds are known to exhibit very strong breeding philopatry (Wernham et al., 2002), total site-faithfulness is assumed, an assumption later relaxed and assessed in Chapter 5.

## 3.2 An Integrated Model for CES Data

We are interested in making inference regarding the population of adult birds breeding in the UK and how their abundance is influenced by changes in the numbers of juveniles they produce (i.e. productivity) and the survival rates of the adult and juvenile birds. Strictly, the CES scheme provides data on adult and juvenile birds at the sample of sites surveyed; however these can be assumed to be representative of the wider UK population (Peach et al., 1991).

In Section 2.2 Bayesian models indexing adult abundance and juvenile abundance, expressed in Equations (2.1) and (2.2) respectively, were presented. These models, which adopt CES data with complete coverage, assume that the adult/juvenile yearly-site counts are independent, random variables from a Poisson distribution whose mean is a log-linear combination of the adult/juvenile specific intercept, site-effect and year-effect terms. Indices of adult and juvenile abundance in year  $t$ , denoted by  $A_t$  and  $J_t$  respectively, are given by the exponential of the appropriate year effect. Note that these indices are unbiased by the presence of transient birds (migrants not residents near the site of capture) in the count data provided that the proportion of transients in the catch remains constant over time.

The adult and juvenile abundance parameters in the Poisson models (Equations (2.1) and (2.2)) are estimated independently. The integrated approach involves creating a meaningful functional relationship between these parameters and, assuming independence, multiplying the adult abundance likelihood ( $L_a$ ) with the juvenile abundance likelihood ( $L_j$ ) to form a single joint likelihood.

For both Sedge Warblers and Reed Warblers we assume that all birds breed after their first winter, aged around one year, and every year thereafter (Cramp, 1992). Furthermore, we assume that there is no net immigration or emigration from the UK population (cf Besbeas et al., 2002, 2003; Brooks et al., 2004).

Such an assumption is reasonable as although birds are mobile organisms, it is rare for an adult bird once found breeding in the UK to subsequently breed abroad, and for juveniles to disperse abroad (Wernham et al., 2002). Let  $\phi_{a,t}$  and  $\phi_{j,t}$  denote the probabilities of *true* survival for an adult and a juvenile bird from year  $t$  to  $t + 1$  respectively. Recalling that  $k$  denotes an unknown positive scaling factor that relates the ratio of juvenile to adult indices of abundance to the true value, we define the deterministic population model for the index of adult abundance:

$$\begin{aligned} A_t &= A_{t-1}\phi_{a,t-1} + kJ_{t-1}\phi_{j,t-1} \\ &= A_{t-1}\phi_{a,t-1} + kP_{t-1}A_{t-1}\phi_{j,t-1}, \end{aligned} \quad (3.1)$$

noting that the index of juvenile abundance for year  $t - 1$  is a function of the indices of adult abundance and productivity in that year, i.e  $J_{t-1} = P_{t-1}A_{t-1}$  (as given in Equation (2.4)).

Equation (3.1) expresses the abundance index of adult birds in year  $t$ , presumed to be breeding, as a function of the abundance index, productivity index and survival probabilities of juvenile and adult birds in the previous year (see Siriwardena et al., 2001; Freeman and Crick, 2003), thereby linking the abundance and demographic parameters of the population under study.

Recursively:

$$A_t = g(A_1, k, \mathbf{J}_{1..t-1}, \boldsymbol{\phi}_{a,1..t-1}, \boldsymbol{\phi}_{j,1..t-1}). \quad (3.2)$$

Substituting Equation (3.2) into the adult abundance likelihood,  $L_a$ , we write a joint likelihood for the adult and juvenile yearly-site counts as:

$$\begin{aligned} L_{joint} &= L_a \times L_j \\ &= f(\mathbf{n}^a | \mathbf{A}_{1..T}, \beta^a, \mathbf{s}_{1..G}^a) f(\mathbf{n}^j | \mathbf{J}_{1..T}, \beta^j, \mathbf{s}_{1..G}^j) \\ &= f(\mathbf{n}^a | A_1, k, \mathbf{J}_{1..T-1}, \boldsymbol{\phi}_{a,1..T-1}, \boldsymbol{\phi}_{j,1..T-1}, \beta^a, \mathbf{s}_{1..G}^a) f(\mathbf{n}^j | \mathbf{J}_{1..T}, \beta^j, \mathbf{s}_{1..G}^j), \end{aligned} \quad (3.3)$$

where  $T$  and  $G$  denote the number of study years and sites respectively.

There is no direct information in either the juvenile or adult CES count data to estimate the productivity constant  $k$ , the scalar relating  $J_t/A_t$  to the ratio of true juvenile to adult abundances; instead it is estimated indirectly via Equation (3.3) and its role in the deterministic population model, Equation (3.1).

Fully time-varying demographic parameters clearly cannot be estimated from Equation (3.3) alone as it is over-parametrised. Thus we consider additional

data, which provide direct information on rates of survival over time, augmenting the joint likelihood given by Equation (3.3) with further likelihoods for data on live-recaptures of birds during the CES itself ( $L_{live}$ ) or recoveries after death of largely different birds from the BTO's National Ring-Recovery (NRR) database ( $L_{dead}$ ). We consider these additional CES live-recapture and ring-recovery data sets in turn.

### 3.2.1 CES Live-Recapture Data

Adult survival rates can be estimated from CES live-recapture data, which consist of live resighting histories for individual adult birds (see Section 2.4). However, transient birds must be accounted for otherwise survival estimates are negatively biased. To this end an extension to the Pradel et al. (1997) model, that estimates the proportion of transient birds in the data by making use of the specific advantage in this respect provided by CES scheme, namely the within-year recaptures, was developed in Chapter 2. This model is referred to as the extended Pradel model.

Recall that estimates of adult survival derived from CES live-recapture data will underestimate “true” survival ( $\phi_a$ ) if there is permanent emigration of birds away from the CES site (Pratt and Peach, 1991; Peach, 1993). Rather, estimates of “apparent” survival ( $S_a$ ) are obtained, which in addition to mortality also incorporates a component due to permanent emigration (Peach et al., 1995b; Thaxter et al., 2006). As adult Sedge and Reed Warblers are highly site faithful (Catchpole, 1974; Peach, 2002; Wernham et al., 2002), we assume for the moment that all losses at a site can be attributed to mortality, and set  $\phi_{a,t} = S_{a,t}$ . This assumption is later re-visited, and assessed, in Chapter 5. Setting  $\phi_{a,t} = S_{a,t}$  in the likelihood for the extended Pradel model, given by Equation (2.13), the likelihood for the CES live-recapture data becomes:

$$L_{live} = f(\mathbf{F}, \mathbf{X} | \tau, \boldsymbol{\varepsilon}_{1..G}, \mathbf{c}_{1..G}, \boldsymbol{\phi}_{a,1..T-1}) \times f(\mathbf{R}, \mathbf{Z} | \mathbf{c}_{1..G}, \boldsymbol{\phi}_{a,1..T-1}) \\ \times f(\mathbf{M}' | \boldsymbol{\varepsilon}_{1..G}, \mathbf{M}' + \mathbf{M}''). \quad (3.4)$$

### 3.2.2 Ring-Recovery Data

The BTO co-ordinates the British and Irish ringing scheme under which approximately 800,000 birds are ringed annually by over 2,000 trained volunteer

ringers (Coiffait et al., 2008). Each uniquely numbered leg-ring bears a return address enabling members of the public to report recoveries of birds found dead, and the identity of the dead bird to be reliably determined in the NRR database.

Models for recoveries of dead birds have a long history (Seber, 1971; North and Morgan, 1979; Brownie et al., 1985; Burnham, 1990). Over a set of study years, ring-recovery data typically consist of the number of birds ringed each year (a cohort) and for each ringed cohort, the corresponding number of rings recovered from birds found dead each recovery year. The likelihood for ring-recovery data in two age classes (juveniles and adults) is well documented (Freeman and Morgan, 1992; Vounatsou and Smith, 1995; Brooks et al., 2000a; King et al., 2009). Here, annual totals of recoveries in year  $t$  from each ringed cohort  $i$ , separately for birds ringed as adults and juveniles, denoted by  $D_{it}^a$  and  $D_{it}^j$  respectively, are assumed to form a collection of independent Multinomial distributions. However, as is frequently the case in European bird studies, the number of Sedge Warblers and Reed Warblers ringed in the BTO's NRR database are not available (North, 1990; Mazzetta, 2009) so we adopt a parameter-reduced form, conditional upon the numbers of birds recovered (North and Morgan, 1979; Burnham, 1990; Mazzetta, 2009).

Conditional on recovery, sometime during the study, the probability that a bird ringed in year  $i$  is recovered in year  $t$  is given by:

$$\frac{\Pr(\text{a bird ringed in year } i \text{ is recovered in year } t)}{\Pr(\text{a bird ringed in year } i \text{ is recovered during the study})}$$

For example, given  $T = 3$  recovery years the conditional probability associated with a juvenile ringed in year 2 being recovered (as an adult) in year 3 is:

$$\frac{\phi_{j,2}(1 - \phi_{a,3})\lambda_{a,3}}{(1 - \phi_{j,2})\lambda_{j,2} + \phi_{j,2}(1 - \phi_{a,3})\lambda_{a,3}} \quad (3.5)$$

where  $\lambda_{j,t}$  and  $\lambda_{a,t}$  are the recovery probabilities for juvenile and adult birds in year  $t$  respectively, i.e. the probability a ringed juvenile/adult bird that *has* died in year  $t$  is reported dead.

To estimate survival under the conditional ring-recovery model, where ringing totals are unknown, it is necessary to assume that recovery rates ( $\lambda$ ) are constant across time and age groups within each cohort. Note that although recovery probabilities of passerines are widely believed to have declined over time (Baillie and Green, 1987), as both Sedge and Reed Warblers are short-lived bird species

those ringed in the early years will have died long before the lower rates applicable to recent cohorts become prevalent, for example. Cohort-specific variation in  $\lambda$  is still accounted for; this may be realistic if, for example, there has been a geographic shift over time in the distribution of ringing locations.

Assuming a constant  $\lambda$  it is easily verified that expression (3.5) reduces to:

$$\frac{\phi_{j,2}(1 - \phi_{a,3})}{(1 - \phi_{j,2}\phi_{a,3})}$$

In general, the conditional probability of recovery in year  $t$  from cohort  $i$  is given by the probability of a bird dying in year  $t$  divided by the probability of it dying before the end of the recovery period. For example (given 3 recovery years) the probabilities of recovery from juvenile cohorts are conveniently expressed in the matrix:

$i/t$	1	2	3
1	$\frac{1 - \phi_{j,1}}{1 - \phi_{j,1}\phi_{a,2}\phi_{a,3}}$	$\frac{\phi_{j,1}(1 - \phi_{a,2})}{1 - \phi_{j,1}\phi_{a,2}\phi_{a,3}}$	$\frac{\phi_{j,1}\phi_{a,2}(1 - \phi_{a,3})}{1 - \phi_{j,1}\phi_{a,2}\phi_{a,3}}$
2		$\frac{1 - \phi_{j,2}}{1 - \phi_{j,2}\phi_{a,3}}$	$\frac{\phi_{j,2}(1 - \phi_{a,3})}{1 - \phi_{j,2}\phi_{a,3}}$
3			$\frac{1 - \phi_{j,3}}{1 - \phi_{j,3}}$

An analogous matrix can be formed for adult birds in which all juvenile survival terms ( $\phi_{j,t}$ ) are replaced by the corresponding adult survival terms ( $\phi_{a,t}$ ). The likelihood for the ring-recovery data, which is simply a product of Multinomials for both juvenile and adult recoveries, is given by

$$L_{dead} = f(\mathbf{D}^j | \phi_{j,1...T}, \phi_{a,2...T}) \times f(\mathbf{D}^a | \phi_{a,1...T}). \quad (3.6)$$

Analysis is restricted to recoveries of birds ringed in Britain during the summer months (May - August), only a small number of which were ringed on CES sites. Those few recoveries of birds ringed in the nest are excluded. In line with the CES scheme, years are measured from May 1<sup>st</sup>, the start of the CES season and the time that Sedge and Reed Warblers return from their west African wintering quarters. As is the case for small songbirds in general, the NRR database contains only a limited number of recoveries of Sedge and Reed Warblers ringed as fledged birds from 1987 to 2005, see Table 3.1. For both species, there are recoveries in all study years.

Estimates of the adult and juvenile survival probability parameters, in the model for the ring-recovery data, are readily obtainable using Bayesian MCMC tech-

**Table 3.1** Number of recoveries of Sedge and Reed Warblers ringed as fledged birds (adult or juvenile) from 1987 to 2005 in the BTO's NRR database.

		Sedge Warbler	Reed Warbler
Age Ringed	Adult	37	81
	Juvenile	257	305

niques. The survival parameters are assigned independent Uniform[0,1] priors. We use a random walk, single update, Metropolis-Hastings algorithm with a Uniform proposal distribution, appropriately truncated. For each survival parameter, the random walk step length is tuned *a priori*. The MCMC simulations, run for 200,000 iterations with the first 100,000 iterations discarded as burn-in, are repeated with different overdispersed starting points. As these independent runs give essentially identical posterior estimates we assume the chain has converged.

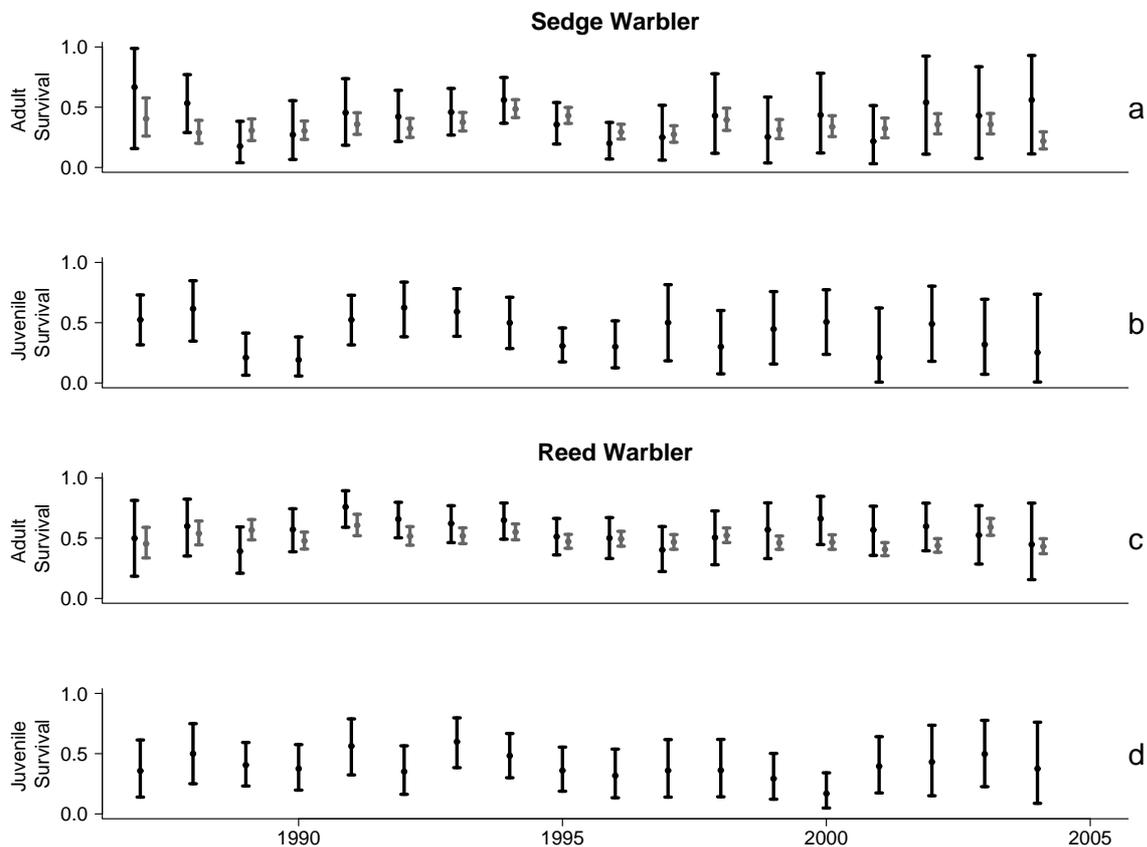
Note that the survival probability for juvenile birds in the final recovery year,  $T$ , occurs only in the bottom right-hand cell of the juvenile probability matrix. As this cell entry is always equal to 1, and thus disappears from the likelihood,  $\phi_{j,T}$  is not estimable from the data - its posterior distribution is simply given by its prior, a Uniform[0,1] in this case. As survival rates in the final year of study ( $\phi_{j,T}$  and  $\phi_{a,T}$ ) are not part of the population model, Equation (3.1), the inability to obtain data driven estimates of  $\phi_{j,T}$  is not an issue, and we shall only present ring-recovery survival estimates from the first study year (1987) to the penultimate study year (2004).

The Bayesian analysis is readily conducted in WinBUGS (Spiegelhalter et al., 2003), but in preparation for its integration with  $L_a$ ,  $L_j$  and  $L_{live}$ , bespoke code was written in C. Classical estimates are readily obtained using Program MARK (White and Burnham, 1999), which has a menu option directly offering this conditional model. For both Sedge and Reed Warblers the classical maximum likelihood estimates were virtually identical to the Bayesian posterior means.

Posterior means of  $\phi_{a,t}$  and  $\phi_{j,t}$ , for both Sedge and Reed Warblers, are shown in Figure 3.1. For adults, the estimates of  $\phi_{a,t}$  obtained from ring-recovery data are compared to those obtained from CES live-recapture data, assuming total site-faithfulness, presented in Section 2.4.5 (Figures 3.1a, c).

As is true for many songbird species there are insufficient ring-recoveries to estimate annual adult and juvenile survival rates with any precision (Baillie et al., 1986). This lack of precision is clearly illustrated in Figure 3.1. Further,

for adults, it is obvious that NRR data are very impoverished compared to CES live-recapture data, with the latter providing considerably more precise estimates (Figure 3.1a, c). Nevertheless, the overall trend between the CES and NRR estimates of  $\phi_{a,t}$  are fairly similar.



**Figure 3.1** Posterior means, and the 95% symmetric credible intervals, for the estimates of adult and juvenile survival using Sedge Warbler (SW) and Reed Warbler (RW) ring-recovery data from the BTO’s NRR database (1987-2005). In plots a) and c) estimates of adult survival from the “best” sites CES live-recapture data are also plotted, offset on the right, in grey. a) SW:  $\phi_{a,t}$ , b) SW:  $\phi_{j,t}$ , c) RW:  $\phi_{a,t}$ , d) RW:  $\phi_{j,t}$ .

Figures 3.1a and c do not provide evidence against the assumption  $\phi_{a,t} = S_{a,t}$ . Neither the Sedge nor Reed Warbler estimates of adult survival from CES data are consistently lower than those from NRR data, as would be the case if there was substantial permanent emigration of adult birds away from their CES site of ringing.

Electronic submission of annual ringing data is becoming increasingly common (Coiffait et al., 2008). With the full computerisation of the NRR database, the availability of cohort ringing numbers in the future is a distinct possibility. Should the ringing totals be known the standard (unconditional) model

for ring-recovery data (see Freeman and Morgan, 1992; Vounatsou and Smith, 1995; Brooks et al., 2000a) should be adopted. Furthermore, the assumption of constant recovery rates across time and age groups (within a cohort) should be evaluated, and the recovery rates modelled appropriately, avoiding the potential to induce bias that violations of the restrictive constant recovery rate assumption may cause (Mazzetta, 2009). Under such circumstances  $L_{dead}$  is modified, but the integrated model described in this chapter (with this new  $L_{dead}$ ) is still valid. Indeed it may be that an integrated approach enables the modelling of complex variation in reporting rates that is not feasible from models for the ring-recoveries alone. For instance, ring-recovery models with age dependence in  $\phi$  and  $\lambda$  can be especially problematical, and this is an ecologically realistic scenario (Catchpole et al., 1995).

### 3.2.3 Joint Analysis of Ring-Recovery Data and CES Live-Recapture Data

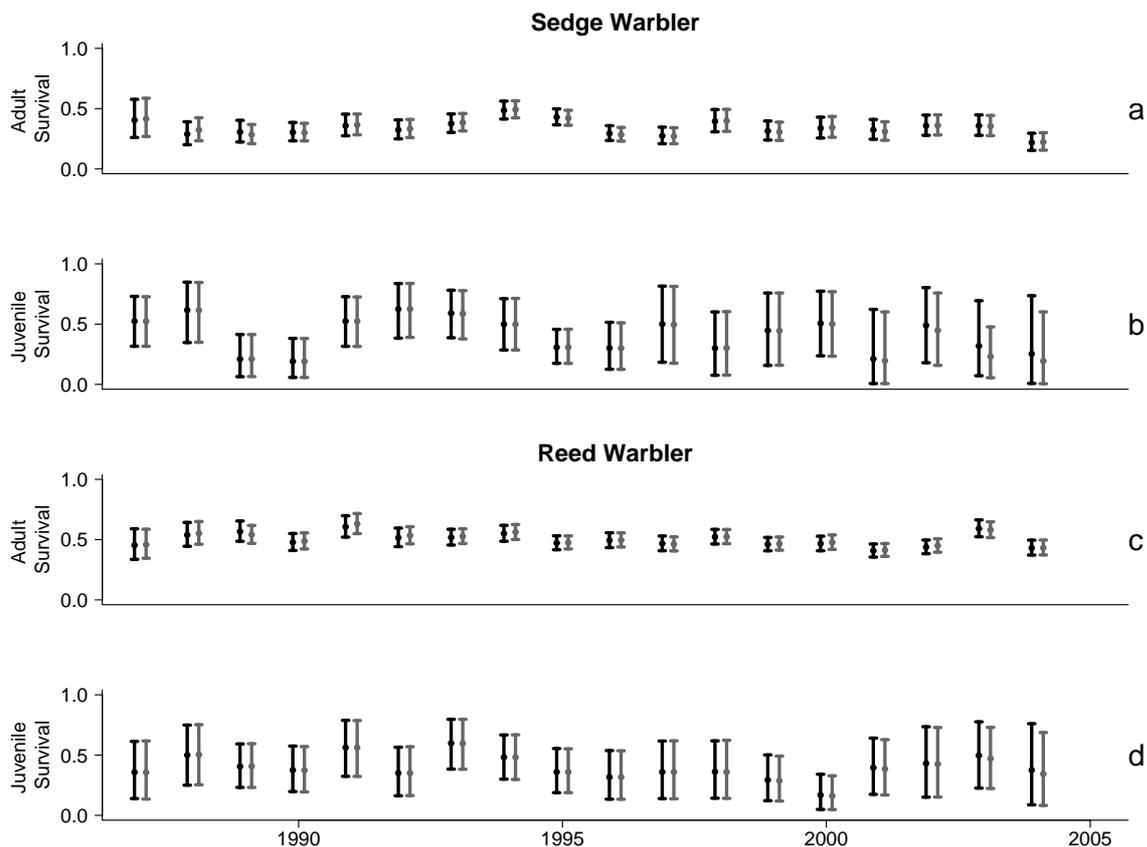
In the previous section, despite the likelihoods for the ring-recovery ( $L_{dead}$ ) and CES live-recapture ( $L_{live}$ ) data sets clearly containing adult survival parameters ( $\phi_{a,1..T-1}$ ) in common, their estimation was conducted independently. As the ring-recovery data consist of dead recoveries from all birds ringed in the UK, the vast majority of which are not ringed at CES sites,  $L_{dead}$  and  $L_{live}$  can be considered independent, and their single joint likelihood formed by straightforward multiplication, i.e.

$$L_{joint} = L_{dead} \times L_{live}.$$

Bayesian estimates of the survival parameters, under this joint likelihood, are readily obtained using a random walk, single update, Metropolis-Hastings algorithm, assigning Uniform[0,1] priors to all parameters. For Sedge and Reed Warbler, the resulting posterior means, and 95% symmetric credible intervals, for  $\phi_{a,t}$  and  $\phi_{j,t}$  are given in Figure 3.2. Shown too are “baseline” survival estimates - for adults, estimates obtained from the independent analysis of CES data (which are more precise than those from NRR data), and for juveniles, estimates obtained from the independent analysis of NRR data.

As NRR data are very impoverished compared to CES data,  $\phi_{a,t}$  estimates from the joint analysis are very similar to their baseline estimates, i.e. the estimates obtained from the independent analysis of CES data. There have, however, been a few minor changes in the posterior means of  $\phi_{a,t}$ , as now these estimates

describe both the NRR and CES data sets. For example, the small increase in the estimate of Sedge Warbler survival in 1988 (Figure 3.2a) on account of the high NRR estimate of  $\phi_{a,t}$  for this year (Figure 3.1a). Estimates of  $\phi_{j,t}$  are also adjusted, albeit slightly, under the joint analysis due to the greater amount of information on  $\phi_{a,t}$  arising from the CES data.



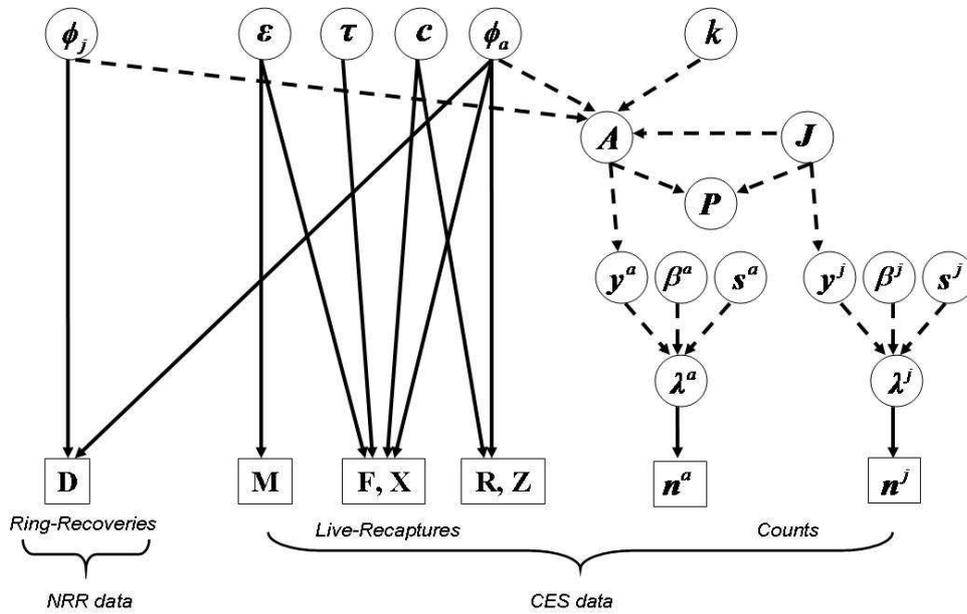
**Figure 3.2** Posterior means, and the 95% symmetric credible intervals, of the baseline estimates (left-hand, black) and the estimates from the joint analysis (right-hand, grey) of the “best” sites CES live-recapture data and the NRR data, for Sedge Warbler (SW) and Reed Warbler (RW). Baseline values are obtained from the CES data alone for adults, and from the NRR data alone for juveniles. a) SW:  $\phi_{a,t}$ , b) SW:  $\phi_{j,t}$ , c) RW:  $\phi_{a,t}$ , d) RW:  $\phi_{j,t}$ .

Generally there is an improvement in precision for both  $\phi_{j,t}$  and  $\phi_{a,t}$ , albeit typically very slight, due to the sharing of information on  $\phi_{a,t}$  between  $L_{live}$  and  $L_{dead}$ . This improvement is greatest for  $\phi_{j,t}$  in the terminal years, 2003 and 2004 (Figures 3.2b, d). In these years both  $\phi_{j,t}$  and  $\phi_{a,t}$  are estimated with particularly low precision from ring-recovery data alone (Figure 3.1). By augmenting the limited information on  $\phi_{a,t}$  in the ring-recovery data with the much greater amount of information in the live-recapture data,  $\phi_{a,t}$  is more accurately determined in  $L_{dead}$  enabling more precise estimation of  $\phi_{j,t}$ .

### 3.2.4 Bayesian Fitting of the Integrated Population Model

The combination of the underlying abundance and survival models in the fully integrated population model is summarised by the Directed Acyclic Graph given in Figure 3.3. Using Bayesian methodology parameter estimates under this model are readily obtained. Bespoke code was written in C for this purpose.

Vague, independent, priors are specified for all parameters. In particular, for the real-valued parameters in the abundance models, and the log-transformed productivity constant  $k$ , we specify a Normal prior with mean 0 and variance 10,000. For the survival, transience, recapture and evasion probabilities we assume a Uniform[0,1] prior. We use a random walk, single update, Metropolis-Hastings algorithm, with a Normal proposal for real-valued parameters, and a Uniform proposal, appropriately truncated, for the probability parameters. Proposal variances and step lengths are tuned *a priori*. The MCMC algorithm is run for 200,000 iterations, of which the first half are discarded as burn-in. Further, the MCMC chain simulations are repeated using different overdispersed starting points. As these independent replications give essentially identical posterior estimates we assume convergence of the MCMC chain.



**Figure 3.3** Directed Acyclic Graph corresponding to the fully integrated model for complete coverage data. Known values are represented by squares and unknown values by circles. Continuous arrows denote stochastic dependencies and dashed arrows deterministic dependencies. Note that live-recaptures and adult counts typically arise from the same data. In such circumstances they are not independent.

### 3.2.5 “Baseline” Models

Independent estimates for the trends in adult abundance,  $A_t$ , and juvenile abundance,  $J_t$ , are obtained via the separate fitting of the Poisson abundance models, described by Equations (2.1) and (2.2), for adults and juveniles respectively, as in Chapter 2. These estimates are not restricted by any demographic assumptions, thus they serve as a “baseline” for similar estimates derived under population models such as those of Equation (3.1). Likewise, as in Chapter 2, baseline estimates for the productivity indices,  $P_t$ , are derived by calculating the ratio of  $J_t$  to  $A_t$  using samples drawn from their posterior distributions.

Baseline estimates of true adult survival,  $\phi_{a,t}$ , are obtainable from CES live-recapture data under the assumption of total site-faithfulness via the extended Pradel et al. (1997) transient model as in Chapter 2. In addition, an alternative baseline for  $\phi_{a,t}$  is also estimable from the independent analysis of the ring-recovery data using the conditional model of Section 3.2.2.

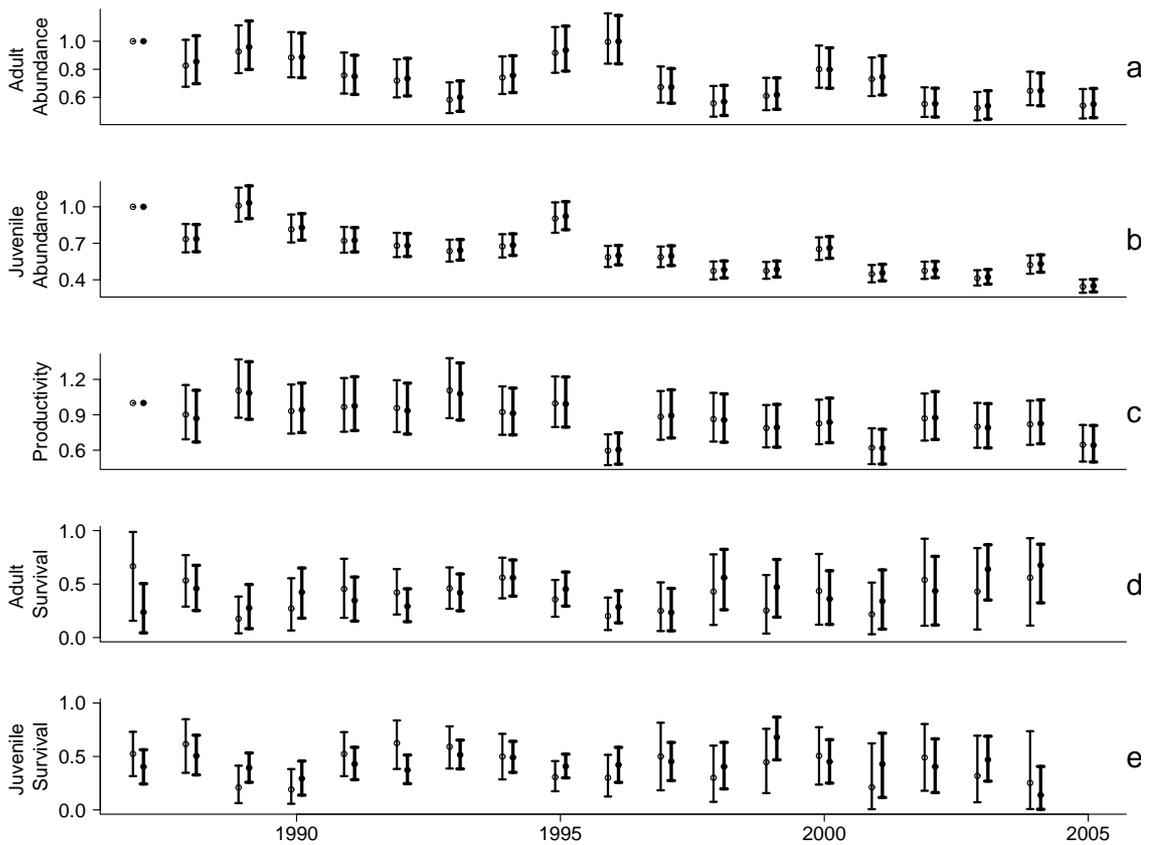
As estimates of  $\phi_{j,t}$  are not obtainable using CES data alone, baseline values are estimated from the independent analysis of the ring-recovery data.

## 3.3 Analysis of Sedge Warbler Data

In the analyses that follow we integrate CES complete coverage data from the “best” sites, as defined in Chapter 2 (Section 2.1), with ring-recovery data from the NRR database. Both data sets cover the years 1987 to 2005. We begin by applying the methods to Sedge Warbler data. An integrated analysis of Reed Warbler data is considered in Section 3.4.

### 3.3.1 Integrated Model with Ring-Recovery Data

A wide range of analyses are possible in the integrated context. Initially we form a model by multiplying the likelihoods  $L_a$  and  $L_j$  for the adult and juvenile yearly-site counts from the CES data with the information from the ring-recoveries via  $L_{dead}$ . These two data sets, and the three component likelihoods, are independent. Posterior means of  $J_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$  and the derived  $A_t$  and  $P_t$  parameters are shown in Figure 3.4 along with their baseline estimates. For  $\phi_{a,t}$  the appropriate baseline is from the ring-recovery data.



**Figure 3.4** Posterior means, and the 95% symmetric credible intervals, from the “baseline” models, denoted by thin left-hand lines, and the integrated model which uses ring-recovery data only in the estimation of adult survival, denoted by bold right-hand lines, for a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$  using Sedge Warbler data. For  $\phi_{a,t}$  the ring-recovery baseline is used.

Comparing the baseline and integrated models reveals that although the derived adult abundance indices from the integrated model are very similar to their baseline (Figure 3.4a) there are some differences in the estimates of adult survival (Figure 3.4d) and juvenile survival (Figure 3.4e). This would suggest that the adult abundance data, through the recursive population model (Equation (3.1)), is driving the estimation of adult and juvenile survival, which is to be expected as the ring-recovery data are relatively sparse. Conversely, as the derived adult abundance indices and their baseline are almost exactly the same, it would seem that the limited ring-recovery data has practically no influence on the recursive population model.

Figure 3.4 illustrates changes in the estimates of adult and/or juvenile survival in the integrated model, not reflected in their baseline models, to produce derived adult abundance indices that correspond to the observed adult count data. Note that productivity, and consequently juvenile abundance, are much reduced

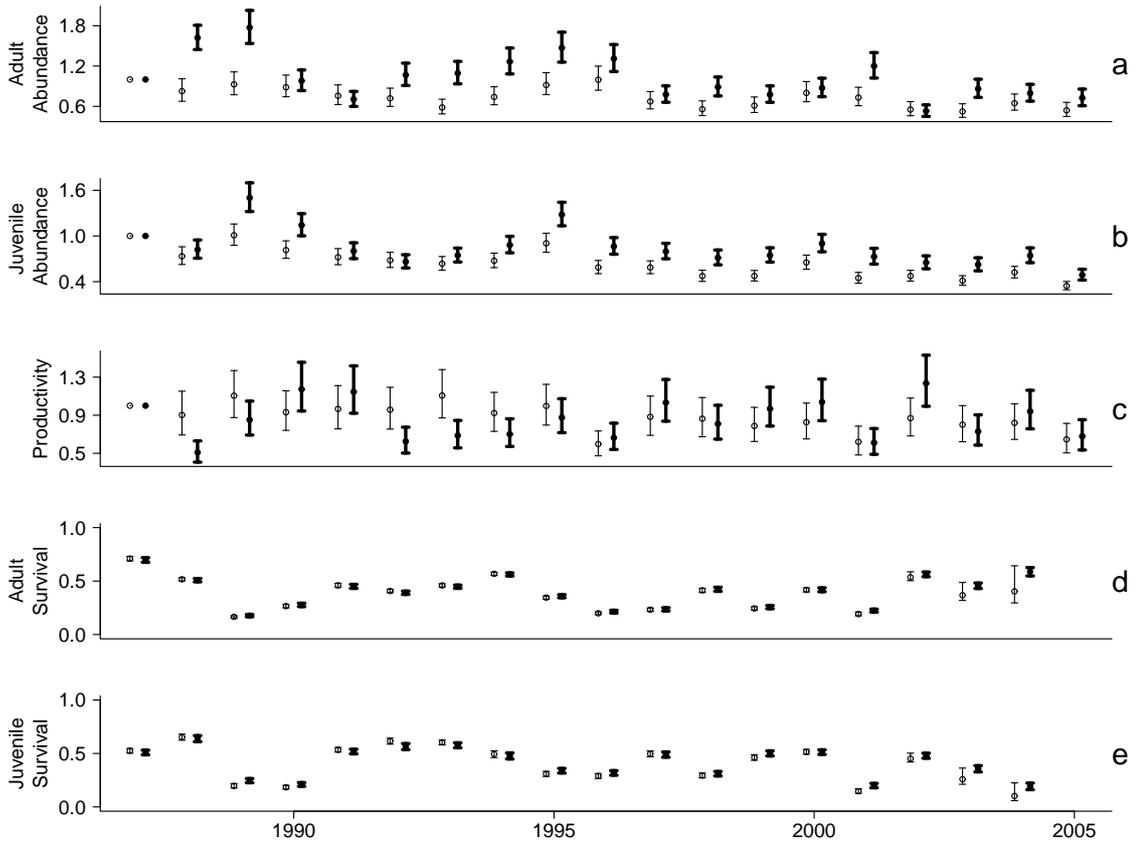
after 1995, and that since then only three years (1998, 1999 and 2003) have been followed by an increase in adult abundance (Figure 3.4a). These years (along with the imprecise terminal years) are also those in which adult and juvenile survival is greatly increased under the integrated analysis (Figure 3.4d, e). Estimates of  $J_t$  and  $P_t$  arise from considerably more data, the CES count data, than those of adult and juvenile survival from the ring-recovery data, thus the former are more resistant to change once the component models are integrated. Estimates of  $J_t$  and/or  $P_t$  in these three years are not sufficient so the integrated model increases the estimates of  $\phi_{a,t}$  and  $\phi_{j,t}$  to account for the greater number of adult birds caught. Therefore, it is the paucity of ring-recovery data that explains the discrepancies between the baseline values and the estimates from the integrated analysis in Figures 3.4d and 3.4e.

Integration also noticeably improves the precision in the estimates of adult and juvenile survival (Figure 3.4d, e). In the baseline  $\phi_{a,t}$  and  $\phi_{j,t}$  are freely estimated from the sparse ring-recovery data, unrestricted by any assumptions relating them to adult abundance. In the integrated case, however, the limited information on  $\phi_{a,t}$  and  $\phi_{j,t}$  from the ring-recovery data is augmented by the population model, given by Equation (3.1), which relates the number of adults in consecutive years to productivity and survival. This extra information improves the precision in the estimates of  $\phi_{a,t}$  and  $\phi_{j,t}$ .

#### *Simulation of a Rich Ring-Recovery Data Set*

To make the observed CES adult count data match the underlying population model, given by Equation (3.1), the integrated analysis alters estimates of  $\phi_{a,t}$  and  $\phi_{j,t}$  for which there is limited direct information available. By simulating a large ring-recovery data set, with parameters consistent with the observed NRR data but that provides precise inference on  $\phi_{a,t}$  and  $\phi_{j,t}$ , we illustrate how the other demographic parameters ( $A_t$ ,  $J_t$ , and  $P_t$ ) are now likewise affected.

Cohort recovery totals in the observed Sedge Warbler NRR data range from 0-37. Selecting recovery totals to range between 1000-2000, and setting the survival parameters equal to the posterior means from the independent analysis of the Sedge Warbler NRR data, a large ring-recovery data set is simulated. Posterior means of the key demographic parameters ( $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ ) from the integrated model which adopts this large simulated ring-recovery data set are shown in Figure 3.5.



**Figure 3.5** Integrated analysis with a large, simulated, ring-recovery data set. Posterior means, and the 95% symmetric credible intervals, from the “baseline” models, denoted by thin left-hand lines, and the integrated model denoted by bold right-hand lines, for a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$  using Sedge Warbler data. For both  $\phi_{a,t}$  and  $\phi_{j,t}$  the simulated ring-recovery data are used as the baseline.

The baseline estimates of  $\phi_{a,t}$  and  $\phi_{j,t}$ , from the simulated ring-recovery data, are now extremely precise (Figure 3.5d, e). The integrated model now responds to discrepancies between the baseline estimates of the demographic parameters and the underlying population model, Equation (3.1), by adjusting the estimates of  $A_t$ ,  $J_t$ , and  $P_t$  for which uncertainty, in comparison, is greater (Figure 3.5a, b, c). For example, the integrated model estimates a higher  $A_t$  in 1988 and 1989, compared to the baseline (Figure 3.5a), in response to the high, but precisely estimated,  $\phi_{a,t}$  and  $\phi_{j,t}$  values in the preceding years (Figure 3.5d, e).

### 3.3.2 Including CES Live-Recapture Data

The far greater amount of adult survival information in the live-recapture likelihood,  $L_{live}$ , makes it appealing to further multiply this into the joint likelihood. However, as  $L_{live}$  is drawn from the same data as the adult abundance likeli-

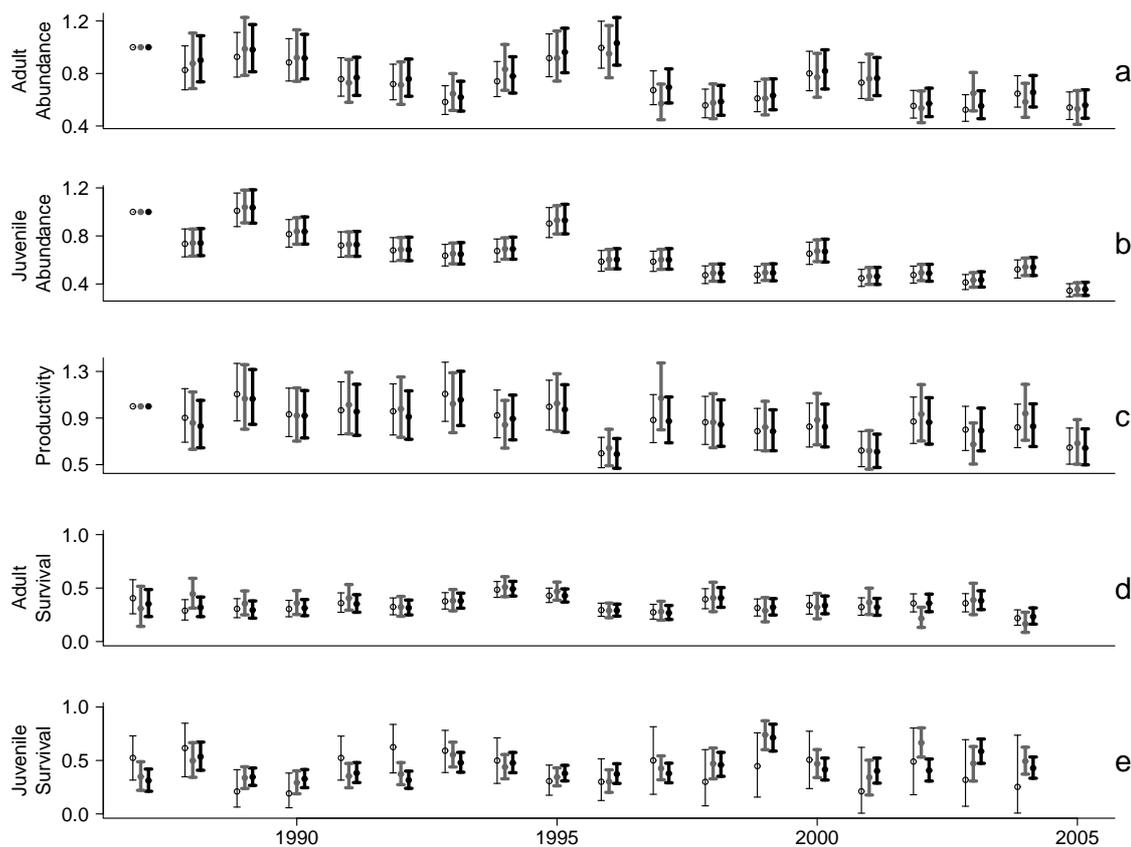
hood,  $L_a$ , these likelihoods are not independent. Indeed, birds contribute data to both  $L_{live}$  and  $L_a$  on every between-year encounter. To accommodate this lack of independence we initially split the data into two groups (one of 35 sites and the other of 36 sites); one is then used to derive  $L_{live}$  and the other independent set to form  $L_a$ . To ensure both split data sets provide good coverage over the duration of the study (which covered few sites in the early years), sites are initially stratified according to when they first joined the CES scheme, and within each of these strata half the sites are randomly assigned to each group. Secondly, we ignore the issue of non-independence and derive  $L_a$  and  $L_{live}$  from the full 71-site data. In both cases the likelihood for the ring-recovery data,  $L_{dead}$ , is also incorporated to provide more direct information on  $\phi_{j,t}$  so that:

$$L_{joint} = L_a \times L_j \times L_{dead} \times L_{live}. \quad (3.7)$$

The juvenile abundance likelihood,  $L_j$ , which is independent of  $L_a$  and  $L_{live}$ , is derived from the full 71-site data in each case.

The posterior means of key parameters,  $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ , and  $\phi_{j,t}$ , from the full and split data integrated analyses are compared to each other, and their baseline estimates, in Figure 3.6. As there are considerably more data on adult survival from the live-recaptures than the limited ring-recoveries, independent estimates of  $\phi_{a,t}$  via  $L_{live}$  are used as the baseline.

Initially ignoring the non-independence between  $L_{live}$  and  $L_a$ , and concentrating on the full 71-site integrated analysis, we note that estimates of  $A_t$ ,  $J_t$ ,  $P_t$ , and  $\phi_{a,t}$  are all very similar to their baseline estimates. Conversely there have been large changes to  $\phi_{j,t}$ . This is to be expected, since we have limited direct information on  $\phi_{j,t}$ , and substantially more on the other demographic parameters; the integrated model thus alters  $\phi_{j,t}$  to make the derived  $A_t$  correspond to the adult count data. For example, in Figure 3.4a the integrated model (without  $L_{live}$ ) has explained the increase in adult abundance from 1998, 1999 and 2003 by increasing both adult and juvenile survival (Figure 3.4d, e), for which little is otherwise known. Now, because the baseline likelihood  $L_{live}$  is also a part of the integrated model, bringing in substantially more information about adult survival,  $\phi_{a,t}$  is not adjusted (Figure 3.6d); rather the integrated model responds to account for the increase in adult abundance by just increasing  $\phi_{j,t}$  in these years (Figure 3.6e).



**Figure 3.6** Posterior means, and the 95% symmetric credible intervals, from the “baseline” models, denoted by thin left-hand lines, and the integrated model which uses ring-recovery data and CES data in the estimation of adult survival. The analysis in which all 71 sites provide information to  $L_{live}$  and  $L_a$ , is represented by bold right-hand lines. The bold grey centre line represents the split data analysis. a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ .

The integrated approach also tends to result in an improved precision from the sharing of information between the component models. This is most noticeable for juvenile survival (Figure 3.6e). Here the very limited direct information on  $\phi_{j,t}$  from the ring-recovery data is augmented in the integrated model by the information from the population model, Equation (3.1).

The full and split data integrated analyses produce consistent results (Figure 3.6). Using the CES data as two independent groups of 36 sites and 35 sites, we find  $A_t$ ,  $P_t$  and  $\phi_{a,t}$ , resemble their baseline, though slightly less so than previously when the full data were used (Figure 3.6a, c, d). As fewer data are now used to form  $L_{live}$  and  $L_a$  this is to be expected. For  $\phi_{j,t}$  both the full and split data integrated analyses estimate trends more similar to each other than either is to the imprecise baseline (Figure 3.6e). As the full data are used to derive  $L_j$  in both the integrated models  $J_t$  is virtually unchanged (Figure 3.6b).

For the integrated models, the 95% symmetric credible intervals for  $A_t$ ,  $P_t$ ,  $\phi_{a,t}$ , and  $\phi_{j,t}$  from the split data analysis are wider, but not dramatically so, than those from the full data analysis (Figure 3.6a, c, d, e). An increase in uncertainty is to be expected, as when the data are split the parameters are being estimated from less information.

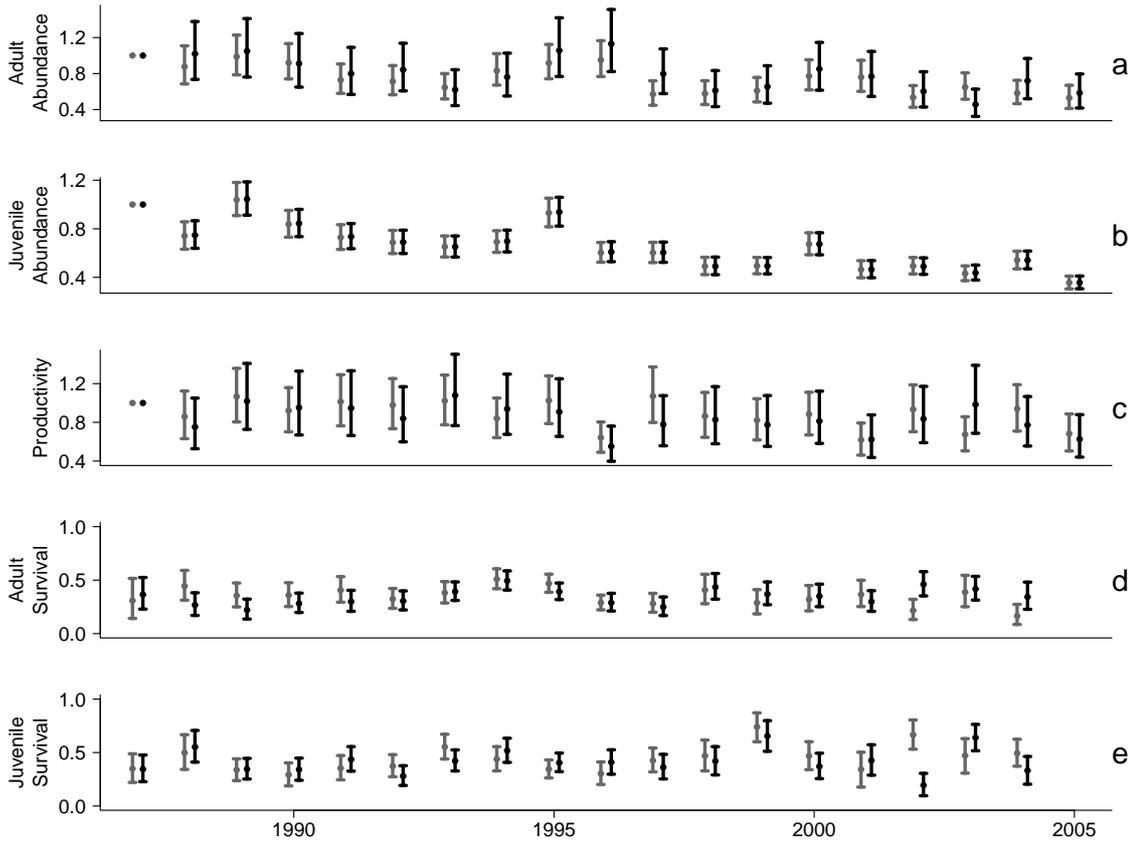
The proportion of transients  $\tau$ , a parameter of  $L_{live}$ , also requires estimation under the fully integrated model. The posterior mean (standard deviation) of  $\tau$  from the full 71-site data analysis is 0.371 (0.0134), and the split data analysis is 0.359 (0.0200), with the 95% symmetric credible intervals overlapping. As expected, the estimate is more precise under the full data analysis due to the greater amount of data.

Interestingly the integrated model has estimated  $\phi_{j,t}$  to be greater than  $\phi_{a,t}$  in several years - 1999, 2003 and 2004 (Figure 3.6d, e). This could be the result of permanent emigration causing negative bias in the estimates of  $\phi_{a,t}$  derived from the CES live-recapture data. We explore this further in Chapter 5.

A nice feature of the integrated model is that it produces estimates of “true seasonal” productivity ( $P_t^s$ ), the number of fledged juveniles per breeding pair per year, through the combination of  $k$  and  $P_t$  (see Equation 2.4). These estimates are consistent with published analyses of Nest Record Scheme (NRS) data. Assuming an equal proportion of female and male birds, the integrated model, using the full 71-site data, estimates on average 3.0 juveniles per pair per year. The split data analysis estimates, on average, 2.9 juveniles per pair per year. Nest failure probabilities, the proportion of nests failing completely due to predation or desertion for example, are approximately 0.015 per day at the egg stage and 0.018 per day at the chick stage for Sedge Warblers (Baillie et al., 2009b). As the duration of laying/incubation and fledging are typically 17-19 and 13-14 days respectively (Cramp, 1992), this gives a proportion of successful nests of about 0.60. With an average clutch size of 5 eggs (Robinson, 2005), and typically 1 brood per year (Cramp, 1992), this corresponds to 3.0 juveniles per pair per year. Though this figure ignores the (unknown) number of second broods, the losses of individual chicks up to (and shortly after) fledging in otherwise successful nests and the uncertainty in the estimates, the close agreement with the integrated model is reassuring.

*Sensitivity to the Split*

To investigate the sensitivity to the split of the data used to create separate data sets for  $L_{live}$  and  $L_a$ , the split data integrated analysis is repeated with the data sets used to form  $L_{live}$ , or  $L_a$ , switched. Posterior means of  $J_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$  and the derived  $A_t$  and  $P_t$  parameters from these two split data integrated analyses are compared in Figure 3.7.



**Figure 3.7** Posterior means, and the 95% symmetric credible intervals, from the split data integrated analyses where the data is split into two groups; one is then used to derive  $L_{live}$  and the other independent set to form  $L_a$ , (grey left-hand lines) and then the data sets are switched (black right-hand lines), for a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ .

The estimated indices,  $A_t$ ,  $J_t$ , and  $P_t$ , are similar between both split data integrated analyses (Figure 3.7a, b, c) as are the estimated survival probabilities,  $\phi_{a,t}$  and  $\phi_{j,t}$  (Figure 3.7d, e). Further, switching the data sets also produces similar estimates of  $\tau$  (posterior mean = 0.383, standard deviation = 0.0173) and seasonal productivity (on average 3.2 young per pair per year).

The main discrepancy between the two split data integrated analyses occurs in 2002; under the first split analysis an increase in  $\phi_{j,t}$  and a decrease in  $\phi_{a,t}$

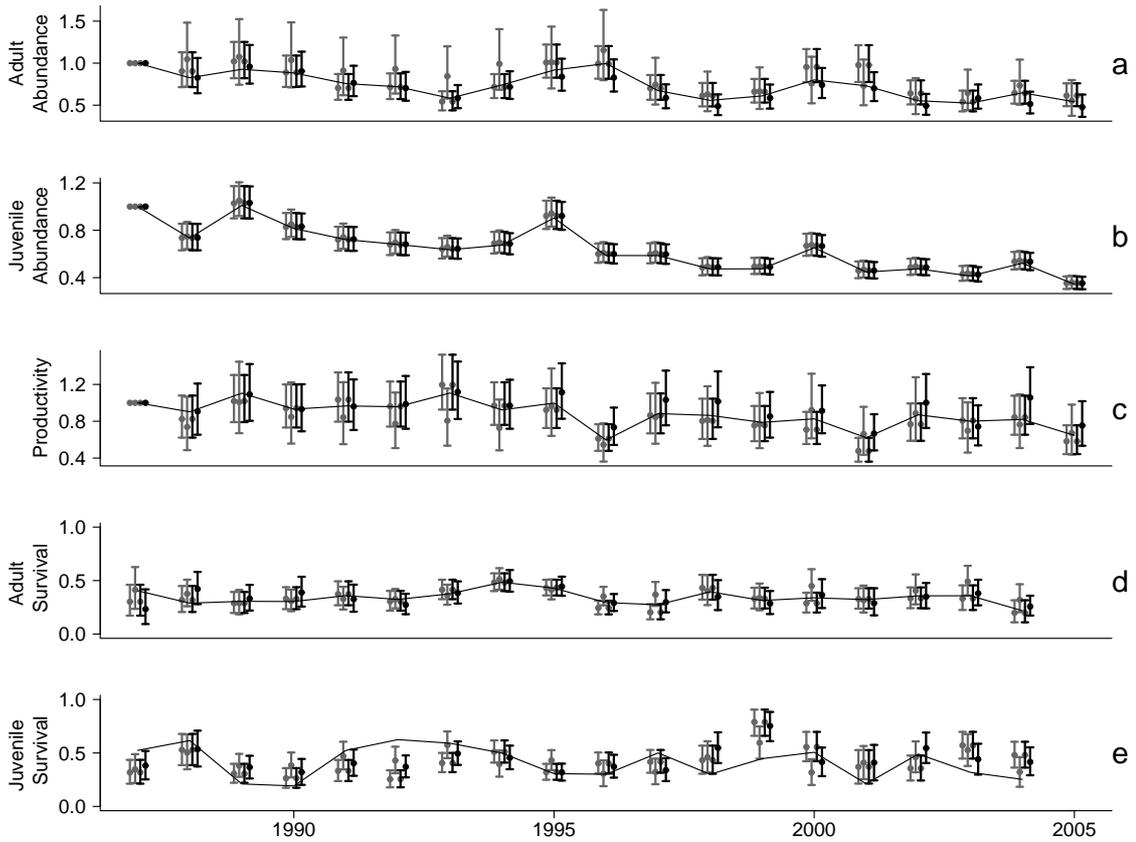
is estimated, whereas the converse is the case when the data sets are switched (Figure 3.7d, e). However, the majority of the differences are minor, and the integrated analysis seems relatively insensitive to the split.

Interestingly the 95% symmetric credible intervals of  $A_t$  and  $P_t$  from the first split analysis are substantially narrower than those when the data sets are switched (Figure 3.7a, c). This is because the indices are relative to a reference year (in this case 1987) and site. The two split analyses were required to have a different reference site, and this change affects the posterior variation of  $\beta^a$  and consequently the precision of  $A_t$  and  $P_t$ . Note that posterior variance of  $\beta^a$  was lower in the first split analysis (0.0100) than in the second (0.0783) resulting in the observed increase in precision.

To further verify the insensitivity of the split, the analysis was repeated with multiple other random splits, albeit all of 36 and 35 sites, respectively. Analogous results were obtained. In Figure 3.8 the results from two of these further random splits, both with the data sets used to form  $L_{live}$ , or  $L_a$ , also switched, are presented. There is general agreement in the estimated annual trends, but precision is highly affected by the split. This is particularly true for  $A_t$  and  $P_t$  (Figure 3.8a, c) as a result of changes in the reference site used. A random split (prior to analysis) will not give biased or misleading results, but as clearly demonstrated in Figure 3.8, it can be sub-optimal in terms of precision.

### *The Importance of Ring-Recovery Data*

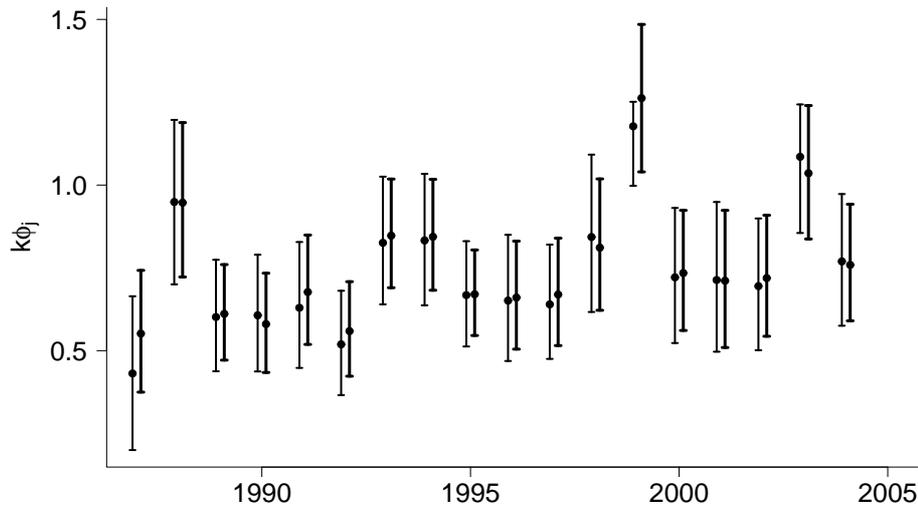
Despite ring-recovery data being limited, their presence is crucial to the integrated analysis. In the absence of ring-recovery data, or informative prior information,  $\phi_{j,t}$  would be estimated entirely indirectly via the joint likelihood and its role in the deterministic population model, Equation (3.1), as is the other “freely” estimated parameter, the productivity constant  $k$ . In such circumstances  $\phi_{j,t}$  and  $k$  are confounded, and we obtain a parameter redundant model. Although their product,  $k\phi_{j,t} \in [0, \infty]$ , is estimable integration becomes less meaningful - the baseline estimates for  $A_t$ ,  $J_t$ ,  $P_t$  and  $\phi_{a,t}$  are simply returned with  $k\phi_{j,t}$  taking the values required to match the population model, Equation (3.1), to the observed data. This is readily verified by removing  $L_{dead}$  from the integrated analysis, that is setting  $L_{joint} = L_a \times L_j \times L_{live}$ . The estimable quantity  $k\phi_{j,t}$ , however, does provide an index for juvenile survival which can still be used for detecting whether juvenile survival rates are decreasing for example, a useful quantity in the absence of any direct information pertaining to  $\phi_{j,t}$ .



**Figure 3.8** Posterior means, and the 95% symmetric credible intervals, from two further split data integrated analyses where the data is split into two groups; one is then used to derive  $L_{live}$  and the other independent set to form  $L_a$ . On the left, (grey lines) are the results from the first additional analysis, giving the initial split and the reversal of the split respectively. On the right, (black lines) are the results from the second additional analysis. The solid line denotes the posterior means from the baseline model. For  $\phi_{a,t}$  the CES data is adopted as the baseline. a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ .

Estimates of  $k\phi_{j,t}$  from the integrated analysis with, and without,  $L_{dead}$  included are not exactly the same (Figure 3.9) demonstrating that even the sparse ring-recoveries have some limited influence on the population model (Equation (3.1)). In addition, an increase in precision is generally realised by including the ring-recovery data, albeit sparse.

In an analogous situation to the absence of direct information on  $\phi_{j,t}$ , if  $k$  were allowed to be time-dependent in the population model, given by Equation (3.1), the baseline estimates of  $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$  and  $\phi_{j,t}$  would be returned, with  $k_t$  taking the values required to match the population model to the observed data. However, time-invariance in  $k$  can be assumed under the constant effort protocol that underpins the CES scheme and the estimation of abundance and productivity indices from the resulting count data.



**Figure 3.9** Posterior means, and the 95% symmetric credible intervals of  $k\phi_{j,t}$ , from the integrated analysis with (bold right-hand lines), and without (thin left-hand lines) ring-recovery data. In both cases all 71-sites contribute to  $L_{live}$  and  $L_a$ .

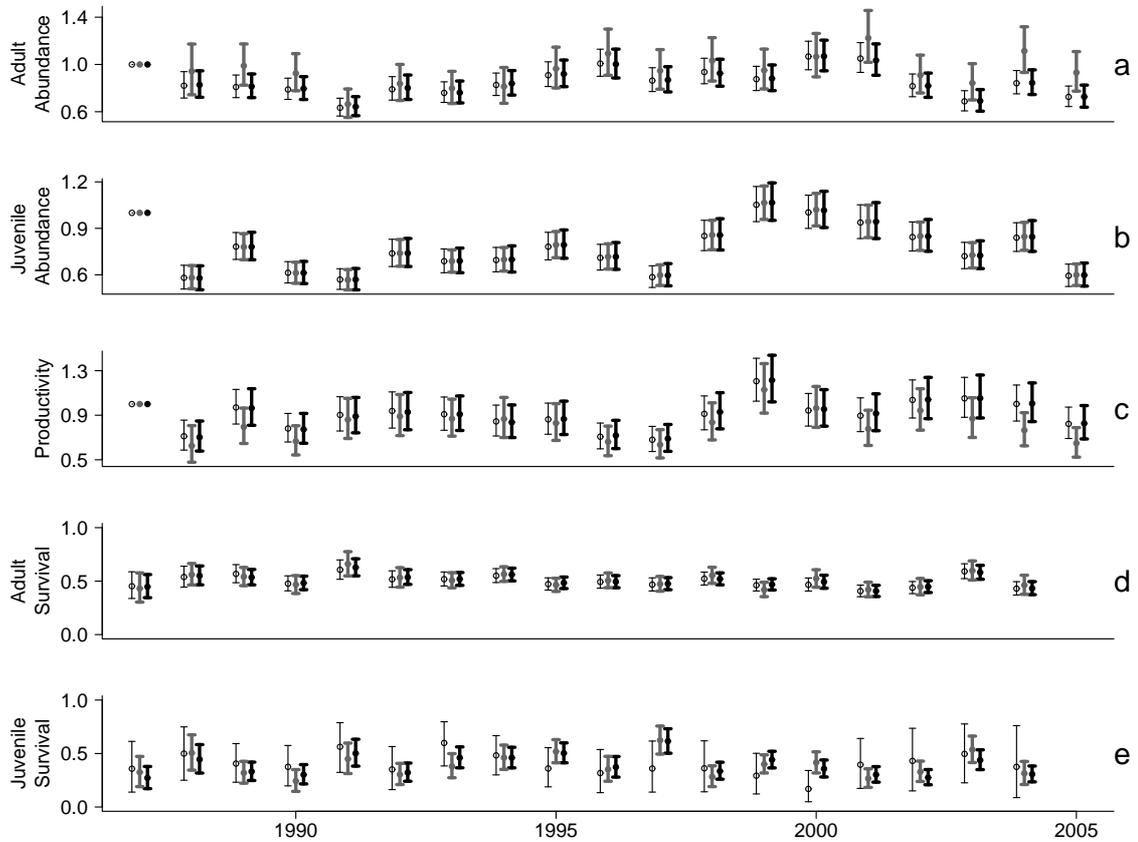
### 3.4 Analysis of Reed Warbler Data

As with the analysis of the Sedge Warbler data, for Reed Warbler data the analysis below demonstrates that integration of the individual demographic models,  $L_a$ ,  $L_j$ ,  $L_{live}$ , and  $L_{dead}$ , can lead to a substantial improvement in precision. This is particularly true for juvenile survival in which there is very sparse direct information available. Further, the estimation of  $\phi_{j,t}$  is driven by the underlying demographic model, given by Equation (3.1), since direct information is limited. In addition, we again demonstrate that the issue of non-independence between  $L_a$  and  $L_{live}$  can be overcome by randomly splitting the CES data at only a small cost to precision, and that the integrated analysis is relatively insensitive to the split chosen.

The fully integrated analysis of the CES ( $L_a$ ,  $L_j$ ,  $L_{live}$ ) and NRR ( $L_{dead}$ ) Reed Warbler likelihoods proceeds as for the Sedge Warbler data, see Equation (3.7). That is, to accommodate the lack of independence between  $L_a$  and  $L_{live}$ , we randomly split the data into two groups (one of 28 sites and the other of 27 sites) stratifying according to when the sites first joined the CES scheme. One set is then used to derive  $L_{live}$  and the other independent set to form  $L_a$ . To investigate the sensitivity of the split the analysis is repeated with the two

data sets used to form  $L_a$ , or  $L_{live}$ , switched. Secondly, we ignore the issue of dependence and derive  $L_a$  and  $L_{live}$  from the full 55-site data.

In Figure 3.10 posterior means of the key parameters ( $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ ) from the full 55-site integrated analysis and the initial split data integrated analysis are compared to each other, and their baseline estimates. Independent estimates of  $\phi_{a,t}$  via  $L_{live}$  are used as the baseline, rather than via  $L_{dead}$ , as there is considerably more information in the live-recaptures than the limited ring-recoveries.



**Figure 3.10** Posterior means, and the 95% symmetric credible intervals, from the “baseline” models, denoted by thin left-hand lines, and the integrated model which uses ring-recovery data and CES data in the estimation of adult survival. The analysis in which all 55 sites provide information to  $L_{live}$  and  $L_a$ , is represented by bold right-hand lines. The bold grey centre line represents the split data analysis. a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ .

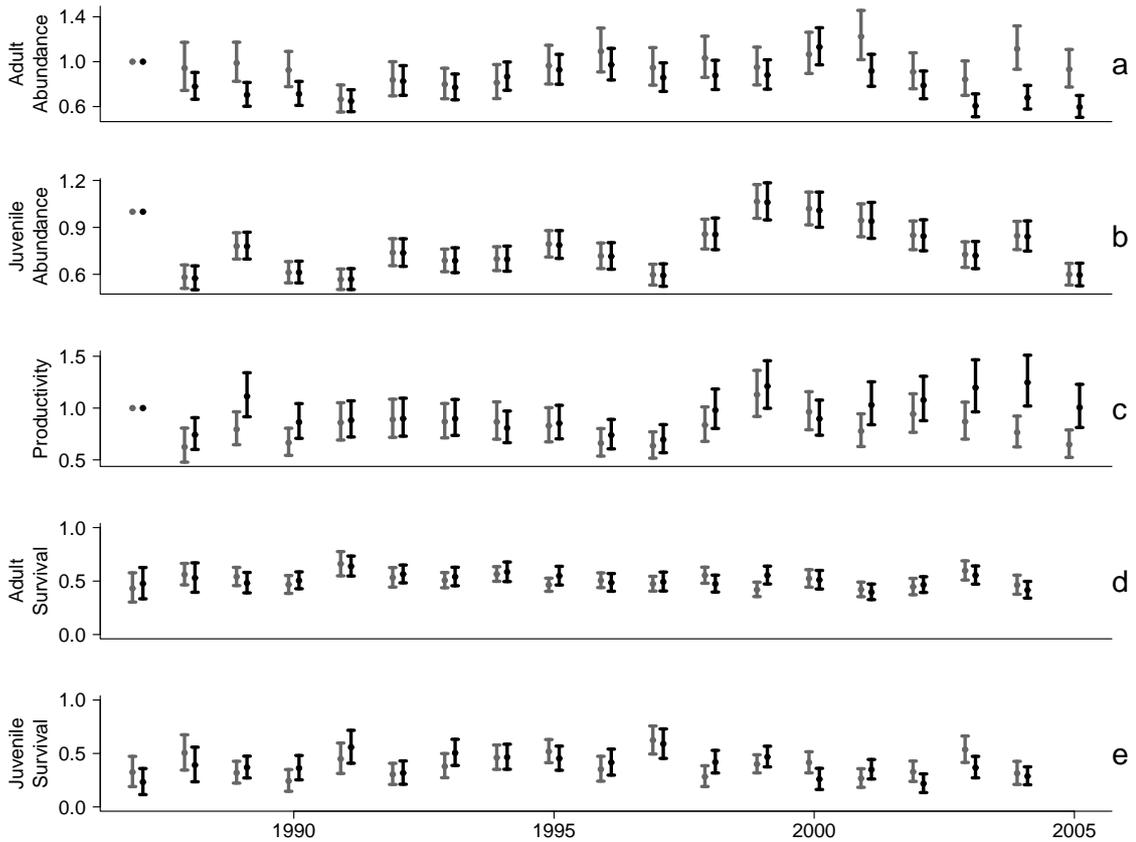
Focusing initially on the full 55-site integrated analysis, analogous to the Sedge Warbler analysis (see Figure 3.6), we note that  $A_t$ ,  $J_t$ ,  $P_t$ , and  $\phi_{a,t}$  are all very similar to their baseline estimates whereas there have been some large adjustments to  $\phi_{j,t}$  (Figure 3.10). Once again it is the paucity of the ring-recovery data that explains the discrepancies in Figure 3.10e. The integrated model adjusts

$\phi_{j,t}$ , for which little is otherwise known, to make the estimates of  $A_t$  derived from the population model (Equation (3.1)) match the observed adult count data. For example, in four years the integrated model estimates  $\phi_{j,t}$  greater than its baseline - 1995, 1997, 1999, 2000. The 1995 and 1999 adjustments to  $\phi_{j,t}$  occur to accommodate the increase in  $A_t$  in the year immediately following, as this increase in adult abundance does not correspond to large enough increases in either  $J_t$  or  $\phi_{a,t}$ , both of which are more precisely estimated in the baseline than  $\phi_{j,t}$ . The 1997 and 2000 adjustments to  $\phi_{j,t}$  counteract the estimated decrease in  $J_t$  which does not correspond to a decrease in  $A_t$  the following year.

As concluded for the analysis of Sedge Warbler data, integration leads to improved precision, particularly for  $\phi_{j,t}$  (Figure 3.10e). Further, the full and split data integrated analyses produce reasonably consistent results (Figure 3.10), although there is a loss of precision in the estimation of  $A_t$ ,  $P_t$ , and  $\phi_{a,t}$  (Figure 3.10a, c, d) due to the smaller amount of data contributing to  $L_a$  and  $L_{live}$ .

In Figure 3.11 the posterior means of  $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ , and  $\phi_{j,t}$  from the initial split are compared to the analysis in which the data sets used to form  $L_{live}$ , or  $L_a$ , are switched. The estimated indices and survival parameters are similar, suggesting the analysis is relatively insensitive to the split. As the full 55-site data are used to derive  $L_j$ , the estimates of  $J_t$  are virtually unchanged (Figure 3.11b). The only major discrepancy between the two split data integrated analyses occurs at the end of the time series (2003-2005) for the estimates of  $A_t$  and  $P_t$ . Although both estimate the same trend in  $A_t$  their absolute values differ (Figure 3.11a), and hence so do the estimates of  $P_t$  (Figure 3.11c). There is also a slight discrepancy in the estimates of  $A_t$  (and consequently  $P_t$ ) at the start of the time series, although the 95% symmetric credible intervals do overlap (Figure 3.11a, c). Here the first split data integrated analysis estimates a small increase in  $A_t$  from 1988 to 1989 followed by a small decrease in 1990, whereas the opposite trend is estimated when the data sets used to form  $L_{live}$ , or  $L_a$ , are switched.

The estimate of the proportion of transients,  $\tau$ , is also reasonably consistent between the full 55-site integrated analysis and the two split data integrated analyses. The posterior mean (standard deviation) of  $\tau$  from the full data analysis is 0.413 (0.0095), and the two split data analyses is 0.391 (0.0130) and 0.445 (0.0134) respectively, with the 95% symmetric credible intervals overlapping. As expected, the estimate is more precise under the 55-site analysis due to the greater amount of data used to derive  $L_{live}$ .



**Figure 3.11** Posterior means, and the 95% symmetric credible intervals, from the split data integrated analyses where the data is split into two groups; one is then used to derive  $L_{live}$  and the other independent set to form  $L_a$ , (grey left-hand lines) and then the data sets are switched (black right-hand lines), for a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ .

Assuming an equal proportion of female and male birds the estimate of true seasonal productivity,  $P_t^s$ , provided by the integrated model through the combination of  $k$  and  $P_t$  is, on average, 2.6 juveniles per breeding pair per year (from both the full and two split data integrated analyses). These estimates are in line with published analyses of NRS data. Nest failure probabilities are approximately 0.017 per day at the egg stage and 0.015 per day at the chick stage for Reed Warblers (Baillie et al., 2009b). As the duration of laying/incubation and fledging are typically 12-15 and 10-12 days respectively (Cramp, 1992), this gives a proportion of successful nests of about 0.67. With an average clutch size of 4 eggs (Robinson, 2005) and approximately 24% of pairs attempting a second brood (Cramp, 1992), this corresponds to an estimate of 3.3 juveniles per pair per year. However, we note that as clutch size diminishes as the season progresses (Cramp, 1992), and given that NRS data are biased in favour of early nests (Crick and Baillie, 1996), the estimate of 4 eggs per nest is likely too high. Further, the daily failure rates reported by Baillie et al. (2009b) correspond to

the proportion of nests failing completely, and hence the calculation here ignores the “partial” losses of individual eggs and chicks up to (and shortly after) fledging in otherwise successful nests. Both these effects will reduce the seasonal productivity. Baillie et al. (2009b) give an average brood size of approximately 3.5, that is each successful nest produces about 3.5 young. After accounting for this 12.5% loss the estimate of true seasonal productivity realised (2.8 juveniles per pair per year) is in close agreement with the integrated model estimate.

### *Stable Population*

As an advisory tool, to highlight declining populations of British birds, the BTO annually calculates “alert” listings, in order of decreasing severity: red, amber or green (Atkinson et al., 2006; Brooks et al., 2008; Baillie et al., 2009b). These alert statuses match categories of decline adopted in the regular “Birds of Conservation Concern” assessments for the UK (e.g. Eaton et al., 2009). The most recent assessment places both Sedge and Reed Warblers on the “green list”, indicating that these species are of little, or no, immediate conservation concern (Eaton et al., 2009). These classifications are consistent with our integrated analysis. The fully integrated model estimates  $\phi_{a,t}$  and  $\phi_{j,t}$ , on average, to be (0.34, 0.43) and (0.50, 0.39) for Sedge and Reed Warblers respectively. Under these conditions a true seasonal productivity of 3.1 and 2.5 juveniles per breeding pair would be required to maintain stable Sedge Warbler and Reed Warbler breeding populations respectively. These estimates are in line with those obtained from the integrated analysis, suggesting that, on average, the long-term trend in the adult Sedge and Reed Warbler populations is stable. There is, however, considerable variability around these averages resulting in the observed short-term fluctuations (Figures 3.6 and 3.10).

## **3.5 Investigating the Relationship between the Demographic Parameters**

The demographic population model, Equation (3.1), specifies a positive relationship between the underlying demographic rates ( $P_t$ ,  $\phi_{a,t}$  or  $\phi_{j,t}$ ) and the abundance of adults in the following year ( $A_{t+1}$ ). For example, for a species with a stable or increasing population, a higher adult survival rate in year  $t$  compared to the previous year is expected to increase adult abundance in year

$t + 1$ . For a species with a declining population, an increase in adult survival is expected to either stop the decline or decrease the rate of decline. However, the importance of any given demographic parameter in driving the size of the adult population will depend on how much it varies and the relative size of this variation (i.e. the “sensitivity” and “elasticity” (Caswell, 2001)), and also its interactions with the other demographic parameters.

Note that under the population model given by Equation (3.1)

$$\frac{A_{t+1}}{A_t} = \phi_{a,t} + kP_t\phi_{j,t},$$

where

$$\frac{A_{t+1}}{A_t} \begin{cases} > 1 & \text{if adult abundance has **increased** from year } t \text{ to year } t+1 \\ < 1 & \text{if adult abundance has **decreased** from year } t \text{ to year } t+1. \end{cases}$$

The correlations between the posterior means of  $\frac{A_{t+1}}{A_t}$  and  $P_t$ ,  $\phi_{a,t}$  or  $\phi_{j,t}$  are calculated, as an exploratory analysis, to investigate the importance of changes in the demographic parameters on adult abundance (Table 3.2).

**Table 3.2** Correlations ( $\rho$ ) between changes in adult abundance ( $\frac{A_{t+1}}{A_t}$ ) and the demographic parameters, calculated using the posterior mean of i) the baseline estimates (for  $\phi_{a,t}$  via *Llive*) and ii) the estimates from the fully integrated model using the full set of sites.

	Sedge Warbler		Reed Warbler	
	Baseline	Integrated	Baseline	Integrated
$P_t$	0.33	0.27	0.21	0.20
$\phi_{a,t}$	0.47	0.69	0.62	0.68
$\phi_{j,t}$	0.25	0.46	0.33	0.77

As expected, the correlations between the underlying demographic rates and the change in adult abundance are positive (Table 3.2). Changes in  $A_t$  seem more highly correlated with survival than productivity. The correlation with adult survival is particularly strong suggesting that this demographic parameter is driving the size of the Sedge and Reed Warbler adult populations.

We note that for both Sedge and Reed Warbler the baseline estimates of  $\phi_{j,t}$  are less correlated with  $\frac{A_{t+1}}{A_t}$  than the estimates under the integrated model. It is possible that the integrated model has contrived a strong correlation with  $\phi_{j,t}$  as it is this parameter, for which direct information is limited, that tends to be adjusted upon integration to make the population model match the observed

adult count data. Conversely, the baseline model may have failed to pick up the relationship due to the sparsity of the data.

Correlations between the posterior means of the underlying demographic parameters are given in Table 3.3. For both Sedge and Reed Warbler a positive correlation between  $\phi_{a,t}$  and  $\phi_{j,t}$  is suggested, and indeed it is biologically reasonable that some similar factors act upon the annual survival rates of both juvenile and adult birds. For Sedge Warbler,  $P_t$  and  $\phi_{a,t}$  appear positively correlated, whereas for Reed Warbler the reverse appears to apply (although this appears fairly negligible). Favourable environmental factors for adult survival that are also favourable for productivity would result in a positive correlation. A negative correlation could be the result of a cost to fitness incurred by adults due to the energy expenditure required to successfully raise a large number of young. The integrated analysis, of both Sedge and Reed Warbler data, suggests a negative correlation between  $P_t$  (or  $J_t$ ) and  $\phi_{j,t}$ . Such a relationship could be due to increased competition as a result of high juvenile abundance. The lack of consistency in the strength and direction of this correlation suggested by the baseline model is due to the limited amount of direct information on  $\phi_{j,t}$ .

**Table 3.3** Correlations ( $\rho$ ) between the demographic parameters calculated using the posterior mean of i) the baseline estimates (for  $\phi_{a,t}$  via  $L_{live}$ ) and ii) the estimates from the fully integrated model (based on the full set of sites).

	Sedge Warbler		Reed Warbler	
	Baseline	Integrated	Baseline	Integrated
$P_t, \phi_{a,t}$	0.32	0.25	-0.09	-0.17
$P_t, \phi_{j,t}$	0.33	-0.26	-0.11	-0.35
$J_t, \phi_{j,t}$	0.04	-0.54	-0.61	-0.49
$\phi_{a,t}, \phi_{j,t}$	0.22	0.18	0.56	0.43

We emphasise that correlation analyses above are entirely exploratory. A much more rigorous investigation would be required to confirm these relationships.

## 3.6 Concluding Remarks

Within this chapter we have described the first integrated model appropriate for CES data, and applied our approach to the analysis of Sedge and Reed Warbler data. Several advantages of this new model are identified, among which is the ability to accurately determine juvenile survival, estimate true seasonal productivity, and produce estimates of the demographic parameters that are both consistent with the totality of the data and each with one another according to the relationship imposed by the underlying population dynamics.

### 3.6.1 Integrated Modelling of UK Bird Populations

The use of integrated modelling with an aim to monitoring UK bird populations is not new. Previous research has combined data from several of the BTO's long-running bird monitoring schemes. The earlier approaches, including Peach et al. (1999), Siriwardena et al. (2001) and Freeman and Crick (2003), fitted deterministic population models to indices of abundance derived beforehand from census data, for example from the Common Bird Census (CBC), while fixing the demographic parameters in these population models to values estimated from *a priori* analyses of nest record and ring-recovery data. Consequently these earlier approaches failed to account fully for the uncertainty in the estimation of either the indices of abundance or the demographic rates.

Besbeas et al. (2002) furthered the deterministic approaches above, although their method still modelled an index of abundance derived previously from census data rather than the raw counts. They used a state-space model for this index in combination with the usual Multinomial-based models for the ring-recovery data, forming a combined likelihood that integrates both census and ringing information. Such an approach enables accurate treatment of the imprecision and correlation in the demographic estimates, and permits stochastic, rather than deterministic, population evolution. Brooks et al. (2004, 2008) and King et al. (2008b) demonstrated how this model could be re-cast into a Bayesian framework.

Freeman et al. (2007b) fitted deterministic population models directly to the raw census data but again fixed the underlying demographic parameters to previously estimated values, ignoring their associated uncertainty. Besbeas and Freeman (2006) essentially combine the methods of Besbeas et al. (2002) and

Freeman et al. (2007b) thereby correctly accounting for both the correlation and sampling variance in both the derived abundance indices and the demographic parameters. However the deterministic population model is retained.

The integrated model, appropriate for CES data, that we have developed combines for the first time the Bayesian approach of Brooks et al. (2004, 2008) and King et al. (2008b), and the fitting to site-specific data arrays of Besbeas and Freeman (2006). The various integrated analyses outlined above all use partial census data, i.e. territory counts at a sample of sites, to gain information about abundance, whereas our approach to Integrated Population Modelling adopts survey data from the CES ringing scheme for this purpose.

The CES scheme is particularly amenable to an integrated approach to population modelling as it provides a cheap, a practical, and a high quality source of information on several of the key demographic parameters: adult abundance, juvenile abundance, productivity and adult survival. Further, for the majority of common songbirds monitored under this scheme it provides the best, if not only, source of information on some of these parameters. Hence the development of an integrated model for CES data is of particular value.

### 3.6.2 An Integrated Model for CES Data - advantages

In general, CES data alone will not enable the estimation of juvenile survival, as these birds, unlike adults, are not sufficiently site faithful (Baillie et al., 1986; Wernham et al., 2002). We have adopted an entirely separate source of data, the recoveries of dead birds from the NRR database, to provide information on this parameter. However, for most small songbird species, especially long-distance migrants, these data are too few to provide precise survival estimates when analysed in isolation (Baillie et al., 1986). We have shown that the integrated approach has the advantage of providing precise inference on juvenile survival even with only limited ring-recovery data. By integrating the ring-recovery data, albeit sparse, in the integrated models for Sedge and Reed Warbler we have produced the first precise estimates of juvenile survival for these two species with any practical value.

Furthermore, being able to estimate the productivity scaling factor,  $k$ , also means that for the first time measures of productivity over an entire season (i.e. uniting multiple broods, and allowing for immediate post-fledging mortality)

are available, rather than an index merely proportional to it. Typically Sedge Warblers lay one clutch per year, but occasionally two. The integrated model enables estimation of the true seasonal productivity which includes these multiple attempts, as will be the norm for many other species such as Reed Warblers. Hitherto productivity rates in integrated models have been previously estimated either as a free parameter in the population model without direct data describing it (Besbeas et al., 2002, 2003; Brooks et al., 2004; Besbeas and Freeman, 2006; King et al., 2008b), or by including estimates of breeding success (at least on a per nesting attempt basis) from nest record data (Freeman et al., 2007b; Reynolds et al., 2009).

Previous integrated models combine demographic information from entirely separate sources of data, requiring the need to ensure that these relate to the same population. We too combine two separate sources of data (CES and NRR), but unlike the previous approaches we use the same data set to provide direct information on abundance, productivity and adult survival (when adopting CES data for this). By using CES data, albeit augmented with additional information on juvenile and adult survival from the NRR database, we ensure more consistency in the sources of information employed.

A key advantage to the integrated approach, over the independent analyses, is that it provides estimates of abundance, productivity and survival consistent with both the CES and NRR data sets, and according to the relationship imposed by the population model, such as Equation (3.1), with one another. For example, if the CES live-recapture data imply a particular value for  $\phi_{a,t}$  with little precision, the integrated model mitigates against such a value if it is not compatible with the more precise estimates of abundance and productivity from the CES count data, and the population model relating these terms.

We have successfully applied the integrated model to Sedge Warbler and Reed Warbler data, two of the most frequently encountered birds under the CES scheme. However the integrated approach presented here has a much wider applicability. Not only does the CES scheme provide sufficient data to routinely monitor 23 other species (Grantham and Robinson, 2008), but there are many other “Constant Effort” ringing programmes directly modelled on the BTO’s CES scheme in Europe, and the MAPS scheme in the USA (DeSante et al., 1999), that generate similar data. These data sets are also amenable to modelling using our integrated approach.

### 3.6.3 Non-Independence of Likelihoods

Captures of individual adult birds contribute to both the adult CES count data and the CES live-recapture data. This complete overlap in the data used to derive the adult abundance likelihood ( $L_a$ ) and the live-recapture likelihood ( $L_{live}$ ) means that they cannot be considered independent. Using the full data to derive  $L_{live}$  and  $L_a$  seems to result in unbiased estimates of the demographic parameters. However we note that the variability estimates are likely to be slightly under-reported as a result of, essentially, using the same data twice (Borysiewicz, 2008).

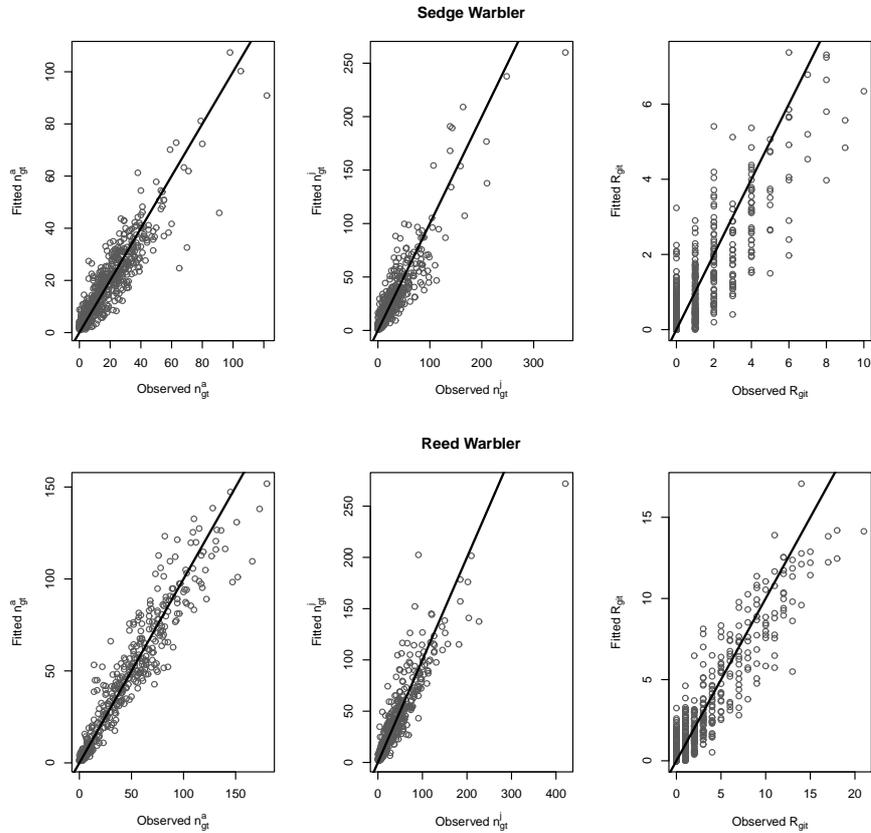
As adult birds are highly site-faithful, contributing data to only one site, the issue of non-independence between  $L_{live}$  with  $L_a$  can be overcome by randomly splitting the CES data according to site into two sets, prior to analysis, at the cost of a small increase of variance. Such an approach will not introduce bias, but the resulting split may be sub-optimal in terms of precision. Finding the optimal split (i.e. should it be 25/75 say), or developing a more rigorous approach in this respect, is a potential avenue of research. We also note that stratification of the sites based on the habitat type and location, in addition to years in which the site contributed data (as used here), may further improve the insensitivity to the split.

A disadvantage of a split data integrated analysis is that it can incur a cost to precision for some demographic parameters compared to their baseline (see Figures 3.6, 3.10). However, the integrated approach still has the advantage of providing estimates of the demographic parameters that are consistent with one another (according to the relationship imposed by the population model). Moreover, the integrated approach enables total seasonal productivity to be estimated, and is the only means of obtaining usefully precise estimates of juvenile survival.

### 3.6.4 Goodness-of-Fit

To assess the goodness-of-fit of the integrated model to the observed data, we follow the graphical approach of Besbeas et al. (2002), and compare the fitted counts and recaptures with their observed values in Figure 3.12. Note that the fitted values are derived using the posterior means of the demographic parameters. Overall, there is no serious indication of a lack of fit, with the

diagnostic plots in Figure 3.12 suggesting good agreement between the model and the CES count and live-recapture data for both Sedge and Reed Warbler.



**Figure 3.12** Graphical checks of the goodness-of-fit of the integrated model (based on the full set of sites). For both Sedge and Reed Warbler, from left to right, are plotted i) the observed versus fitted adult count ( $n_{gt}^a$ ), ii) the observed versus fitted juvenile count ( $n_{gt}^j$ ), and iii) the observed number of recaptures versus the expected number in the previously caught component of the live-recapture data ( $R_{git}$ ).

### 3.6.5 Potential Developments

The current analysis provides wide scope for development. To date we have restricted our analysis to the subset of complete capture records. The ability to improve precision by including additional data for years with missed visits is very appealing, however for these years the assumption of constant effort is violated and requires special treatment. In the following chapter (Chapter 4) the independent baseline models for abundance, productivity and adult survival are modified so that CES data with incomplete coverage can be included. In Chapter 5 the current integrated analysis is extended to include such data.

At present our integrated analysis assumes total site-fidelity of adult birds, enabling the estimates of adult survival from the CES live-recapture data to be considered that of “true survival”. However, this assumption is almost certainly false, although the extent to which it is violated is unknown. As breeding philopatry of Sedge and Reed Warblers is high (Wernham et al., 2002) departures from this assumption are believed to be minor; it is this that makes the CES scheme a viable proposition for monitoring. Extensions to the current integrated analysis that enables the possibility of permanent emigration away from the CES site, and the magnitude and importance of permanent emigration to be formally assessed, are developed in Chapter 5.

## 3.7 Chapter Summary

CES data enables abundance, productivity and adult survival to be monitored, and as such are particularly amenable to an integrated approach to population modelling. Efficient and independent models for each of these demographic parameters, adopting CES data with complete coverage, were developed in Chapter 2. Building on this methodology, and using a deterministic demographic model, in this chapter we have produced the first fully integrated population model based on CES data. Using Bayesian methods, this integrated model has been successfully applied to both Sedge Warbler and Reed Warbler data.

Within this chapter we have identified specific advantages of the integrated approach, key of which is the ability to provide estimates of abundance, productivity and survival that are consistent with data, and according to the relationship imposed by the underlying population dynamics, with one another. The methodology also enables juvenile survival, for which data are sparse, to be accurately determined. Furthermore, estimates of true seasonal productivity are readily obtained.

The integrated model presented here adopts only a subset of the CES data, that from years with complete coverage. The potential to improve precision by including all data is appealing. With this in mind, in following chapter we extend the current methodology, developing parsimonious, although independent, models for the demographic parameters (abundance, productivity and adult survival) appropriate for CES data containing years with incomplete coverage.



## Chapter 4

# Independent Statistical Models for CES Data with Incomplete Coverage

The fundamental assumption underpinning the CES scheme, and the independent demographic models presented in Chapter 2, from which the integrated population model of Chapter 3 is formed, is that of “constant effort”. In practice, there is potentially considerable deviation from this ideal arising from the absence of data from isolated individual visits within a site’s period of operation, some years receiving fewer than the full 12 mist-netting sessions. Reasons for missed visits may include, for example, bad weather, or the volunteer ringers’ personal circumstances rendering them unavailable. Ignoring missed visits, and assuming all visits were made, fails to account for the reduced capture rates and underestimates the number of birds at a site. As a result of missed visits the constant effort assumption is therefore violated. One *ad hoc* approach to address this, used previously in Chapters 2 and 3, would be to omit data from years with “incomplete coverage”. However, this wastes valuable information, and potentially decreases the precision of the parameter estimates. In both the Sedge and Reed Warbler CES data sets nearly half of the annual records have incomplete coverage and are thus inappropriate for analysis via the independent models of Chapter 2 or the integrated models of Chapter 3.

When one or more within-year visits are missed a “censored” yearly-site count results. Such a count may have been higher had all 12 visits been made. Peach et al. (1998) proposed a non-parametric approach to address this problem in the

modelling of abundance, correcting censored yearly-site counts for missed visits prior to model fitting. This *a priori* correction was also adopted by Robinson et al. (2007) in the modelling of productivity. An important shortcoming of Peach et al.'s (1998) approach is that, by failing to account for the uncertainty in the corrected counts, the precision in the parameter estimates is over-reported.

In applying the Peach et al. (1998) *a priori* correction method to CES data, Miles et al. (2007) showed that for 7 species the estimated ratios of juvenile to adult birds (a measure of productivity) were similar to those based only upon data without missed visits (i.e. the subset of “complete coverage” data) and hence are unbiased by any such correction. Otherwise, this important step in the analysis of the CES count data has received little attention, nor has any alternative been proposed. Such an alternative must be unaffected by the sampling imbalance invariably observed - both the numbers of birds caught, and the likelihood of a visit being missed, varies across the mist-netting season. We propose a Bayesian correction method in which imputation is simultaneous with model fitting. By incorporating the additional uncertainty arising from the correction for missed visits within the model fitting algorithm, the Bayesian approach more accurately reports the precision in the parameter estimates.

Under the CES protocol, parameters associated with the site-specific recapture rates in the survival models of Chapter 2 are assumed to be constant across time, simplifying the model structure. However, a year with fewer than 12 visits is expected to have a lower recapture rate than one where all 12 are made. Further, this reduction is likely to depend upon which visits were missed. To ensure time-invariance, previous studies have analysed only the subset of CES data with complete coverage (i.e. analyses of Chapter 2), or set *a priori* criteria for including data with a minimum number of visits made, for example at least 5 of the first 6 visits (Peach et al., 1995a). We provide the first attempt to explicitly account for missed visits in the modelling of survival, from CES data, that negates the need to allow for full time dependence in the recapture rates.

In this chapter we describe in detail novel abundance, productivity, and survival models, cast in a Bayesian framework, for CES data in which missed visits have resulted in non-constant sampling effort. Using simulation studies we assess the models, comparing the results to the underlying “truth” and to those obtained using the current approach of the BTO. The models are applied to Sedge Warbler and Reed Warbler CES data, and compared to the analyses of Chapter 2 in which only complete coverage data was employed.

The demographic models for CES data with incomplete coverage described in this chapter are independent of one another. We return to an integrated approach to population modelling in Chapter 5, combining the methodology presented here with that developed in Chapter 3.

## 4.1 Data

As sites with sparse data contribute little information to the analysis at the cost of extra site-specific parameters, following Chapter 2, such sites are excluded even for this incomplete coverage data analysis. We restrict the Sedge and Reed Warbler data sets to include only those sites operated for 4 years or more, of which at least 2 years had complete coverage, that caught appreciable numbers of Sedge/Reed Warblers (at least 10 adults, and at least 10 juveniles over the duration of operation). The reduced “best” sites, Sedge and Reed Warbler data sets contain 103 and 82 sites respectively (Table 4.1). By including data from years with incomplete coverage, the total number of captures (juvenile and adult birds combined) in the full CES data now represented in the reduced “best” sites data set has increased from 49% to 79% for Sedge Warblers, and from 54% to 85% for Reed Warblers in comparison with the analyses of Chapter 2 (Tables 2.1, 4.1).

In the “best” sites data set, for both Sedge and Reed Warbler, 33% of the yearly-site records are incomplete (Table 4.1). In general only a few visits, if any, tend to be missed, with only  $\approx 9\%$  of the years with incomplete coverage having 4 or more missed visits.

Incomplete coverage generally occurs at random, or essentially so, as a consequence of bad weather, net damage, volunteer unavailability etc. However, the probability that any one of the 12 within-year visits are missed is not equal, for example later visits are more prone to being missed than earlier visits (Table 4.2). Furthermore, if a visit is missed then it is more likely that several other visits will also be missed (Table 4.3). Such a scenario may be caused by a long spell of bad weather, or result from some ringers being less dedicated than others.

**Table 4.1** Sedge and Reed Warbler data sets which include years with incomplete coverage: the number of adult and juvenile captures (excluding multiple within-year encounters), sites included, yearly-site records, and those yearly-site records with incomplete coverage (percentage) in the reduced CES data set of “best” sites and the full CES data set, 1987-2005. The right-hand column gives the percentage in the full data represented in the “best” sites data.

	Best sites	Full data	Percentage
<b>Sedge Warbler</b>			
Adult captures	16862	22085	76%
Juvenile captures	23269	28605	81%
Sites	103	283	36%
Yearly-site records	963	1848	52%
Incomplete yearly-site records	317 (33%)	821 (44%)	39%
<b>Reed Warbler</b>			
Adult captures	27902	34792	80%
Juvenile captures	30179	33347	90%
Sites	82	241	34%
Yearly-site records	766	1564	49%
Incomplete yearly-site records	253 (33%)	678 (43%)	37%

**Table 4.2** Percentage of times each individual within-year visit is missed. Sedge and Reed Warbler “best” sites data, 1987-2005.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>Sedge Warbler</b>	2.9	3.7	3.9	4.4	4.3	5.7	4.5	5.0	4.7	5.7	5.2	8.4
<b>Reed Warbler</b>	3.0	4.3	4.4	4.7	3.8	5.7	4.2	4.4	4.7	5.1	4.8	7.4

**Table 4.3** Observed and expected percentages of missed within-year visits in the Sedge and Reed Warbler “best” sites data, 1987-2005. The expected percentage is calculated using a Binomial distribution, where the probability a visit is missed is given by the observed proportion of missed visits (0.049 and 0.047 for Sedge and Reed Warblers respectively), and it is assumed that visits are missed independently.

number missed	0	1	2	3	4+	
<b>Sedge Warbler</b>						
Observed	67.1	19.9	6.7	3.4	2.9	$\chi_4^2 = 421.1$
Expected	54.7	33.8	9.6	1.6	0.3	
<b>Reed Warbler</b>						
Observed	67.0	21.0	6.4	2.7	2.9	$\chi_4^2 = 371.4$
Expected	56.1	33.2	9.0	1.5	0.2	

## 4.2 Estimating Abundance from CES Data Containing Censored Counts

Using count data from the CES scheme the BTO regularly produces indices of abundance, enabling the status of UK breeding bird populations to be monitored. However, a feature of these data is that some yearly-site counts are “censored” due to visits missed within certain years. Peach et al. (1998) developed an intuitive, non-parametric method that corrects data for missed visits prior to model-fitting in the form of a Poisson regression model with an additive offset. This classical method, currently routinely employed by the BTO, assumes the offset is known *a priori* and without error. By ignoring the uncertainty in the correction, the precision in the estimated abundance indices is over-reported. We propose an alternative Bayesian approach that produces annual indices of abundance whose uncertainty also incorporates a component due to the correction for missed visits. The novel Bayesian approach (first presented in Cave et al., 2009a), and the classical approach by Peach et al. (1998), are both described in detail and compared using simulated data, before being applied to Sedge Warbler and Reed Warbler CES data from the “best” sites. We discuss these methods before considering a more computationally intensive approach in which, by modelling the full visit-level count data rather than simply the annual count totals, missing visits do not pose a problem.

### 4.2.1 Classical Model of Peach et al. (1998)

The method of Peach et al. (1998), employed routinely by the BTO, *a priori* corrects censored yearly-site counts for missed visits by estimating the number of additional individual birds that would have been caught had the full complement of 12 visits been made. Underlying this approach is the assumption that while there is variation throughout the mist-netting season in the numbers of birds caught, this variation is consistent between years, otherwise bias is likely to be introduced.

For the adult CES count data we define the following quantities:

$n_{gt}^a$  denotes the number of unique adult birds caught at site  $g$  in year  $t$  given that all 12 visits are made. If less than 12 visits are made  $n_{gt}^a$  is unknown but represents the total number of individuals that would have been caught had all 12 visits been made.

$\ell_{gt}^a$  denotes the observed number of unique adult birds caught at site  $g$  in year  $t$ . When one or more visits are missed this is a “lower bound” for  $n_{gt}^a$ , and referred to as a “censored count”. In years of complete coverage, that is when all 12 visits are made,  $\ell_{gt}^a = n_{gt}^a$ .

$N_g^a$  denotes the total number of unique adult birds caught at site  $g$  for all years of complete coverage.

$N_{gt}^a$  denotes the number of adult birds that belong to  $N_g^a$  that are caught at visits corresponding to those made in year  $t$ . Thus, for example, if visits 2 and 5 at site  $g$  in year  $t$  were missed then  $N_{gt}^a$  is equal to  $N_g^a$  minus the number of those birds caught only at visits 2 and/or 5. When all visits are made to site  $g$  in year  $t$ ,  $N_{gt}^a = N_g^a$ .

In a year of missed visits, the corrected adult yearly-site count,  $E_{gt}^a$ , proposed by Peach et al. (1998) for site  $g$  in year  $t$  is

$$E_{gt}^a = \ell_{gt}^a \frac{N_g^a}{N_{gt}^a}, \quad (4.1)$$

which, it is readily verified, reduces to  $n_{gt}^a$  in a year of complete coverage. The number of birds caught varies between visits due to migratory behaviour, thus the ratio of  $N_g^a$  to  $N_{gt}^a$  accounts not only for the number of missed visits but also their precise timing. However, the approach is conservative in its overall effect, mitigating against extremely high or low corrected values.

Recall that the Poisson model for adult abundance, based on complete coverage data, described in Chapter 2, is given by:

$$n_{gt}^a \sim \text{Poisson}(\lambda_{gt}^a),$$

where  $\ln(\lambda_{gt}^a) = \beta^a + s_g^a + y_t^a$ . (Equation (2.1))

The corrected adult counts will not, in general, have integer values as required by the Poisson model described above. However, noting that the difference between the log of the observed and corrected adult counts can be expressed as

$$\begin{aligned} \ln(\ell_{gt}^a) - \ln(E_{gt}^a) &= \ln(\ell_{gt}^a/E_{gt}^a) \\ &= \ln(N_{gt}^a/N_g^a), \end{aligned}$$

we assume the observed (possibly censored) adult counts have a Poisson distribution, i.e.  $\ell_{gt}^a \sim \text{Poisson}(\lambda_{gt}^a)$ , where the natural logarithm of the mean is given by:

$$\begin{aligned}\ln(\lambda_{gt}^a) &= \beta^a + s_g^a + y_t^a + \ln\left(\frac{\ell_{gt}^a}{E_{gt}^a}\right) \\ &= \beta^a + s_g^a + y_t^a + \ln\left(\frac{N_{gt}^a}{N_g^a}\right).\end{aligned}\quad (4.2)$$

Without missing visits, the expected count at site  $g$ , in year  $t$ , is simply derived from an intercept ( $\beta^a$ ), the  $g^{\text{th}}$  site effect ( $s_g^a$ ) and  $t^{\text{th}}$  year effect ( $y_t^a$ ), additive on the logarithmic scale, and is identical to the model for data with complete coverage of Chapter 2, described by Equation (2.1).

The right-most term in Equation (4.2) is an offset, the log-transformed ratio of the observed count (albeit not necessarily from 12 visits) and the corrected value, used to adjust for the missed visits where appropriate. This model is readily fitted by maximum likelihood in any Generalized Linear Modelling package, for example, as in the current research, using the `glm` function with the `offset` argument in R (R Development Core Team, 2007).

Analogous quantities,  $n_{gt}^j$ ,  $\ell_{gt}^j$ ,  $N_g^j$ , and  $N_{gt}^j$ , are readily defined for the juvenile CES count data, from which the corrected juvenile yearly-site counts, denoted by  $E_{gt}^j$ , are derived. Likewise, the juvenile abundance model assumes that the  $\ell_{gt}^j$  have a Poisson distribution with mean  $\lambda_{gt}^j$  given by:

$$\ln(\lambda_{gt}^j) = \beta^j + s_g^j + y_t^j + \ln\left(\frac{N_{gt}^j}{N_g^j}\right).\quad (4.3)$$

For identifiability, the first year effect and an arbitrary reference site effect in both Equations (4.2) and (4.3) are constrained to zero. Analogous to the models in Chapter 2, indices of adult abundance and juvenile abundance are given by  $A_t = \exp(y_t^a)$  and  $J_t = \exp(y_t^j)$  respectively.

Sites without any years of complete coverage cannot be accommodated quite this way, as  $N_g^a = N_{gt}^a = 0$  and  $N_g^j = N_{gt}^j = 0$ . An analogous correction is thus employed based upon catches at all sites, rather than catches only from the site in question (Peach et al., 1998) - a “global” rather than a “local” correction factor. Note, in the reduced data set of “best” sites we consider in this chapter this situation does not arise as, by definition, all sites contributing data have at least 2 years of complete coverage.

From Equation (4.1) it is clear that observed censored counts of zero, i.e.  $\ell_{gt} = 0$ , are given a corrected value of zero. Accordingly, in a situation where the data

contain a higher number of censored counts at zero, as a result of a large number of missed visits within each year, the method of Peach et al. (1998) will introduce bias. Thus, criteria for correction are imposed: an observed count,  $\ell_{gt}$ , is omitted from the analysis altogether if at site  $g$  in year  $t$ :

1.  $< 8$  visits are made,
2. for adult counts,  $< 4$  of the first 6 visits are made,
3. for juvenile counts,  $< 4$  of the final 6 visits are made.

Additional criteria for correction, set by Peach et al. (1998) and employed by the BTO, impose that if, at site  $g$ , fewer than 10 individuals are caught during all years with complete coverage (i.e.  $N_g^a < 10, N_g^j < 10$ ) the global correction is used.

#### 4.2.2 Alternative Bayesian Model for Censored Counts

As the Peach et al. (1998) approach above corrects censored counts, and treats them as known before the model parameters are estimated, no account is taken of their inherent uncertainty. We now describe an alternative Bayesian approach, in which the imputation and the model fitting are combined in a single process, and the associated parameter uncertainty is more accurately determined (Cave et al., 2009a). We consider modelling the full visit-level data in Section 4.2.6, another alternative approach, but here we still consider the annual count data in which censoring is an issue. The Poisson distribution is retained, however the lower bound  $\ell_{gt}$  is explicitly incorporated into the likelihood for those cases with fewer than 12 visits. A full, non-censored count for these is estimated but this is not required prior to model fitting. The approach is identical for both adult and juvenile data, hence the adult ( $a$ ) and juvenile ( $j$ ) superscripts have been omitted in the description that follows.

Recall that  $\ell_{gt}$  denotes the observed yearly-site count for site  $g$  in year  $t$ . If all 12 visits were made, then  $\ell_{gt} = n_{gt}$ . We denote such “fully observed” counts by  $n_{gt}^{obs}$ . If less than 12 visits were made  $n_{gt}$  is unknown, with  $\ell_{gt}$  providing partial information that  $n_{gt} \geq \ell_{gt}$ . We denote such counts by  $n_{gt}^{cen}$ , and their observed lower bounds by  $\ell_{gt}^{obs}$ . We refer to  $n_{gt}^{cen}$  as the “imputed censored count”.

We assume that  $n_{gt}^{obs}$  and  $n_{gt}^{cen}$  are independent, random variables from a Poisson distribution with parameter  $\lambda_{gt} = \exp(\beta + s_g + y_t)$ . The distribution of  $\ell_{gt}^{obs}$  is unknown.

## i Auxiliary Variables

Given the observed data,  $(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs})$ , the posterior distribution for the unknown model parameters (denoted by  $\boldsymbol{\theta}$ , where  $\boldsymbol{\theta} = (\beta, \mathbf{s}, \mathbf{y})$ ) is given by:

$$\pi(\boldsymbol{\theta} | \mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}) \propto f(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs} | \boldsymbol{\theta}) \times p(\boldsymbol{\theta}).$$

However  $\pi(\boldsymbol{\theta} | \mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs})$ , up to the proportionality constant, is not obtainable as the distribution for  $\boldsymbol{\ell}_{gt}^{obs}$  is not known. We therefore modify the posterior distribution by introducing ‘‘auxiliary variables’’ to act as surrogates for the unobserved counts,  $n_{gt}^{cen}$ , assumed to be Poisson distributed. These auxiliary variables are treated as random variables to be estimated.

Let  $\eta_{gt}$  denote the level to which  $n_{gt}^{cen}$  is censored such that  $n_{gt}^{cen} = \ell_{gt}^{obs} + \eta_{gt}$ . That is,  $\eta_{gt}$  corresponds to the unknown number of additional unique birds that would have been caught had the visits missed been made. Clearly  $\eta_{gt}$  is an integer value in the range  $[0, \infty)$ .

Treating all  $\eta_{gt}$  values as parameters to be estimated, i.e. auxiliary variables, the modified posterior distribution is given by:

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\eta}_{gt} | \mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}) &\propto f(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt} | \boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \\ &= f(\mathbf{n}_{gt}^{obs} | \boldsymbol{\theta}) \times f(\boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt} | \boldsymbol{\theta}) \times p(\boldsymbol{\theta}). \end{aligned} \quad (4.4)$$

Note that  $f(\boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt} | \boldsymbol{\theta}) \equiv f(\mathbf{n}_{gt}^{cen} | \boldsymbol{\theta})$  with  $\mathbf{n}_{gt}^{cen} = \boldsymbol{\ell}_{gt}^{obs} + \boldsymbol{\eta}_{gt}$ . Both  $f(\mathbf{n}_{gt}^{obs} | \boldsymbol{\theta})$  and  $f(\boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt} | \boldsymbol{\theta})$  are the product of independent Poisson distributions with  $\boldsymbol{\theta} = (\beta, \mathbf{s}, \mathbf{y})$ . Samples of the unknown model parameters and the auxiliary variables are readily drawn from  $\pi(\boldsymbol{\theta}, \boldsymbol{\eta}_{gt} | \mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs})$  using a Metropolis-Hastings algorithm (Gelman et al., 2004).

## ii Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is employed to i) update the model parameters  $\{\beta, \mathbf{s}, \mathbf{y}\}$  and ii) update the imputed censored counts  $(\mathbf{n}_{gt}^{cen} = \boldsymbol{\ell}_{gt}^{obs} + \boldsymbol{\eta}_{gt})$ , in two steps as described below.

Recall that at iteration  $h$  of the Metropolis-Hastings algorithm, a proposed value for parameter  $\theta_j$  (denoted by  $\theta_j^*$ ) is accepted with probability:

$$\alpha(\theta_j^h, \theta_j^*) = \min \left( 1, \frac{\pi(\theta_j^* | \mathbf{x}) q(\theta_j^h | \theta_j^*)}{\pi(\theta_j^h | \mathbf{x}) q(\theta_j^* | \theta_j^h)} \right), \quad (4.5)$$

where  $\theta_j^h$  denotes the current value,  $\mathbf{x}$  the data,  $\pi$  the posterior distribution, and  $q$  the proposal distribution (see Section 1.5).

*Step One: Updating the Model Parameters*

Assuming that the yearly-site counts, denoted by  $n_{gt}$ , are independent, random variables from a Poisson distribution with parameter  $\lambda_{gt} = \exp(\beta + s_g + y_t)$ , the “likelihood” at the end of iteration  $(h - 1)$  is given by:

$$f(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt}^h | \beta^h, \mathbf{s}^h, \mathbf{y}^h) = \prod_{g,t} \frac{\exp(-\exp(\beta^h + s_g^h + y_t^h)) \times \exp(\beta^h + s_g^h + y_t^h)^{n_{gt}^{obs}}}{n_{gt}^{obs}!} \\ \times \prod_{g,t} \frac{\exp(-\exp(\beta^h + s_g^h + y_t^h)) \times \exp(\beta^h + s_g^h + y_t^h)^{n_{gt}^{cen(h)}}}{n_{gt}^{cen(h)}!} \quad (4.6)$$

where  $n_{gt}^{cen(h)} = \ell_{gt}^{obs} + \eta_{gt}^h$ .

Vague, independent, Normal priors,  $p$ , with mean 0 and variance 10,000 are given to all model parameters. Normal proposal distributions are used for updating all parameters, with the proposal variance for each parameter tuned *a priori*. Recall that  $s_1$  and  $y_1$  are constrained to 0.

At iteration  $h$  the algorithm begins by updating  $\beta$ , then one by one the site and year effects. For example, to update  $\beta$ , where the current value is denoted by  $\beta^h$ , a new value,  $\beta^*$ , is proposed from the  $\text{Normal}(\beta^h, \sigma_\beta^2)$  distribution, which is accepted with probability:

$$\alpha(\beta^h, \beta^*) = \min \left( 1, \frac{f(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt}^h | \beta^*, \mathbf{s}^h, \mathbf{y}^h) \times p(\beta^*)}{f(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt}^h | \beta^h, \mathbf{s}^h, \mathbf{y}^h) \times p(\beta^h)} \right),$$

noting that, as the proposal distribution is symmetric it cancels out in the expression for the acceptance probability. If the move is accepted  $\beta^{h+1} = \beta^*$ , else  $\beta^{h+1} = \beta^h$ .

An update for the  $g^{th}$  site effect,  $s_g$ , is proposed from the  $\text{Normal}(s_g^h, \sigma_{s_g}^2)$  distribution which is accepted with probability:

$$\alpha(s_g^h, s_g^*) = \min \left( 1, \frac{f(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt}^h | \beta^{h+1}, s_2^{h+1}, \dots, s_{g-1}^{h+1}, s_g^*, s_{g+1}^h, \dots, s_G^h, \mathbf{y}^h) \times p(s_g^*)}{f(\mathbf{n}_{gt}^{obs}, \boldsymbol{\ell}_{gt}^{obs}, \boldsymbol{\eta}_{gt}^h | \beta^{h+1}, s_2^{h+1}, \dots, s_{g-1}^{h+1}, s_g^h, s_{g+1}^h, \dots, s_G^h, \mathbf{y}^h) \times p(s_g^h)} \right).$$

If the move is accepted  $s_g^{h+1} = s_g^*$ , else  $s_g^{h+1} = s_g^h$ .

Analogous updates are performed for the year effects,  $y_t$ .

*Step Two: Imputing Censored Counts*

By treating the censoring amounts,  $\eta_{gt}$ , similarly to the model parameters, their posterior distributions are estimated using a Metropolis-Hastings algorithm. As  $n_{gt}^{cen} = \ell_{gt}^{obs} + \eta_{gt}$ , where  $\ell_{gt}^{obs}$  is known, in practice we shall update the imputed counts  $n_{gt}^{cen}$ , corrected for censoring, directly.

To update an imputed censored count, where the current value is denoted by  $n_{gt}^{cen(h)}$ , a new value  $n_{gt}^{cen(*)}$  is proposed from the discrete Uniform[A,B] distribution. Here  $A = \max(\ell_{gt}^{obs}, n_{gt}^{cen(h)} - \delta_{gt})$  and  $B = n_{gt}^{cen(h)} + \delta_{gt}$ , where the random walk jump,  $\delta_{gt}$ , is a positive integer. We consider this non-symmetric proposal distribution to ensure that  $n_{gt}^{cen(*)} \geq \ell_{gt}^{obs}$ .

Under this scheme the proposal distribution is not necessarily symmetric, i.e. when  $n_{gt}^{cen(h)} - \delta_{gt} < \ell_{gt}^{obs}$  or  $n_{gt}^{cen(*)} - \delta_{gt} < \ell_{gt}^{obs}$ , and it must be incorporated explicitly into the expression for the acceptance probability. The proposal distribution is given by:

$$q(n_1|n_2) = \begin{cases} \frac{1}{n_2 + \delta_{gt} + 1 - \ell_{gt}^{obs}} & \text{if } n_2 - \delta_{gt} < \ell_{gt}^{obs} \\ \frac{1}{2\delta_{gt} + 1} & \text{otherwise,} \end{cases} \quad (4.7)$$

where  $n_1 = n_{gt}^{cen(*)}$  and  $n_2 = n_{gt}^{cen(h)}$ , or  $n_1 = n_{gt}^{cen(h)}$  and  $n_2 = n_{gt}^{cen(*)}$ .

Note that although the length of the random walk,  $\delta$ , can be allowed to differ between imputed censored counts, for convenience a common  $\delta$ , tuned *a priori*, is used in the updating step for all imputed censored counts. This proved to be adequate.

All likelihood terms in expression (4.6), aside from the contribution by the  $n_{gt}^{cen}$  being updated, cancel out in the expression for the acceptance probability. Noting that  $n_{gt}^{cen(*)} = \ell_{gt}^{obs} + \eta_{gt}^*$  and  $n_{gt}^{cen(h)} = \ell_{gt}^{obs} + \eta_{gt}^h$ , and denoting the likelihoods for  $n_{gt}^{cen(*)}$  and  $n_{gt}^{cen(h)}$ , evaluated at the updated parameter values, by  $f(\ell_{gt}^{obs}, \eta_{gt}^* | \beta^{h+1}, s_g^{h+1}, y_t^{h+1})$  and  $f(\ell_{gt}^{obs}, \eta_{gt}^h | \beta^{h+1}, s_g^{h+1}, y_t^{h+1})$  respectively, the proposed update is therefore accepted with probability:

$$\alpha(n_{gt}^{cen(h)}, n_{gt}^{cen(*)}) = \min \left( 1, \frac{f(\ell_{gt}^{obs}, \eta_{gt}^* | \beta^{h+1}, s_g^{h+1}, y_t^{h+1}) \times q(n_{gt}^{cen(h)} | n_{gt}^{cen(*)})}{f(\ell_{gt}^{obs}, \eta_{gt}^h | \beta^{h+1}, s_g^{h+1}, y_t^{h+1}) \times q(n_{gt}^{cen(*)} | n_{gt}^{cen(h)})} \right).$$

If the move is accepted  $n_{gt}^{cen(h+1)} = n_{gt}^{cen(*)}$ , else  $n_{gt}^{cen(h+1)} = n_{gt}^{cen(h)}$ .

Note that  $\eta_{gt}^{(h+1)} = n_{gt}^{cen(h+1)} - \ell_{gt}^{obs}$ .

*Gibbs Sampler*

As the posterior conditional distribution for the imputed censored count,  $n_{gt}^{cen}$ , being a Poisson distribution truncated at  $\ell_{gt}^{obs}$ , is of “semi-standard” form, a Gibbs sampler (Gelman et al., 2004) could be implemented. That is, using Rejection Sampling,  $n_{gt}^{cen(h)}$  is updated by sampling directly from the Poisson distribution with mean  $\lambda_{gt}^{h+1} = \exp(\beta^{h+1} + s_g^{h+1} + y_t^{h+1})$ . If  $n_{gt}^{cen(*)} < \ell_{gt}^{obs}$  the sampling process is repeated until a suitable update is drawn.

In practice, as we are frequently attempting to sample from within a very small tail of the Poisson distribution, it will often take many draws to find an acceptable update. Due to the inefficiency of Rejection Sampling the random walk Metropolis-Hastings algorithm, described in Step Two where updates are proposed according to Equation (4.7), is preferred.

*Implementation*

Bespoke code for implementing the Bayesian model, described above, is written in C. A total of 100,000 samples from the posterior distributions of the parameters and imputed censored counts were drawn after an initial burn-in of 100,000 updates was discarded. Point estimates for the model parameters are given by their posterior means, whereas for the imputed counts, as the posterior distributions tend to be skewed to the right, posterior medians are used. For all analyses described below, several runs using overdispersed starting points gave essentially identical results. Furthermore, examination of the trace-plots indicated good movement and mixing of the MCMC chain in all simulations.

For a site without complete coverage in any year, counts corrected for censoring cannot be reliably imputed as the data imposes no constraint on the estimate of the site effect. Recall that  $\ln(\lambda_{gt}) = \beta + s_g + y_t$ . If, for a particular site  $g$ , all count data are censored then the upper limit for  $s_g$  is only restricted by the prior on  $s_g$ . Data from such sites should be omitted from the analysis. This feature motivated the requirement for at least 2 years of complete coverage in the selection of the “best” sites CES data (see Section 4.1). An analogous argument holds for years in which complete coverage was not achieved at any site in operation, although in CES data collected by the BTO, for Sedge and Reed Warbler, this has not arisen.

*Overdispersion*

The Bayesian method described here can readily be adapted to accommodate different, possibly more appropriate models. For example, overdispersion is likely in count data from flocking birds that tend to be caught in groups. Although Peach et al. (1998) noted that overdispersion was negligible for the majority of adult CES data sets they considered, most juvenile CES data sets were overdispersed. When count data are overdispersed the Negative Binomial distribution provides an alternative to the Poisson distribution since it allows the variance to exceed the mean. The Bayesian method is readily formulated under a Negative Binomial model by simply replacing the Poisson likelihood, Equation (4.6), with the Negative Binomial likelihood and including an extra Metropolis-Hastings updating step for the overdispersion parameter. Note, this is equivalent to the Poisson-Gamma hierarchical model (see King et al., 2008b).

**4.2.3 Model Assumptions**

Both the classical and Bayesian approaches use the information provided by the non-censored counts, augmented with the extra information from the observed lower bounds of the censored counts, to impute counts corrected for missed visits. Both approaches assume the non-censored counts ( $\mathbf{n}_{gt}^{obs}$ ) are representative of censored counts ( $\mathbf{n}_{gt}^{cen}$ ), although the ways this is done are very different.

The classical Peach et al. (1998) approach uses the pattern of within-year catches, observed in the non-censored count data, to estimate the number of additional birds that would have been caught had all 12 visits been made. If non-censored and censored count data differ with respect to the pattern of within-year catches then such a correction is not reliable.

The Bayesian approach estimates posterior distributions for the imputed censored counts,  $n_{gt}^{cen}$ , by assuming that they arise from the Poisson distribution estimated using the non-censored counts, supplemented with information provided by the observed lower bounds. In the absence of a lower bound, the induced posterior for the imputed count is simply  $n_{gt}^{cen} = \exp(\beta + s_g + y_t)$ , derived from the posterior distributions of the model parameters estimated entirely from the non-censored count data. The lower bounds provide additional information, in particular that  $n_{gt}^{cen} \geq \ell_{gt}$ . Although this information is used to adjust the posterior distributions of the model parameters  $(\beta, \mathbf{s}, \mathbf{y})$ , such ad-

justments are modulated by the non-censored counts. If the underlying Poisson model differed between these two sets then it would not be possible to construct reliable posterior distributions for the imputed censored counts from the non-censored count data.

Hypothetical scenarios that violate the crucial assumption of representative non-censored count data can be constructed. However, in practice, assuming the non-censored counts are representative of the censored counts is sound, as the causes of censoring (ringer unavailability, net-damage, bad weather etc) are unrelated to the within-year pattern of catches or to the number of birds caught.

As the assumptions underlying the two approaches differ, the methods may perform better in different circumstances. For example, the Bayesian approach, which makes no assumptions regarding the pattern of within-year catches, is expected to perform better if the inter-annual variation is large, or if years with missed visits experience atypical catch patterns. On the other hand, when the underlying model is incorrectly specified, the model-based Bayesian approach may perform poorly, and the non-parametric Peach et al. (1998) approach may be preferable (Cave et al., 2009a).

#### 4.2.4 Testing the Bayesian Model - a Simulation Study

We now describe two analyses of simulated data. In the first, artificial data from a Poisson distribution with known parameter values are generated, and in which yearly-site counts are censored at random. We consider both a realistic level of censoring (based on that in the Sedge Warbler data), and one considerably more extreme than is likely to be seen in practice. Trends in abundance are then estimated via the Bayesian model to investigate its performance. As these simulated data sets are not determined at the individual within-year visit level, the *a priori* correction of Peach et al. (1998) is not possible. Using count data not predetermined to have arisen from the Poisson distribution, a second simulation then introduces an element of lack of fit in the Poisson model. Here visits are selected, at random, to be missed from a subset of the Sedge Warbler data with complete coverage. This allows a test of the Bayesian method in circumstances where the Poisson assumption is violated to a degree, and furthermore, as the data provide a complete record from all visits to a site in a given year, the method of Peach et al. (1998) can also be applied, and the two approaches compared.

*Artificial Poisson Data*

To obtain Poisson data the conventional Peach et al. (1998) abundance model, given by Equation (4.2), is first fitted to the entire adult Sedge Warbler CES count data, from 1986-2005, using all available 178 sites. The resulting maximum likelihood parameter estimates  $(\hat{\beta}^a, \hat{s}_g^a, \hat{y}_t^a)$  are then used to simulate Poisson yearly-site counts,  $n_{gt}$ , i.e.

$$n_{gt} \sim \text{Poisson}(\lambda_{gt}) \text{ where } \lambda_{gt} = \exp(\hat{\beta}^a + \hat{s}_g^a + \hat{y}_t^a).$$

A balanced data set (20 years of data for each of the 178 sites) is thus produced, and a set of estimates for the year effects, unaffected by censoring, obtained.

Each simulated yearly-site count,  $n_{gt}$ , is then either selected for censoring (with probability 0.35, approximately the proportion of censored counts in the Sedge and Reed Warbler CES data sets) or retained at its existing value. Selected counts are censored by subtracting an integer censoring amount, denoted by  $\eta_{gt}$ , from  $\{0, \dots, n_{gt}\}$  with probability:

$$\Pr(\eta_{gt} = i) = \begin{cases} \frac{1}{2^{i+1}} & \text{for } i \in [0, n_{gt} - 1] \\ \frac{1}{2^i} & \text{when } i = n_{gt}. \end{cases}$$

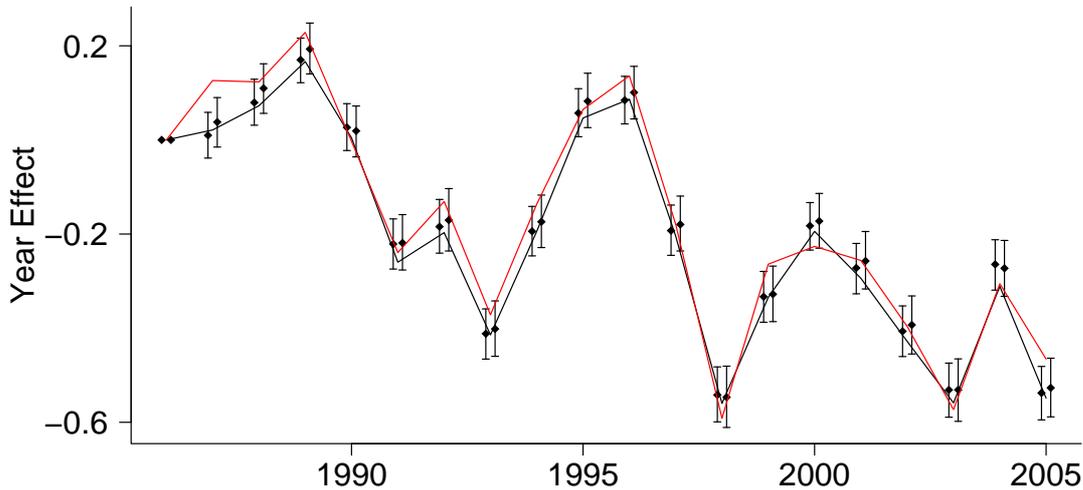
The “observed” censored count in the simulated data is given by  $\ell_{gt} = n_{gt} - \eta_{gt}$ .

Based on the Peach et al. (1998) corrected count, the above censoring mechanism produces data sets in which the proportion of censored counts, and the magnitude of censoring, approximates that predicted to be present in the Sedge Warbler data (Table 4.4). Note, a similar level of censoring is predicted to occur in the Reed Warbler data. The Bayesian approach is then used to estimate the year effects (posterior means) from this censored data for comparison with those from the full, pre-censored data (Figure 4.1).

**Table 4.4** Proportion of counts censored by the amount  $\eta$  in the simulated data under the two censoring mechanisms. The predicted level of censoring, in the observed adult Sedge Warbler CES data from 1986-2005, is obtained by the Peach et al. (1998) corrected count minus the “observed” censored count, i.e.  $E_{gt} - \ell_{gt}$ .

Level ( $\eta$ )	0	1	2	3	4	5	6	7	8+
<b>Predicted</b>	.51	.23	.12	.05	.03	.02	.02	.01	.02
<b>Simulated 1</b>	.57	.24	.10	.05	.02	.01	.002	.002	.005
<b>Simulated 2 (more severe)</b>	.27	.12	.08	.05	.06	.04	.04	.03	.29

The estimated year effects from the censored data are in close agreement with the true underlying values from the pre-censored data, with the 95% symmetric credible intervals including the true values in all years (Figure 4.1). The posterior medians of the imputed counts themselves are also in close agreement with their true (pre-censoring) values, though they are inflated, on average, by about 1.8. This apparent over-inflation occurs as the posterior distributions for the imputed counts tend to be skewed to the right (Figure 4.2). We note however, under this conservative censoring mechanism, simply ignoring the presence of censored counts, treating them as if they were observed in full and setting  $n_{gt} = \ell_{gt}$  regardless of whether censored or not, does not cause pronounced bias. Due to the limited level of censoring this lack of bias is to be expected.



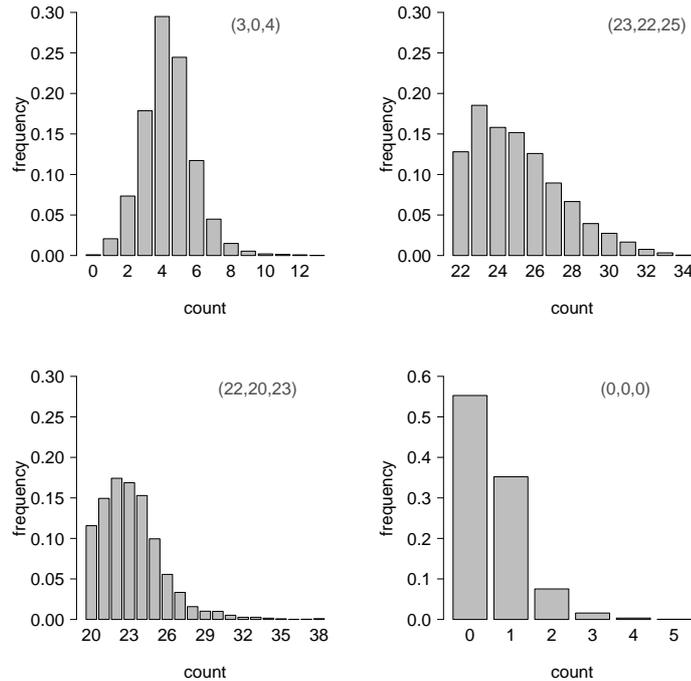
**Figure 4.1** Posterior means, and 95% symmetric credible intervals, for the estimated year effects in the Bayesian model. Estimates under the first and second (more severe) censoring mechanisms are on the left and right respectively. The black line, corresponding to estimates from the pre-censored data, represents the underlying truth. Estimates under the analysis which ignores the presence of censoring, as determined by the second (more severe) censoring mechanism, are shown by the red line.

Given the relatively limited effect of the censoring mechanism above, we repeat these analyses but choose integer censoring amounts,  $\eta_{gt}$ , with probability

$$\Pr(\eta_{gt} = i) = \frac{1}{n_{gt} + 1} \text{ for } i \in [0, n_{gt}].$$

That is, all possible censorings are chosen with equal probability resulting in selected counts being more severely censored (Table 4.4). The resulting censored counts are less strongly correlated ( $\rho = 0.81$ ) with the true (pre-censoring)

values than were the set from the previous simulations ( $\rho = 1.0$ ). In addition, ignoring censoring now causes bias in the estimated abundance trend (Figure 4.1). Nevertheless, in spite of this larger level of censoring the Bayesian approach performs well (Figure 4.1). The posterior medians of the imputed counts again are in agreement with their true values, albeit with some over-inflation due to the skewed posterior distributions.



**Figure 4.2** Four, arbitrarily chosen, posterior distributions for corrected censored counts imputed using the Bayesian approach. Data were initially simulated from a Poisson distribution and censored to a level comparable to the Sedge Warbler data. Given in brackets are the true (pre-censored) count, the observed lower bound, and the posterior median.

### *Introducing a Lack of Fit*

In the second scenario, to simulate data not predetermined to arise from a Poisson distribution, we randomly censor yearly-site counts in the Sedge Warbler complete coverage data, from the “best” sites, of Chapter 2. Censoring is achieved as follows:

1. Yearly-site counts are selected for censoring with probability 0.35.
2. For each selected count, 1 visit is chosen at random to be missed.
3. Additional visits, for each censored count, are then chosen to be missed with probability 0.05.
4. Captures recorded at the visits selected to be missed are removed and the now censored counts calculated.

Note that the same pattern of missed visits is applied to censor both adult and juvenile data, as will clearly occur in practice.

The level of censoring achieved is summarised in Table 4.5. The percentage of missed visits in the censored counts (13%) is similar to that observed in the Sedge Warbler “best” sites incomplete coverage data (14%). Censoring resulted in 239 (2.4%) adult birds, and 499 (3.4%) juvenile birds, no longer being “caught” within a given year. The censored counts are highly correlated with the true pre-censored values (adults:  $\rho = 0.99$ , juveniles:  $\rho = 0.98$ ).

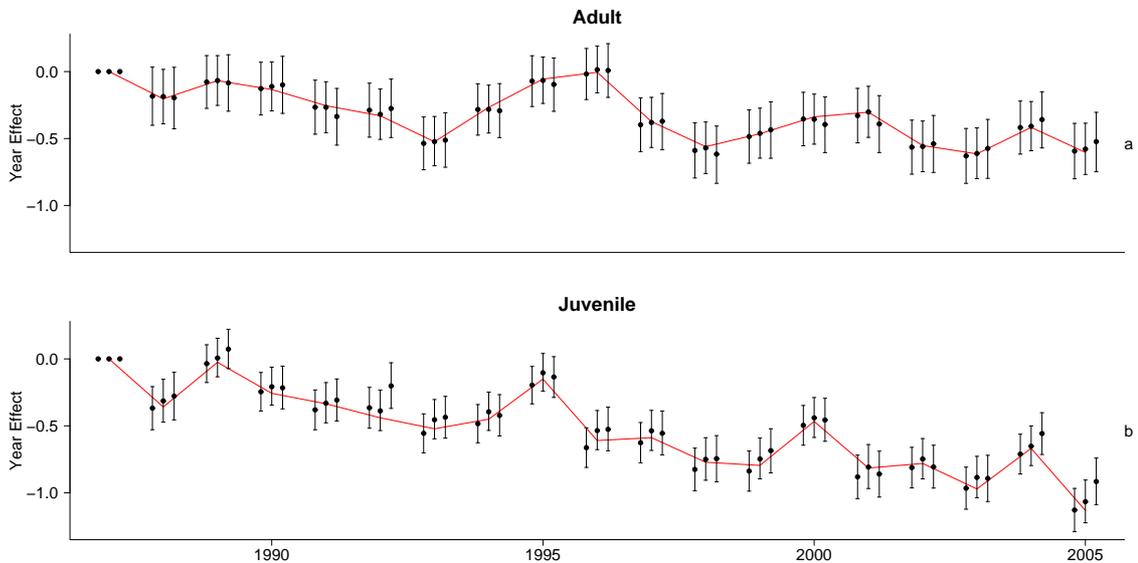
**Table 4.5** Level of censoring,  $\eta_{gt} = n_{gt} - \ell_{gt}$ , imposed upon the adult and juvenile Sedge Warbler complete coverage CES data.

Level ( $\eta$ )	0	1	2	3	4	5	6	7	8+
<b>Adult</b>	78 45.9%	43 25.3%	17 10.0%	9 5.3%	8 4.7%	4 2.4%	4 2.4%	1 0.6%	6 3.5%
<b>Juvenile</b>	91 53.5%	17 10.0%	16 9.4%	8 4.7%	6 3.5%	5 2.9%	2 1.2%	4 2.4%	21 12.4%

As these data are produced at the level of individual within-year visits, the method of Peach et al. (1998) can also be applied, and the results compared to those from the Bayesian approach. Adjustment in the former required the “global”, rather than “local”, correction for 4.1% and 7.1% of the censored adult and juvenile counts respectively. Furthermore, censored counts failed to meet the correction criteria on 4 occasions (2.4%), so were omitted altogether. Of the censored juvenile counts, 11 were censored at zero, but only 2 of these had a non-zero true (pre-censored) count (true values: 1 and 3). For the adult censored counts, 5 were censored at zero, but only 1 of these had a non-zero true count (true value = 1). Consequently, the bias caused by the enforced zero correction will be minimal. For every site, and for every year, there is at least 1 record with complete coverage, hence the Bayesian method can be applied to all censored counts.

In Figure 4.3 the estimated year effects obtained from the censored data, when either the classical method of Peach et al. (1998) or the Bayesian method are applied, are compared to those from the pre-censored data. Figure 4.3 shows excellent agreement between the estimated annual trend under the Peach et al. (1998) method and that obtained using the data prior to censoring. The trend in the posterior means from the Bayesian truncated Poisson method also matches well. Both methods appear to perform marginally worse for the juvenile data,

however the years in which performance is poorer differ. For example, in 1992 the Bayesian posterior mean is too high, whereas in 1996 the classical maximum likelihood estimate is too low. In both these years the alternative approach seems to be performing well. Nevertheless, on the whole both methods do well.



**Figure 4.3** Estimated year effects from artificially censored Sedge Warbler data using the classical Peach et al. (1998) correction method (left) and the Bayesian approach (right). In the centre are estimates from the pre-censored data. The red line corresponds to the estimates when censoring is ignored.

As the classical Peach et al. (1998) model does not incorporate the uncertainty due to the correction for missed visits, the estimated year effects are reported with slightly higher precision than those under the Bayesian approach. For both the adult and juvenile analyses, the Bayesian 95% symmetric credible interval is approximately 0.02 wider than the classical 95% confidence interval.

To assess the importance of correcting data for missed visits, year effects are calculated using the models of Chapter 2. That is, the presence of censoring is ignored and censored counts are treated as if they were fully observed. The resulting year effects, denoted by a red line in Figure 4.3, indicate that missed visits have little impact on the trends in adult and juvenile abundances. This is unsurprising due to the small level of censoring in the counts (Table 4.5).

The correlation between the Bayesian posterior median of the imputed censored count ( $n_{gt}^{cen}$ ) with the true (pre-censored) count is high (adults:  $\rho = 0.95$ , juveniles:  $\rho = 0.89$ ), though that with the Peach et al. (1998) corrected count

( $E_{gt}$ ) is somewhat higher (adults:  $\rho = 0.99$ , juveniles:  $\rho = 0.99$ ). The Bayesian posterior medians, on average, are 4.8 and 7.4 higher than the true count for adult and juvenile data respectively.

#### 4.2.5 Analysis of Sedge and Reed Warbler Data

Using the reduced Sedge and Reed Warbler CES incomplete coverage data from the “best” sites, the trends in juvenile and adult abundances are estimated. We consider both the classical Peach et al. (1998) *a priori* correction method, currently employed by the BTO, and our novel Bayesian approach that conditions on the lower bounds of the censored counts.

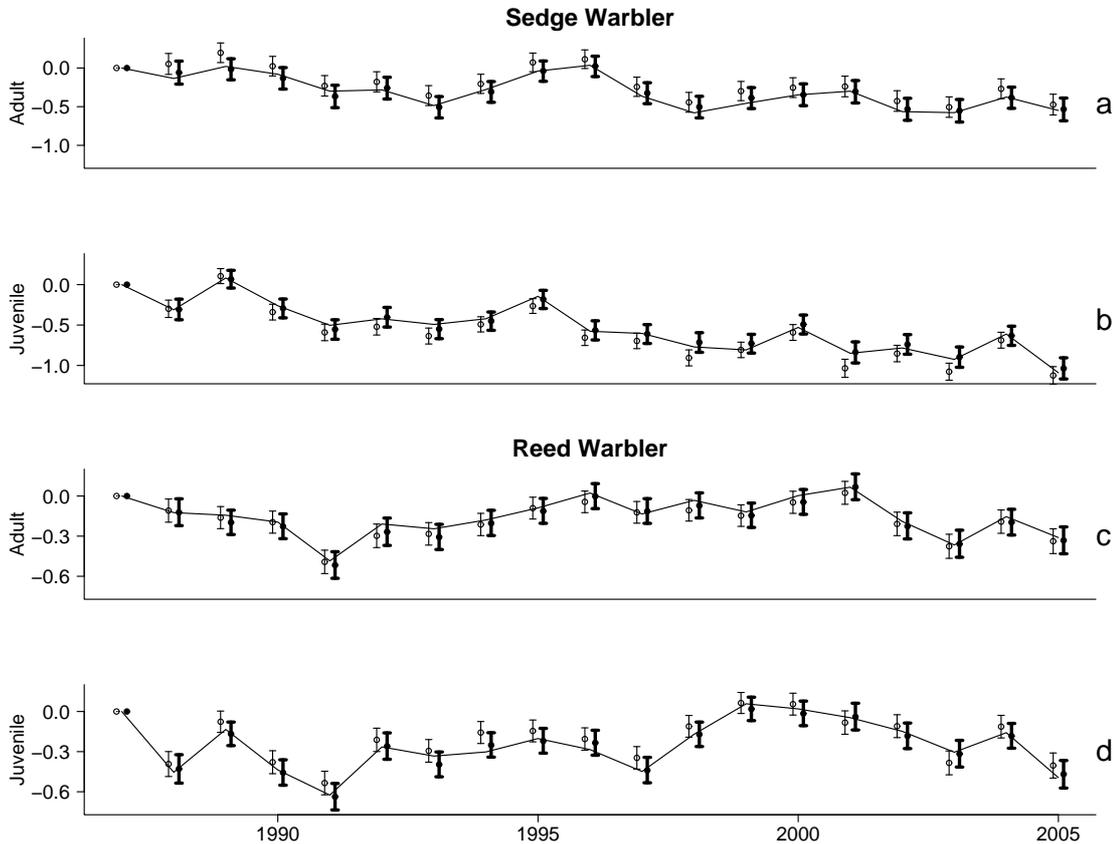
As criteria for selection of these “best” sites require that they must have at least 2 years of complete coverage, the Bayesian approach is suitable for all censored counts. Criteria set for the Peach et al. (1998) correction, however, require that a small percentage of censored counts are corrected using the global correction factor, whilst others are omitted from the analysis altogether (see Table 4.6).

Recall that under the method of Peach et al. (1998) censored counts of zero can only be corrected to zero. The correction criteria ensure that observed counts for years with many missed visits are eliminated, however the data still contain a small percent of counts censored at zero (Table 4.6). If the true unobserved yearly-site count is non-zero bias will be introduced into the estimates of the model parameters. Given the small number of such counts, such bias is likely to be negligible. Note that the Bayesian model will enable these counts to be corrected to non-zero values.

**Table 4.6** Sedge and Reed Warbler data from the “best” sites: censored yearly-site counts corrected using the Peach et al. (1998) approach. The total number (percentage) of censored counts able to be corrected, the number (percentage) of these requiring a global correction, and the number (percentage) of counts censored at zero.

	corrected	global	zero
<b>Sedge Warbler</b>			
<i>Adult Data</i>	300 (94.6%)	29 (9.7%)	15 (5.0%)
<i>Juvenile Data</i>	289 (91.1%)	34 (11.8%)	20 (6.9%)
<b>Reed Warbler</b>			
<i>Adult Data</i>	240 (94.9%)	18 (7.5%)	7 (2.9%)
<i>Juvenile Data</i>	233 (92.1%)	21 (9.0%)	6 (2.6%)

Estimates of the year effects from the adult and juvenile abundance models ( $y_t^a$  and  $y_t^j$ ) from both methods, maximum likelihood estimates from the classical approach of Peach et al. (1998) and posterior means from the Bayesian approach, are compared in Figure 4.4.



**Figure 4.4** Classical maximum likelihood estimates and the 95% confidence intervals, denoted by thin lines (left), and Bayesian posterior means and the 95% symmetric credible intervals, denoted by thick lines (right), of  $y_t^a$  and  $y_t^j$  from Sedge Warbler (SW) and Reed Warbler (RW) CES data that include years with incomplete coverage. The analysis of the reduced subset of data with censored counts omitted is shown as a line. a) SW:  $y_t^a$ , b) SW:  $y_t^j$ , c) RW:  $y_t^a$ , d) RW:  $y_t^j$ .

Despite a relatively high number of censored counts (33%, see Table 4.1), the estimated year effects under both methods, for adults and juveniles, are similar (Figure 4.4). However, the reported uncertainty in these estimates is slightly greater (but likely more reliable) under the Bayesian model than the classical Peach et al. (1998) model. This is to be expected, for the classical approach does not incorporate the uncertainty due to the correction of counts for missed visits.

Also shown in Figure 4.4 are estimates derived from only those counts based on a full set of 12 visits, for which no correction is required. Note, in such circum-

stances the Bayesian and classical models give essentially identical estimates (see Section 2.2). They bear a great resemblance to the estimates under both correction methods, suggesting that both methods produce reliable estimates that are unbiased due to the correction for missed visits.

In years of missed visits, predictions of the unknown true counts via the *a priori* correction of Peach et al. (1998) (denoted by  $E_{gt}$ ), and those arising from the Bayesian model (denoted by  $n_{gt}^{cen}$ ) are highly correlated with each other (Table 4.7). Both are also highly correlated with their observed lower bounds,  $\ell_{gt}$ , however the Bayesian approach is less so, and in addition corrects the censored counts by a much greater amount (Table 4.8). The likely reason for this is two fold. In part it arises because the Peach et al. (1998) approach is conservative, mitigating against extremely high or low imputed values, in part because the posterior distribution for  $n_{gt}^{cen}$  tends to be right skewed. Also note that as a consequence of the higher number of juveniles caught per visit, corrected juvenile counts are inflated, on average, by a greater amount than adult counts.

**Table 4.7** Correlations between the corrected adult and juvenile censored counts, in the Sedge and Reed Warbler CES data, under the classical Peach et al. (1998) approach and the Bayesian approach. The Bayesian corrected count is given by the posterior median.

	<i>adult counts</i>	<i>juvenile counts</i>
<b>Sedge Warbler</b>	0.91	0.83
<b>Reed Warbler</b>	0.95	0.90

**Table 4.8** Adult and juvenile corrected counts for Sedge and Reed Warbler CES data sets. Comparisons of the corrected count, calculated using the Peach et al. (1998) classical approach or the Bayesian approach (posterior median), to their observed lower bounds. Correlation ( $\rho$ ) between the corrected count and its observed lower bound ( $\ell_{gt}$ ), and the mean, and standard deviation (sd), of the increase in the corrected count from its lower bound.

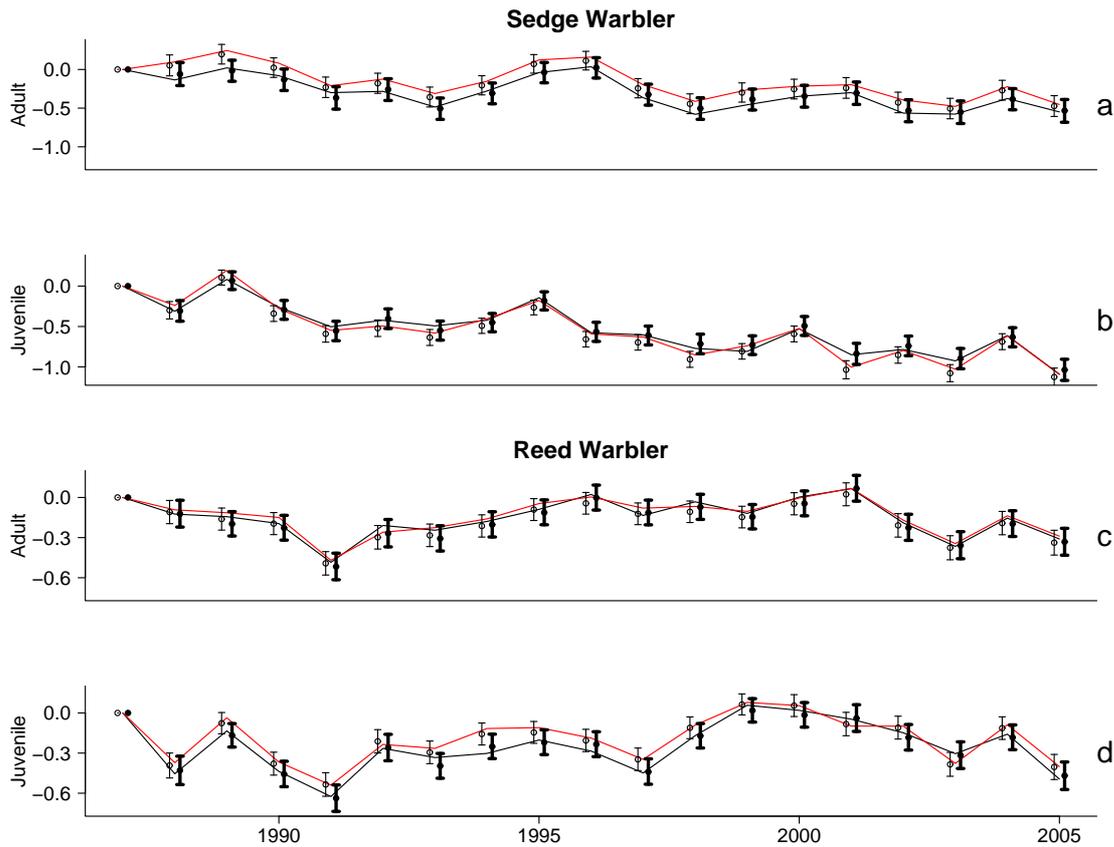
		<b>Sedge Warbler</b>		<b>Reed Warbler</b>	
		<i>Classical</i>	<i>Bayesian</i>	<i>Classical</i>	<i>Bayesian</i>
	$\rho$	1.00	0.90	1.00	0.93
<b>adult</b>	mean increase	1.2	6.2	2.8	10.6
	sd of increase	1.61	6.07	3.64	12.62
	$\rho$	0.97	0.84	0.99	0.86
<b>juvenile</b>	mean increase	2.6	13.4	3.3	15.8
	sd of increase	6.27	20.62	5.90	21.74

*Assessing the Importance of Correcting Data for Censored Counts*

To assess the importance of correcting data for non-constant effort, as a result of missed visits, year effects are calculated using the models of Chapter 2 but applied to the data that include years with incomplete coverage. That is, the presence of missed visits is ignored and we set  $n_{gt} = \ell_{gt}$  regardless of whether the count was censored or not. The resulting year effects, which are essentially identical under the Bayesian and classical approaches, are plotted in Figure 4.5, denoted by a red line. Estimates under the classical Peach et al. (1998) approach closely match those of the analysis when censoring is ignored (this is especially evident in Figure 4.5a, d). That this is the case is not surprising due to the small level to which censored counts are inflated by the Peach et al. (1998) correction (Table 4.8). However, that they are more similar to the analysis which ignores censoring, than to the analysis that omits censored counts, and therefore is unbiased by any correction, suggests that the Peach et al. (1998) correction is too conservative. Conversely, estimates under the Bayesian approach are more similar to those from the analysis which omits censored counts. This is to be expected - the strong distributional assumptions force the imputed censored counts to follow the trend in abundance estimated from non-censored counts, albeit augmented by the additional information from the observed lower bounds of the censored counts and gaining precision as a consequence.

The parallel shift evident in Figure 4.5a is due to a slightly higher proportion of missed visits in the base year, 1987 (Table 4.9). The Bayesian model inflates these censored counts to a higher level than the conservative Peach et al. (1998) approach does. Consequently in subsequent years, with less frequent censored counts, the year effects under the Bayesian method become smaller relative to those in the base year than they do following the Peach et al. (1998) approach.

Nonetheless, the trends in adult and juvenile abundances are relatively unaffected by ignoring censoring, indicating that the level to which constant effort is violated in the Sedge and Reed Warbler CES data, due to missed visits, results in little bias (Figure 4.5). That this is the case could be attributable to a consistent effect of missed visits across years. If the reduction in the observed count, caused by missed visits, is consistent across years, then the relative size of the year effects is unchanged. For Sedge and Reed Warbler CES data this appears to be the case as the frequencies of censored counts and missed visits are approximately equal between years (Table 4.9). Further, incomplete coverage seems to result in only a small reduction in the observed count (Table 4.8).



**Figure 4.5** Estimates of  $y_t^a$  and  $y_t^j$ , from Sedge Warbler (SW) and Reed Warbler (RW) CES data that include years with incomplete coverage, obtained using the classical Peach et al. (1998) approach (left) and the Bayesian approach (right). The analysis of the reduced subset of data with censored counts deleted is shown as a black line. The analysis when the presence of missed visits is ignored, and censored counts are treated as fully observed counts, is shown as a red line. a) SW:  $y_t^a$ , b) SW:  $y_t^j$ , c) RW:  $y_t^a$ , d) RW:  $y_t^j$ .

#### *The Precision in the Estimates of Abundance*

In Table 4.10 the average width of the 95% symmetric credible intervals for the key parameters of interest  $y_t$  are compared when either censored counts are i) omitted from the analysis altogether, ii) incorporated into the analysis using the Bayesian imputation method, or iii) assumed to be fully observed, i.e. we set  $n_{gt} = \ell_{gt}$ . As is to be expected, simply ignoring the presence of missed visits and assuming all counts are fully observed results in parameter estimates with the most precision. The uncertainty arising from the imputation of counts corrected for missed visits, using the Bayesian method, increases the uncertainty in the year effects, however utilising the information in the censored counts does improve precision compared to omitting such counts from the analysis altogether.

**Table 4.9** Annual number of yearly-site counts, percentage of censored counts, and the percentage of missed visits in Sedge and Reed Warbler CES data from the “best” sites, 1987-2005.

year	Sedge Warbler			Reed Warbler		
	counts	censored	visits missed	counts	censored	visits missed
1987	28	42.9%	8.3%	26	42.3%	8.0%
1988	31	41.9%	6.5%	28	39.3%	5.7%
1989	35	17.1%	1.7%	32	21.9%	2.1%
1990	41	24.4%	2.4%	34	26.5%	2.7%
1991	44	27.3%	4.7%	35	28.6%	4.5%
1992	46	41.3%	5.6%	35	42.9%	5.7%
1993	54	27.8%	3.4%	41	29.3%	3.3%
1994	57	28.1%	2.8%	41	31.7%	2.6%
1995	58	32.8%	3.9%	44	34.1%	4.0%
1996	64	35.9%	4.3%	48	41.7%	4.9%
1997	65	41.5%	6.4%	49	32.7%	4.9%
1998	69	42.0%	7.7%	54	40.7%	7.6%
1999	69	39.1%	5.0%	55	34.5%	4.7%
2000	54	25.9%	3.5%	44	25.0%	3.4%
2001	45	33.3%	7.0%	35	40.0%	8.3%
2002	52	26.9%	4.6%	43	25.6%	4.3%
2003	53	37.7%	5.2%	42	40.5%	4.8%
2004	53	20.8%	3.5%	42	16.7%	2.4%
2005	45	33.3%	6.3%	38	34.2%	6.6%
TOTAL	963	32.9%	4.9%	766	33.0%	4.7%

**Table 4.10** The average width of the 95% symmetric credible interval for the abundance year effects from the Bayesian analysis which i) omits censored counts, ii) imputes counts corrected for missed visits, or iii) ignores the presence of censored counts and treats them as if they were fully observed.

	Omit	Impute	Ignore
<b>Sedge Warbler</b>			
<i>Adult</i>	0.3466	0.2792	0.2562
<i>Juvenile</i>	0.2709	0.2394	0.1991
<b>Reed Warbler</b>			
<i>Adult</i>	0.2206	0.1913	0.1692
<i>Juvenile</i>	0.2202	0.1902	0.1699

### 4.2.6 Modelling CES Count Data at the Visit Level

An alternative, viable, approach for dealing with missed visits is to model the counts at the visit level. That is, instead of determining the unique number of birds caught across all 12 visits, and modelling the yearly-site counts, the unique number caught at each visit is modelled, and an extra factor, the “visit” effect, is added to the abundance models of Chapter 2.

We let  $n_{gtk}^a$  and  $n_{gtk}^j$  denote the number of unique adult and juvenile birds, respectively, caught at visit  $k$ , to site  $g$ , in year  $t$ . For the adult data we assume that the visit-level counts,  $n_{gtk}^a$ , are independent, random variables from a Poisson distribution with parameter  $\lambda_{gtk}^a$ . We express  $\lambda_{gtk}^a$  as a log-linear combination of an intercept term,  $\beta^a$ , plus site-, year- and visit- specific effects,  $s_g^a$ ,  $y_t^a$  and  $v_k^a$  respectively. Mathematically,

$$\ln(\lambda_{gtk}^a) = \beta^a + s_g^a + y_t^a + v_k^a, \quad (4.8)$$

where  $g \in [1, G]$ ,  $t \in [1, T]$ ,  $k \in [1, K = 12]$ . For identifiability the first site, year and visit effects in Equation (4.8) are constrained to zero.

Similarly, for juveniles, we assume that the  $n_{gtk}^j$  are independent, random variables from a  $\text{Poisson}(\lambda_{gtk}^j)$  distribution and form the analogous log-linear model to Equation (4.8) for  $\lambda_{gt}^j$ :

$$\ln(\lambda_{gtk}^j) = \beta^j + s_g^j + y_t^j + v_k^j. \quad (4.9)$$

However, as the first Sedge and Reed Warbler clutches tend to be laid at the end of May (Bibby, 1978), juveniles are yet to have fledged by the start of the CES season. Thus we consider juvenile visit-level counts from visit 5 onwards, i.e. in Equation (4.9)  $k \in [5, K = 12]$ .

Following Chapter 2, estimated indices of adult abundance and juvenile abundance in year  $t$  are given by  $A_t = \exp(y_t^a)$  and  $J_t = \exp(y_t^j)$  respectively.

Adopting visit-level counts to estimate abundance has several important advantages over the standard approaches, Bayesian or classical, that adopt yearly-site counts. Firstly, the problem of missed visits is overcome, with data from visits missed assumed to follow the estimated trend. Secondly, as missed visits contribute no information to the analysis, the visit-level models, defined by Equations (4.8) and (4.9), can be readily fitted using standard classical and Bayesian methods. Thirdly, a greater level of precision in the estimates may arise simply due to the greater amount of data. For Sedge Warbler, the number

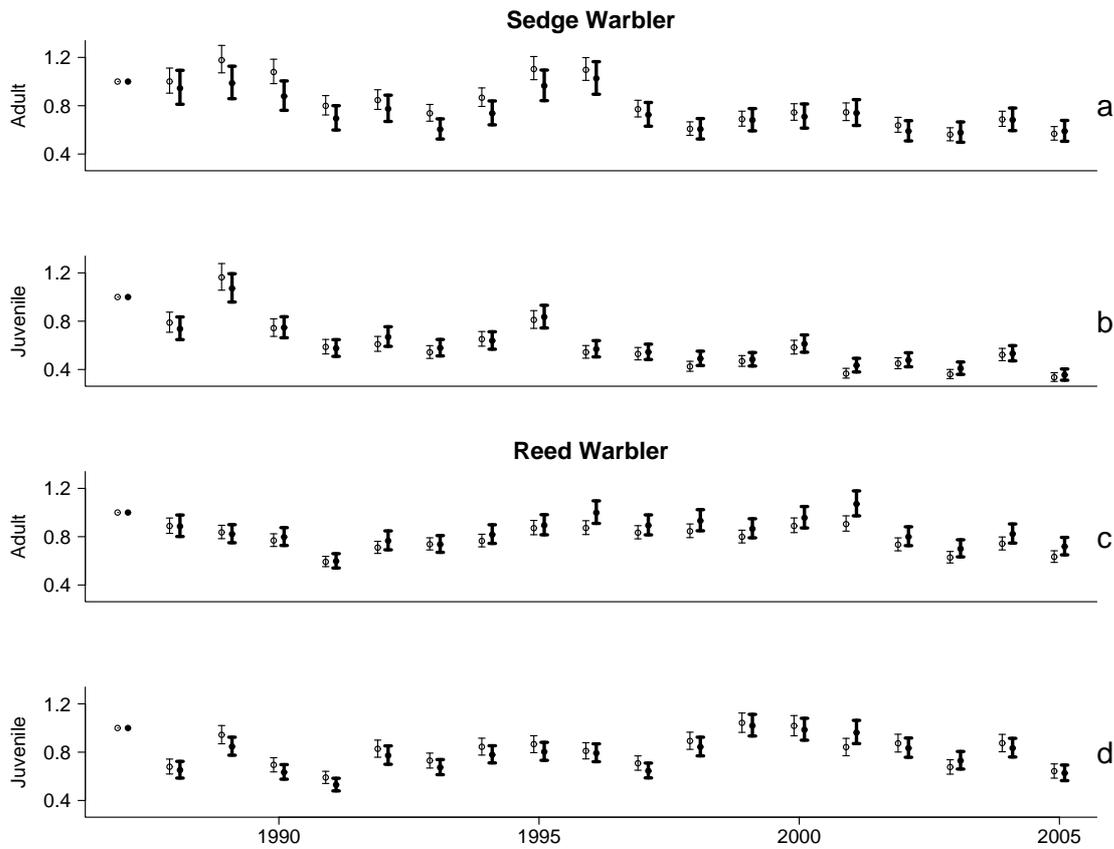
of counts modelled increases from 963 to 10994 in the adult analysis, and to 7286 in the juvenile analysis. For Reed Warbler, the number of counts modelled increases from 766 to 8758 in the adult analysis, and to 5820 in the juvenile analysis. Fourthly, the visit effects may be of interest in their own right, providing indices of seasonal change in abundance. Lastly, the interaction between visit and site may provide an insight into spatial differences in arrival times to, and departure times from, the breeding grounds. For example, it might be anticipated that birds return to northern sites later, or that birds arrive at preferred habitats earlier. Similarly, the interaction between visit and year may highlight temporal changes in migration times, which would be particularly interesting for studies of phenology.

In Figure 4.6 indices of adult and juvenile abundance, from the “best” sites Sedge and Reed Warbler incomplete coverage data, estimated from the visit-level models are compared to those of Section 4.2.5, in which the yearly-site counts are adopted and the Bayesian model is applied. Note that the Bayesian and classical estimates under the visit-level model are essentially identical. The estimated trends in adult and juvenile indices of abundance, for both species, are similar despite different data and different models clearly being used. As is to be expected, precision in the estimates is improved under the visit-level model, simply due to the greater number of counts.

Despite the advantages of the visit-level model, we retain the standard model for abundance, which adopts yearly-site counts, as the assumptions underlying this model, and its ability to provide reliable indices of abundance, has been more thoroughly tested for the bird species monitored under the CES scheme (see, for example, Baillie et al., 1986; Peach et al., 1996, 1998).

### 4.2.7 Concluding Remarks

Accurately quantifying the uncertainty in parameter estimates is particularly important when the findings will be used for identifying and monitoring populations of species at risk. CES count data are routinely used by the BTO to produce indices of abundance. Under the assumption of constant effort, any variability between years in the number of birds caught is attributed to changes in the population level and to stochastic variation, and not to varying intensity of capture effort. Failure in this latter, crucial assumption, as a result of missing within-year visits, requires explicit consideration in the analysis.



**Figure 4.6** Bayesian posterior means and the 95% symmetric credible intervals, of the indices of abundance when the CES count data are modelled at i) the visit level (left) and ii) the annual level (right). a) SW:  $A_t$ , b) SW:  $J_t$ , c) RW:  $A_t$ , d) RW:  $J_t$ .

Two alternative approaches have been considered to formally deal with the presence of missed visits - the classical Peach et al. (1998) model and our novel Bayesian model (Cave et al., 2009a). The simulation studies presented here indicate that when the underlying Poisson model is correct, both methods perform well, even in the presence of extreme censoring. Furthermore, both methods appear robust to the level of model failure likely to be present in the CES data.

#### *Advantages of the Bayesian Method over the Classical Approach*

One of the most important distinctions between the routinely used classical approach and our Bayesian method is the accuracy in the reported uncertainty. Whereas the Bayesian approach produces parameter estimates whose uncertainty also incorporates a component due to the correction for missed visits, the Peach et al. (1998) approach underestimates the standard errors - by ignoring the uncertainty arising from the *a priori* correction of censored counts,

model parameters are reported with more precision than they merit. By falsely reporting the precision of the abundance indices, inference becomes unreliable, and declines might be inferred where a population is in fact stable.

Not all censored counts can be corrected using the Peach et al. (1998) approach, as failure to satisfy the correction criteria means that some data are omitted from the analysis altogether. The correction is also conservative in its overall effect, mitigating against extremely high or low corrected values. In addition, by default all censored counts of zero are corrected to zero. This can introduce extreme bias, particularly for species with typically low yearly-site counts where censored values of zero are more common, although the criteria for correction do help to alleviate this. The Bayesian method is not conservative, and will enable the correction of zeros to non-zero values. Different correction criteria need be applied however as censored counts for sites without complete coverage in any year, or from years without complete coverage at any site, cannot be reliably imputed. In the CES data this rarely occurs, and in general the Bayesian method is expected to be able to deal with more censored counts than the Peach et al. (1998) approach. The correction of counts to integer values, by using the posterior median as a point estimate, is arguably another advantage of the Bayesian approach over the Peach et al. (1998) for which correction of typically non-integer.

#### *Importance of Accounting for Censoring*

Our studies indicate that for the level and pattern of censoring present in the Sedge and Reed Warbler “best” sites CES data sets, the trends in adult and juvenile abundances are relatively unaffected by the missed visits. Simply treating censored counts as if they were fully observed appears to provide an adequate, and fairly unbiased, estimated index for most years for the Sedge and Reed Warbler data. This lack of bias is due to i) the small effect of censoring on the observed count and ii) the reduction in the observed count caused by missed visits being consistent between years. The latter should be the case if visits were missed at random between years. Further studies using CES data from other species, and a more rigorous understanding of the censoring process, is required before advocating the simple approach of ignoring censoring in the general modelling of abundance from CES data. Using our Bayesian method, the presence of censored counts in the data is more appropriately dealt with.

### 4.3 Estimating Productivity from CES Data Containing Censored Counts

Recall that in Section 2.3 an index of productivity in year  $t$ , denoted by  $P_t$ , was defined to be the ratio of the indices of juvenile to adult abundance in year  $t$ , i.e.

$$P_t = \frac{J_t}{A_t}. \quad (\text{Equation (2.4)})$$

Indices of productivity are therefore readily estimated using both the classical Peach et al. (1998) correction method (Section 4.2.1), and the Bayesian approach (Section 4.2.2). Denoting  $\hat{y}_t^a$  and  $\hat{y}_t^j$  to be the maximum likelihood estimators for  $y_t^a$  and  $y_t^j$  under the adult and juvenile abundance models respectively, obtained using the approach of Peach et al. (1998), the classical maximum likelihood estimator of  $P_t$  is given by  $\exp(\hat{y}_t^j - \hat{y}_t^a)$ . A Bayesian estimate for the productivity index is derived by calculating the ratio of  $J_t$  to  $A_t$  using samples drawn from the posterior distributions of these parameters, obtained using the alternative Bayesian model for censored counts.

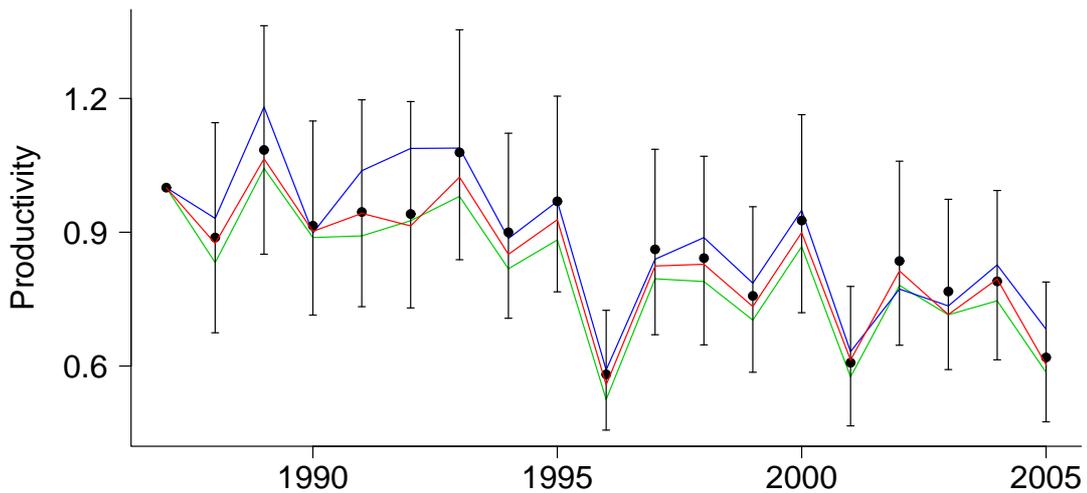
Despite the presence of missing visits in CES data having little impact on the estimated trends of adult or juvenile abundances for Sedge and Reed Warblers (Section 4.2.5), missing visits may bias the estimates of productivity as the within-year catch patterns of adult and juvenile birds differ. In particular, adult capture rates are expected to be highest during the early mist-netting visits and to decline near the end of the season as adult birds begin their winter migration. Conversely, capture rates of juvenile birds are expected to be extremely low during the first few visits but to increase during the middle of summer as increasing numbers of juveniles fledge. According to which visits are missed, the degree of censoring in  $n_{gt}^a$  and  $n_{gt}^j$  may be quite different, changing the observed proportion of juveniles in the catch.

#### 4.3.1 Simulation Study

In the second simulation study of Section 4.2.4, in which Sedge Warbler CES data were artificially censored, the same pattern of missed visits was applied to censor both adult and juvenile yearly-site counts. Therefore, using this artificially censored Sedge Warbler data, the ability of the Bayesian method to

produce reliable estimates of the productivity indices ( $P_t$ ), in circumstances where the Poisson assumption for adult and juvenile counts is violated to a degree, can be assessed. Furthermore, as the method of Peach et al. (1998) can be used to obtain *a priori* corrected counts, estimates of  $P_t$  under the classical and Bayesian approaches can be compared (see Figure 4.7).

The estimated trend in productivity is in good agreement with that from the pre-censored data, which represents the truth, for both the classical and Bayesian approaches (Figure 4.7). This suggests that both methods produce good estimates of productivity given the number and pattern of missed visits typically present in CES data. Furthermore, both methods appear to be robust to the degree of model failure present in CES data.



**Figure 4.7** Estimated productivity indices,  $P_t = J_t/A_t$ , from counts simulated by artificially censoring Sedge Warbler CES data. Posterior means and 95% symmetric credible intervals from the pre-censored data are plotted in black. The lines denote estimates from the censored data - green: maximum likelihood estimates using the classical Peach et al. (1998) correction method, and blue: posterior means using the Bayesian approach. The red line corresponds to estimates under the analysis where the presence of censoring is ignored.

The red line in Figure 4.7 corresponds to the analysis where the presence of censoring is ignored, and all counts, censored or not, are regarded as if they were fully observed. The simulation analysis suggests that, for the severity of censoring likely to be present in CES data (see Tables 4.4, 4.5), little bias is introduced by simply ignoring censoring.

### 4.3.2 Analysis of Sedge and Reed Warbler Data

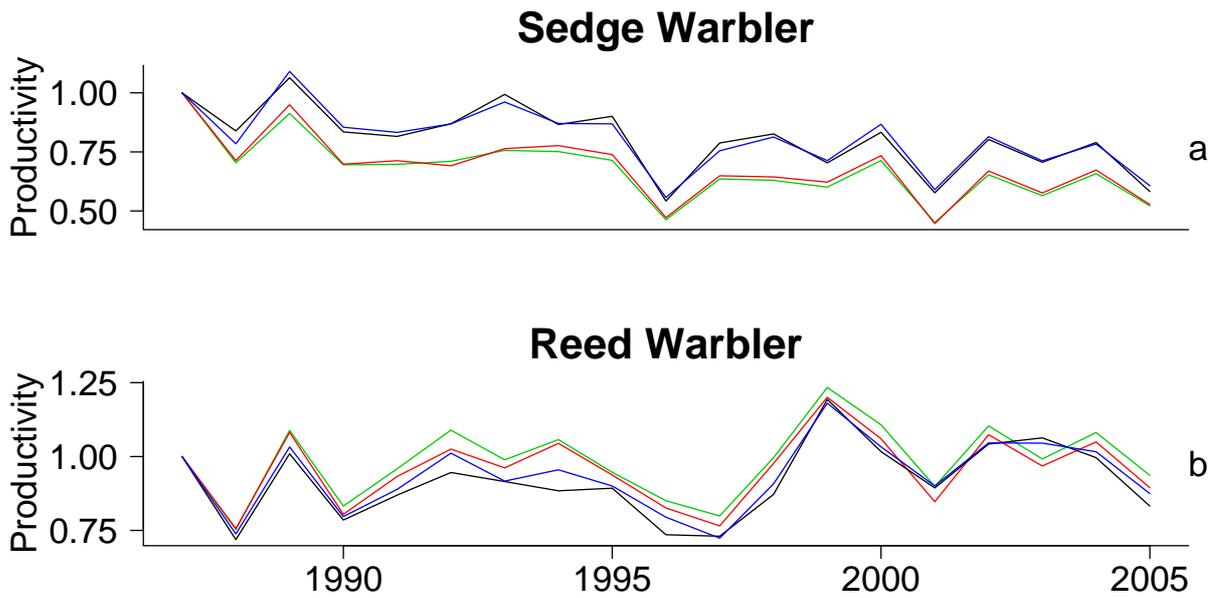
Adopting the reduced CES data from the “best” sites, but including years with incomplete coverage, the productivity trends for Sedge and Reed Warblers are estimated. To address the problem of censored counts, we consider both the classical Peach et al. (1998) *a priori* correction method (Section 4.2.1) and our novel Bayesian approach (Section 4.2.2).

Note that criteria set for the Peach et al. (1998) correction result in some data being omitted from the classical analysis (see Table 4.6). In particular, count data from site  $g$  in year  $t$  can only be included if at least 4 of the first 6, and 4 of the last 6, visits are made. This corresponds to 90.5% and 91.3% of the Sedge and Reed Warbler “best” sites CES data respectively. Conversely, no data are omitted from the Bayesian analysis as criteria for selection of these “best” sites require that they must have at least 2 years of complete coverage.

Indices of productivity, derived from the ratio of juvenile to adult abundance indices, are plotted in Figure 4.8. As for the analysis of abundance (Section 4.2.5) the classical approach of Peach et al. (1998) results in indices more closely related to the analysis which ignores censoring, and hence is subject to bias as a result of missing visits, than to the analysis that omits censored counts altogether. This is most evident for Sedge Warbler. Conversely, estimates under the Bayesian model more closely match the analysis which excludes censored counts, and thereby is known to be unbiased by missed visits or any correction for them. We note, however, that both methods will be biased if the non-censored count data are not representative of the censored count data.

That the classical method produces estimates of  $P_t$  most similar to the analysis that ignores censoring can be attributed to the conservative nature of the Peach et al. (1998) correction. By contrast the strong distributional assumptions of the Bayesian approach forces the corrected counts to follow the trend estimated for the non-censored counts.

Aside from small discrepancies, the overall trend in productivity estimated using the Bayesian approach is, however, similar to that from the analysis that ignores censoring, treating all counts as fully observed. Assuming that the Bayesian approach is performing well, this would suggest that the pattern and frequency of missed visits in CES data causes little bias in the estimates of  $P_t$ . This could arise because missed visits i) result in a negligible level of censoring for both adult and juvenile counts and/or ii) have a consistent effect between years.



**Figure 4.8** Indices of productivity from the Sedge and Reed Warbler CES “best” sites data containing years with missed visits. The black line corresponds the analysis which omits all censored counts, the red line to the analysis which ignores censoring, the blue line to Bayesian analysis, and the green line to the classical analysis following Peach et al. (1998).

### 4.3.3 Productivity Model Employed by the BTO

The BTO routinely produces indices of productivity for common songbirds in the UK, employing a classical logistic regression model, under which the number of juvenile birds caught is assumed to have a Binomial distribution conditional on the total number of birds (adults and juveniles) captured. Recall that the Binomial based productivity model for complete coverage data, described in Section 2.3.1, is given by:

$$n_{gt}^j \mid (n_{gt}^j + n_{gt}^a) \sim \text{Binomial}(n_{gt}^j + n_{gt}^a, \theta_{gt}),$$

where  $\text{logit}(\theta_{gt}) = \beta^p + s_g^p + y_t^p$ . (Equation (2.5))

Further, recall it was shown that this Binomial based model produces productivity indices,  $P_t = \exp(y_t^p)$ , analogous to those derived from the Poisson models for abundance via Equation (2.4), as in the section above.

Robinson et al. (2007) have extended the methods of Peach et al. (1998), for estimating productivity from CES data with incomplete coverage under this Binomial based model. Following the approach of Peach et al. (1998), Robinson et al. (2007) correct adult and juvenile yearly-site counts for missed visits prior

to model fitting, as described by Equation (4.1). The corrected adult and juvenile yearly-site counts, denoted by  $E_{gt}^a$  and  $E_{gt}^j$  respectively, in general do not take integer values that are required by the Binomial based productivity model. Robinson et al. (2007) therefore construct an offset that accounts for the reduction in the observed adult and juvenile counts ( $\ell_{gt}^a$  and  $\ell_{gt}^j$ ) as a result of missed visits. By noting that the difference between the log odds of the observed and corrected proportion of juvenile birds in the annual catch can be expressed as:

$$\text{logit} \left( \frac{\ell_{gt}^j}{\ell_{gt}^j + \ell_{gt}^a} \right) - \text{logit} \left( \frac{E_{gt}^j}{E_{gt}^j + E_{gt}^a} \right) = \ln \left( \frac{\ell_{gt}^j E_{gt}^a}{\ell_{gt}^a E_{gt}^j} \right),$$

Robinson et al. (2007) assume that, at site  $g$  in year  $t$ , the observed number of juvenile birds caught ( $\ell_{gt}^j$ ) have a Binomial distribution conditional on the total number of observed birds caught ( $\ell_{gt}^j + \ell_{gt}^a$ ), with parameter  $\theta_{gt}$  given by:

$$\text{logit}(\theta_{gt}) = \beta^p + s_g^p + y_t^p + \ln \left( \frac{\ell_{gt}^j E_{gt}^a}{\ell_{gt}^a E_{gt}^j} \right). \quad (4.10)$$

Clearly if all 12 visits are made, the right-most term of Equation (4.10), the offset, reduces to zero and the expected proportion of juveniles in the annual catch is identical to that given by the model for complete coverage data, Equation (2.5). Furthermore, analogous to the model for complete coverage data, an index of productivity in year  $t$ , denoted by  $P_t$ , is given by  $\exp(y_t^p)$ .

This model is readily fitted by maximum likelihood in any Generalized Linear Modelling package, for example, as in the current research, using the `glm` function with the `offset` argument in R (R Development Core Team, 2007).

The criteria for correction follow those specified for the abundance models of Section 4.2.1. In particular, a “global” rather than “local” correction is used for sites without complete coverage in any year, or that caught fewer than 10 individuals across all years with complete coverage. Further, since the majority of adult birds tend to be caught early in the mist-netting season, whereas juveniles tend to be caught later, data for a given site, from years in which fewer than 4 of the first 6 visits were completed, and/or fewer than 4 of the last 6 visits were completed, are omitted from the analysis altogether.

Observed adult or juvenile yearly-site counts of zero pose a problem under Equation (4.10) as the offset is undefined. Simply excluding such data may

lead to bias in the estimated productivity indices. For example, consider a year in which juvenile abundance is low. By omitting data from this year when  $\ell_{gt}^j = 0$ , a disproportionate number of higher juvenile yearly-site counts would be retained, causing  $P_t$  to be overestimated. To prevent such bias Robinson et al. (2007) modify Equation (4.10) to incorporate such data, modelling

$$\text{logit}(\theta_{gt}) = \begin{cases} \beta^p + s_g^p + y_t^p + \ln\left(\frac{E_{gt}^a}{\ell_{gt}^a}\right) & \text{if } \ell_{gt}^j = 0, \ell_{gt}^a > 0 \\ \beta^p + s_g^p + y_t^p + \ln\left(\frac{\ell_{gt}^j}{E_{gt}^j}\right) & \text{if } \ell_{gt}^j > 0, \ell_{gt}^a = 0. \end{cases} \quad (4.11)$$

If no adult or juvenile birds are caught, i.e.  $\ell_{gt}^j = \ell_{gt}^a = 0$ , then no information regarding productivity is obtained and such data are excluded from the analysis without bias.

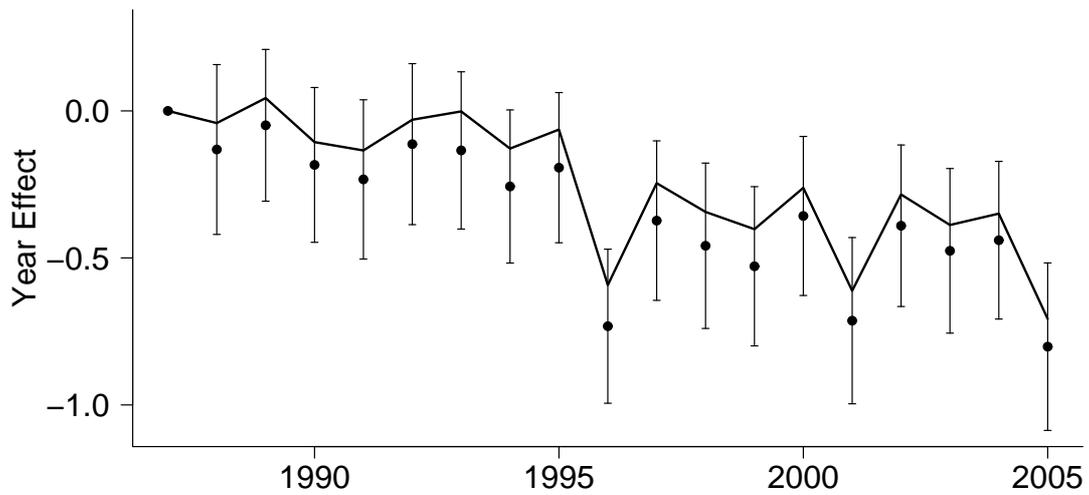
As with the Peach et al. (1998) classical models for abundance, the estimated  $E_{gt}^a$  and  $E_{gt}^j$  values (the corrected adult and juvenile counts) are assumed to be known without error, thus the uncertainty in the resulting parameter estimates is under-reported.

#### *Simulation Study*

Studies by Miles et al. (2007), that compared parameter estimates obtained via the approach of Robinson et al. (2007) to those based upon the subset of data in which years with incomplete coverage were excluded, suggest that, for the number and pattern of missed visits typically present in CES data, estimates of productivity are unbiased by the correction for missed visits. Our simulation studies, for example the analysis of artificially censored Sedge Warbler CES data (Section 4.2.4), also indicate that the approach of Robinson et al. (2007) works reasonably well (see Figure 4.9).

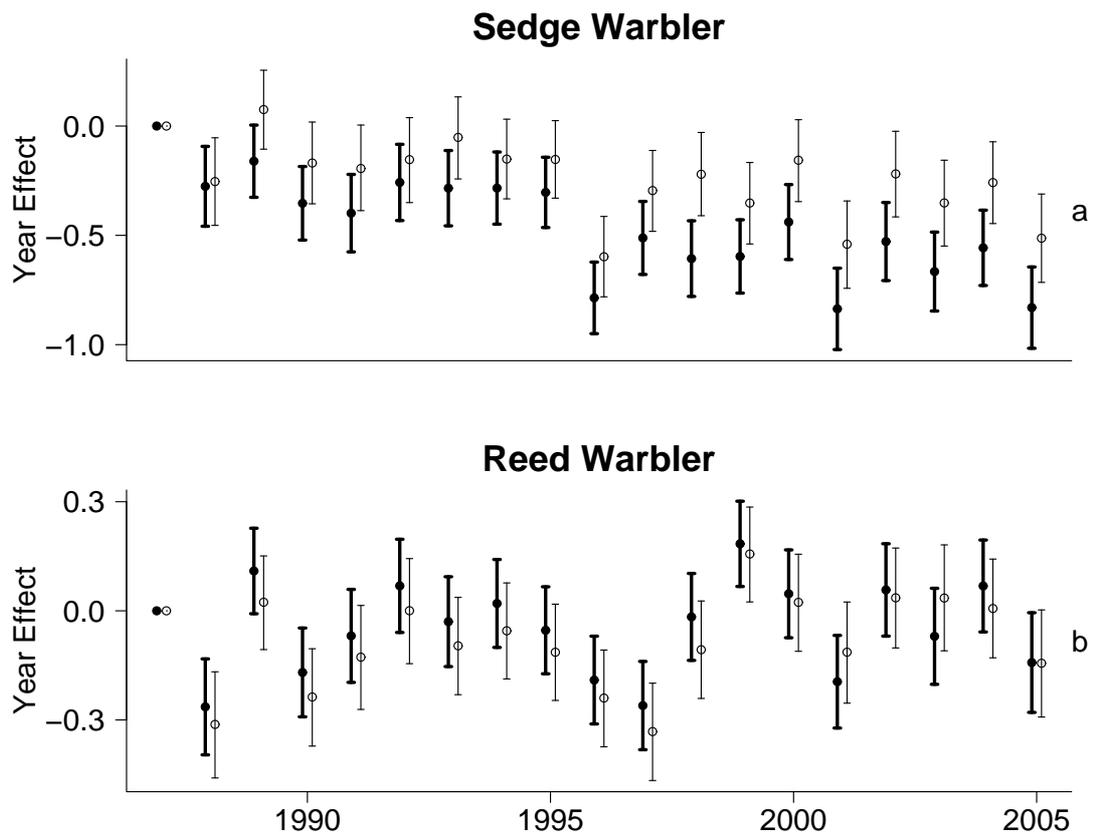
#### *Sedge and Reed Warbler Data*

In Figure 4.10 estimated productivity year effects from the reduced, “best” sites, CES data, for Sedge and Reed Warbler, that include years with incomplete coverage are given. Estimates from our novel Bayesian approach (Section 4.2.2) are compared to those from the approach of Robinson et al. (2007), the method routinely employed by the BTO for such data. There are only slight differences



**Figure 4.9** The maximum likelihood estimates, and 95% confidence intervals, of the productivity year effects under the model of Robinson et al. (2007) from the analysis of artificially censored Sedge Warbler data. The line corresponds to estimates from the pre-censored data.

between the productivity estimates under the two methods for Reed Warbler (Figure 4.10b). Although the differences are more pronounced for Sedge Warbler, the overall trend in productivity is very similar (Figure 4.10a). Indeed, the apparent discrepancy can be attributed to a reference year (in this case 1987) effect. If another year, for example 2005, was used as the reference the differences would not be as pronounced. As is to be expected, the reported precision is higher using the classical approach as the uncertainty in the corrected counts has not been accounted for.



**Figure 4.10** Trend in productivity calculated using the BTO’s classical approach (left) and our novel Bayesian approach (right), from Sedge and Reed Warbler CES “best” sites data with incomplete coverage. Left: maximum likelihood estimates, and 95% confidence intervals, of  $y_t^p$  calculated using the classical approach of Robinson et al. (2007). Right: posterior means and 95% symmetric credible intervals of  $y_t^p = y_t^j - y_t^a$ , under the novel Bayesian approach.

## 4.4 Estimating Adult Survival from CES Data with Missing Visits

An asset of the Constant Effort design is that probabilities of recapture (at a site) can be considered constant over time, simplifying the model structure (Peach, 1993; DeSante et al., 1999). Indeed, the findings of Chapter 2 have provided support for this assumption when CES live-recapture data with complete coverage were analysed (see Section 2.4.8). The ability to apply such a parsimonious model is particularly desirable - by reducing the number of parameters requiring estimation, the precision in the survival estimates, the biologically interesting parameters, is substantially improved (Peach et al., 1990).

However for a given site, when one or more visits are missed, the “effort” spent mist-netting is reduced, and the recapture rate for that year is lower compared to one in which the full complement of visits was made. Consequently, when data from years with incomplete coverage are included into the analysis, the site-specific probabilities of recapture,  $c_g$  as defined in Chapter 2, can no longer be assumed time-invariant.

Model EP, the extended Pradel et al. (1997) transient model presented in Section 2.4.1.iv, which provides annual estimates of apparent adult survival,  $S_{a,t}$ , from CES data with complete coverage, is readily adapted to allow for this time-variance by replacing each  $c_g$  parameter with its appropriate time varying value,  $c_{gt}$ , increasing the number of recapture parameters from  $G$  up to  $G \times (T - 1)$ , noting that time-varying values are only required for years with missed visits and that not all sites are operational in all  $T$  years. Furthermore, each  $\varepsilon_g$  parameter of model EP must also be replaced with its appropriate time varying value,  $\varepsilon_{gt}$ , increasing the number of evasion parameters from  $G$  up to  $G \times T$ . In total, allowing for time-variance in the rates of recapture, due to missed visits, results in a net gain of up to  $G \times (2T - 3)$  parameters. In the analyses of Sedge and Reed Warbler incomplete coverage data, from the “best” sites, upon accounting for years in which a site was not operated, or in which all 12 visits were made, an extra 317 and 253 parameters respectively are required, motivating the need to develop an alternative, more parsimonious model.

Following Peach et al. (1998), we assume that the pattern of within-year catch rates is consistent between years, and present a Bayesian model that adjusts  $c_g$  and  $\varepsilon_g$  for missed visits by fitting a parametric model to this predictable,

within-year, seasonal change in catchability. Such a modification enables both the number and timing of missed visits to have an effect on recapture rates. We fit this model to simulated data, and to Sedge and Reed Warbler CES data, and demonstrate how ignoring the presence of missed visits and/or transient birds in the data biases the estimation of apparent survival.

#### 4.4.1 Effect of Missed Visits and Transient Birds on Survival Estimates from CES Data

For live-recapture data collected under the CES scheme, by design, the probability of recapture is assumed to be constant across years but allowed to vary between sites (denoted by  $c_g$ ). Further, as the CES sites are assumed to sample with varying catch effort the same wider population of adult birds, the same annual apparent survival rates are adopted across all sites (denoted by  $S_{a,t}$ ). However, CES live-recapture data contain both records from transient birds and records from years with missed visits. Failure to adequately account for these nuances in the analysis of CES live-recapture data will result in biased survival estimates.

Consider CES live-recapture data from a single site,  $g$ . When visits are missed the overall catch effort for that year is reduced and as a consequence the average  $c_g$  across all years is decreased. To counteract the reduced  $c_g$  in years with unusually few missed visits, estimates of  $S_{a,t}$  are expected to be positively biased when the presence of missed visits is ignored. Conversely, in years with many missed visits estimates of  $S_{a,t}$  are expected to be negatively biased. However, if the effect of missed visits is consistent between years bias will not be introduced into the estimates of  $S_{a,t}$ .

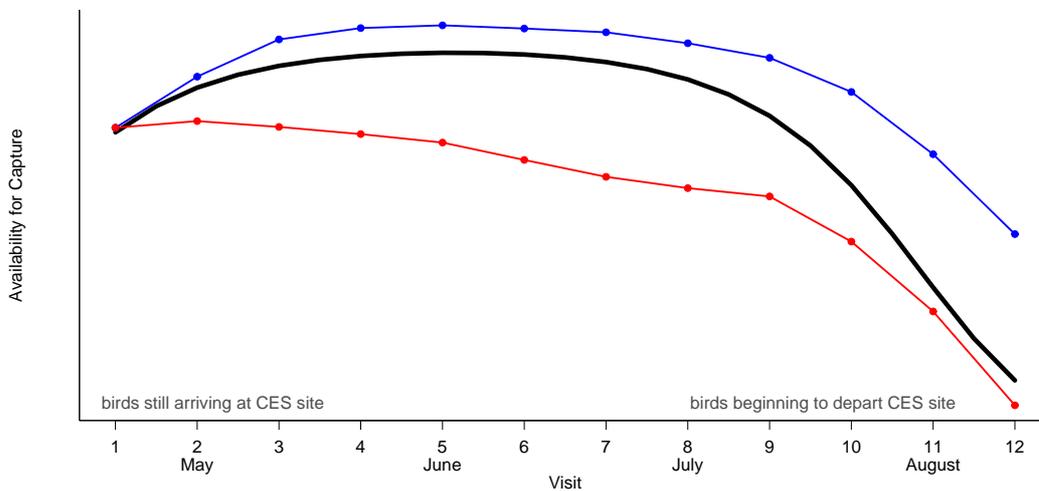
The presence of transients (birds not breeding in the vicinity of the CES site) in the data set will increase the number of birds only seen once. Since these birds can be thought of as having a zero probability of apparent survival after initial recapture (as they permanently emigrate from the study area), estimates of  $S_{a,t}$  are expected to be negatively biased when their presence in the data is ignored (Pradel et al., 1997), see Section 2.4 for further discussion.

The combined occurrence of missed visits and transient birds in the data, when not adequately accounted for, has a complicated effect on the magnitude and direction of the bias in estimates of  $S_{a,t}$ . In years with many missed visits, the

presence of transients is expected to magnify the negative bias. However in years with few missed visits, their presence counteracts the bias induced by the assumption of time-invariant  $c_g$ , and the direction of resulting bias is unknown.

#### 4.4.2 Pattern of Within-Year Captures

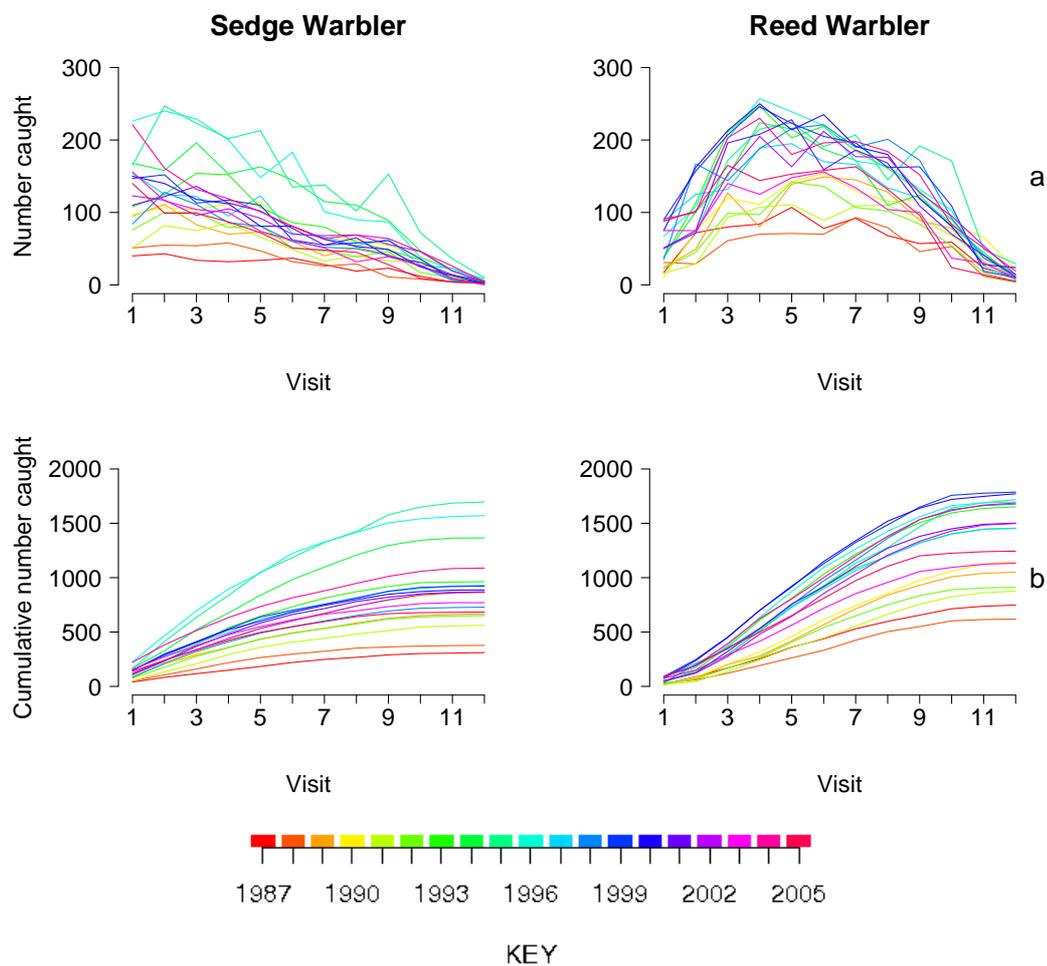
The availability for capture of adult birds, over the 12 visit mist-netting season, is expected to vary in a predictable manner as a result of individual heterogeneity in the arrival times to, and departures times from, the CES sites, for example following the hypothetical pattern given in Figure 4.11. Here, as many of the sites will not have their full complement of birds at the start of the study, the availability for capture increases during the first few visits (early May). As the breeding season draws to a close (mid July), availability for capture declines as birds begin leaving the breeding grounds. Helping to justify this hypothetical pattern, superimposed on Figure 4.11 are the estimated adult visit effects,  $v_k^a$ , from the Bayesian analysis of the CES visit-level count data for Sedge (red) and Reed (blue) Warbler as described in Section 4.2.6. Subtle variants of this broad picture can be expected for other migrant species.



**Figure 4.11** Hypothetical pattern in the availability for capture of adult birds over the 12 visit CES mist-netting season. Changes in availability for capture are caused by individual heterogeneity in arrival times at, and departure times from, CES sites. Superimposed are the posterior means of the adult visit effects,  $v_k^a$ , from the analysis of the CES visit-level count data for Sedge (red) and Reed (blue) Warbler.

The exact pattern in the availability for capture across the mist-netting season varies between species as a result of differences in their breeding and migration

strategies. However, under the assumption of constant effort, for a given species the pattern is expected to be consistent between years. For Sedge Warbler and Reed Warbler such consistency is evident in Figure 4.12. Plotted separately for each year, using the complete coverage data from the “best” sites of Chapter 2, is the total number of adult birds caught at each visit (Figure 4.12a) and the cumulative number (Figure 4.12b), proxies for the availability for capture. Note that changes in yearly catch sizes reflect both temporal variability in abundance, and annual differences in the set of CES sites in operation.



**Figure 4.12** Annual within-year catch sizes, plotted separately for each year, using Sedge and Reed Warbler CES complete coverage data from the “best” sites. Plotted for each visit are a) the total number of adult birds caught, and b) the cumulative number of adults caught. The 12 visits span the British summer months, May–August.

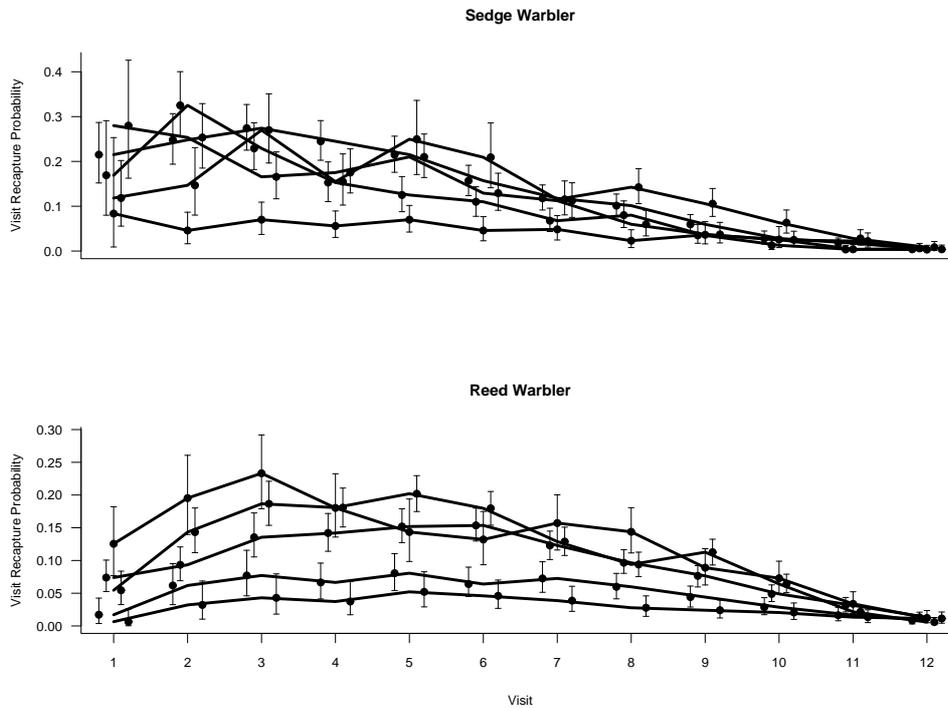
Figures 4.11 and 4.12 illustrate differences in the overall trends in the number of birds caught across the mist-netting season between Sedge Warbler and Reed Warbler. The shape of the overall trend, and the inter-specific differences, have a biological explanation. The first Sedge Warblers begin to arrive at their

breeding sites in early April, with arrivals reaching their peak in mid May (Simms, 1985; Peach, 2002), whereas Reed Warblers arrive slightly later, and over a more protracted time period, with birds first arriving in mid April and continuing through to mid June (Simms, 1985; Redfern and Alker, 2002). Thus in the Reed Warbler plots the increase in the number of adult birds caught is much greater, and occurs across more visits, than that observed in the Sedge Warbler plots. During the mid summer visits adult Sedge and Reed Warblers (both female and male birds) incubating eggs on the nest are unavailable for capture resulting in the steady decline observed in the numbers caught. Sedge and Reed Warblers begin leaving their breeding sites in late July for their pre-migratory feeding grounds (Couzens, 2006) resulting in the levelling out of the cumulative numbers caught from visit 9 onwards (Figure 4.12b).

We propose that such a consistent change in catchability across the mist-netting season can be modelled using a parametric regression equation, such as a quadratic polynomial. As the following analysis demonstrates, this regression equation must also take into account the inherent, but approximately consistent, differences in the within-year recapture rates between the individual CES sites.

We formulate a simple survival model for the full capture histories at the individual visit level in which, under the assumption of constant effort, the probability of recapture at visit  $k$  is constant between years but allowed to vary between sites (denoted by  $v_{gk}$ ). Let  $S_{a,tk}$  denote the probability that an adult bird survives from visit  $k$  in year  $t$  until the following visit which, if  $k = K$ , is the first visit in year  $t + 1$ , or otherwise is visit  $k + 1$  in year  $t$ . We assume complete survival within the mist-netting season, i.e.  $S_{a,tk} = 1$  for  $k \in [1, K - 1]$ , but allow survival to vary “annually” between visit  $K$  in year  $t$  and visit 1 in year  $t + 1$ , i.e.  $S_{a,tk} < 1$  for  $k = K$ . That is, we assume most mortality occurs over the migration and wintering period, which seems reasonable (Peach, 1993). For simplicity, the presence of transient birds in the data is ignored, biasing the survival estimates, but assuming a constant probability of transience,  $\tau$ , the overall trend is unaffected. The model is fitted using Bayesian techniques. We use random walk Metropolis-Hastings updates, with Uniform proposal distributions, appropriately truncated, for the survival and recapture probability parameters. Proposal step lengths are tuned *a priori*. The simulations are run for 200,000 iterations with the first 100,000 iterations discarded as burn-in. As essentially identical posterior estimates were obtained from multiple runs with different overdispersed starting points we assume the chain has converged. The within-

year recapture probabilities,  $v_{gk}$ , for 5 randomly selected sites, are examined in Figure 4.13. Given the uncertainty in the estimates, assuming a predictable trend in  $v_{gk}$  across the mist-netting season, with consistent differences between individual sites, seems reasonable.



**Figure 4.13** Estimated within-year (visit) recapture probabilities,  $v_{gk}$ , for 5 randomly chosen sites, from the analysis of Sedge and Reed Warbler complete coverage data. Points denote posterior means and bars denote 95% symmetric credible intervals.

### 4.4.3 Survival Model for CES Data with Missed Visits - Model MV

In Chapter 2, Section 2.4.1.i, an ordinary Cormack-Jolly-Seber model, referred to as “model CJS”, was described. Model CJS is appropriate for CES live-recapture data with complete coverage, assuming that the data contain only records from resident birds. Initially, ignoring the presence of transient birds in the data, we formulate a parsimonious model to adjust the site-specific recapture probabilities of model CJS in years with missed visits. We refer to this adjusted model as “model MV”. We return to the issue of accounting for transient birds in Section 4.4.4, the necessity of which has previously been demonstrated (see Section 2.4).

### i Model MV

Recall that for CES data, under model CJS, it is assumed that the probabilities of recapture (at a site) are constant over time (denoted by  $c_g$ ), and that the apparent adult survival rates vary annually but are common across all sites (denoted by  $S_{a,t}$ ) (see Section 2.4.1.i for details). We now describe model MV, a parsimonious approach for adjusting the  $c_g$  parameters of model CJS in years with missed visits that requires only two extra parameters.

For visit  $k \in [1, \dots, K]$  let  $\kappa$  denote the standardised visit number. That is

$$\kappa_k = \frac{k - \bar{k}}{\sigma_k}$$

where  $\bar{k}$  and  $\sigma_k$  are the mean and standard deviation of  $\{1, \dots, K\}$  respectively. For example, for CES data where  $K=12$ ,  $\bar{k}=6.5$  and  $\sigma_k = \sqrt{13}$ . Assuming a constant seasonal change in catchability between years, we model the probability of recapture at visit  $k$  in site  $g$  (denoted by  $v_{gk}$ ) using the quadratic regression equation

$$\text{logit}(v_{gk}) = s_g^r + \beta_1^r \kappa + \beta_2^r \kappa^2. \quad (4.12)$$

Here the site effects,  $s_g^r$ , enable consistent differences in  $v_{gk}$  between individual CES sites.

The annual probability of recapture at site  $g$  in year  $t$ , denoted by  $c_{gt}$ , is equal to the probability that a bird is caught at least once during any of the visits made that year. Assuming that there are no deaths within the mist-netting season, and that the within-year recaptures of an individual bird occur independently, we can write:

$$c_{gt} = 1 - \prod_{k=1}^K (1 - v_{gk})^{I_{gtk}}, \quad (4.13)$$

where  $I_{gtk}$  is an indicator function such that:

$$I_{gtk} = \begin{cases} 1 & \text{if visit } k \text{ to site } g \text{ in year } t \text{ is made} \\ 0 & \text{if visit } k \text{ to site } g \text{ in year } t \text{ is missed.} \end{cases} \quad (4.14)$$

It is readily verified that Equation (4.13) gives  $c_{gt} = 0$  when site  $g$  is not operated at all in year  $t$ . Further, at a given site,  $g$ , the annual recapture probabilities,  $c_{gt}$ , are identical in years for which the same number and set of visits are made, an appropriate assumption under the constant effort design.

Recall that if the annual recapture probabilities are assumed to be constant across time, the m-array cell probabilities ( $p_{git}$  and  $\chi_{gi}$ ; the probabilities that a bird captured and released in year  $i$  is, respectively, seen again in year  $t$ , or never seen again, for site  $g$ ) under model CJS are given by Equation 2.7. For  $T$  study years, if time dependence in the annual recapture probabilities is permitted, the m-array cell probabilities under model CJS are modified accordingly:

$$p_{git} = \begin{cases} S_{a,i}c_{gt} & \text{for } t = i + 1 \leq T, \\ S_{a,t-1}c_{gt} \prod_{h=i}^{t-2} S_{a,h}(1 - c_{g(h+1)}) & \text{for } t = i + 2, \dots, T \end{cases}$$

and  $\chi_{gi} = 1 - \sum_{t=i+1}^T p_{git} \quad i \leq T. \quad (4.15)$

Model MV is readily formed by replacing the recapture probabilities in Equation (4.15) with the values derived from expression (4.13) via the parametric regression given by Equation (4.12).

Note that the linear regression,  $v_{gk} = s_g^r + \beta_1^r \kappa + \beta_2^r \kappa^2$ , is more readily interpretable than the logistic regression of Equation (4.12). However, a distinct advantage of the logistic model is that the logit link ensures that the probabilities  $v_{gk}$ , and hence  $c_{gt}$ , are restricted to the allowable range  $[0,1]$ . When a linear regression is applied constraints must be imposed on the regression parameters ( $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$ ) to ensure that the derived values of  $v_{gk} \in [0, 1]$ . For example, an update for  $s_g^r \in [-\infty, \infty]$  can only be accepted if  $v_{gk} \in [0, 1] \forall k, g$ . Whether this condition is met depends on the current values of  $\beta_1^r$  and  $\beta_2^r$ . In the Bayesian model fitting framework, imposing such constraints made tuning the proposal distributions for the regression parameters extremely difficult. Furthermore, as the range of acceptable updates was frequently limited, poor mixing of the MCMC chain often resulted.

The information in the data pertaining to the estimation of the recapture regression parameters ( $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$ ) arises from years with missed visits, as it is these years in which  $c_{gt}$  changes. Without missed visits, estimation of the regression parameters is controlled entirely by their priors and the overall annual recapture probabilities, and not by the underlying trend in seasonal catchability. In order to obtain reliable estimates of these regression parameters, data with a sufficiently large number of years with incomplete coverage, and in which each visit is missed a sufficiently large number of times, are required.

*Implementation*

Model MV is readily fitted using Bayesian techniques, however care must be taken when specifying priors. In the absence of *a priori* information regarding the parameter values uninformative priors are desired (Gelman et al., 2004). For the survival parameters we use independent Uniform[0,1] priors. For the underlying regression parameters  $(s_g^r, \beta_1^r, \beta_2^r)$  uninformative prior specification is more difficult - priors uninformative about  $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$  may induce informative priors on  $v_{gk}$  or  $c_{gt}$ . For example, vague Normal( $0, \sigma^2$ ) priors on  $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$  do not present a Uniform prior over the interval [0,1] for the recapture probabilities ( $v_{gk}$  and  $c_{gt}$ ). Rather, as  $\sigma^2 \rightarrow \infty$  the induced priors on  $v_{gk}$  and  $c_{gt}$  tend to point masses on 0 and 1 (i.e. “bathtub” shaped priors are induced).

In the case of the simple logistic model,  $\text{logit}(v_{gk}) = s_g^r$ , a “logistic prior” on  $s_g^r$ , that is

$$p(s_g^r) = \frac{\exp(s_g^r)}{(1 + \exp(s_g^r))^2}, \quad (4.16)$$

does present a Uniform[0,1] prior on the recapture probabilities (King and Brooks, 2008). However the specification of similarly uninformative priors under the regression model for  $v_{gk}$  (given in Equation (4.12)) is complicated by the presence of the additional  $\beta^r$  terms.

Furthermore, the shape of the logit link function means that large (or small) values for the regression parameters have similar  $v_{gk}$  values, i.e.  $\text{logit}^{-1}(5) \approx \text{logit}^{-1}(6)$ . In the absence of sufficient data-based information regarding  $s_g^r$ ,  $\beta_1^r$  or  $\beta_2^r$ , the use of uninformative priors may cause a convergence issue. When neither the prior nor the likelihood are altered much by changes to  $s_g^r$ ,  $\beta_1^r$  or  $\beta_2^r$ , the MCMC chain drifts off to larger and larger (positive or negative) values. In such circumstances restricting the priors to “sensible” ranges is advisable.

These observations highlight the importance of specifying sensible priors, and conducting a prior sensitivity analysis (Ghosh et al., 2006). When a wealth of data are available prior sensitivity is not an issue - the likelihood will outweigh the prior. However, as CES data are likely to contain only limited information on the regression parameters, in all analyses model MV was fitted with a range of different sets of independent priors to assess potential prior sensitivity. In particular, the analyses are repeated with a selection of priors; for  $\beta_1^r$  and  $\beta_2^r$  either Normal(0,10000), Normal(0,10), Uniform[-10,10] or Uniform[-5,5] priors, and for the  $G$  site-effects,  $s_g^r$ , either the priors listed above, or the logistic prior.

We use random walk Metropolis-Hastings updates, with Normal proposal distributions for the real-valued regression parameters, and with Uniform proposal distributions, appropriately truncated, for the survival probability parameters. Proposal variances and step lengths are tuned *a priori*. The simulations are run for 200,000 iterations with the first 100,000 iterations discarded as burn-in. For all analyses, independent replications with different overdispersed starting points produced essentially identical posterior estimates. Further, examination of the trace-plots demonstrated good movement and mixing through the parameter space. The bespoke code was written in C.

## ii Testing Model MV - a Simulation Study

We now describe two simulation studies conducted to assess model MV. In both cases live-recapture data with known survival probabilities ( $S_{a,t}$ ), known recapture regression parameters ( $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$ ) and arising from years with incomplete coverage, are artificially generated. In the first study a simple data set is simulated, with a large number of ringed birds released annually and a large number of visits missed completely at random. In the second study a more realistic data set is simulated that has lower release numbers, consistent with the number present in CES data, and far fewer missed visits chosen according to the pattern observed in CES data.

For both studies, data from  $G$  sites over  $T$  years with (potentially)  $K = 12$  visits per year are simulated according to the following scheme. For each site  $g$ , and for each year  $t$  it is operated:

1. Select the within-year visits to be missed. If a visit is deemed missed set  $I_{gtk} = 0$ , else  $I_{gtk} = 1$ .
2. Using Equations (4.12) and (4.13) calculate  $c_{gt}$  for the pre-specified values of  $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$ .
3. Using the derived values of  $c_{gt}$ , and the pre-specified values of  $S_{a,t}$ , determine the m-array cell probabilities as given by Equation (4.15).
4. Generate the number of ringed birds released in year  $i$ , denoted by  $R_{gi\bullet}$ , from a discrete Uniform distribution.
5. For the  $R_{gi\bullet}$  birds released in year  $i$  simulate the number that are next recaptured in year  $t$  (denoted by  $R_{git}$ ) and the number that are never seen again (denoted by  $Z_{gi}$ ) using the Multinomial distribution with cell probabilities given in Step 3.

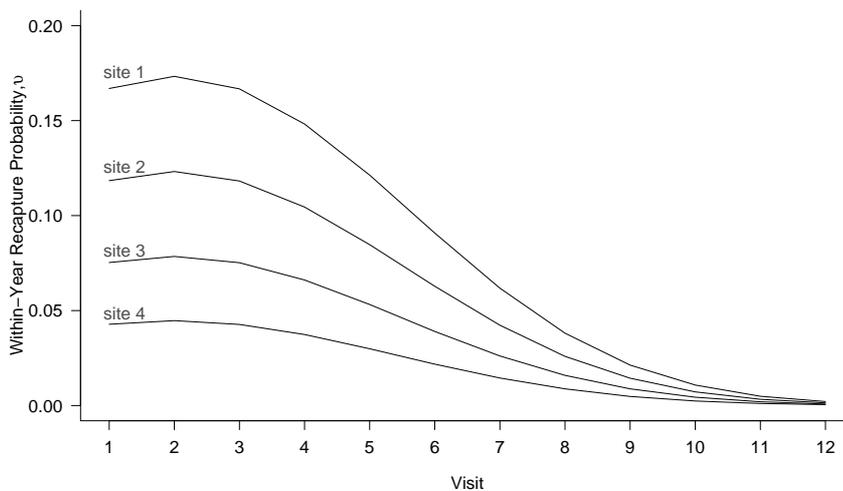
For both simulation studies, parameter estimates from the models that i) ignore the presence of missed visits (model CJS of Section 2.4.1.i) and ii) adjusts recapture probabilities for missed visits (model MV) are compared to each other, and to the true underlying values. These comparisons enable the effect of missed visits on the estimates of  $S_{a,t}$ , and the reliability of model MV, to be assessed.

The sensitivity of model MV to the priors for the recapture regression parameters is investigated by specifying different sets of priors for these parameters, and comparing the resulting posterior estimates. Further, the robustness of the survival estimates to these different prior specifications is examined.

### *Simple Data Set*

A simple, fully balanced live-recapture data set (i.e. data for all sites over all years) from  $G = 4$  sites and  $T = 10$  years is simulated following the 5 step scheme outlined above, and in particular where:

- visits are selected to be missed, at random, with probability 0.4 (Step 1),
- recapture regression parameters are chosen such that i) there are large differences in the within-year recapture probabilities,  $v_{gk}$ , between visits and between sites (see Figure 4.14), ii) the trend in  $v_{gk}$  across visits is consistent with that expected to be observed in CES data (i.e. Figure 4.11), and iii) the annual recapture probabilities,  $c_{gt}$ , are similar in magnitude to those estimated from the analyses of Sedge and Reed Warbler data with complete coverage (see Table 4.11),

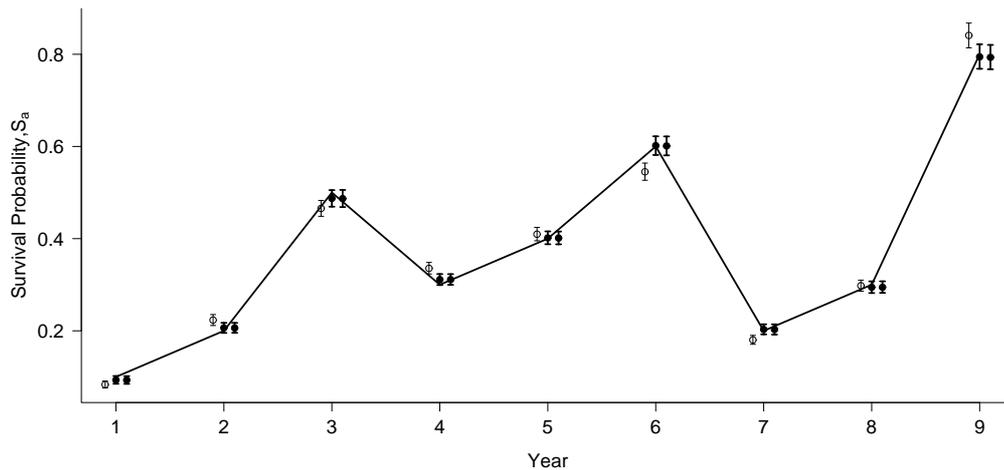


**Figure 4.14** The within-year recapture probabilities,  $v_{gk}$ , used to simulate a “simple” live-recapture data set.

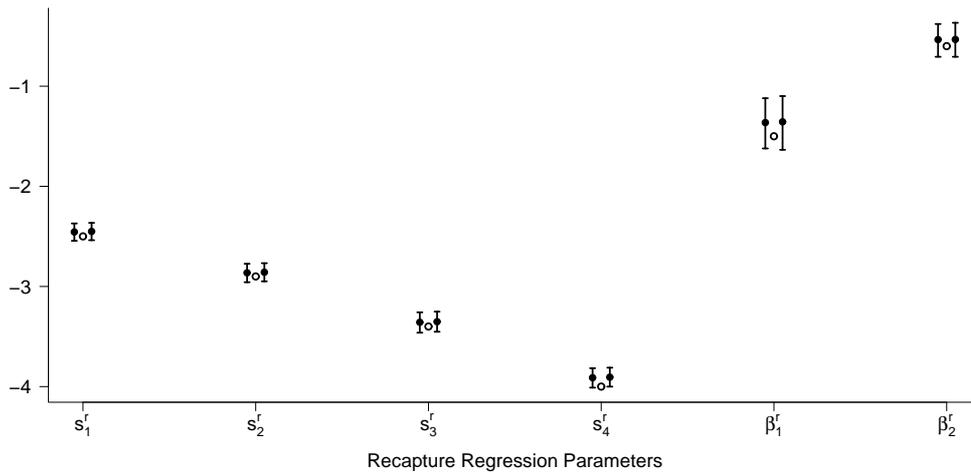
- annual survival probabilities are chosen (mean = 0.38, range = 0.1 - 0.8) that are comparable to those estimated from the analyses of Sedge and Reed Warbler data with complete coverage,
- $R_{gi\bullet} \sim \text{Uniform}[3000,4000]$ , that is a large number of birds are released at each site in every year (Step 4).

Two different sets of independent Normals priors are given to the recapture regression parameters to assess prior sensitivity. Initially very flat Normal priors with variance 10000 are specified, then these Normal priors are made considerably tighter by reducing the variance to 10.

Estimates of survival from model MV, with both sets of priors, are compared to those from model CJS, and to their true underlying values, in Figure 4.15. Ignoring the presence of missed visits, by fitting the standard CJS model, results in some bias in the estimates of survival (Figure 4.15). This bias is eliminated when model MV is used to adjust  $c_g$  for missed visits. Indeed, model MV performs very well with estimates of  $S_{a,t}$ ,  $s_g^r$ ,  $\beta_1^r$ , and  $\beta_2^r$  all in agreement with their true underlying values (Figures 4.15 and 4.16). Furthermore, regardless of the prior specification, estimates under model MV are virtually unchanged, indicating that the posterior distributions of the regression parameters are data-driven with little prior sensitivity.



**Figure 4.15** Annual survival probabilities estimated from artificially simulated “simple” data under model CJS (left) and model MV with either i) flat Normal(0,10000) priors (centre) or ii) tighter Normal(0,10) priors (right) for the recapture regression parameters. The solid line denotes the true underlying  $S_{a,t}$  values used to simulate the data.



**Figure 4.16** Recapture regression parameters of model MV estimated from artificially simulated “simple” data. Points denote posterior means and bars 95% symmetric credible intervals. Two different sets of priors for the regression parameters are used: left - flat Normal(0,10000) priors, right - tighter Normal(0,10) priors. The empty circles (centre) denote the true underlying values from which the data are simulated.

Table 4.11 gives the mean annual recapture probability, across all years and sites, estimated under models CJS and MV. For the latter both the mean recapture probability that would have been achieved had all 12 visits been made, and that observed given the presence of missed visits, is presented.

**Table 4.11** Comparing the mean annual recapture probability,  $c$ , under models CJS and MV with the true mean  $c$  used in the simulation of the “simple” data set. For model MV and the truth, the mean  $c$  had all 12 visits been made, and the mean  $c$  given that some visits were missed, are presented.

	Model CJS	Model MV	Truth
	0.31	0.30	0.30
If all 12 visits made		0.45	0.45

Recall that for model CJS catch effort (at a site) is assumed to be constant over time, implying that the probability of recapture is likewise. When visits are missed the overall capture effort for that year is reduced and the mean  $c_g$  across all years decreases. Under model CJS the estimate of  $c_g$  is essentially a weighted mean of the individual  $c_{gt}$  values (which vary as a result of missed visits) across all years site  $g$  is operated. That is, model CJS “adjusts” for missed visits by reducing  $c_g$  in all years.

By constraining  $c_g$  to be constant across time, when missed visits cause it to vary temporally, is expected to bias the estimates of  $S_{a,t}$ . For example, survival

probabilities that precede years where the estimated value of  $c_g$  is less than the true underlying  $c_{gt}$  values (from the set of sites surveyed in year  $t$ ) will be positively biased. Such a phenomenon causes the positive bias in the estimate of  $S_{a,2}$  and  $S_{a,9}$  when model CJS is applied (Figure 4.15). Conversely, survival estimates are negatively biased in years 6 and 7. Here, model CJS has estimated  $c_g$  too high relative to the true underlying values of  $c_{g7}$  and  $c_{g8}$ , respectively, for at least the majority of the sites. In this simulation study visits are missed completely at random. Were this not the case, say if years with missed visits tended to have several, such bias is likely to be greater.

That the recapture regression parameters, and therefore the survival estimates, from this simple simulation study are insensitive to the different priors is not surprising (Figures 4.15 and 4.16). With such a high number of releases, and with many missed visits causing large changes in  $c$  (see Table 4.11), there is a considerable amount of information in the data regarding the regression parameters. Therefore in this particular simulation study, the information in the data, summarised in the Multinomial likelihood function, overwhelms the prior information. However, the release numbers and frequency of missed visits in the CES data are smaller. As the CES data are likely to be considerably less informative with regards to the regression parameters, here prior sensitivity could be an issue. A second data set is simulated to test the performance of model MV when the data are more limited.

#### *A More Realistic Data Set*

In the second simulation study a more realistic data set, with  $G = 40$  sites and  $T = 22$  study years, is simulated. Note that each site need not contribute data in all 22 years, as in the CES scheme. Here the release numbers, pattern and frequency of missed visits, and annual recapture and survival probabilities are based on those observed in the Sedge Warbler CES data (1983-2005).

In particular:

- visits are selected to be missed according to the pattern observed in Sedge Warbler CES live-recapture data from a subset of 40 sites (Step 1),
- recapture regression parameters are chosen such that i) the trend in the resulting within-year recapture probabilities,  $v_{gk}$ , over the mist-netting season, and ii) the scale of the derived annual recapture probabilities,  $c_{gt}$ , are comparable to those estimated from the analysis of Sedge Warbler CES data (see Figure 4.18),

- annual survival probabilities are chosen (mean = 0.30, range = 0.14 - 0.49) that are comparable to those estimated from the analysis of Sedge Warbler CES data,
- $R_{gi\bullet} \sim \text{Uniform}[5,80]$ , which is consistent with the Sedge Warbler CES data (Step 4).

Five different sets of prior distributions for the recapture regression parameters of model MV are considered.

For  $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$ :

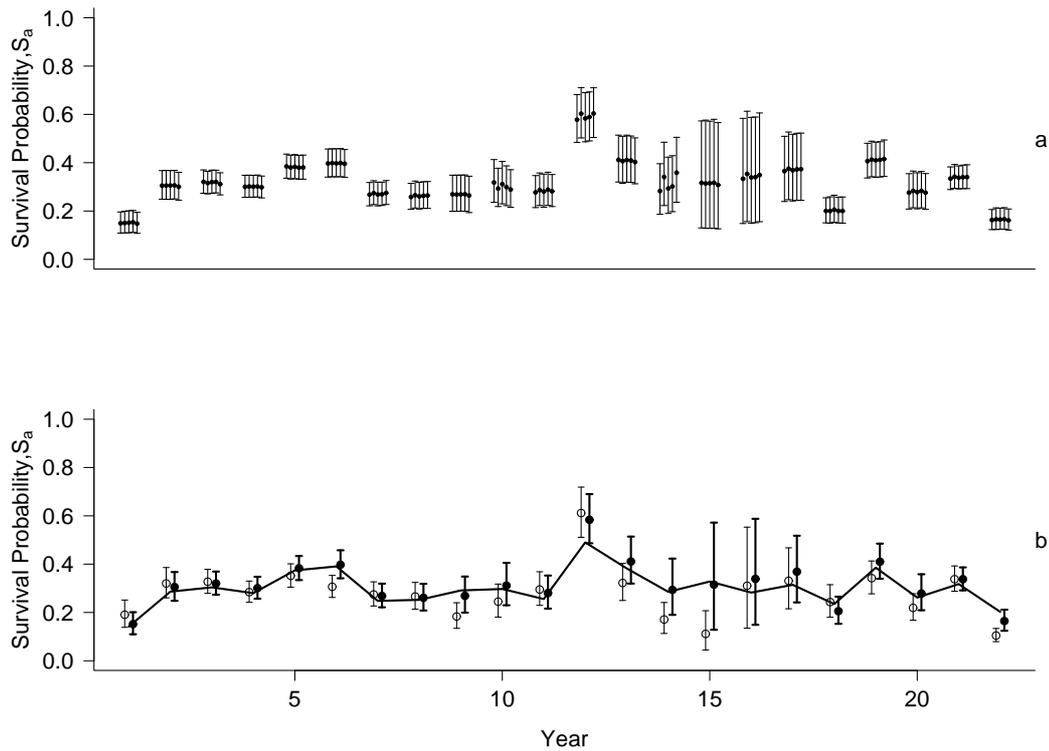
1. Normal(0,10000),
2. Normal(0,10),
3. Uniform[-10,10],
4. Uniform[-5,5],

and

5. Logistic for  $s_g^r$  (Equation (4.16)), Uniform[-5,5] for  $\beta_1^r$  and  $\beta_2^r$ .

The estimates of survival, under model MV, are insensitive to these different priors (Figure 4.17a), and are in general agreement with their true underlying values (Figure 4.17b). Ignoring the presence of missed visits, and using model CJS to estimate survival, results in some bias (Figure 4.17b). Model MV has successfully eliminated this bias, but there is a trade-off - the cost to precision in some years is high, for example year 15. There is little precision in the estimate of  $S_{a,15}$  as, in the simulated data, year 15 and the years immediately before and after it have unusually few sites contributing information. Further, for those few sites that are contributing data, only one per year has complete coverage. This lack of information, which when compounded by the uncertainty in the recapture probabilities, results in the loss of precision under model MV. In a data set with many missed visits, this implies reduced precision for model MV, but biased estimates for model CJS. In general, the loss of precision should not be as pronounced as that for  $S_{a,15}$ , since CES data typically does not have these odd years with so few sites contributing information.

The paucity of information on  $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$  in this more realistic data set results in prior sensitivity on these parameters - the shape and width of their posterior distributions differ considerably depending on the prior specified (Table 4.12). As such sensitivity results in considerable variability in the derived within-year recapture probabilities,  $v_{gk}$ , the estimated trend may not reliably predict the true underlying seasonal change in catchability from which the data are



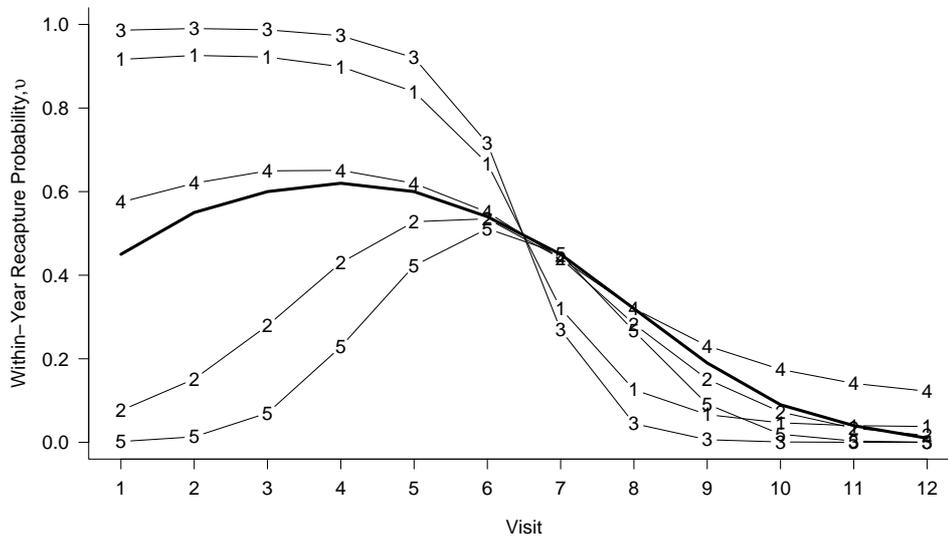
**Figure 4.17** Estimated survival probabilities from artificially simulated “realistic” data. Bars denote 95% symmetric credible intervals, and points denote posterior means. a: Estimates from model MV in which five different sets of priors are specified for the recapture regression parameters. b: Comparing the survival estimates from model CJS (left) and model MV (with Uniform[-10,10] priors for the regression parameters) (right) with their true underlying values from which the data are simulated (line).

generated (Figure 4.18). In general this is not a concern, for typically it is the survival parameters that are of interest, with the recapture probabilities being viewed as nuisance parameters.

Despite the seasonal change in catchability being poorly and inconsistently estimated, a robustness of the survival estimates to the changes in the regression parameters is evident in Figure 4.17a. Such robustness is due to a wide range of regression parameters producing nearly identical estimates of  $c_{gt}$  given the number and pattern of missed visits present in the data. In spite of large differences in the estimates of  $s_g^r$ ,  $\beta_1^r$  and  $\beta_2^r$ , the derived values of  $c_{gt}$  are very similar (the 95% symmetric credible intervals overlap) regardless of the priors specified. Further, these estimates are consistent with the true underlying values of  $c_{gt}$  from which the data are simulated. Thus the information in the CES data regarding  $c_{gt}$  is driving the estimation of the regression parameters, selecting values consistent with their priors.

**Table 4.12** Posterior means (95% symmetric credible intervals) for the recapture regression parameters ( $\beta_1^r$ ,  $\beta_2^r$ ) estimated from model MV, using from the simulated “realistic” data, under different prior specifications.

Prior	$\beta_1^r$	$\beta_2^r$
Normal(0,10000)	-4.43 (-8.755, -0.806)	-0.81 (-2.897, 1.105)
Normal(0,10)	-1.11 (-3.014, 0.339)	-2.48 (-5.368, -0.747)
Uniform[-10,10]	-4.28 (-8.054, -0.146)	-1.55 (-6.159, 0.831)
Uniform[-5,5]	-2.49 (-4.850, -0.134)	-1.68 (-4.464, 0.466)
Logistic, Uniform[-5,5]	-0.84 (-2.285, 0.569)	-3.94 (-4.956, -2.296)
<b>Truth</b>	<b>-1.40</b>	<b>-1.00</b>



**Figure 4.18** Estimated trends in the within-year recapture probabilities from the simulated “realistic” data, using model MV, when different priors are given to the recapture regression parameters: 1 - Normal(0,10000), 2 - Normal(0,10), 3 - Uniform[-10,10], 4 - Uniform[-5,5], and 5 - Logistic/Uniform[-5,5]. Under each, the posterior mean of  $v_{gk}$  is calculated from Equation (4.12) when  $s_g^r = 0$ . The true trend, from which the data are simulated, is denoted by the thick line.

### iii Brief Concluding Statement

In conclusion, the simulation studies indicate that model MV performs better than model CJS in producing reliable estimates of  $S_{a,t}$  from CES style data in which the presence of missed visits causes temporal variation in  $c_g$ .

Model MV performed particularly well in the first simulation study. Here a large number of missed visits causes considerable variation in  $c_g$  between the years. In the second simulation study, when far sparser data more realistically

considered to have arisen from the CES scheme are studied, model MV still appears to produce unbiased estimates of  $S_{a,t}$ . The pattern and frequency of missed visits in this data set, however, provides very little information on the recapture regression parameters, and thus their estimation is driven by their priors and the information in the data regarding  $c_{gt}$ . Unbiased estimates of  $c_{gt}$  and  $S_{a,t}$  seem to result despite the discrepancy between the estimated regression parameters and their true values.

Model MV only considers missed visits, and assumes all birds are “residents”. As no account has yet been made for the temporary presence of transient birds at the CES sites, estimates of apparent survival obtained from model MV will be negatively biased when CES live-recapture data are adopted. We now extend model MV to account explicitly for transient birds. The resulting model, dealing with both missed visits and transient birds, is referred to as “model EPMV”.

#### 4.4.4 Survival Model for CES data with Missed Visits and Transient Birds - Model EPMV

We begin by demonstrating how the individual models for CES style data containing either i) missed visits (model MV) or ii) transient birds (model EP of Section 2.4.1.iv) can be combined into a single model for data where both feature. This model is then modified to let the within-year recapture probability increase after the initial capture that year, an effect we refer to as the “encounter effect”. Such an effect was observed in the analyses of Sedge and Reed Warbler CES data in Chapter 2, and was hypothesised to be the consequence of individual heterogeneity in the capture rates (see Section 2.4.5). The implications of failing to adjust for this encounter effect are discussed. The models which exclude, or include, the encounter effect are denoted by EPMV(0) and EPMV( $\xi$ ) respectively.

##### i Model EPMV(0) - *No Encounter Effect*

The presence of transient birds in the CES live-recapture data (i.e. birds not resident near the CES site during breeding) produces a negative bias in the estimates of apparent adult survival ( $S_{a,t}$ ) if not adequately accounted for (Pradel et al., 1997). In Chapter 2 we efficiently extended the standard transient model of Pradel et al. (1997) to utilise the ancillary within-year recapture information,

in addition to the between-year recaptures, to identify residents. The resulting model, referred to as “model EP”, produces unbiased estimates of adult survival with improved precision over the standard Pradel et al. (1997) approach. However, to model the within-year recaptures, the parametrisation of model EP involves additional site-specific “evasion” parameters, denoted by  $\varepsilon_g$ .

Crucial to the derivation of model EP was the assumption of constant effort (at any given site). Violations of this assumption will cause time dependence in the site-specific recapture ( $c_g$ ) and evasion ( $\varepsilon_g$ ) probabilities. For example, the occurrence of missed visits in year  $t$  is expected to decrease  $c_g$  and increase  $\varepsilon_g$  in that year, i.e. if fewer visits are made the chance of being caught is reduced, and the chance of “evading” future recapture within the same year is increased.

Recall that the full likelihood for model EP, given by Equation (2.13), is the product of three independent likelihoods (see Section 2.4.1:iv):

1.  $f(\mathbf{F}, \mathbf{X} | \tau, \boldsymbol{\varepsilon}_{1\dots G}, \mathbf{c}_{1\dots G}, \mathbf{S}_{a,1\dots T-1})$ , a Multinomial based likelihood from the “first recapture” segment of the capture histories, i.e. arising from the m-array for “newly-marked” birds,
2.  $f(\mathbf{R}, \mathbf{Z} | \mathbf{c}_{1\dots G}, \mathbf{S}_{a,1\dots T-1})$ , a Multinomial based likelihood from the “subsequent recapture” segment of the capture histories, i.e. arising from the m-array for “previously-marked” birds,
3.  $f(\mathbf{M}' | \boldsymbol{\varepsilon}_{1\dots G}, \mathbf{M}' + \mathbf{M}'')$ , a Binomial based likelihood from the “multiple within-year encounters” in the years subsequent to ringing.

Recall that, for site  $g$ ,  $f_{git}$  denotes the probability that a bird ringed and released in year  $i$  is first recaptured in year  $t$ , and  $x_{gi}$  denotes the probability that it is never seen again. Assuming time-invariance in the recapture ( $c$ ) and evasion ( $\varepsilon$ ) probabilities, the cell probabilities corresponding to the “newly-marked” m-array ( $f_{git}$  and  $x_{gi}$ ) are given by Equation (2.12). When time dependence in  $c$  and  $\varepsilon$  is permitted, these cell probabilities are modified as follows:

$$f_{git} = \begin{cases} (1 - \tau)(1 - \varepsilon_{gi}) & \text{if } t = i \leq T, \\ (1 - \tau)\varepsilon_{gi}S_{a,i}c_{gt} & \text{if } t = i + 1 \leq T, \\ (1 - \tau)\varepsilon_{gi}S_{a,t-1}c_{gt} \prod_{h=i}^{t-2} S_{a,h}(1 - c_{g(h+1)}) & \text{if } t \in [i + 2, T], \end{cases}$$

and  $x_{gi} = 1 - \sum_{t=i}^T f_{git} \quad i \leq T. \quad (4.17)$

Allowing for time dependence in  $c$ , the cell probabilities corresponding to the “previously-marked” m-array,  $p_{git}$  and  $z_{gi}$ , are given by Equation (4.15).

The “multiple within-year encounters” after the year of ringing are modelled by

$$M'_{gt} | M'_{gt} + M''_{gt} \sim \text{Bin}(M'_{gt} + M''_{gt}, \varepsilon_{gt}),$$

for  $t \in [2, T]$ , where  $M'_{gt}$  and  $M''_{gt}$  are respectively the number of ringed resident birds recaptured in year  $t$ , at site  $g$ , exactly once or repeatedly.

As with model MV, an expression for  $c_{gt}$  is derived from Equation (4.13) via the parametric regression model, Equation (4.12). We now describe how an analogous expression for  $\varepsilon_{gt}$  is obtained.

Recall that for a particular site  $g$ :

$$\begin{aligned} \varepsilon_{gt} &= \Pr(\text{a resident bird is caught exactly once within year } t | \text{recaptured in year } t) \\ &= \frac{\Pr(\text{a resident bird is caught exactly once within year } t)}{\Pr(\text{a resident bird is recaptured in year } t)}. \end{aligned}$$

Assuming that individual birds are caught independently at successive within-year visits, and that all birds (at site  $g$  in year  $t$ ) experience identical within-year recapture probabilities, we derive:

$$\begin{aligned} \varepsilon_{gt} &= \frac{\sum_{k=1}^K \left( I_{gtk} v_{gk} \times \prod_{h=1, h \neq k}^K (1 - v_{gh})^{I_{gth}} \right)}{1 - \prod_{k=1}^K (1 - v_{gk})^{I_{gtk}}} \\ &= \frac{\sum_{k=1}^K \left( I_{gtk} v_{gk} \times \prod_{h=1, h \neq k}^K (1 - v_{gh})^{I_{gth}} \right)}{c_{gt}}, \end{aligned} \tag{4.18}$$

where  $I_{gtk}$  is the indicator function (given in Equation (4.14)) accounting for missed visits, and the  $v_{gk}$  (and  $v_{gh}$ ) defined here are modelled using the parametric regression, Equation (4.12).

## ii Model EPMV( $\xi$ ) - *Encounter Effect*

The analyses of Sedge and Reed Warbler CES data in Chapter 2 (Section 2.4.5) gave rise to lower than expected  $\varepsilon_g$  estimates had i) within-year recaptures been independent, and ii) all birds experienced the same recapture rates. This suggests that, within a year, a bird once caught has a higher probability of being seen at a subsequent visit than a bird as yet unseen. Individual heterogeneity in the recapture rates, caused say by territories being at variable proximity to the mist-net, was hypothesised to be the cause of this “encounter effect”.

To adjust for the encounter effect, the expression for  $\varepsilon_{gt}$  given in Equation (4.18) is modified as follows:

$$\begin{aligned} \varepsilon_{gt} &= \frac{\sum_{k=1}^K \left( I_{gtk} \nu_{gk} \times \prod_{h=1}^{k-1} (1 - \nu_{gh})^{I_{gth}} \times \prod_{h=k+1}^K (1 - \nu_{gh}^*)^{I_{gth}} \right)}{1 - \prod_{k=1}^K (1 - \nu_{gk})^{I_{gtk}}} \\ &= \frac{\sum_{k=1}^K \left( I_{gtk} \nu_{gk} \times \prod_{h=1}^{k-1} (1 - \nu_{gh})^{I_{gth}} \times \prod_{h=k+1}^K (1 - \nu_{gh}^*)^{I_{gth}} \right)}{c_{gt}} \end{aligned} \quad (4.19)$$

where:

$$\begin{aligned} \text{logit}(\nu_{gk}) &= s_g^r + \beta_1^r \kappa + \beta_2^r \kappa^2 \\ \text{logit}(\nu_{gk}^*) &= s_g^r + \beta_1^r \kappa + \beta_2^r \kappa^2 + \xi. \end{aligned} \quad (4.20)$$

That is, within the current year, after the initial capture, the log odds of recapture at a subsequent visit is altered by  $\xi \in [-\infty, +\infty]$ . This response only applies to the within-year recapture probabilities for the current year, and does not affect those in later years, when a bird may occupy a nearby, but different, territory.

Under the assumption of constant effort,  $\xi$  remains constant between years, however  $\xi$  may be site dependent, for example if the length, position or type of the mist-net affected recapture heterogeneity. When  $\xi$  is site dependent we denote the model by  $\text{EPMV}(\xi_g)$ .

Mathematically, the modification to  $\varepsilon_{gt}$  given by Equation (4.20) is equivalent to including a temporary “trap response” (Kendall et al., 1995) that lasts (only) the duration of the current mist-netting season, and where

$$\xi \begin{cases} > 0 & \text{“trap-happy” response,} \\ = 0 & \text{no trap response,} \\ < 0 & \text{“trap-shy” response.} \end{cases}$$

For the CES data this is unlikely to be the biological definition of  $\xi$ , rather  $\xi$  should be thought of as a constant that enables the discrepancy between  $c$  and  $\varepsilon$ , hypothesised to be the consequence of individual heterogeneity in recapture rates, to be corrected. If individual heterogeneity were indeed responsible for the observed discrepancy then  $\xi > 0$ . Although such a constraint can be readily imposed on the prior for  $\xi$  we allow  $\xi \in [-\infty, \infty]$  as almost certainly a combination of mechanisms is causing the observed discrepancy between  $c$  and  $\varepsilon$ , some of which work in the opposite direction. For example, birds learning to avoid

the mist-nets (a “trap-shy” response) and temporal variation in the within-year recapture probabilities,  $v$ , both result in a negative encounter effect, i.e.  $\xi < 0$ .

### *Implementation*

The EPMV models (EPMV(0), EPMV( $\xi$ ), EPMV( $\xi_g$ )) are readily fitted using Bayesian techniques. We use random walk Metropolis-Hastings updates, with Normal proposal distributions for the real-valued parameters ( $s_g^r$ ,  $\beta_1^r$ ,  $\beta_2^r$ ,  $\xi_g$ ), and Uniform proposal distributions, appropriately truncated, for the probability parameters ( $S_{a,t}$ ,  $\tau$ ). Proposal variances and step lengths are tuned *a priori*.

Vague, independent Uniform[0,1] priors are given to the survival and transient probabilities. However, as with model MV, specifying vague priors for the underlying recapture regression parameters ( $s_g^r$ ,  $\beta_1^r$ ,  $\beta_2^r$  and  $\xi_g$ ) that induce uninformative priors on the recapture rates ( $v_{gk}$ ,  $v_{gk}^*$  and  $c_{gt}$ ) is problematical. As a default,  $s_g^r$  and  $\xi_g$  are given Logistic priors (Equation (4.16)), whereas  $\beta_1^r$  and  $\beta_2^r$  are given Uniform[-10,10] priors. However, to assess potential prior sensitivity all analyses were repeated with a range of different sets of independent priors for these parameters (e.g., Normal(0,10000), Normal(0,10) and Uniform[-5,5]).

Using bespoke code written in C, the simulations are run for 200,000 iterations, with the first 100,000 iterations discarded as burn-in. The resulting trace-plots indicated good mixing and movement through the parameter space, and that the chain had converged before the burn-in period had ended. Further, a variety of different starting points gave essentially identical posterior results.

### **iii Testing the EPMV Models - a Simulation Study**

To assess the EPMV models three CES style live-recapture data sets at the individual visit level are simulated, and in which records from transient birds and from years with missed visits are both present. The simulations differ only in the values specified for  $\xi_g$ . In the first study  $\xi_g = 0$ , consistent with the assumption of recapture homogeneity. Then the effect of heterogeneity was introduced by setting  $\xi_g > 0$ , either constant ( $\xi_g = 0.2 \forall g$ ) or varying ( $\xi_g \in [0.1, 0.4]$ ) across sites.

Values for the survival and transient probabilities are chosen to reflect those observed in Sedge and Reed Warbler CES data. Recapture regression parameters

are selected to provide large seasonal and site-specific changes in the within-year recapture probabilities ( $v_{gk}$ ) but that produce annual site-specific recapture ( $c_{gt}$ ) and evasion ( $\varepsilon_{gt}$ ) probabilities comparable to those observed in CES data.

Large, fully-balanced data sets, from  $G = 4$  sites over  $T = 10$  years with (potentially)  $K = 12$  visits per year, are simulated according to the following scheme.

For each site, and for all  $T$  years:

1. Select at random the within-year visits to be missed, in which each visit has a probability of 0.4 of being missed. If a visit is deemed missed set  $I_{gtk} = 0$ , else  $I_{gtk} = 1$ .
2. For the pre-specified values of  $s_g^r$ ,  $\beta_1^r$ ,  $\beta_2^r$ , and  $\xi_g$  calculate  $c_{gt}$  and  $\varepsilon_{gt}$ , via Equation (4.12), using Equations (4.13) and (4.19) respectively.
3. Generate the number of “unringed” birds caught,  $N_{gt}$ , from a discrete Uniform distribution, Uniform[1000,4000]. Let  $N_{gt}^{trans} \sim \text{Bin}(N_{gt}, \tau)$  be transients and  $N_{gt}^{res} = N_{gt} - N_{gt}^{trans}$  be residents.

#### 4. Simulate Transient Data

- (a) For each transient bird caught select the single visit it was seen at. That is, sample  $k \in [1, K]$  with probability weights proportional to  $v_{gk}I_{gtk}$ .
- (b) Record the within-year live-recapture history for each transient bird.

#### 5. Simulate Resident Data

Using the derived values of  $c_{gt}$ , and the pre-specified values of  $S_{a,t}$ , determine the m-array cell probabilities as given by Equation (4.15).

For each resident bird in  $N_{gt}^{res}$ :

- (a) Using the Multinomial distribution, with cell probabilities calculated above, determine which year, if any, it was first recaptured.
  - (b) If the bird was recaptured in year  $j < T$ , “re-release” in year  $j$  and repeat above step until the bird is “never seen again”.
  - (c) For every year caught (including the inaugural ringing year) generate the set of within-year visits the bird was encountered. That is from the set of all possible visit capture histories, given seen during at least one visit, select one history at random weighted according to the probability determined by  $s_g^r$ ,  $\beta_1^r$ ,  $\beta_2^r$ ,  $I_{gtk}$  and  $\xi_g$ .
  - (d) Record its full within-year live-recapture history.
6. Combine the transient and resident within-year live-recapture data.

Models EPMV(0), EPMV( $\xi$ ) and EPMV( $\xi_g$ ) are fitted to all three simulated data sets. Each analysis is repeated several times with different prior distributions, Normal(0,10000), Normal(0,10), Uniform[-10,10], Uniform[-5,5] or Logistic as appropriate, specified for the recapture regression parameters ( $s_g^r$ ,  $\beta_1^r$ ,  $\beta_2^r$ ,  $\xi_g$ ).

All three simulated data sets consist of capture histories from a large number of birds, and have a high proportion of missed visits that result in large deviations from the assumption of constant effort (within a site). Therefore, as is to be expected, the data sets are informative regarding  $s_g^r$ ,  $\beta_1^r$ ,  $\beta_2^r$ , and (when included in the model)  $\xi$  or  $\xi_g$ , and the estimation of these parameters was not sensitive to the priors specified. The results from the simulation study are summarised in Table 4.13.

**Table 4.13** Summary of the parameter estimates obtained under the different EPMV adult survival models using artificially simulated data. For the sets of parameters  $S_{a,t}$ ,  $c_{gt}$ , and  $\varepsilon_{gt}$ , the average absolute difference between the posterior means and the true underlying values are given, and in brackets the average width of the 95% symmetric credible intervals. For  $\tau$  the absolute difference between the posterior mean and the true underlying value is given, and in brackets the width of the 95% symmetric credible interval.

	EPMV(0)	EPMV( $\xi$ )	EPMV( $\xi_g$ )
<b>Study 1, <math>\xi_g = 0 \forall g</math></b>			
$S_{a,t}$	0.015 (0.047)	0.015 (0.050)	0.015 (0.051)
$c_{gt}$	0.010 (0.027)	0.009 (0.033)	0.012 (0.045)
$\varepsilon_{gt}$	0.005 (0.016)	0.005 (0.017)	0.005 (0.020)
$\tau$	0.003 (0.038)	0.004 (0.037)	0.004 (0.038)
<b>Study 2, <math>\xi_g = 0.2 \forall g</math></b>			
$S_{a,t}$	0.033 <sup>†</sup> (0.046)	0.015 (0.052)	0.014 (0.052)
$c_{gt}$	0.040 <sup>‡</sup> (0.029)	0.006 (0.035)	0.009 (0.043)
$\varepsilon_{gt}$	0.005 (0.018)	0.004 (0.018)	0.005 (0.020)
$\tau$	0.025 (0.042)	0.016 (0.038)	0.013 (0.040)
<b>Study 3, <math>\xi_g \in [0.1, 0.4]</math></b>			
$S_{a,t}$	0.030 <sup>†</sup> (0.036)	0.019 (0.041)	0.018 (0.040)
$c_{gt}$	0.054 <sup>‡</sup> (0.022)	0.023 (0.030)	0.007 (0.039)
$\varepsilon_{gt}$	0.009 (0.017)	0.005 (0.019)	0.007 (0.021)
$\tau$	0.018 (0.025)	0.005 (0.025)	0.004 (0.025)

<sup>†</sup> Negatively biased.

<sup>‡</sup> Positively biased.

In the first simulation study ( $\xi_g = 0$  for all sites) model EPMV(0) seems to produce good estimates for all the model parameters ( $S_{a,t}$ ,  $\tau$ ,  $s_g^r$ ,  $\beta_1^r$ , and  $\beta_2^r$ ) and for the derived parameters ( $c_{gt}$  and  $\varepsilon_{gt}$ ). Models EPMV( $\xi$ ) and EPMV( $\xi_g$ ) produce virtually identical estimates to model EPMV(0) albeit with slightly

less precision. The reduction in precision is to be expected on account of the greater number of nuisance parameters requiring estimation. Further, the 95% symmetric credible intervals for  $\xi$  (model EPMV( $\xi$ )) and  $\xi_g$  (model EPMV( $\xi_g$ )) encompass the true value, 0, in all cases.

In the second simulation study ( $\xi_g = 0.2$  for all sites) models EPMV( $\xi$ ) and EPMV( $\xi_g$ ) both seem to produce good estimates for the parameters. However when homogeneity is incorrectly assumed, and model EPMV(0) is applied, survival estimates are negatively biased. Further, bias introduced into the estimates of the regression parameters has resulted in estimates of  $c_{gt}$  that are positively biased, although the estimates of  $\varepsilon_{gt}$  appear relatively unaffected.

When  $\xi_g$  differs between sites ( $\xi_g = \{0.1, 0.2, 0.3, 0.4\}$ ), the third simulation study, model EPMV( $\xi_g$ ) seems to produce accurate estimates for all the parameters. Failure to account for this site dependence, or constraining  $\xi_g = 0$ , introduces bias into some of the parameters. Analogous to the second study, incorrectly imposing the constraint  $\xi_g = 0 \forall g \in [1, G]$  results in negatively biased  $S_{a,t}$  estimates and positively biased  $c_{gt}$  estimates, whereas in comparison the estimates of  $\tau$  and  $\varepsilon_{gt}$  are relatively unaffected. Ignoring the site dependence of  $\xi_g$  introduces bias into the estimates of  $c_{gt}$ . In particular, at sites for which  $\xi_g$  is less than the constant  $\xi$  estimated under model EPMV( $\xi$ ),  $c_{gt}$  is underestimated. Conversely when  $\xi_g > \xi$ ,  $c_{gt}$  is over-estimated. However, the survival estimates, the quantities of ecological interest, seem robust to departures from  $\xi$  site-invariance.

In conclusion, the simulation studies indicate that model EPMV performs well, producing reliable estimates of  $S_{a,t}$  from CES style data in which the presence of missed visits and transient birds may introduce bias if not adequately accounted for. However, incorrectly imposing a restrictive constraint on  $\xi$ , for example setting  $\xi_g = 0$  or  $\xi_g = \xi \forall g \in [1, G]$ , may introduce bias into the parameter estimates. In particular, applying model EPMV(0) when  $\xi > 0$ , consistent with individual recapture heterogeneity being present, has large consequences: estimates of  $S_{a,t}$  are negatively biased and estimates of  $c_{gt}$  are positively biased. In comparison estimates of  $\tau$  and  $\varepsilon_{gt}$  are relatively unaffected, for which we offer an explanation. Model EPMV( $\xi$ ) enables a decrease ( $\xi > 0$ ) or an increase ( $\xi < 0$ ) in  $\varepsilon_{gt}$  relative to  $c_{gt}$ . Note that  $S_{a,t-1}$  and  $c_{gt}$  always occur together in the likelihood for the live-recapture data, whereas  $\varepsilon_{gt}$  appears separately of these terms and of  $\tau$  (see Equations (4.15) and (4.17)). Changes in  $c_{gt}$  can thus be “fixed” by adjusting  $S_{a,t-1}$ . Conversely, any changes in  $\varepsilon_{gt}$  will alter both

the “first recapture” likelihood (see Equation (4.17)) and the extra Binomial likelihood. It is resistance of  $\varepsilon_{gt}$  to change that causes estimates of  $c_{gt}$ , and therefore  $S_{a,t-1}$ , to be biased when model EPMV(0) is applied, although model EPMV( $\xi$ ) or model EPMV( $\xi_g$ ) is more appropriate.

#### iv Sedge Warbler and Reed Warbler CES Live-Recapture Data

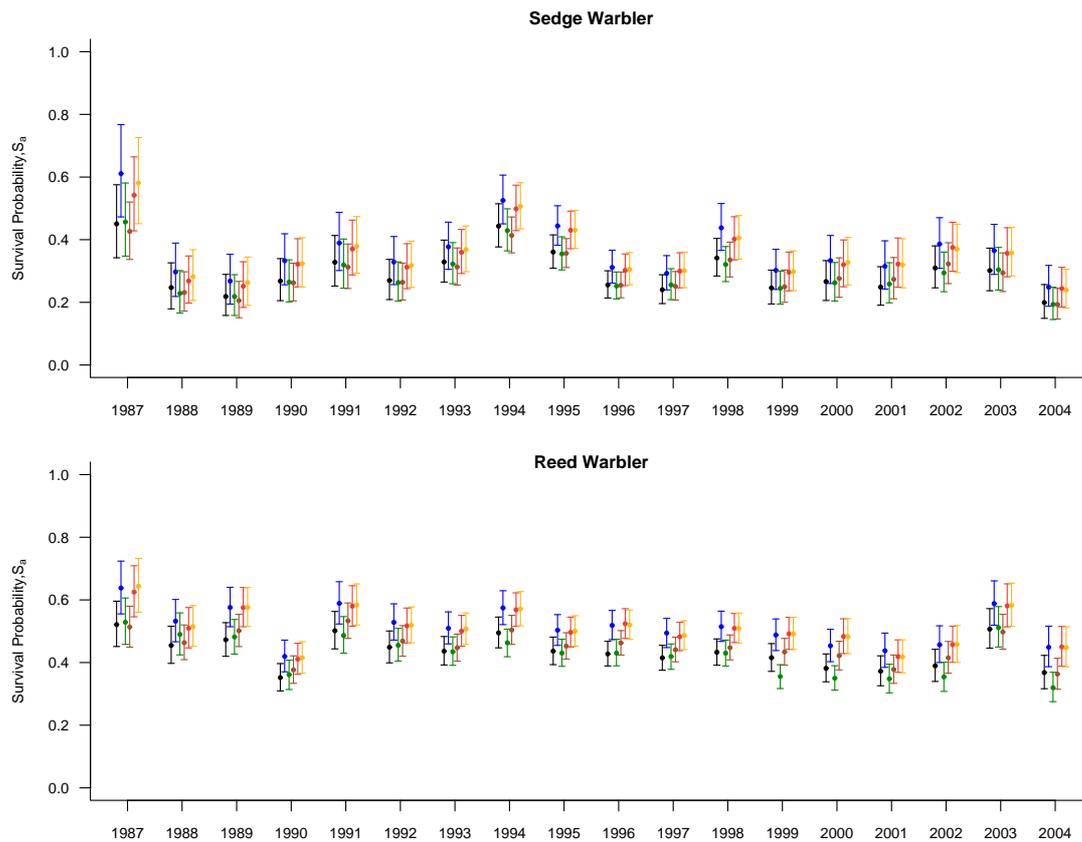
We now present a study of the Sedge Warbler and Reed Warbler CES live-recapture data, comparing the adult survival models that account for, or fail to account for, missed visits and/or transient birds. The analyses described here are based on incomplete coverage data from the subset of the “30 best” sites only, 1987-2005. That is, the 30 sites where the most adult Sedge/Reed Warblers were caught.

The following models are fitted to the Sedge and Reed Warbler CES data:

Model	Missed Visits	Transient Birds	Encounter Effect
CJS	ignored	ignored	none
EP	ignored	modelled	implicit
MV	modelled	ignored	none
EPMV(0)	modelled	modelled	none
EPMV( $\xi$ )	modelled	modelled	modelled (constant)
EPMV( $\xi_g$ )	modelled	modelled	modelled (site dependent)

In Figure 4.19 estimates of adult survival under these models are compared. Although the annual trends in the estimated survival probabilities are similar across all models, and the 95% symmetric credible intervals overlap, systematic changes in the estimates are evident. The posterior means from models EP, EPMV( $\xi$ ), and EPMV( $\xi_g$ ) are consistently higher than those of models CJS, MV, and EPMV(0).

That models CJS and MV produce such similar  $S_{a,t}$  estimates is due to the relatively low percentage of missed visits (5%) causing only minor departures from a constant  $c_g$  over time. Likewise, that model EP gives near identical results to models EPMV( $\xi$ ) and EPMV( $\xi_g$ ) can also be attributed to the small percentage of missed visits in the data. Conversely, the large proportion of transients in the data (approximately 0.39 for Sedge Warblers and 0.42 for Reed Warblers, Table 4.14) increases estimates of  $S_{a,t}$  under models EP, EPMV( $\xi$ ), and EPMV( $\xi_g$ ) compared to models CJS and MV.



**Figure 4.19** Comparing the adult survival estimates under the different models which account for, or fail to account for, missed visits and/or transient birds. Bars denote 95% symmetric credible intervals, and points denote posterior means. From left to right, the models are CJS (black), EP (blue), MV (green), EPMV(0) (brown), EPMV( $\xi$ ) (red), and EPMV( $\xi_g$ ) (yellow). Analyses are based on the Sedge and Reed Warbler CES data from the “30 best” sites, 1987-2005.

For both Sedge and Reed Warbler, model EPMV( $\xi$ ) provides strong evidence that  $\xi$  is greater than zero, an encounter effect consistent with individual recapture heterogeneity. The posterior mean (95% symmetric credible interval) for  $\xi$  from the Sedge and Reed Warbler analyses are 0.918 (0.7409, 1.1023) and 0.781 (0.6968, 0.8783) respectively. These equate to an increase in the odds of recapture at subsequent visits, after the initial visit capture, by approximately 2.5 and 2.2 times respectively.

Figure 4.19 and Table 4.14 indicate that allowing for site dependence in  $\xi$  is not important. Models EPMV( $\xi$ ) and EPMV( $\xi_g$ ) produce very similar results. Further, for both the Sedge Warbler and Reed Warbler analyses 24/30 of the 95% symmetric credible intervals for  $\xi_g$  (under model EPMV( $\xi_g$ )) encompass the posterior mean of the site-invariant value estimated from model EPMV( $\xi$ ). We also note that for all sites, and for both species, the posterior means of  $\xi_g$  are greater than zero.

**Table 4.14** Comparing the posterior means obtained under the different adult survival models, using Sedge and Reed Warbler CES data from the “30 best” sites, 1987-2005. For  $S_{a,t}$ ,  $c_{gt}$  and  $\varepsilon_{gt}$  the average of the posterior means for these sets of parameters are given.

<i>model</i>	CJS	EP	MV	EPMV(0)	EPMV( $\xi$ )	EPMV( $\xi_g$ )
<b>Sedge Warbler</b>						
$S_{a,t}$	0.296	0.365	0.291	0.291	0.348	0.354
$c_{gt}$	0.400	0.483	0.411	0.705	0.525	0.509
$\varepsilon_{gt}$	n/a	0.492	n/a	0.530	0.499	0.497
$\tau$	n/a	0.386	n/a	0.365	0.386	0.387
<b>Reed Warbler</b>						
$S_{a,t}$	0.435	0.515	0.425	0.451	0.510	0.513
$c_{gt}$	0.356	0.445	0.347	0.587	0.452	0.450
$\varepsilon_{gt}$	n/a	0.592	n/a	0.641	0.595	0.595
$\tau$	n/a	0.415	n/a	0.390	0.416	0.416

Due to the failure to allow for the encounter effect,  $S_{a,t}$  estimates under model EPMV(0) differ from those of models EPMV( $\xi$ ) and EPMV( $\xi_g$ ). The under-estimation of  $S_{a,t}$ , and the over-estimation of  $c_{gt}$ , relative to models EPMV( $\xi$ ) and EPMV( $\xi_g$ ) apparent in Table 4.14 are explained by the bias introduced by failing to account for the observed positive encounter effect.

The estimates of  $\beta_1^r$ ,  $\beta_2^r$ ,  $s_g^r$ , and their precision, differ between models MV, EPMV(0), EPMV( $\xi$ ), and EPMV( $\xi_g$ ). The EPMV models, which utilise the additional information on these parameters contained in the within-year recaptures, have improved precision. Furthermore, for any given model the estimation of  $\beta_1^r$ ,  $\beta_2^r$ , and  $s_g^r$  is sensitive to the prior specified. This sensitivity is a consequence of the paucity of information, arising from the limited number of missed visits, regarding these parameters. Despite this prior sensitivity estimates of  $S_{a,t}$ ,  $\tau$ ,  $c_{gt}$ , and (when included)  $\varepsilon_{gt}$  are virtually unaffected by the choice of priors. Such robustness is due to a wide range of regression parameters producing nearly identical estimates of  $c_{gt}$  and  $\varepsilon_{gt}$  given the number and pattern of missed visits present in the data. Conversely, estimates of  $\xi$  were insensitive to their prior. This is to be expected due to the large amount of information in the data with regards to  $c_{gt}$  and  $\varepsilon_{gt}$ , and the discrepancy between them.

In conclusion, for the Sedge and Reed Warbler CES data from the “30 best” sites, adjusting for transients is considerably more important than accounting for the few missed visits. The large proportion of transients in the data causes survival estimates under models CJS and MV to be negatively biased. The failure to incorporate the positive encounter effect results in model EPMV(0)

estimates also being negatively biased. As the number and effect of missed visits is minimal, and as  $\xi$  appears site-invariant, models EP, EPMV( $\xi$ ), and EPMV( $\xi_g$ ) give virtually identical results, believed to be unbiased. Further, the similarity between estimates arising from these models suggests that  $\xi$  does indeed deal with the discrepancy between  $c$  and  $\varepsilon$ . Finally, it is also noted that the precision in the estimation of  $S_{a,t}$  is very similar across all models - the extra site-specific parameters of models EP and EPMV( $\xi_g$ ) cause only a small increase in posterior variation (Figure 4.19).

### *Model Reduction*

We note that the EPMV models differ from their simplified versions, model MV and EP, in a number of important aspects. In particular:

1. Model EP estimates  $c_g$  and  $\varepsilon_g$  independently of one another, whereas under the EPMV models they are constrained to be a function of the underlying within-year recapture probabilities,  $v_{gk}$ .
2. The EPMV models require priors to be specified on the recapture regression parameters, which in turn induce priors on  $c_g$  and  $\varepsilon_g$ . Model EP specifies priors directly on  $c_g$  and  $\varepsilon_g$ .
3. The EPMV models utilise the within-year recaptures to estimate the recapture regression parameters unlike model MV.

Such differences mean that the EPMV models will not strictly reduce to MV, if no transients are present, nor to model EP, when all visits are made; however, if the EPMV models are adequately accounting for transient birds and missed visits the results are expected to be very similar when these conditions apply.

Reassuringly when the restriction  $\tau = 0$  is imposed and the within-year recapture information is removed, the EPMV models produce results almost identical to those obtained from model MV in the analyses of both Sedge and Reed Warbler data. Further, when we assume all missed visits were made but with no birds observed, models EPMV( $\xi$ ) and EPMV( $\xi_g$ ) produce very similar results to model EP. Unsurprisingly, this similarity is increased when  $\xi$  is site dependent - a less restrictive constraint on the association between  $c_g$  and  $\varepsilon_g$ . Conversely, models EPMV(0) and EP give very different results when the missed visits were assumed to be made but with no birds observed. Model EPMV(0) produces estimates of  $S_{a,t}$  that are lower, and estimates of  $c_g$  that are higher, than those of model EP. This discrepancy is attributable to the positive encounter effect. These observations indicate that model EPMV( $\xi$ ), and its extension to allow

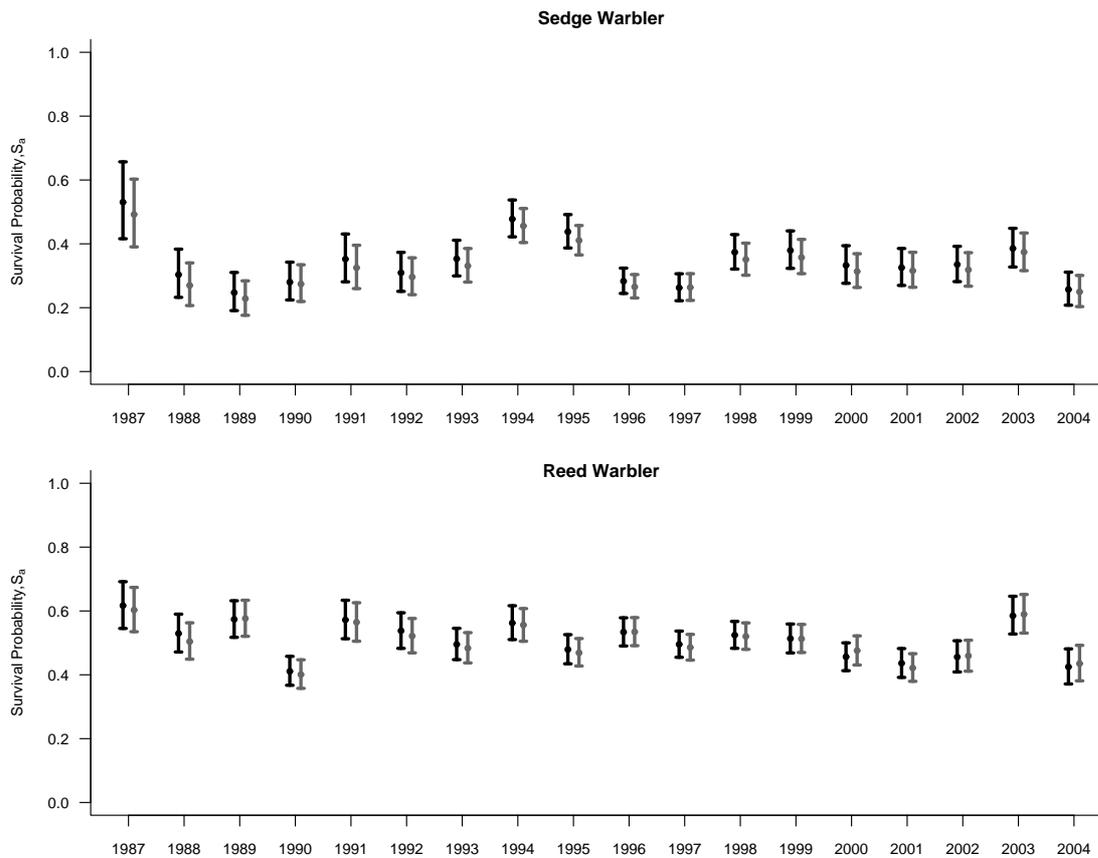
for site dependence in the encounter effect, model  $\text{EPMV}(\xi_g)$ , are adequately accounting for transient birds and/or missed visits.

#### 4.4.5 Analysis of Sedge and Reed Warbler Data

Models  $\text{EPMV}(\xi)$  and  $\text{EPMV}(\xi_g)$  were fitted to the Sedge and Reed Warbler incomplete coverage data from the full set of “best” sites (103 and 82 sites respectively). A prior sensitivity analysis of the recapture regression parameters  $(\beta_1^r, \beta_2^r, s_g^r, \xi/\xi_g)$  revealed that the estimation of  $S_{a,t}$ ,  $\tau$ ,  $c_{gt}$  and  $\varepsilon_{gt}$  were robust despite the prior sensitivity of  $\beta_1^r$ ,  $\beta_2^r$ , and  $s_g^r$ . Note that the estimation of  $\xi$  (and  $\xi_g$ ) was data driven with little prior sensitivity. Further, model  $\text{EPMV}(\xi_g)$  provided no strong evidence of site dependence in  $\xi_g$ , producing virtually identical estimates of  $S_{a,t}$ ,  $\tau$ ,  $c_{gt}$ , and  $\varepsilon_{gt}$  as model  $\text{EPMV}(\xi)$ . We therefore select model  $\text{EPMV}(\xi)$  as an appropriate model for both the Sedge and Reed Warbler data.

In Figure 4.20 estimates of adult survival from model  $\text{EPMV}(\xi)$  are given. These are compared to the estimates under model EP in which the presence of missed visits in the data is ignored. Analogous to the findings above, accounting for missed visits has very little impact on the estimates of survival. This is to be expected since the number of missed visits is small ( $\approx 5\%$ ). Furthermore, as between years visits are missed at random, or essentially so, their effect on the annual recapture rate (at a given site) will be fairly consistent between years.

In addition to the survival ( $S_{a,t}$ ) and transient ( $\tau$ ) probabilities, model EP requires the estimation of 2 “nuisance” parameters per site (an annual recapture probability parameter,  $c_g$ , and an evasion probability parameter,  $\varepsilon_g$ ). As the encounter effect,  $\xi$ , is constrained to be site-invariant, model  $\text{EPMV}(\xi)$  is more parsimonious, requiring the estimation of only 1 “nuisance” parameter per site ( $s_g^r$ ) plus 3 further parameters ( $\beta_1^r$ ,  $\beta_2^r$ ,  $\xi$ ) in the recapture regression, Equation (4.20), from which  $c_g$  and  $\varepsilon_g$  are derived. The fewer parameters of model  $\text{EPMV}(\xi)$  result in improved the precision, albeit marginally, in the estimation of  $S_{a,t}$ . The average width of the 95% symmetric credible interval for  $S_{a,t}$ , under models EP and  $\text{EPMV}(\xi)$ , is 0.121 and 0.112 respectively for Sedge Warbler, and 0.103 and 0.101 for Reed Warbler. Due to this small improvement in precision, and the slight changes in the estimates of  $S_{a,t}$  upon accounting for missed visits, model  $\text{EPMV}(\xi)$  is preferred.



**Figure 4.20** Estimates of adult survival from CES “best” sites data with incomplete coverage. On the left (black) are estimates under model EP in which the presence of missed visits in the data is ignored, and on the right (grey) are estimates under model EPMV( $\xi$ ). Points denote posterior means, bars 95% symmetric credible intervals.

For both species, estimates of  $S_{a,t}$  obtained under model EPMV( $\xi$ ) are reasonably precise, and are in agreement with those obtained from the analysis of complete coverage data (see Figure 4.21, Section 4.5). The posterior mean (95% symmetric credible interval) for  $\tau$  is 0.370 (0.3499, 0.3885) and 0.405 (0.3896, 0.4192) for Sedge and Reed Warblers respectively, again in agreement with those estimated from the complete coverage data, which gave posterior means for  $\tau$  of 0.370 and 0.411 for Sedge and Reed Warblers respectively. The posterior mean (95% symmetric credible interval) for  $\xi$  is 1.011 (0.8667, 1.1638) and 0.878 (0.7912, 0.9672) for Sedge and Reed Warblers respectively. These positive values are consistent with the effect that individual recapture heterogeneity has on  $c_g$  and  $\varepsilon_g$ . Should such individual heterogeneity exist, additional analyses have provided no evidence that it arises due to differences in female and male recapture rates. Further, the independent analyses of the female and the male live-recapture data still produce estimates of  $\xi > 0$ .

### 4.4.6 Concluding Remarks

The CES scheme is designed such that a constant amount of effort is spent catching birds, at a particular site, during all years in which it is operated. However, the presence of missed within-year visits reduces the overall catch effort and violates the assumption of time-invariant recapture probabilities. Furthermore, CES live-recapture data are known to contain records from transient birds, whose presence will negatively bias estimates of survival if not adequately accounted for. We have successfully developed a model for CES data that accounts for both of these nuances to produce unbiased estimates of apparent adult survival,  $S_{a,t}$ .

The BTO have recently begun developing a survival model for CES live-recapture data that takes into account transient birds, the c-dash method of Chapter 2 (Freeman, 2008), but their model is strictly only applicable to data from years with a complete set of 12 visits. Other attempts to estimate survival from CES data have set *a priori* criteria for including data with a minimum number of visits made, for example at least 5 of the first 6 visits, to deal with the problem of missed visits (e.g. Peach et al., 1995a). Our approach (the EPMV models) represents the first formal attempt to deal with both missed visits and transient birds in the modelling of CES live-recapture data.

We have developed a parsimonious approach that circumvents the time consuming need to model data at the visit level, or to allow complete time dependence in the parameters associated with recapture rates ( $c_g$  and  $\varepsilon_g$ ). By fitting a parametric model to the seasonal change in catchability over the mist-netting season, we form expressions for  $c_g$  and  $\varepsilon_g$  adjusted for missed visits. Underpinning our model is the key assumption that the trend in the within-year recapture probabilities ( $v_{gk}$ ) is consistent across years and between sites; an analogous assumption to that employed by Peach et al. (1998) and Robinson et al. (2007) to correct CES yearly-site counts for missed visits in the modelling of abundance and productivity, now routinely used by the BTO.

Our novel model has been successfully applied to both Sedge and Reed Warbler CES incomplete coverage data. These analyses indicate that failure to account for transient birds results in large biases in the estimates of survival. On the other hand, the pattern and number of missed visits typically present seem to result in only small biases. We note that under the CES scheme each site is visited up to 12 times per year, and typically only a few visits (1 or 2) are

missed (Table 4.3). It is therefore not surprising that missed visits have such little impact. Our extension to Pradel et al.'s (1997) transient model (model EP), or the *ad hoc* approach of the BTO, in many cases will be sufficient for providing estimates of survival relatively unbiased by missed visits. However, for potentially a minor cost to precision (for example, if the encounter effect,  $\xi$ , is site dependent) the presence of missed visits is more correctly dealt with by fitting the appropriate EPMV model. Indeed, if a common  $\xi$  across all sites can be assumed, model EPMV( $\xi$ ), being more parsimonious than model EP, has improved precision.

Our analyses also indicate that estimates of  $S_{a,t}$  are unaffected by the sensitivity of the recapture regression parameters,  $\beta_1^r$ ,  $\beta_2^r$ , and  $s_g^r$ , to their priors. When the number (or range) of visits missed is sparse, the data contain insufficient information to estimate these parameters reliably. However, if the data contains rich information on  $c_g$  and  $\varepsilon_g$  (as is typically the case with CES data), a wide range of regression parameters will produce estimates consistent with the observed data, and therefore produce unbiased estimates of  $S_{a,t}$ .

### **i Potential Extensions**

All analyses to date have modelled the trend in within-year recapture probabilities ( $v_{gk}$ ) using a quadratic regression, i.e. Equation (4.20). Although the simple quadratic model accommodates a biologically reasonable trend, a model of another order, for example a third order polynomial, may provide a better fit to the data. Alternatively the more flexible model,  $\text{logit}(v_{gk}) = s_g^r + v_k^r$  in which  $v_k^r$  is the  $k^{\text{th}}$  visit effect, could be more appropriate. A Reversible-Jump MCMC algorithm can be used to select between such competing models (Green, 1995) but due to the paucity of information in the data regarding the regression parameters, such an analysis has not been conducted.

The regression site effects ( $s_g^r$ ), which enable the within-year recapture probabilities to fluctuate in a consistent manner between sites, are expected to depend on site-specific covariates, for example net-length, habitat type and location. Simplifying the current model for  $v_{gk}$ , given in Equation (4.20), by regressing  $s_g^r$  upon such covariates however has not been attempted once again due to the paucity of information regarding the regression parameters.

To allow the discrepancy between  $c_g$  and  $\varepsilon_g$ , observed in Chapter 2 and hypothesised to be consequence of individual heterogeneity in the recapture probabil-

ities, an encounter effect  $\xi$  was incorporated into the model. For both Sedge and Reed Warbler, allowing for site dependence in  $\xi$  did not seem important. However, for other species should there be evidence for site dependence in  $\xi$ , the random effects model  $\xi_g \stackrel{iid}{\sim} \text{Normal}(0, \sigma_\xi^2)$ , where a  $\Gamma^{-1}(a, b)$  prior is given to  $\sigma_\xi^2$ , should be considered as a more parsimonious alternative to model  $\text{EPMV}(\xi_g)$ .

In determining the most appropriate model for CES live-recapture data no formal approach to model selection has been undertaken. For example, model  $\text{EPMV}(\xi)$  was chosen in preference to model  $\text{EPMV}(\xi_g)$  on the basis that both models produced virtually identical estimates of  $S_{a,t}$  and that model  $\text{EPMV}(\xi)$  is the more parsimonious of the two. To formally test whether or not site dependence in  $\xi$  should be incorporated, model selection based on DIC, the Bayesian equivalent to the AIC (Gelman et al., 2004), could be used. Alternatively, Reversible-jump MCMC (Green, 1995) can be used for model selection. Moreover, currently goodness-of-fit has not been properly assessed. Although we offer no formal goodness-of-fit test, a plot of the observed numbers of recaptures against the fitted numbers should provide an indication of any serious lack of fit.

## ii Pledger et al. (2003) Mixture Model

We have hypothesised that the encounter effect is due to individual heterogeneity in the recapture rates, with some birds being inherently more catchable than others. Pledger et al. (2003) have provided a framework for allowing heterogeneity in recapture probabilities in the analysis of live-recapture data. Their approach assumes that there are  $C$  classes, each of which comprises animals with relatively homogeneous recapture probabilities. Importantly, these classes need not actually exist, they are essentially an artifact to allow for wider heterogeneity. Each marked animal has probability  $\omega_c$  of belonging to class  $c$  (class membership is unknown). The probability of an individual animal  $i$ 's capture history ( $\text{CH}_i$ ) conditional on first capture ( $f_i$ ) is therefore:

$$\Pr(\text{CH}_i|f_i) = \sum_{c=1}^C \Pr(\text{CH}_i|f_i \text{ and class}=c) \times \Pr(\text{class} = c)$$

By taking this individual based approach to forming the likelihood, individual heterogeneity is allowed for.

Fitting such a model to Sedge and Reed Warbler data is complicated by the

presence of missed visits and transients. In particular, transients will look like a class with a very low recapture probability. As an exploratory analysis, data until the first recapture (within or between years) are removed (the remaining data known to be from resident birds), and we work with only complete coverage data. Vague, independent Uniform[0,1] priors are given to the survival, class-specific recapture and class membership probabilities. For simplicity, models for individual sites are fitted independently, and only models with 2 classes are considered. In all cases there was no evidence that recapture probabilities differed between the 2 classes (the 95% symmetric credible intervals overlapped).

That exploratory analyses with the Pledger et al. (2003) mixture model failed to detect any heterogeneity in the recapture probabilities for either Sedge or Reed Warblers is explained. The Pledger et al. (2003) model assumes that animals retain the same group membership over successive years; this is unlikely to be the case for CES data. Biologically individual heterogeneity (in CES data) is likely to result as a consequence of birds having nests and territories situated at variable proximity to the mist-net. For example, birds with nests closer to the mist-net are expected to have a higher  $c$  (and a lower  $\varepsilon$ ) than those birds with nests further away. However, there is no guarantee that a bird with a high  $c$  one year will have a similarly high  $c$  in the following year. Although adult birds are highly faithful to their breeding sites they do not occupy exactly the same position within the site during consecutive years. We would expect CES data to contain 2 levels of heterogeneity in the recapture probabilities - within a year recapture rates are expected to vary between birds, and across years recapture rates of an individual bird are expected to vary. Fitting the Pledger et al. (2003) model to CES data, in an attempt to deal with capture heterogeneity, is therefore inadequate for this data.

## 4.5 Comparing the Analyses which Exclude or Include Incomplete Coverage Data

Developing models for CES data that enable years with incomplete coverage to be included in the analysis is motivated by a potential improvement in the precision of estimation. Including such years in the “best” sites data set nearly doubles the number of yearly-site counts contributing to the analysis, and substantially increases the total number of observed captures (in the full CES data) now represented in the reduced “best” sites data set (Table 4.15).

It is also noted that the recruitment and retention of CES volunteers is aided if the maximal amount of data, within reason, can be analysed and contribute to feedback in the form of reports etc.

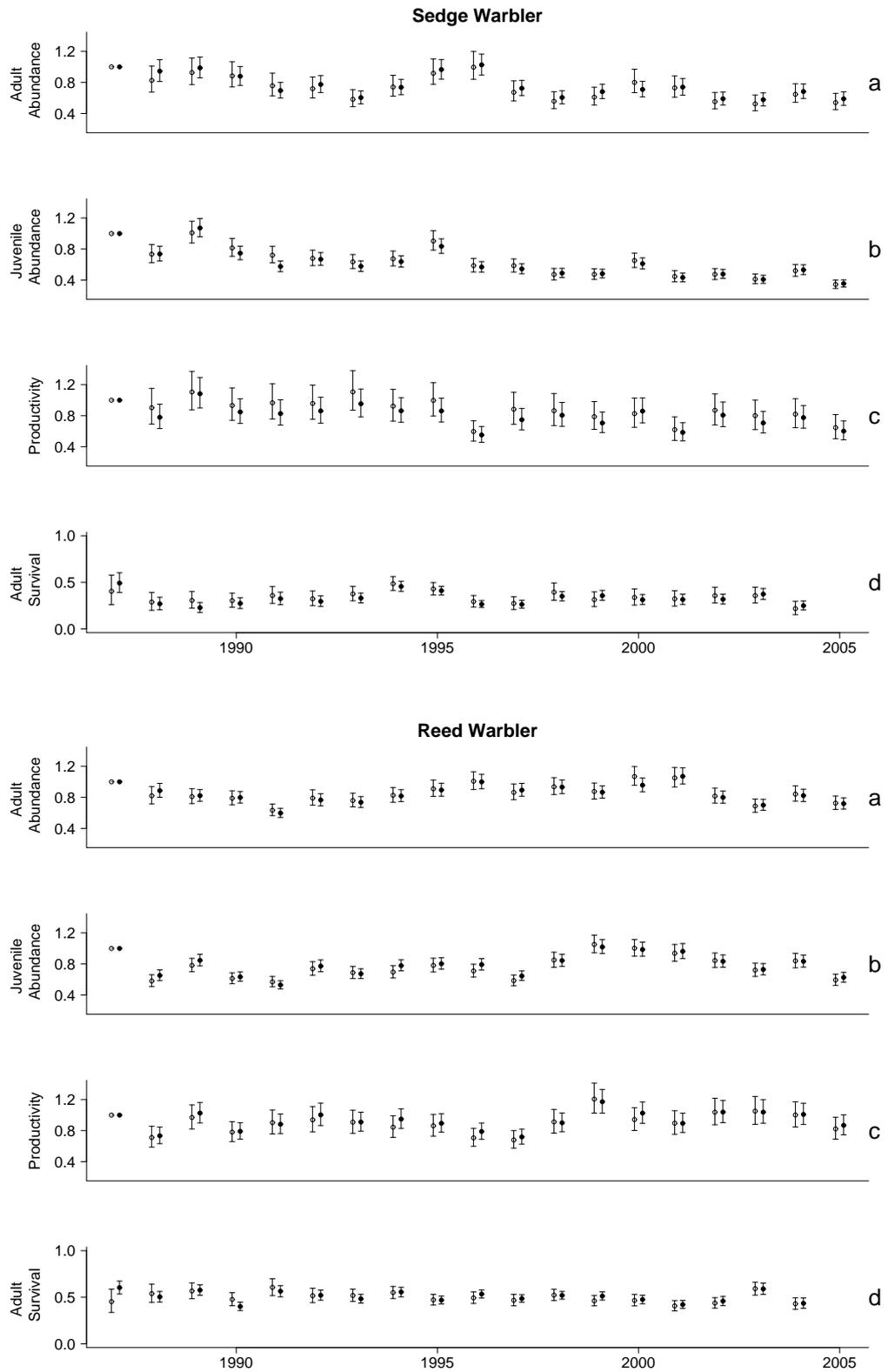
Extension to include data with incomplete coverage in the modelling of abundance, productivity, and adult survival requires the estimation of (potentially) additional parameters. Thus there is a trade-off between the additional parameters to be estimated and the increase in the amount of data. Such a consideration led to the development of a parsimonious model for adult survival.

In Figure 4.21 estimated indices of adult abundance, juvenile abundance and productivity, and estimates of adult survival, from the subset of “best” sites data which excluded (analyses of Chapter 2), or included, the additional years with incomplete coverage are compared. In all cases the appropriate (independent) Bayesian model for complete or incomplete data is applied. For both Sedge and Reed Warbler, Figure 4.21 demonstrates that the trade-off is worthwhile.

Despite the need to impute over 250 censored counts in the modelling of abundance and productivity, the ability to include, and appropriately model, years with incomplete coverage is advantageous - in the analyses of both Sedge and Reed Warbler CES data including these additional years has improved the precision in the estimation of  $A_t$ ,  $J_t$ , and  $P_t$ . Estimates of  $S_{a,t}$  also have improved precision, due to both the increase in information used, and the application of a more parsimonious model (as the encounter effect is assumed site-invariant). Furthermore, the estimated  $A_t$ ,  $J_t$ ,  $P_t$  and  $S_{a,t}$  trends are similar regardless of whether years with incomplete coverage are included in the analysis. This is to be expected: as incomplete coverage occurs essentially at random, omitting data from such years will not bias the results.

**Table 4.15** Comparing Sedge and Reed Warbler “best” sites data sets which either exclude or include years with incomplete coverage: adult and juvenile captures (excluding multiple within-year encounters), sites, and yearly-site records included in the reduced CES data sets of “best” sites expressed as a percentage of the full CES data set.

	Complete Coverage	Incomplete Coverage
<b>Sedge Warbler</b>		
Adult captures	46%	76%
Juvenile captures	51%	81%
Sites	25%	36%
Yearly-site records	28%	52%
Incomplete yearly-site records	0%	29%
<b>Reed Warbler</b>		
Adult captures	54%	80%
Juvenile captures	54%	90%
Sites	23%	34%
Yearly-site records	27%	49%
Incomplete yearly-site records	0%	28%



**Figure 4.21** Estimates of the demographic parameters from the Bayesian models which exclude (left) or include (right) years with incomplete coverage into the analysis of the “best” sites Sedge and Reed Warbler data (1987-2005). Points denote posterior means, bars 95% symmetric credible intervals. a:  $A_t$ , b:  $J_t$ , c:  $P_t$ , d:  $S_{a,t}$ .

## 4.6 Chapter Summary

In this chapter we have extended the models for abundance, productivity and adult survival, presented in Chapter 2, to enable CES data from years with incomplete coverage to contribute to their separate analysis, improving the precision in parameter estimation.

We have proposed a Bayesian alternative to the classical models for abundance (Peach et al., 1998) and productivity (Miles et al., 2007; Robinson et al., 2007) routinely employed by the BTO, that has the key advantage of more accurately reporting the uncertainty associated with the parameter estimates. Simulation studies have indicated that the alternative Bayesian approach performs well.

Our model for adult survival represents the first formal attempt to deal parsimoniously with the effect missing visits have on the recapture rates, as well as accounting for transient birds in the data. The method is readily adapted to incorporate the “encounter effect”, hypothesised to arise as the result of individual heterogeneity in the capture rates. Simulation studies suggest our model provides estimates of apparent adult survival unbiased by missed visits or transient birds.

Our novel models have been successfully applied to Sedge Warbler and Reed Warbler CES data, in which missed visits have resulted in non-constant sampling effort, and estimates of the demographic parameters (abundance, productivity and adult survival) obtained. At present the abundance, productivity and adult survival analyses for CES data with incomplete coverage have been conducted independently. By sharing information between these independent models, that is undertaking an integrated approach to population modelling as in Chapter 3, a further improvement in precision is likely. We develop such an integrated model for incomplete coverage data in the following chapter, combining the methodology presented in Chapter 3 with that of the present chapter.

## Chapter 5

# Extensions to the Integrated Population Model

An integrated population model, appropriate for CES data with complete coverage in all years, and in which adult birds exhibit extreme site-fidelity, was developed and successfully fitted to Sedge Warbler and Reed Warbler data in Chapter 3. In this chapter we explore two major modifications to the current integrated model, namely:

- 1) the inclusion of CES data from years with incomplete coverage,
- 2) allowing for a lack of site-faithfulness in adult birds.

In Chapter 4 it was demonstrated how the individual component baseline models for adult abundance, juvenile abundance and adult survival, that adopted CES data, could be extended to include those years with incomplete coverage. Despite these models requiring the estimation of additional “nuisance” parameters, the inclusion of such data proved beneficial - an improvement in the precision of estimation was realised.

In the first part of this chapter, Section 5.1, we extend the integrated population model of Chapter 3 to incorporate CES data from years with incomplete coverage using the models presented in Chapter 4. We fit this extended integrated model to Sedge and Reed Warbler data. By comparing the results to those from the integrated model for complete coverage data only, which necessitates omitting a large fraction of the CES data, we demonstrate the improved precision achieved by enabling data from years with missed visits to contribute to the analysis.

In the second part of this chapter, Section 5.2, we address the potential for permanent emigration from the CES sites. Despite adult Sedge and Reed Warblers exhibiting strong breeding philopatry (Wernham et al., 2002) some permanent emigration from the site of ringing, but within the domain of the UK population, is expected, the extent of which is typically unknown. It might be, as we initially assumed, insignificant having virtually no impact on the estimates of “true” adult survival ( $\phi_{a,t}$ ) from CES live-recapture data, or it could be important resulting in negative bias in  $\phi_{a,t}$ . We demonstrate how the integrated model, for either complete or incomplete coverage data, can be modified to allow for permanent emigration thus removing the bias, and apply this modified model to both Sedge and Reed Warbler data. Further, we show how a Reversible-jump MCMC algorithm can be used to assess whether there is significant permanent emigration of adult birds from their CES site of ringing.

Within this chapter we use CES data with complete coverage or incomplete coverage, as appropriate, from the “best” sites as defined in Sections 2.1 and 4.1 respectively. Further, note that throughout this chapter we assume the population monitored is closed with respect to the UK, with all emigration and transient movements between breeding sites occurring within the UK, but not necessarily between sites sampled within the study.

## 5.1 An Integrated Model for CES Data that Include Years with Incomplete Coverage

In Chapter 2 models for CES data with complete coverage were presented. Recall that the likelihoods for the observed adult and juvenile CES count data,  $L_a$  and  $L_j$ , respectively, provide information on the indices of adult and juvenile abundance,  $A_t$  and  $J_t$ , and via Equation (2.4) information on the productivity indices,  $P_t$  (Sections 2.2 and 2.3). Under the assumption of total site-fidelity of adult birds, estimates of “true” survival rates for adults,  $\phi_{a,t}$ , are obtainable from the model for adult CES live-recapture data, for which the likelihood is denoted by  $L_{live}$  (Section 2.4). The likelihood for the National Ring-Recovery (NRR) data,  $L_{dead}$ , presented in Chapter 3, provides estimates of “true” survival rates for juvenile birds,  $\phi_{j,t}$ , in addition to further information on  $\phi_{a,t}$  (Section 3.2.2). Furthermore, recall that, assuming independence of the individual likelihoods from which the demographic parameters ( $A_t$ ,  $J_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ )

are estimated, the joint likelihood is simply the product of these component likelihoods (Equation (3.7)), i.e.

$$L_{joint} = L_a \times L_j \times L_{dead} \times L_{live}.$$

In Chapter 3 an integrated model was formulated by creating a meaningful functional relationship between the demographic parameters of  $L_{joint}$ . This was achieved using a deterministic population model, given by Equation (3.1), that expressed the adult abundance in year  $t$  as a function of the adult and juvenile abundances in year  $t - 1$  and the intermediate survival rates. That is:

$$\begin{aligned} A_t &= A_{t-1}\phi_{a,t-1} + kJ_{t-1}\phi_{j,t-1} \\ &= A_{t-1}\phi_{a,t-1} + kP_{t-1}A_{t-1}\phi_{j,t-1}, \end{aligned} \quad (\text{Equation (3.1)})$$

or recursively

$$= g(A_1, k, \mathbf{J}_{1..t-1}, \phi_{a,1..t-1}, \phi_{j,1..t-1}). \quad (\text{Equation (3.2)})$$

The component likelihoods of  $L_{joint}$  were therefore integrated by substituting Equation (3.1) into the likelihood for adult counts,  $L_a$ .

Extending this integrated model to incorporate CES data from years of incomplete coverage is straightforward. Here the likelihoods based on CES data,  $L_a$ ,  $L_j$  and  $L_{live}$ , are simply derived from the appropriate model of Chapter 4 which accounts for missed visits. In particular, the likelihoods for the adult and juvenile CES counts,  $L_a$  and  $L_j$ , respectively, are derived from the Bayesian model of Section 4.2.2 which accounts for the presence of censored counts in the observed data. To modify the recapture rates ( $c_g$  and  $\varepsilon_g$ ) for non-constant effort as a result of missed visits, the likelihood for the CES live-recapture data,  $L_{live}$ , is derived from the extended Pradel et al. (1997) transient model adjusted for missed visits with a site-invariant encounter effect as appropriate, i.e. model EPMV( $\xi$ ) of Section 4.4.4. Due to the high degree of site-fidelity exhibited by adult Sedge and Reed Warblers (Catchpole, 1974; Peach, 2002; Wernham et al., 2002), we initially assume that all losses at a site can be attributed to mortality, and set  $\phi_{a,t} = S_{a,t}$  in  $L_{live}$ . Note that  $L_{dead}$  is unchanged as it is formulated from ring-recovery data. Integration is once again achieved by substituting the population equation for  $A_t$ , given by Equation (3.1), into  $L_a$ . That is:

$$\begin{aligned} L_a &= f(\mathbf{n}^{a,obs}, \boldsymbol{\ell}^{a,obs}, \boldsymbol{\eta}^a | \mathbf{A}_{1..T}, \beta^a, \mathbf{s}_{1..G}^a) \\ &= f(\mathbf{n}^{a,obs}, \boldsymbol{\ell}^{a,obs}, \boldsymbol{\eta}^a | A_1, k, \mathbf{J}_{1..T-1}, \phi_{a,1..T-1}, \phi_{j,1..T-1}, \beta^a, \mathbf{s}_{1..G}^a) \end{aligned} \quad \text{via Equation(3.1),}$$

$$L_j = f(\mathbf{n}^{j,obs}, \ell^{j,obs}, \boldsymbol{\eta}^j | \mathbf{J}_{1...T}, \beta^j, \mathbf{s}_{1...G}^j),$$

$$L_{live} = f(\mathbf{F}, \mathbf{X} | \tau, \beta_1^r, \beta_2^r, \mathbf{s}_{1...G}^r, \xi, \boldsymbol{\phi}_{a,1...T-1}) \times f(\mathbf{R}, \mathbf{Z} | \beta_1^r, \beta_2^r, \mathbf{s}_{1...G}^r, \boldsymbol{\phi}_{a,1...T-1}) \\ \times f(\mathbf{M}' | \beta_1^r, \beta_2^r, \mathbf{s}_{1...G}^r, \xi, \mathbf{M}' + \mathbf{M}''),$$

$$L_{dead} = f(\mathbf{D}^j | \boldsymbol{\phi}_{j,1...T}, \boldsymbol{\phi}_{a,2...T}) \times f(\mathbf{D}^a | \boldsymbol{\phi}_{a,1...T}).$$

Note that the parameter constraints imposed on the integrated model are based on those determined from the independent analyses of the component data sets. For example, the encounter effect,  $\xi$ , is constrained to be site invariant as determined by the independent analysis of the CES live-recapture data (Section 4.4.5). Although not considered here, model selection with integrated data can be achieved in a Bayesian framework using DIC (Gelman et al., 2004) or Reversible-jump MCMC (Green, 1995) for example.

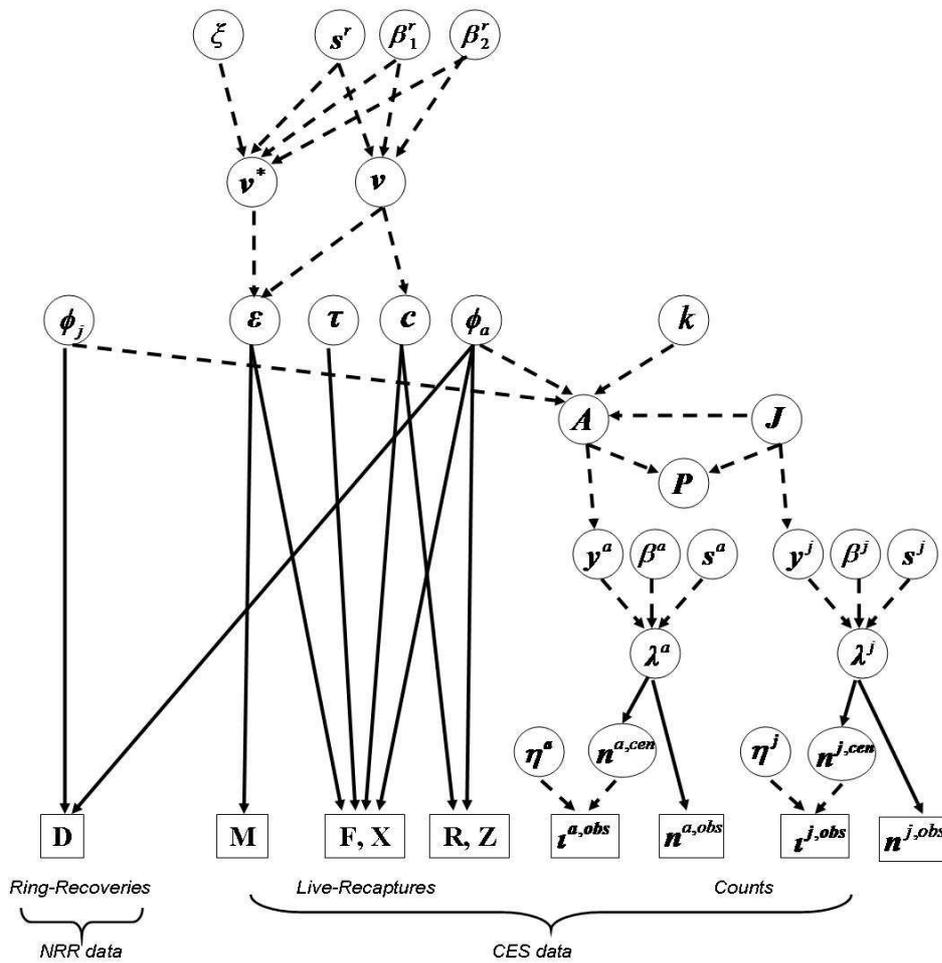
### 5.1.1 Bayesian Model Fitting

The combination of the underlying abundance and survival models in the fully integrated analysis is summarised by the Directed Acyclic Graph given in Figure 5.1. Using Bayesian methodology this integrated model is readily fitted.

Vague, independent, priors are specified for all parameters. In particular, for the real-valued parameters in the abundance models, and the log-transformed productivity constant  $k$ , we specify a Normal prior with mean 0 and variance 10,000. For the survival and transience probabilities we assume a Uniform[0,1] prior. For the underlying recapture regression parameters, the site ( $s_g^r$ ) and encounter ( $\xi$ ) effects are given logistic priors, whereas the visit coefficients ( $\beta_1^r$  and  $\beta_2^r$ ) are given Uniform[-10,10] priors. We use a random walk, single update, Metropolis-Hastings algorithm, with a Normal proposal for the real-valued parameters, a Uniform proposal, appropriately truncated, for the probability parameters, and a discrete Uniform proposal, appropriately truncated, for the imputed censored yearly-site counts ( $n_{gt}^{cen}$ ). Proposal variances and step lengths are tuned *a priori*.

The simulations are run for 200,000 iterations using bespoke code written in C. The first 100,000 iterations are discarded as burn-in. For both the Sedge Warbler and Reed Warbler analyses, essentially identical posterior estimates are obtained from independent replications with different overdispersed starting

points, so that we assume the chain has converged. Furthermore, examination of the trace-plots indicated good movement and mixing through the parameter space, and that the chain had converged before the end of the burn-in period. Additional simulations with a range of different sets of independent priors for the recapture regression parameters ( $s_g^r$ ,  $\beta_1^r$ ,  $\beta_2^r$ ,  $\xi$ ) indicated that the posterior distributions for the induced recapture probabilities ( $c_{gt}$ ,  $\varepsilon_{gt}$ ) are data-driven with little prior sensitivity, and the adult survival probabilities ( $\phi_{a,t}$ ) are robust.



**Figure 5.1** Directed Acyclic Graph corresponding to the fully integrated model for CES data that include years with incomplete coverage data. Known values are represented by squares and unknown values by circles. Continuous arrows denote stochastic dependencies and dashed arrows deterministic dependencies. Note that live-recaptures and adult counts typically arise from the same data. In such circumstances they are not independent.

### 5.1.2 “Baseline” Models

Independent estimates of the key demographic parameters ( $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ ), that are not restricted by any demographic assumptions imposed by the deterministic population model, Equation (3.1), serve as a “baseline” for similar estimates derived under the integrated model. Baseline estimates for the trends in adult abundance,  $A_t$ , and juvenile abundance,  $J_t$ , are obtained via the independent fitting of the Bayesian model to the adult and juvenile CES incomplete coverage count data, as in Chapter 4 (Section 4.2.2). From these independent analyses, baseline estimates for the productivity indices,  $P_t$ , are readily derived by calculating the ratio of  $J_t$  to  $A_t$  using samples drawn from their posterior distributions. Baseline estimates of true adult survival,  $\phi_{a,t}$ , are obtainable from CES live-recapture data alone, under the assumption of total site-faithfulness, via model EPMV( $\xi$ ), the extended Pradel et al. (1997) transient model adjusted for missed visits, as in Chapter 4 (Section 4.4.4). An alternative baseline for  $\phi_{a,t}$  is also estimable from the independent analysis of the ring-recovery data using the conditional model (Chapter 3, Section 3.2.2), however as there are considerably more live-recaptures than ring-recoveries the former baseline, being more precise, is preferred. Baseline estimates of  $\phi_{j,t}$  are obtained from the independent analysis of the ring-recovery data.

### 5.1.3 Analysis of Sedge and Reed Warbler Data

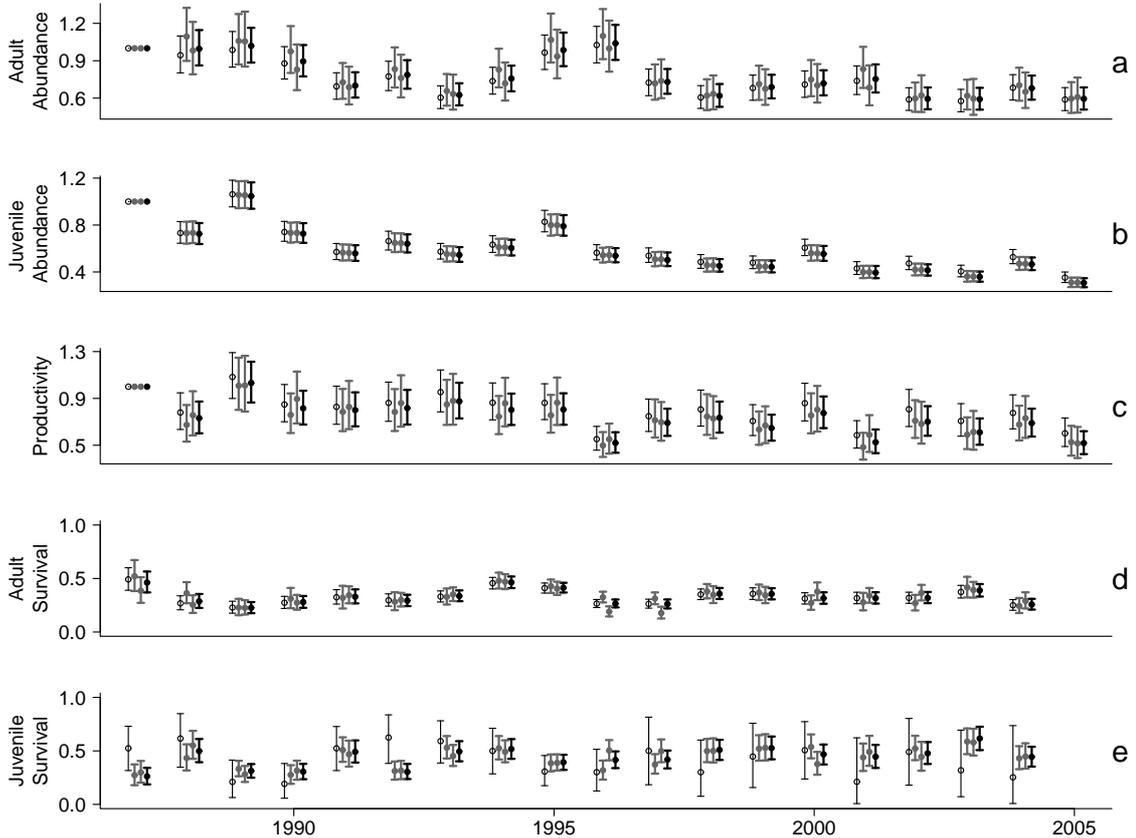
A fully integrated analysis of the “best” sites CES data that include years with incomplete coverage, as defined in Chapter 4 (Section 4.1), and the limited ring-recovery data, is carried out for both Sedge and Reed Warbler. As the live-recapture likelihood,  $L_{live}$ , provides a far greater amount of information on  $\phi_{a,t}$  than the ring-recovery likelihood,  $L_{dead}$ , it is incorporated into the integrated analysis so that:

$$L_{joint} = L_a \times L_j \times L_{dead} \times L_{live}. \quad (\text{Recall Equation (3.7)})$$

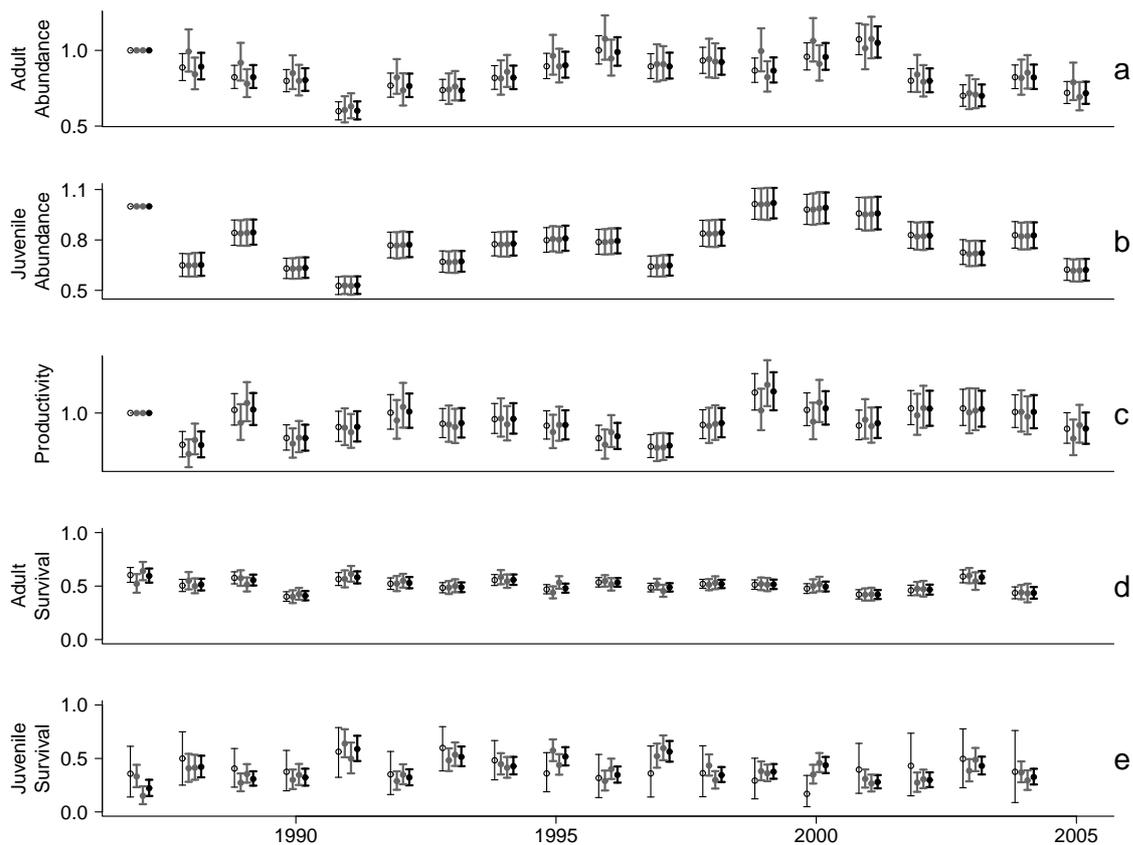
To deal with the lack of independence between  $L_{live}$  and  $L_a$ , which are drawn from the same data, split data analyses are conducted. Here the Sedge Warbler CES data from all 103 sites are split into two groups (one of 52 sites and the other of 51 sites); one is used to derive  $L_{live}$  and the other independent set to form  $L_a$ . Likewise, the 82 sites contributing to the Reed Warbler CES data are split into two groups (both of 41 sites). In each case, to ensure the split data sets provide good coverage over the duration of the study (which covered

few sites in the early years), sites are initially stratified according to when they first joined the CES scheme, and within each of these strata half the sites are randomly assigned to each group. To investigate the sensitivity of the split, the split data analyses are repeated with the two data sets used to form  $L_a$ , or  $L_{live}$ , switched. Secondly, we ignore the issue of non-independence and derive  $L_a$  and  $L_{live}$  using data from all sites.

For Sedge and Reed Warblers, posterior means of the key demographic parameters ( $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ ) from the full and split data integrated analyses are compared to each other, and their baseline estimates in Figures 5.2 and 5.3 respectively. Note as there are considerably more live-recaptures than ring-recoveries, baseline estimates of  $\phi_{a,t}$  from  $L_{live}$  are used.



**Figure 5.2** Estimates of the Sedge Warbler demographic parameters from CES data that include years with incomplete coverage. Posterior means, and the 95% symmetric credible intervals, from the “baseline” models, denoted by thin left-hand lines, and the fully integrated model. The integrated analysis in which all 103 sites provide information to  $L_{live}$  and  $L_a$ , is represented by bold right-hand lines. The split data analysis, and its reversal, are represented by the respective grey centre lines. a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ .



**Figure 5.3** Estimates of the Reed Warbler demographic parameters from CES data that include years with incomplete coverage. Posterior means, and the 95% symmetric credible intervals, from the “baseline” models, denoted by thin left-hand lines, and the fully integrated model. The integrated analysis in which all 82 sites provide information to  $L_{live}$  and  $L_a$ , is represented by bold right-hand lines. The split data analysis, and its reversal, are represented by the respective grey centre lines. a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ .

From Figures 5.2 and 5.3 it is clear that the important conclusions are consistent between the two species, and with the integrated analysis of the complete coverage CES data only (see Chapter 3). In particular:

1. An integrated approach leads to improved precision, particularly for  $\phi_{j,t}$ , due to the extra information arising from the demographic population model, given by Equation (3.1).
2. As direct information on  $\phi_{j,t}$  is so limited, the underlying demographic population model drives its estimation.
3. The split and full data integrated analyses produce reasonably consistent results with only minor discrepancies (e.g.  $A_{1999}$  in Figure 5.3a).
4. The issue of non-independence between  $L_a$  and  $L_{live}$  can be overcome by randomly splitting the data at only a small cost to precision.
5. The split data integrated analysis seems relatively insensitive to the split.

For Sedge and Reed Warbler, the split data integrated analyses were also repeated using multiple other random splits. As was noted in Chapter 3 (see Figure 3.8) the necessary changes in the reference site used affects the precision of the intercept term,  $\beta^a$ , and therefore the posterior variance of  $A_t$  and  $P_t$ . For example, if the posterior variance of  $\beta^a$  is increased then the observed precision in  $A_t$  and  $P_t$  decreases.

The proportion of transients ( $\tau$ ) and the encounter effect ( $\xi$ ) also require estimation under the fully integrated model. For both species, the posterior estimates under the full and two split data integrated analyses are very similar to their baseline values, and to one another (Table 5.1). Due to the greater amount of data being employed to derive  $L_{live}$  their estimates are more precise under the full data integrated analysis than either of the split data integrated analyses.

**Table 5.1** Posterior mean (standard deviation) of  $\tau$  and  $\xi$  from the “baseline” model (model EPMV( $\xi$ )), the full data integrated analysis, and the two split data integrated analyses, for Sedge and Reed Warblers using CES data that include years with incomplete coverage.

<b>Sedge Warbler</b>				
	Baseline	Integrated	Split	Split Reversed
$\tau$	0.370 (0.0100)	0.371 (0.0100)	0.350 (0.0148)	0.391 (0.0130)
$\xi$	1.011 (0.0758)	1.030 (0.0791)	1.096 (0.1102)	1.021 (0.1019)
<b>Reed Warbler</b>				
	Baseline	Integrated	Split	Split Reversed
$\tau$	0.405 (0.0077)	0.406 (0.0077)	0.415 (0.0105)	0.398 (0.0110)
$\xi$	0.878 (0.0448)	0.885 (0.0444)	0.961 (0.0664)	0.844 (0.0646)

The imputed censored counts ( $n_{gt}^{cen}$ ), the derived recapture probabilities ( $c_{gt}$ ) and the derived evasion probabilities ( $\varepsilon_{gt}$ ) from the integrated analyses (split or full) are also very similar to their baseline estimates, as indicated by the high correlation between their posterior medians ( $\rho \approx 1.0$  in all cases).

From published analyses of Nest Record Scheme (NRS) data, we obtain respective estimates of true seasonal productivity ( $P_t^s$ ) for Sedge and Reed Warblers of approximately 3.0 and 2.8 juveniles per breeding pair per year (see Chapter 3). These are consistent with the estimates given by the integrated model. For Sedge Warblers the integrated model, using the full 103-site data, estimates, on average, 3.0 juveniles per pair per year. The two split data integrated analyses estimate, on average, 3.0 and 3.1 juveniles per pair per year respectively. For Reed Warblers the integrated model, using the full 82-site data, estimates, on

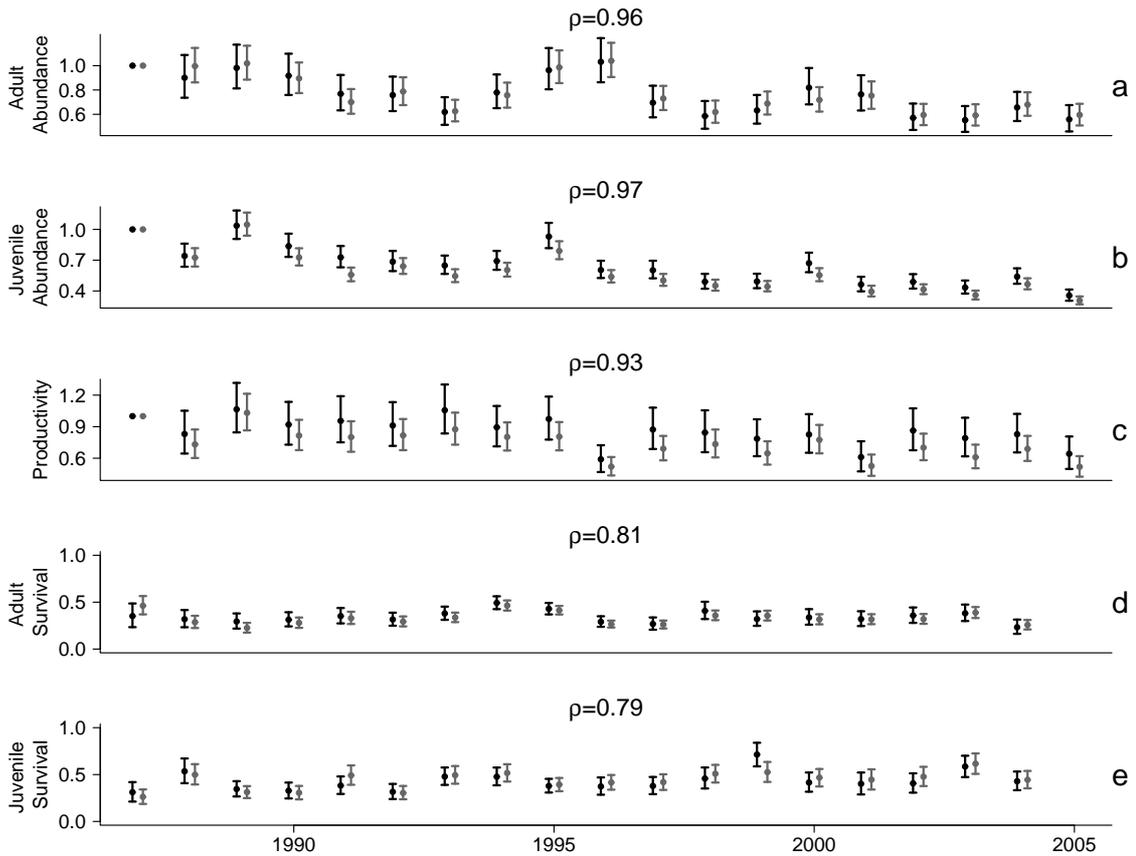
average, 2.5 juveniles per pair per year. The two split data integrated analyses also give an estimate of 2.5 juveniles per pair per year.

Conducting a split data integrated analysis can lead to reduced precision in the estimates of the demographic parameters compared to their baseline (Figures 5.2, 5.3, Table 5.1). However, the integrated approach still has the advantage of providing estimates of the demographic parameters that are consistent with both the CES and NRR data sets, and according to the relationship imposed by the population model, with one another. Moreover, the integrated approach enables total seasonal productivity to be estimated, rather than an index merely proportional to it, and it is only means of obtaining usefully precise estimates of juvenile survival.

#### 5.1.4 Advantage of Including Years with Incomplete Coverage into the Integrated Analysis

In Figure 5.4 estimates of the key demographic parameters ( $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ ), for Sedge Warbler, from the fully integrated model which excluded years with incomplete coverage (analysis of Chapter 3) are compared to those from the fully integrated model of the current chapter which enables data from these years to be included. The estimated trends in the demographic parameters are very similar between both integrated analyses, as indicated by the high correlations in their posterior means ( $0.79 \leq \rho \leq 0.97$ ). Further, the 95% symmetric credible intervals of the estimates under both analyses overlap, even those of the precisely estimated  $\phi_{a,t}$  values.

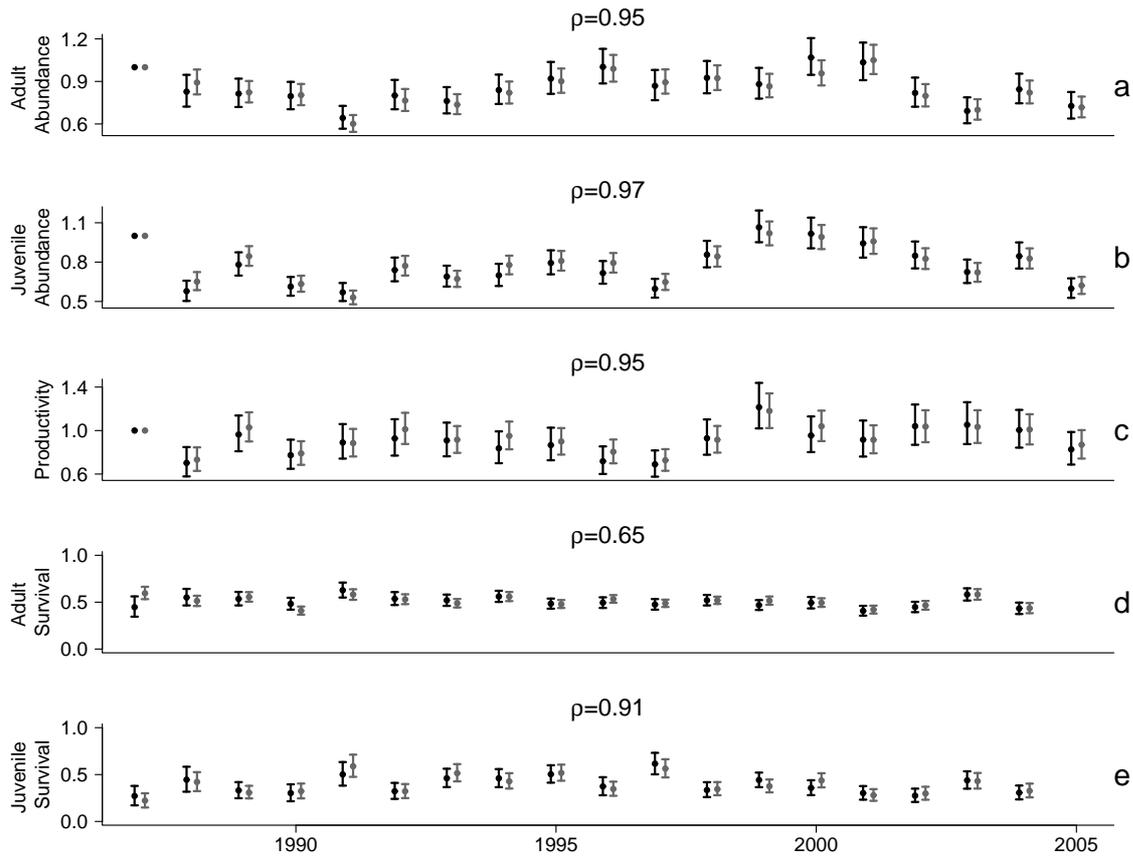
We note however some minor differences when additional data from years with incomplete coverage are included into the integrated analysis of Sedge Warbler data. For example, in 1991 the estimate (posterior mean) of  $J_t$  has decreased, thus to make the derived adult abundance index in 1992 match the adult count data, the estimate of  $\phi_{j,t}$  in 1991 has been increased (Figure 5.4a, b, e). Likewise, a reduction in the estimate of  $A_t$  in 2000 has been accommodated by a decrease in the estimate of  $\phi_{j,t}$  in 1999 (Figure 5.4a, e). Further, estimates of juvenile abundance, and consequently productivity, have become slightly lower relative to the reference year (Figure 5.4b, c).



**Figure 5.4** Posterior means, and the 95% symmetric credible intervals, from the fully integrated model of Sedge Warbler data which either exclude (left-hand, black lines), or include (right-hand, grey lines) CES data with incomplete coverage for a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ . For each parameter the correlation ( $\rho$ ) between the posterior means is given. The integrated analysis of the complete and incomplete coverage data are based on the full set of 71 and 103 sites respectively.

The similarity in the estimates of the demographic parameters between the integrated analyses which either exclude or include years of complete coverage is also evident in the plot for Reed Warbler, Figure 5.5. We observed a slightly lower correlation for  $\phi_{a,t}$  due to small discrepancies in the 1987 and 1990 posterior means, but we note that the two analyses produce estimates of these parameters with overlapping 95% symmetric credible intervals (Figure 5.5d).

That the integrated analyses which exclude, or include, years with incomplete coverage give analogous results is to be expected. Incomplete coverage occurs at random, or essentially so, due to bad weather, net damage, ringer unavailability etc, thus omitting data without complete coverage will not cause a systematic bias in the results.



**Figure 5.5** Posterior means, and the 95% symmetric credible intervals, from the fully integrated model of Reed Warbler data which either exclude (left-hand, black lines), or include (right-hand, grey lines) CES data with incomplete coverage for a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $\phi_{a,t}$ , e)  $\phi_{j,t}$ . For each parameter the correlation ( $\rho$ ) between the posterior means is given. The integrated analysis of the complete and incomplete coverage data are based on the full set of 55 and 82 sites respectively.

The motivation behind including additional data from years with incomplete coverage into the integrated model was to improve precision, in particular the precision in the estimates of the key demographic parameters,  $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ , and  $\phi_{j,t}$ . However, the inclusion of such data necessitates the estimation of extra nuisance parameters, resulting in a trade-off between the precision gained from additional data and the precision lost from fitting a less parsimonious model in which there is uncertainty in the extra parameters being estimated. When years of incomplete coverage are included into the integrated analysis, the number of sites contributing data increases from 71 to 103 for Sedge Warbler, and 55 to 71 for Reed Warbler. Each additional site comes at the cost of extra site-specific parameters:  $s_g^a$  in  $L_a$ ,  $s_g^j$  in  $L_j$ ,  $s_g^r$  in  $L_{live}$ . Furthermore, the adult and juvenile abundance models also require the imputation of corrected values for the censored counts (317 and 253 additional “parameters” in the analyses

of Sedge and Reed Warbler data respectively). In the adult survival model, EPMV( $\xi$ ), the  $G$  site-specific recapture parameters ( $c_g$ ) and the  $G$  site-specific evasion parameters ( $\varepsilon_g$ ) are replaced by expressions involving  $G + 3$  parameters ( $s_g^r, \beta_1^r, \beta_2^r, \xi$ ), see Chapter 4, Section 4.4.4. Note, in this case there is actually a reduction in the number of parameters required per site as the encounter effect,  $\xi$ , which determines the relationship between  $c_{gt}$  and  $\varepsilon_{gt}$ , can be assumed site-invariant. This may not always be the case. In total, the number of parameters requiring estimation in the fully integrated model increases from 339 to 896 for Sedge Warbler, and 275 to 726 for Reed Warbler on the inclusion of years with incomplete coverage, i.e. a 164% increase for both species. In contrast the data used to derive  $L_a$  and  $L_j$ , the yearly-site counts, increases by approximately 80%, and the data used to derive  $L_{live}$ , the annual adult recaptures, increases by approximately 60% (see Tables 2.1, 4.1).

From Figures 5.4 and 5.5 it is clear that the integrated analyses of Sedge Warbler and Reed Warbler data benefit from including years with incomplete coverage despite the additional parameters - the 95% symmetric credible intervals for  $A_t, J_t, P_t, \phi_{a,t}, \phi_{j,t}$  are narrower. However we note that the time taken to implement the MCMC algorithm, on a standard desktop computer, is considerably longer, with computational time approximately doubling.

### 5.1.5 Investigating the Relationship between the Demographic Parameters

Since songbirds are assumed to become adult (assumed breeding) at around 1 year of age, to investigate the importance of changes in the demographic parameters on adult abundance, the correlation between the posterior means of  $\frac{A_{t+1}}{A_t}$  and  $P_t, \phi_{a,t}$  or  $\phi_{j,t}$  are calculated (see Table 5.2). As is to be expected the correlations are positive. As with the integrated analysis of the complete coverage data only (see Section 3.5), changes in  $A_t$  seem more highly correlated with survival, and in particular adult survival, than productivity, once again suggesting that adult survival could be the demographic parameter that is driving the size of the Sedge and Reed Warbler breeding populations.

As was observed when only CES data from years with complete coverage were used in the integrated analysis, the baseline estimates of  $\phi_{j,t}$  are less correlated with  $\frac{A_{t+1}}{A_t}$  than the estimates under the integrated model. Due to the paucity of data in the baseline for  $\phi_{j,t}$  it is this parameter that tends to be adjusted upon

integration to make the population model match the observed adult count data, and hence it is possible that the integrated model has contrived this stronger correlation.

Correlations between the posterior means of the underlying demographic parameters are given in Table 5.3. The pattern is similar to that observed from the analysis of complete coverage data only (see Table 3.3), which is to be expected since these two analyses produced such similar posterior estimates. In particular, for both Sedge and Reed Warbler,  $\phi_{j,t}$  appears to be positively correlated with  $\phi_{a,t}$  and negatively correlated with  $P_t$  and  $J_t$ . Furthermore, for Sedge Warbler, but now also for Reed Warbler,  $P_t$  and  $\phi_{a,t}$  are positively correlated. Such correlations have biologically reasonable explanations - they suggest that similar environmental factors act on the survival rates of both adult and juvenile birds, that density-dependence/competition affects juvenile survival, and that environmental conditions detrimental to adult fitness lowers both their survival rates and reproductive output.

**Table 5.2** Correlations ( $\rho$ ) between changes in adult abundance ( $\frac{A_{t+1}}{A_t}$ ) and the demographic parameters when data with incomplete coverage are included in the integrated analysis. Correlations are calculated using the posterior mean of i) the baseline estimates (for  $\phi_{a,t}$  via  $L_{live}$ ) and ii) the estimates from the fully integrated model (based on the full set of sites).

	Sedge Warbler		Reed Warbler	
	Baseline	Integrated	Baseline	Integrated
$P_t$	0.29	0.25	0.22	0.23
$\phi_{a,t}$	0.57	0.68	0.49	0.59
$\phi_{j,t}$	0.35	0.61	0.25	0.77

**Table 5.3** Correlations ( $\rho$ ) between the demographic parameters when data with incomplete coverage are included in the integrated analysis. Correlations are calculated using the posterior mean of i) the baseline estimates (for  $\phi_{a,t}$  via  $L_{live}$ ) and ii) the estimates from the fully integrated model (based on the full set of sites).

	Sedge Warbler		Reed Warbler	
	Baseline	Integrated	Baseline	Integrated
$P_t, \phi_{a,t}$	0.19	0.17	0.25	0.21
$P_t, \phi_{j,t}$	0.23	-0.55	-0.22	-0.35
$J_t, \phi_{j,t}$	-0.01	-0.74	-0.66	-0.60
$\phi_{a,t}, \phi_{j,t}$	0.23	0.21	0.22	0.13

## 5.2 Allowing for Permanent Emigration

On account of the high degree of breeding philopatry exhibited by adult birds, the possibility of permanent emigration was not accounted for in the original formulation of the fully integrated model. We assumed total site-faithfulness of adult birds, and interpreted the adult survival probability estimated from CES live-recapture data to be true survival (denoted by  $\phi_a$ ) rather than apparent survival (denoted by  $S_a$ ). This assumption seems to be reasonable for Sedge and Reed Warbler, given the similarity in scale of the survival estimates from CES live-recapture data and ring-recovery data, the latter being unaffected by any lack of site-fidelity (Peach et al., 1991). However, in reality departures from this assumption are likely with some adult birds, albeit a small proportion, not returning to their CES site of ringing in all subsequent breeding seasons (Pratt and Peach, 1991; Peach, 1993). If the rate of permanent emigration is sufficiently high, estimates of true adult survival directly obtained from CES live-recapture data will be negatively biased, as birds who permanently emigrate are indistinguishable from those that have died. We also note that even if site-fidelity can be justly assumed for Sedge and Reed Warbler, for other species monitored under the CES scheme the validity of this assumption may be more questionable, for example Willow Warbler *Phylloscopus trochilus* (Peach, 1993), so model development along these lines is of value.

In this section we develop an alternative integrated model that allows for permanent emigration away from the CES site of ringing, but within the domain of the UK population. Further, we demonstrate how the Reversible-jump MCMC framework (Green, 1995) can be used to assess whether permanent emigration should be explicitly accounted for within the integrated model.

### 5.2.1 An Integrated Population Model for CES Data in the Presence of Permanent Emigration

We assume that there is no net immigration or emigration of adult (or juvenile) birds from the UK population (c.f. Besbeas et al., 2002, 2003; Freeman and Crick, 2003; Brooks et al., 2004), but allow for permanent movements between breeding sites within the UK. Let  $1 - \gamma$  denote the probability that an adult resident bird permanently emigrates away from its CES site of ringing in any given year. In the presence of permanent emigration, true survival ( $\phi_{a,t}$ ) cannot

be estimated from the CES live-recapture data alone, but information from CES data can be used to estimate  $\phi_{a,t}$  in the integrated context by setting

$$S_{a,t} = \gamma\phi_{a,t} \quad (5.1)$$

in the likelihood for the live-recapture data. That is, when data from years with incomplete coverage are excluded or included, we express  $L_{live}$  as

$$L_{live} = f(\mathbf{F}, \mathbf{X} | \tau, \boldsymbol{\varepsilon}_{1...G}, \mathbf{c}_{1...G}, \gamma, \boldsymbol{\phi}_{a,1...T-1}) \times f(\mathbf{R}, \mathbf{Z} | \mathbf{c}_{1...G}, \gamma, \boldsymbol{\phi}_{a,1...T-1}) \\ \times f(\mathbf{M}' | \boldsymbol{\varepsilon}_{1...G}, \mathbf{M}' + \mathbf{M}''),$$

or

$$L_{live} = f(\mathbf{F}, \mathbf{X} | \tau, \beta_1^r, \beta_2^r, \mathbf{s}_{1...G}^r, \xi, \gamma, \boldsymbol{\phi}_{a,1...T-1}) \times f(\mathbf{R}, \mathbf{Z} | \beta_1^r, \beta_2^r, \mathbf{s}_{1...G}^r, \gamma, \boldsymbol{\phi}_{a,1...T-1}) \\ f(\mathbf{M}' | \beta_1^r, \beta_2^r, \mathbf{s}_{1...G}^r, \xi, \mathbf{M}' + \mathbf{M}''),$$

respectively. (5.2)

Note that  $\gamma \in [0, 1]$ , where a value of 1 signifies total site-fidelity (as previously assumed) and a value of 0 corresponds to a somewhat hypothetical population in which individuals breed at different sites each year.

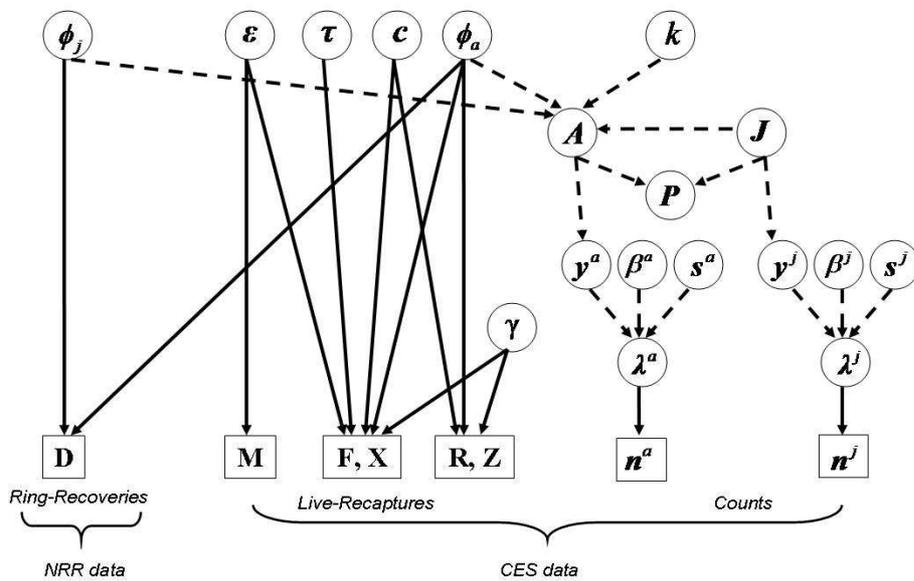
The Directed Acyclic Graph, that summarises the combination of the underlying abundance and survival models in the fully integrated analysis, is modified only slightly by Equation (5.1). Here, stochastic dependencies between the unknown  $\gamma$  parameter and the known  $\mathbf{F}$ ,  $\mathbf{X}$ ,  $\mathbf{R}$ , and  $\mathbf{Z}$  live-recapture data statistics are added. For example, the Directed Acyclic Graph for the fully integrated model that adopts only complete coverage CES data is now given by Figure 5.6. Likewise, an analogous modification to Figure 5.1, the Directed Acyclic Graph for the fully integrated model that also utilises data from years with incomplete coverage, is made.

In the independent analysis of the CES live-recapture data  $\gamma$  and  $\phi_{a,t}$  are confounded. Upon integration these two components are separately identifiable since  $L_{dead}$ , and the deterministic population model given by Equation (3.1), provide information regarding  $\gamma$ .

Using Bayesian methodology, and bespoke code written in C, parameter estimates from the fully integrated model, modified to account for permanent emigration, are readily obtained. For complete coverage data only, the method proceeds as described in Chapter 3, Section 3.2.4. When years of incomplete coverage are included, the method proceeds as in Section 5.1.1 of the current

chapter. In both situations we assume a Uniform[0,1] prior for  $\gamma$ , and use random walk Metropolis-Hastings updates, with a Uniform proposal appropriately truncated and step length tuned *a priori*. The MCMC algorithm is run for 200,000 iterations, with the first half discarded as burn-in. For all analyses, using different starting points had essentially no effect on the posterior estimates.

Baseline estimates, unrestricted by the demographic assumptions imposed by the deterministic population model (Equation (3.1)), are obtained via the independent fitting of the component abundance, productivity and survival models for complete or incomplete coverage data as appropriate (see Sections 3.2.5 and 5.1.2, respectively).



**Figure 5.6** Directed Acyclic Graph, for complete coverage CES data, corresponding to the fully integrated model that allows for permanent emigration. Known values are represented by squares and unknown values by circles. Continuous arrows denote stochastic dependencies and dashed arrows deterministic dependencies. Note that live-recaptures and adult counts typically arise from the same data. In such circumstances they are not independent.

### 5.2.2 Analysis of Sedge and Reed Warbler Data

We fit the fully integrated model, allowing for permanent emigration away from the CES sites, to the Sedge Warbler, and to the Reed Warbler, “best” sites data sets that either exclude (Section 2.1) or include (Section 4.1) years with incomplete coverage. In all integrated analyses both  $L_{live}$  and  $L_{dead}$  are incorporated so that:

$$L_{joint} = L_a \times L_j \times L_{dead} \times L_{live}. \quad (\text{Recall Equation (3.7)})$$

We begin by focusing on the results from the integrated analysis for Sedge Warbler, in which incomplete coverage CES data are adopted. To deal with the issue of non-independence between  $L_{live}$  and  $L_{dead}$ , split data analyses are conducted. To aid comparison, we present results from the same split presented in Section 5.1.3 (Figure 5.2).

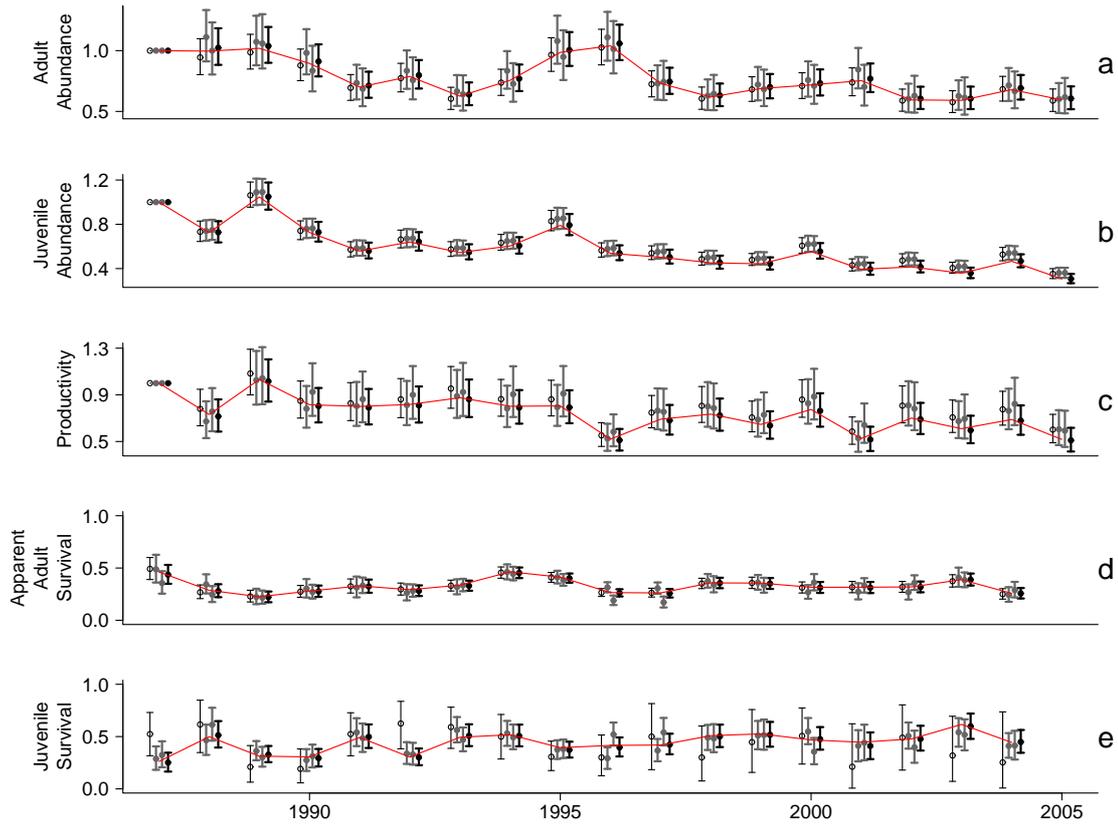
#### *Sedge Warbler Integrated Analysis Adopting Incomplete Coverage Data*

The posterior means of the key demographic parameters,  $A_t$ ,  $J_t$ ,  $P_t$ ,  $S_{a,t}$ , and  $\phi_{j,t}$ , from the full and split data integrated analyses are compared to each other, and their baseline estimates, in Figure 5.7. There are considerably more data on adult survival from the live-recaptures than the limited ring-recoveries, but as the live-recapture data alone cannot provide a baseline for  $\phi_{a,t}$ , we present estimates of  $S_{a,t}$ .

From Figure 5.7 analogous conclusions are drawn to the original integrated model that assumes total site-fidelity of adult birds (refer to Section 5.1.3). Namely i) the integrated approach can lead to a substantial improvement in precision, ii) parameters for which direct information is limited, for example  $\phi_{j,t}$ , the underlying demographic model, Equation (3.1), will drive their estimation, and iii) the full data and split data integrated analyses produce consistent results, with the latter being relatively insensitive to the split randomly chosen.

The posterior mean (95% symmetric credible interval) of  $\gamma$ , the probability of not permanently emigrating, from the integrated analysis of the full 103-site data is 0.76 (0.641, 0.907), and the two split data integrated analyses is 0.80 (0.684, 0.942) and 0.73 (0.612, 0.867) respectively, indicating a high, but importantly not a total, degree of site-faithfulness. Although the upper limit of the 95% symmetric credible interval for  $\gamma$  is close to the maximum allowable value, 1, which corresponds to total site-faithfulness, values indicating reasonably large departures from total site-fidelity are also plausible given the width of the 95% symmetric credible interval.

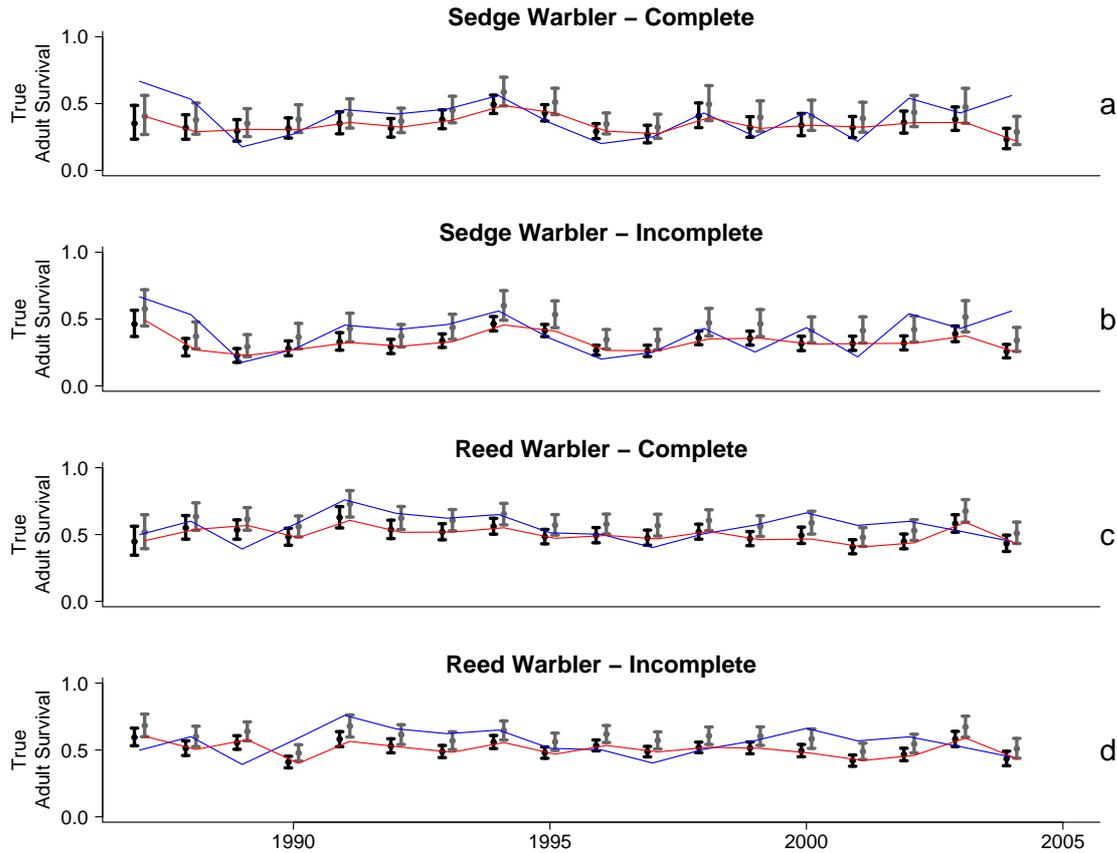
To gauge the impact of failing to accommodate this degree of site-infidelity on the estimates of the key demographic parameters, we also plot (in red) on Figure 5.7 the posterior means estimated by the original integrated model (as in Section 5.1.3). Whether or not an allowance for permanent emigration is incorporated has virtually no effect on the estimated indices,  $A_t$ ,  $J_t$ , and  $P_t$  (Figure 5.7a,



**Figure 5.7** Estimates of the Sedge Warbler demographic parameters from the integrated model that allows for permanent emigration. a)  $A_t$ , b)  $J_t$ , c)  $P_t$ , d)  $S_{a,t}$ , e)  $\phi_{j,t}$ . Posterior means, and the 95% symmetric credible intervals, from the “baseline” models, denoted by thin left-hand lines, and the fully integrated model. The integrated analysis in which all 103 sites provide information to  $L_{live}$  and  $L_a$ , is represented by bold right-hand lines. The split data analysis, and its reversal, are represented by the respective grey centre lines. The red line denotes the posterior means from the original integrated model that assumes total site-fidelity of adult birds. Under this model  $S_{a,t} = \phi_{a,t}$ .

b, c). The posterior means are almost identical under both models, with the correlation between them being 1.0. Juvenile survival,  $\phi_{j,t}$ , is likewise unaffected (Figure 5.7e).

Estimates of apparent adult survival,  $S_{a,t}$ , in the integrated model which allows for permanent emigration are virtually identical to those under the original integrated model which does not, but that sets  $S_{a,t} = \phi_{a,t}$  (Figure 5.7d). Estimates of adult survival under the original model are not therefore a compromise between  $S_{a,t}$  (from live-recapture data) and  $\phi_{a,t}$  (from ring-recovery data), but rather reflect  $S_{a,t}$ . As the ring-recovery data are sparse, this is to be expected. Also as expected, estimates of true adult survival,  $\phi_{a,t}$ , are increased when an allowance for permanent emigration is incorporated into the integrated model (Figure 5.8a). Note,  $L_{live}$  is also a function of  $\beta_1^r$ ,  $\beta_2^r$ ,  $s_g^r$ ,  $\xi$  and  $\tau$ , but these parameters appear unaffected by the adjustment for permanent emigration.



**Figure 5.8** Posterior means, and the 95% symmetric credible intervals, of true adult survival ( $\phi_{a,t}$ ) from the integrated analysis that assumes total site-fidelity of adult birds (left-hand black lines) or allows for permanent emigration (right-hand grey lines). The blue line denotes the baseline estimates (posterior means) of  $\phi_{a,t}$  from the ring-recovery data via  $L_{dead}$ . The red line denotes the baseline estimates (posterior means) from the CES live-recapture data via  $L_{live}$ . Note, the full 103-site data are used to derive both  $L_{live}$  and  $L_a$  in the integrated analyses.

Importantly, we note that estimates of the productivity constant  $k$  change whether or not permanent emigration is allowed for in the formulation of the integrated model. This means that the true seasonal productivity,  $P_t^s$ , is affected (recall Equation 2.4) even if the relative index,  $P_t$ , is not. When estimates of adult survival from  $L_{live}$  are adjusted for permanent emigration, the posterior mean of  $k$  decreases by 13%, from 2.05 to 1.79, with corresponding 95% symmetric credible intervals of (1.677, 2.411) and (1.356, 2.330) respectively.

Such results are to be expected. The CES count data provide a rich source of direct information on  $A_t$ ,  $J_t$ , and consequently  $P_t$ , thus these parameters are unbiased by the presence of permanent emigration, and its effect on the estimates of  $\phi_{a,t}$  via  $L_{live}$ . By contrast, since there is little direct information on  $\phi_{a,t}$  in the ring-recovery data, its estimation is dominated by the much larger

amount of information on  $S_{a,t}$  in the live-recapture data. Assuming that permanent emigration occurs, when it is not accounted for, upon integration the model essentially estimates  $\phi_{a,t}$  with the lower estimate of  $S_{a,t}$  obtained from  $L_{live}$ , thus producing estimates of true adult survival that are negatively biased. In turn  $k$ , which is freely estimated, is scaled up to deal with the discrepancy between true and apparent adult survival rates in the population model (Equation (3.1)). It is therefore not surprisingly that, when the integrated model accounts for permanent emigration, a positive posterior correlation between  $k$  and  $\gamma$  is observed ( $\rho = 0.55$ ).

If substantially more ring-recovery data were available than live-recapture data, so that the information in  $L_{dead}$  overwhelms that in  $L_{live}$ , estimates of  $\phi_{a,t}$  from the integrated model that assumes total site-fidelity would no longer be subject to such extreme negative bias if this assumption were violated. Nonetheless, some negative bias in  $\phi_{a,t}$  would be expected due to the influence of the live-recapture data on the integrated model, albeit limited under such circumstances.

Note that when  $L_{live}$  is not included into the integrated model, i.e.

$$L_{joint} = L_a \times L_j \times L_{dead},$$

so that permanent emigration is not an issue, estimates of  $\phi_{a,t}$  and  $k$  (average  $\phi_{a,t} = 0.43$ ,  $k = 1.70$ ) are closer to those of the fully integrated model that adjusts  $L_{live}$  for permanent emigration (average  $\phi_{a,t} = 0.43$ ,  $k = 1.79$ ) than the integrated model which does not (average  $\phi_{a,t} = 0.33$ ,  $k = 2.05$ ). There is much uncertainty in these estimates, especially when  $L_{live}$  is omitted from the integrated analysis, and the 95% symmetric credible intervals do overlap, however the direction of the movements in  $\phi_{a,t}$  and  $k$  are consistent with the effect of permanent emigration.

#### *Sedge and Reed Warbler Integrated Analyses Adopting either Complete Coverage or Incomplete Coverage Data*

For both Sedge and Reed Warbler, regardless of whether years of incomplete coverage are included in the analysis, the posterior estimates of  $\gamma$ , the probability of not permanently emigrating, indicate a high, but not a total, degree of site-faithfulness (Table 5.4). Note that, for both species, the upper limit of the 95% symmetric credible interval for  $\gamma$  is very close to the maximum permissible value, 1, which specifies total site-faithfulness, but that the range of  $\gamma$  encompassed by the 95% symmetric credible interval includes values corresponding to

large violations in the assumption of total site-fidelity. Furthermore, note that  $\gamma$  is positively correlated with the productivity constant,  $k$  (Table 5.4). This is to be expected as  $k$ , which is freely estimated, mops up the discrepancy between true and apparent survival. That different data are clearly used in the complete and incomplete coverage analyses explains some of the variation between the estimates of  $\gamma$  for Sedge Warbler.

**Table 5.4** Posterior mean and 95% symmetric credible interval of  $\gamma$ , the probability of not permanently emigrating, from the integrated analysis of Sedge and Reed Warbler complete coverage, and incomplete coverage, “best” sites data sets. Also given is the posterior correlation between  $\gamma$  and the productivity constant,  $k$ . Note, these results are from the integrated analysis that uses the full data to derive both  $L_{live}$  and  $L_a$ .

	Posterior mean	95% credible interval	Correlation with $k$
<b>Sedge Warbler</b>			
Complete	0.82	(0.696, 0.956)	0.37
Incomplete	0.76	(0.641, 0.907)	0.55
<b>Reed Warbler</b>			
Complete	0.84	(0.771, 0.914)	0.50
Incomplete	0.85	(0.787, 0.926)	0.57

The effect of adjusting for permanent emigration on the estimates of the demographic parameters ( $A_t$ ,  $J_t$ ,  $P_t$ ,  $S_{a,t}$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ ) is summarised in Table 5.5. Analogous results to those above are obtained from the integrated analyses of the Sedge Warbler complete coverage data, and both the complete and incomplete coverage Reed Warbler data sets. In particular, when the presence of permanent emigration is not accounted for (assuming that it occurs):

1. The estimates of  $A_t$ ,  $J_t$ ,  $P_t$  and  $\phi_{j,t}$  appear unbiased (see Table 5.5), as are the  $c_g$ ,  $\varepsilon_g$ ,  $\tau$ , and in the case of incomplete coverage data,  $\beta_1^r$ ,  $\beta_2^r$ ,  $s_g^r$ , and  $\xi$  parameters of  $L_{live}$ .
2. As there are by far more live-recaptures than ring-recoveries, estimates of  $\phi_{a,t}$  are essentially given by  $S_{a,t}$ , and thus are negatively biased (see Figure 5.8).
3. To counteract the reduced estimates of  $\phi_{a,t}$ , the productivity constant  $k$ , and consequently the estimates of true seasonal productivity, are positively biased.

Given that permanent emigration causes negative bias in  $\phi_{a,t}$ , the population model (Equation (3.1)) will allow fewer adults to survive year  $t$ , and contribute to the abundance in year  $t + 1$ , than do in reality. So that the derived adult

abundance indices match the wealth of adult CES count data, we might expect positive bias to be induced in the parameters relating to the number of juveniles produced in year  $t$  that survive to adulthood. As the CES juvenile count data provide precise inference on  $J_t$ , it is expected that the productivity constant  $k$ , which is freely estimated, and/or  $\phi_{j,t}$ , for which direct information is limited, to be increased. That  $\phi_{j,t}$  seems relatively unaffected by the adjustment for permanent emigration, aside from the occasional year for which there have been some large adjustments, is somewhat surprising (Table 5.5). That this is the case highlights the importance of ring-recovery data. Ring-recovery data not only separate  $k$  and  $\phi_{j,t}$ , but also mitigate against unrealistic values of  $\phi_{j,t}$  being estimated.

**Table 5.5** The average percentage change in the posterior means of the demographic parameters when  $L_{live}$  is adjusted for permanent emigration in the fully integrated population model. Bracketed are the minimum and maximum percentage changes. All calculations are based on the integrated analysis that uses the full data to derive both  $L_{live}$  and  $L_a$ .

	Sedge Warbler		Reed Warbler	
	Complete	Incomplete	Complete	Incomplete
$A_t$	1.2 (0.8, 1.8)	-1.9 (-2.7, -1.3)	1.1 (0.7, 1.6)	-0.8 (-1.1, -0.2)
$J_t$	-1.7 (-2.0, -1.6)	-3.0 (-0.6, 0.0)	-0.2 (-0.5, -0.1)	-1.3 (-2.8, 0.1)
$P_t$	-3.0 (-3.6, -2.4)	1.6 (1.0, 2.3)	-1.2 (-1.8, -0.7)	-0.5 (-2.1, 1.2)
$S_{a,t}$	1.6 (-1.5, 5.8)	1.1 (-0.7, 5.3)	1.6 (-0.6, 3.5)	0.8 (-0.5, 2.2)
$\phi_{a,t}$	-20.1 (-23.9, -15.0)	-30.1 (-32.4, -24.6)	-17.0 (-19.5, -14.7)	-16.5 (-18.0, -14.8)
$\phi_{j,t}$	-0.7 (-4.4, 5.6)	1.1 (-4.6, 8.0)	-1.3 (-7.9, 4.3)	0.9 (-5.6, 10.8)

Focusing once more on Figure 5.8, which compares the estimates of  $\phi_{a,t}$  under the integrated models which do, or do not, allow for permanent emigration, we observe a slight decrease in precision when permanent emigration is allowed. This is to be expected - now  $\phi_{a,t}$  is the product of  $\gamma$  and  $S_{a,t}$ , Equation (5.1), both of which have associated uncertainty. The increased values of  $\phi_{a,t}$ , when permanent emigration is allowed, presented in Figure 5.8 are consistent with the scale of estimates predicted from ring-recovery data. For Sedge Warbler and

Reed Warbler respectively, baseline estimates of  $\phi_{a,t}$  via  $L_{dead}$  are on average 0.40 and 0.56 respectively. When permanent emigration is included in the integrated model, the average  $\phi_{a,t}$  is increased from about 0.33 to 0.41 for Sedge Warblers, and 0.51 to 0.59 for Reed Warblers.

### 5.2.3 Assessing Adult Site-Fidelity

The original formulation of the fully integrated model presented in Chapter 3, and its modification in this current section, constitute two competing models for the same data. Originally  $\gamma$  was constrained to 1, a condition which corresponds to total site-fidelity of adult birds. Subsequently this assumption has been relaxed such that  $\gamma \in [0, 1]$ , allowing for permanent emigration.

The analyses of this section have revealed that the nonrestrictive assumption,  $\gamma \in [0, 1]$ , can have large consequences for the resulting estimates of  $\phi_{a,t}$  and  $k$  - estimates of  $\phi_{a,t}$  increase and  $k$  decrease. It is therefore important to formally assess which of the two alternative models is more appropriate. This can be achieved using posterior model probabilities, which provide a measure of the degree of posterior uncertainty regarding each model. The Reversible-jump MCMC methodology (Green, 1995) is a tool for calculating these posterior model probabilities.

#### *Posterior Model Probabilities*

Within the Bayesian model discrimination framework, for selecting between alternative models for the same data set, the model itself is considered to be an unknown parameter that requires estimation, and the posterior distribution is extended to include both parameter and model uncertainty.

Let the data be denoted by  $\mathbf{x}$ , and the  $m^{th}$  competing model and its parameters be denoted by  $M_m$  and  $\boldsymbol{\theta}_m$  respectively. The joint posterior distribution over both the model and parameter space is denoted by  $\pi(\boldsymbol{\theta}_m, M_m | \mathbf{x})$ . Using Bayes' theorem:

$$\pi(\boldsymbol{\theta}_m, M_m | \mathbf{x}) \propto L(\mathbf{x} | \boldsymbol{\theta}_m, M_m) p(\boldsymbol{\theta}_m | M_m) p(M_m), \quad (5.3)$$

where  $L(\mathbf{x} | \boldsymbol{\theta}_m, M_m)$  is the likelihood of the data under  $M_m$ ,  $p(\boldsymbol{\theta}_m | M_m)$  is the prior on the parameters in model  $M_m$ , and  $p(M_m)$  is the prior probability for

model  $M_m$ . The corresponding posterior model probabilities, the updated support for each competing model upon observing data  $\mathbf{x}$ , given by

$$\pi(M_m|\mathbf{x}) = \int \pi(M_m, \boldsymbol{\theta}_m|\mathbf{x}) d\boldsymbol{\theta}_m,$$

provide a means of discriminating between the competing models.

### *The Reversible-jump MCMC Algorithm*

The Reversible-jump MCMC algorithm, an extension of the Metropolis-Hastings algorithm, constructs a Markov chain with stationary distribution, up to proportionality, equal to  $\pi(\boldsymbol{\theta}_m, M_m|\mathbf{x})$ . The Reversible-jump MCMC algorithm itself essentially consists of two steps:

1. **Within Model Moves**, i.e. updating each parameter of the current model, for example by using a Metropolis-Hastings algorithm.
2. **Between Model Moves**, i.e. updating the model.

Thus, Reversible-jump MCMC enables both the model and the parameter space to be explored simultaneously, and allows competing models to be quantitatively discriminated using the estimated posterior model probabilities or the Bayes factors (see below). Note that as the number of parameters may vary between alternative models this algorithm accommodates moves between parameter spaces of different dimensions.

In the context of the methodology described and employed in this thesis to date, Step 1 is “standard”. Random walk, single update, Metropolis-Hastings algorithms have been routinely employed for this purpose. We focus now on Step 2, the “additional” element of the Reversible-jump algorithm.

### *Step 2: Between Model Moves*

Each between-model move also comprises of two steps:

- A. proposing a move to one of the alternative models,
- B. accepting or rejecting the proposed move.

In the special case of nested competing models, as here, in which one model can be derived from the other by relaxing parameter restrictions, the between model move proceeds as follows.

### Step A: Propose a New Model

We propose a move from the current model (denoted by  $M_m$ ) to a competing nested model (denoted by  $M_m^*$ ) with probability  $\Pr(M_m^*|M_m)$ . Consider first the case when a move to a nested model of a higher dimension is proposed.

At iteration  $h$  of the Reversible-jump MCMC algorithm let the current state (model and parameters) be denoted by  $(M_m, \boldsymbol{\theta}_m)_h$ , and the state of the proposed move be denoted by  $(M_m^*, \boldsymbol{\theta}_m^*)$ . As the models are nested, and  $M_m^*$  is of a higher dimension than  $M_m$ , we can express  $\boldsymbol{\theta}_m^* = (\boldsymbol{\theta}_m, \mathbf{u})$ , where  $\mathbf{u}$  denotes the additional subset of parameters in the proposed nested model, and for the  $d$  parameters of  $M_m^*$  and  $M_m$  in common,  $\theta_{m,j}^* \equiv \theta_{m,j}$  for  $j \in [1, \dots, d]$ . We simulate values for the parameters of  $\mathbf{u}$  from an appropriate proposal distribution, denoted by  $q(\mathbf{u})$ .

### Step B: Accept or Reject

Recall that  $\Pr(M_m^*|M_m)$  is the probability of proposing a move from model  $M_m$  to model  $M_m^*$ . Conversely,  $\Pr(M_m|M_m^*)$  is the probability of proposing the reverse move. We accept the proposed move from model  $M_m$  to model  $M_m^*$  with probability  $\alpha = \min(1, A)$ , where

$$\begin{aligned} A &= \frac{\pi(\boldsymbol{\theta}_m^*, M_m^*|\mathbf{x}) \Pr(M_m|M_m^*)}{\pi(\boldsymbol{\theta}_m, M_m|\mathbf{x}) \Pr(M_m^*|M_m)q(\mathbf{u})} \\ &= \frac{L(\mathbf{x}|\boldsymbol{\theta}_m^*, M_m^*)p(\boldsymbol{\theta}_m^*|M_m^*)p(M_m^*) \Pr(M_m|M_m^*)}{L(\mathbf{x}|\boldsymbol{\theta}_m, M_m)p(\boldsymbol{\theta}_m|M_m)p(M_m) \Pr(M_m^*|M_m)q(\mathbf{u})} \quad \text{via Equation (5.3)}. \end{aligned}$$

$$\text{We set } (M_m, \boldsymbol{\theta}_m)_{h+1} = \begin{cases} (M_m^*, \boldsymbol{\theta}_m^*) & \text{if the move is accepted} \\ (M_m, \boldsymbol{\theta}_m)_h & \text{otherwise.} \end{cases}$$

The alternative case, when a move to a nested model of a lower dimension is proposed (e.g. the reverse move from model  $M_m^*$  with parameters  $\boldsymbol{\theta}_m^*$  to model  $M_m$  with parameters  $\boldsymbol{\theta}_m$ ), it is accepted with probability  $\alpha = \min(1, A^{-1})$ .

Note that in general, where competing models need not be nested, a bijection  $g$  must be defined between the parameter space of the current and proposed models,

$$(\boldsymbol{\theta}_m^*, \mathbf{u}^*) = g(\boldsymbol{\theta}_m, \mathbf{u}),$$

in the model proposal step. Here  $\mathbf{u}^*$  and  $\mathbf{u}$  are sets of random variables with the respective density functions  $q^*(\mathbf{u}^*)$  and  $q(\mathbf{u})$ . The expression for  $A$  in the

accept or reject step is given by

$$A = \frac{\pi(\boldsymbol{\theta}_m^*, M_m^* | \boldsymbol{x}) \Pr(M_m | M_m^*) q^*(\boldsymbol{u}^*)}{\pi(\boldsymbol{\theta}_m, M_m | \boldsymbol{x}) \Pr(M_m^* | M_m) q(\boldsymbol{u})} \times \left| \frac{\partial(\boldsymbol{\theta}^*, \boldsymbol{u}^*)}{\partial(\boldsymbol{\theta}, \boldsymbol{u})} \right|,$$

where the Jacobian term on the right results from the transformation of variables in the bijection.

### *Discriminating Between Competing Models*

The posterior model probabilities,  $\pi(M_m | \boldsymbol{x}) = \int \pi(M_m, \boldsymbol{\theta}_m | \boldsymbol{x}) d\boldsymbol{\theta}_m$ , provide a measure of the degree of posterior uncertainty regarding each model. However, as the integration is usually analytically intractable, we use Reversible-jump MCMC methodology to sample from the posterior distribution,  $\pi(M_m, \boldsymbol{\theta}_m | \boldsymbol{x})$ , directly. The required marginal distribution,  $\pi(M_m | \boldsymbol{x})$ , can be estimated by the proportion of times the chain is in each model. It is important to recognise that posterior model probabilities are conditional on the set of models considered, and that they merely quantify the posterior evidence in favour of each particular model relative to the set of competing models. The “best” model for the data may, or may not, be contained within this set.

Bayes factors (Kass and Raftery, 1995) are frequently used to quantify the evidence in favour of one model versus another. The Bayes factor ( $B_{21}$ ) for two competing models ( $M_1$  and  $M_2$ ) is defined as the ratio of the posterior odds to the prior odds. That is:

$$B_{21} = \frac{\pi(M_2 | \boldsymbol{x}) / p(M_2)}{\pi(M_1 | \boldsymbol{x}) / p(M_1)}, \quad (5.4)$$

where values of  $B_{21} > 1$  indicate that the data more strongly supports  $M_2$  than  $M_1$ . Kass and Raftery (1995) provide following protocol for interpreting the evidence against model  $M_1$  in favour of  $M_2$ :

$$B_{21} = \begin{cases} 1 - 3 & \text{not worth mentioning} \\ 3 - 20 & \text{positive} \\ 20 - 150 & \text{strong} \\ 150 - \infty & \text{very strong} \end{cases}$$

Bayes factors thus provide a simple means of assessing the posterior evidence in

favour of one competing model against other. Such models need not be nested, and multiple comparisons present no problem.

#### *A Reversible-jump MCMC Algorithm for Assessing Site-Fidelity*

We wish to compare the two nested integrated population models:

$$M_1 : \gamma = 1$$

$$M_2 : \gamma \neq 1$$

Both models are given an equal prior weighting, i.e.  $p(M_1) = p(M_2) = 0.5$  and the probability of proposing a move between models is set to 1, i.e.  $\Pr(M_1|M_2) = \Pr(M_2|M_1) = 1$ . Recall that  $\gamma$  is given a Uniform[0,1] prior.

When proposing a move from  $M_1$  to  $M_2$ , after simplification, the acceptance probability  $\alpha$  is given by  $\min(1, A)$  where,

$$A = \frac{L_{live}(\mathbf{x}|\boldsymbol{\theta}, \gamma, M_2)}{L_{live}(\mathbf{x}|\boldsymbol{\theta}, M_1)q(\gamma)},$$

and  $\mathbf{x} = (\mathbf{F}, \mathbf{X}, \mathbf{R}, \mathbf{Z}, \mathbf{M}')$ . When years with incomplete coverage are excluded,  $\boldsymbol{\theta} = (\tau, \boldsymbol{\varepsilon}_{1...G}, \mathbf{c}_{1...G}, \boldsymbol{\phi}_{a,1...T-1})$ , or included,  $\boldsymbol{\theta} = (\tau, \beta_1^r, \beta_2^r, \mathbf{s}_{1...G}^r, \xi, \boldsymbol{\phi}_{a,1...T-1})$ , see Equation (5.2).

The Bayes factor comparing  $M_1$  and  $M_2$ , Equation (5.4), reduces to the ratio of the corresponding posterior model probabilities since the prior model probabilities are equal.

#### *Sedge Warbler and Reed Warbler Data*

To assess the importance of allowing for permanent emigration away from the CES sites in the integrated analyses of Sedge Warbler and Reed Warbler data, a Reversible-jump MCMC algorithm is implemented. As a Uniform[0,1] proposal distribution for  $\gamma$  in the ‘‘between model’’ move (Step 2) proved inefficient (low empirical transition probabilities between models were obtained) a Normal proposal for  $\gamma$ , appropriately truncated and the variance tuned *a priori*, is used. The posterior mean of  $\gamma$ , estimated from the independent fitting of the integrated model adjusted for permanent emigration, was used as the mean of this Normal proposal (see Table 5.4). The algorithm is run for 500,000 iterations with the first 300,000 discarded as burn-in, and uses data from the full set of sites to derive both  $L_{live}$  and  $L_a$ . For all analyses, essentially identical results

were obtained from multiple runs with overdispersed starting points, so that we assume the chain has converged.

In Table 5.6 the Bayes factors summarising the evidence against  $M_1$ , the null assumption of total site-fidelity, in favour of  $M_2$  are given for Sedge Warbler and Reed Warbler. When only complete coverage CES data are adopted, our results provide weak posterior evidence against the assumption of total site-fidelity in adult Sedge Warblers. However, the inclusion of additional information, from years with incomplete coverage, provides strong posterior against the constraint  $\gamma = 1$ . For Reed Warbler our results provide strong posterior support that, due to a lack of site-fidelity in adult birds,  $\gamma \neq 1$ , when only data from complete coverage years are analysed. This posterior support increases when the extra data from years with incomplete coverage are considered as well. In summary, for both species, we conclude that the estimates of survival from CES live-recapture data should be adjusted for permanent emigration. We also note that conducting these analyses with a different prior distribution chosen for  $\gamma$  (either Beta(2,5) or Beta(5,2)), gave the same interpretation.

**Table 5.6** Bayes factors summarising the evidence against  $M_1$  (assumption of total site-fidelity,  $\gamma = 1$ ) in favour of  $M_2$  (allowing for permanent emigration,  $\gamma \neq 1$ ).

	Bayes factor, $B_{21}$	evidence against $M_1$
<b>Sedge Warbler</b>		
Complete	2.3	weak
Incomplete	31.2	strong
<b>Reed Warbler</b>		
Complete	48.4	strong
Incomplete	346.2	very strong

### 5.2.4 True Seasonal Productivity

From published analyses of NRS data, estimates of true seasonal productivity ( $P_t^s$ ) for Sedge and Reed Warbler, respectively, of approximately 3.0 and 2.8 juveniles per breeding pair per year are obtained (see Chapter 3). These estimates are consistent with those from the fully integrated model that assumes total site-fidelity. However, for both species, Reversible-jump MCMC algorithms provide evidence that it is important to adjust adult survival rates from  $L_{live}$  for permanent emigration. Estimates of true seasonal productivity from

the original integrated model are thus positively biased as a consequence of the productivity constant  $k$  mopping up the discrepancy between apparent,  $S_{a,t}$ , and true,  $\phi_{a,t}$ , adult survival in the demographic population model (Equation (3.1)).

For Sedge Warbler, the fully integrated model adjusted for permanent emigration yields estimates of, on average, 2.7 and 2.6 juveniles per pair per year from the analyses of complete coverage data and incomplete coverage data respectively. As to be expected these estimates are slightly lower than those obtained from the original integrated model, but are now also slightly lower than the NRS estimate of 3.0. However, the estimate from NRS data does not take into account the “partial” losses of individual eggs and chicks up to (and shortly after) fledging in otherwise successful nests. Baillie et al. (2009b) give an average brood size of approximately 4.6, that is from an initial clutch size of 5 eggs on average each successful nest produces about 4.6 young. When this 8% loss is accounted for, the estimate from NRS data is reduced to 2.7 juveniles per pair per year, a much closer agreement with the estimates from the integrated model adjusted for permanent emigration.

For Reed Warbler, the fully integrated model adjusted for permanent emigration yields estimates of, on average, 2.1 and 2.0 juveniles per pair per year from the analyses of complete coverage data and incomplete coverage data respectively. These estimates appear low given the estimate of 2.8 from NRS data, an estimate already adjusted for “partial” losses in the egg and nestling stages. Indeed, in only one year (1999) does the integrated model predict true productivity to be greater than 2.7. It is unclear why there is such a discrepancy between the NRS and integrated model estimates. Bias in favour of early nests in the NRS data (Crick and Baillie, 1996) in which clutch size is known to be higher (Bibby, 1978; Cramp, 1992) most likely contributes to the difference. As the breeding season of Reed Warbler is more protracted than Sedge Warbler (Bibby, 1978) NRS data for Reed Warblers are likely to be more prone to such bias. Nevertheless a decrease in the average brood size from 3.5 to approximately 2.5 would be required to achieve estimates in line with the integrated model, too large a reduction compared to that observed in the study by Bibby (1978). Here, at the end of July, average brood size had only decreased to about 3.0. Post-fledging mortality could also explain the discrepancy. Were post-fledging losses higher for Reed Warbler than Sedge Warbler a much lower CES productivity estimate relative to the NRS estimate would be expected as the latter does not account for these losses. Note, if appreciable numbers of

birds were ringed in the nest one could modify  $L_{dead}$  to accommodate 3 age classes (post-fledging, independent juvenile, and adult) and obtain estimates of post-fledging survival (e.g. Thomson et al., 1999).

In addition to the possibility of positive bias in the NRS estimate of true seasonal productivity, the integrated model may be under-estimating  $P_t^s$  if the ring-recovery data are too sparse to produce reliable enough estimates of  $\phi_{a,t}$  and  $\phi_{j,t}$ . Over-estimation in any of these parameters will cause an under-estimation in  $k$ , and in turn estimates of true seasonal productivity. We note that independent estimates of  $\phi_{a,t}$  and  $\phi_{j,t}$  from ring-recovery data are consistent in scale to those obtained from the integrated analyses, whether or not an effect of permanent emigration is included. However, if the average  $\phi_{j,t}$  was closer to 0.3, than to the average ring-recovery estimate of 0.4, after re-scaling  $k$ , which in the integrated model would account for this difference, true seasonal productivity would now be about 2.8. This lower value of  $\phi_{j,t}$  is “consistent” with the ring-recovery data given the level of uncertainty (for most years it is encapsulated by the 95% symmetric credible intervals, see Figure 3.1, Chapter 3), and closer to the estimate used by Bibby (1978) of 0.24. Note that the independent estimates of  $\phi_{j,t}$  and  $\phi_{a,t}$  from ring-recovery data alone are subject to some prior sensitivity. Use of the Uniform[0,1] prior could be causing an over-estimation of  $\phi_{j,t}$  by pushing the posterior mean towards 0.5.

Our integrated analysis relies on the ring-recovery data to separate  $k$  from  $\phi_{j,t}$ . If, as the NRS estimates of true productivity for Reed Warbler leads us to question, the ring-recovery data are too sparse to provide precise enough estimates of  $\phi_{j,t}$ ,  $k$  and  $\phi_{j,t}$  may not be reliably separable. Potentially the estimates of  $\phi_{j,t}$  obtained from the integrated model should be viewed as an index of juvenile survival, still a biologically useful quantity. Ideally one would like a more informative set of data on  $\phi_{j,t}$ , but unfortunately this is not available for this species.

### *Stable Population*

The fully integrated model, allowing for permanent emigration, estimates  $\phi_{a,t}$  and  $\phi_{j,t}$ , on average, to be (0.41, 0.43) and (0.59, 0.40) for Sedge and Reed Warblers respectively. Under these conditions a seasonal productivity of 2.8 and 2.1 juveniles per breeding pair would be required to maintain stable Sedge Warbler and Reed Warbler breeding populations respectively. These estimates are in line with those obtained from the integrated analysis, suggesting that,

as previously concluded (see Chapter 3), the long-term trends in the Sedge and Reed Warbler populations are stable. This conclusion is consistent with their recent assigning of “Green” conservation status, indicating that these species are of little, or no, immediate conservation concern (Eaton et al., 2009).

### 5.3 Concluding Remarks

Despite adult songbirds exhibiting strong breeding philopatry (Wernham et al., 2002) some permanent emigration from the site of breeding, but within the domain of the UK population, is to be expected. In Cormack-Jolly-Seber type models (Cormack, 1964) permanent emigration results in negatively biased estimates of true survival ( $\phi$ ), as birds that leave the study area permanently are not distinguished from those that have died. The extent, and the importance, of departures from total site-fidelity on survival estimates from CES live-recapture data is typically unknown. Further, certain species monitored by the CES scheme may be more, or less, site-faithful than others. For example, Peach’s (1993) study suggested that Willow Warbler are less site-faithful than Reed Warbler.

We have modified our original integrated model (of Chapter 3) to accommodate the effect permanent emigration has on the estimates of adult survival from CES live-recapture data. By utilising both the ring-recovery data and the deterministic population model given by Equation (3.1), this modified integrated analysis enables true adult survival,  $\phi_{a,t}$ , to be separated from permanent emigration,  $1 - \gamma$ , in  $L_{live}$ . Further, we have demonstrated how Reversible-jump MCMC methodology (Green, 1995) can be used to assess whether or not it is appropriate to incorporate an effect of permanent emigration into the integrated analysis. Our results indicate, that for both Sedge and Reed Warbler, estimates of adult survival from CES live-recapture data should be adjusted for permanent emigration.

Estimates of  $\gamma$  provide a means of assessing the degree of site-fidelity of adult birds, however the definition of “site” here is very important. In the context of our analysis “site” refers to the area surveyed by the mist-net, and not to the overall available habitat. For example, Wicken Fen in Cambridgeshire is vast, with the current reserve covering over 770 hectares, much of which is reedbed or wet scrub. The area of Wicken Fen effectively surveyed by CES mist-netting is

small in comparison. The movement of birds relatively short distances to parts of the Fen not surveyed does occur (see Peach et al., 1990).

Some individual movements, within similar habitat at a given location (e.g. Wicken Fen), will constitute temporary emigration, with the bird returning to the area surveyed by the CES scheme in a subsequent year. As such “temporary emigration” is likely to be random, it is not expected to bias survival estimates (Kendall et al., 1997). However, songbirds are short-lived; Robinson (2005) gives a typical life-span for Sedge Warbler of only two years; Reed Warbler with their slightly higher survival rates will live longer, but certainly not considerably more. Thus there is a limit to the amount of “back and forth” movements into, and out of, parts of the overall habitat surveyed by the mist-net. Accordingly, short distance movements can constitute permanent migration. Note that our analyses indicate that about 7% and 14% of Sedge and Reed Warblers respectively will survive sufficiently long to breed in 3 consecutive years, the minimum length of time needed for temporary emigration to occur. However, of course, not all birds will be ringed before, or in, their first breeding year, for example some will be aged 2 or 3 when first captured, further limiting the possibility of temporary emigration. Consequently,  $\gamma$  quantifies both “long” distance movements to an entirely distinct site (which is probably rare), and short distance movements within the same habitat, the latter strictly speaking are not violations of breeding philopatry.

At present our integrated model assumes a constant  $\gamma$ . Rates of permanent emigration might well vary between sites, for example mist-nets might cover all the available habitat of a small reedbed but not of a larger one. Under such conditions apparent survival would be subject to site variation, that is  $S_{a,gt} = \gamma_g \phi_{a,t}$ , noting that true survival is assumed to represent the wider population of UK birds. As independent analyses of the live-recapture data detected no evidence of site-dependence in  $S_a$  this model was not employed. Time invariance in  $\gamma$  was also not considered on the grounds that emigration rates are less likely to change over time than between sites.

## 5.4 Chapter Summary

In this chapter the fully integrated model, developed for CES data in Chapter 3, was extended to enable years with incomplete coverage to contribute to the analysis, and modified to account for the effect permanent emigration has on the estimates of survival from CES live-recapture data. This “improved” integrated model was applied to Sedge Warbler and Reed Warbler data.

Extending the fully integrated model to years with incomplete coverage proved beneficial with regards to precision, although more costly in terms of computation time. Simply excluding years with incomplete coverage does not appear to bias the results.

The importance of incorporating permanent emigration into the fully integrated analysis was assessed using Reversible-jump MCMC methodology. For both Sedge and Reed Warbler, our results provide convincing evidence that permanent emigration should be accounted for within the model. Failure to do so negatively biases the estimates of true adult survival, and positively biases the estimates of true seasonal productivity.

# Chapter 6

## Discussion

The CES scheme, an annual programme of standardised mist-netting across a large number of sites, generally provides the best, and sometimes the only, source of information concerning the key demographic parameters for certain common songbirds in the UK: adult abundance, juvenile abundance, productivity and adult survival. Producing reliable, and usefully precise, estimates of these demographic parameters from models adopting CES data was the primary motivation behind the work presented in this thesis. The methodology developed will help facilitate the reliable monitoring, and effective management, of Britain's common songbird populations.

Using independent models adopting CES data, the BTO routinely produces indices of abundance and productivity, using the methods of Peach et al. (1998) and Robinson et al. (2007) respectively. To a considerably more limited extent, CES data have also been used to obtain estimates of adult survival (for example Pratt and Peach, 1991; Peach, 1993; Freeman, 2008). We have extended the current methodology employed by the BTO, cast the models in a Bayesian framework, and by developing an integrated population model for CES data, undertaken a more holistic approach to population monitoring. With respect to the latter, we have advocated an alternative parametrisation for the productivity model, one that is more amenable to an integrated approach to population modelling. The methodologies described in this thesis have been successfully applied to Sedge Warbler and Reed Warbler CES data collected from 1987 to 2005 inclusive.

Underpinning the CES scheme, and the models for the demographic parameters, is the assumption of constant mist-netting effort at a given site over all years

it is operated. Assuming constant effort, any variability between years in the number of birds caught is attributed to changes in the population level and to stochastic variation, and not to varying intensity of capture effort, thus enabling indices of abundance and productivity to be obtained. To model adult survival constant effort is not required. However, the ability to constrain parameters associated with recapture rates to be time-invariant is desirable as the precision in estimation is improved (Peach et al., 1990).

The absence of data from isolated visits within a year violates the key assumption of constant effort. One ad-hoc approach to address this is to omit data from years with missed visits. Motivated by the potential to improve precision by including such data into the analysis, we have developed novel Bayesian models that produce reliable estimates of abundance, productivity and adult survival from CES data in which missing visits have resulted in non-constant effort.

In the modelling of abundance and productivity, the BTO currently employs the approach of Peach et al. (1998), correcting the observed count data for missed visits prior to model fitting. As the corrected counts are assumed to be known without error, this approach under-reports the uncertainty in the parameter estimates. Taking advantage of the added flexibility afforded by the Bayesian paradigm, we have developed an alternative Bayesian model that combines in a single process the correction for missed visits and the model fitting, thereby more accurately determining the uncertainty in the parameter estimates.

Previously the modelling of adult survival from CES data has received little attention despite these data providing possibly the best source of information on this important demographic parameter for small songbirds. We have developed the first model to explicitly account for missed visits in the modelling of survival. Our parsimonious model negates the need to allow for full time dependence in the recapture rates, or to impose *a priori* criteria for including data to ensure time-invariance in the recapture rates. Further, we have extended efficiently the current methodology of Pradel et al. (1997) to account for transient birds in the live-recapture data, making use of the specific advantages provided by the design of the CES scheme in this respect, namely the within-year recaptures. Our novel model produces estimates of apparent adult survival from CES data, unbiased by either missed visits or transient birds, that have improved precision over the traditional approaches.

The analyses of abundance, productivity, and adult survival that adopt CES data, conducted by the BTO and others, are done independently. The simul-

taneous estimation of these parameters, via an integrated population model, to produce parameter estimates that are consistent and robust with both the data, and according to the underlying population dynamics, with one another, is appealing. To this end, we have produced the first integrated population model for CES data, augmented with additional information, albeit limited for the species considered here, on adult and juvenile survival rates from the National Ring-Recovery (NRR) database. Within this thesis we identify specific advantages of this new integrated approach for the monitoring of songbirds from CES data, among which is the ability to obtain estimates of total seasonal productivity (i.e. uniting multiple broods, and allowing for immediate post-fledging mortality), rather than an index merely proportional to it. Further, as integration enables the sharing of information between the component demographic models, improved precision in parameter estimation typically results. This is particularly relevant for juvenile survival, with this thesis presenting the first estimates of Sedge and Reed Warbler juvenile survival that are usefully precise. CES data alone provides estimates of apparent adult survival, as opposed to true survival, which in addition to mortality also incorporates a component due to permanent emigration (Peach et al., 1995b; Thaxter et al., 2006). Through integration, and the joint analysis with NRR data, the methodology enables true adult survival to be estimated. Furthermore, we have shown how the importance of incorporating permanent emigration into the integrated model can be assessed using Reversible-jump MCMC methodology.

The flexibility of integrated population modelling means that any number of disparate, but comparable, data sources are potentially useful. The integrated population model for CES data presented in this thesis could be extended even further, bringing in independent information from other BTO monitoring schemes. For example, data on abundance from the Common Bird Census, and its successor, the Breeding Bird Survey, and information on productivity from the Nest Record Scheme, could be incorporated if a useful amount of data were available.

The models presented in this thesis have a much wider applicability than for Sedge and Reed Warbler CES data alone. Not only does the CES scheme provide sufficient data to routinely monitor 23 other species (Grantham and Robinson, 2008), but there are many other “Constant Effort” ringing programmes directly modelled on the BTO’s CES scheme in Europe (EURING Website, 2008), and the MAPS scheme in the U.S.A. (DeSante et al., 1999), that generate similar data. These data sets will also be amenable to analysis using the

methodologies presented in this thesis. Furthermore, our novel transient model (referred to as model EP) generalises to any two stage live-recapture study where transients are believed to be present, and inference on the primary sampling period is of interest. As the “Robust” design (Pollock, 1982) is a feature common to many other mark-recapture studies (Kendall and Bjorkland, 2001), this model is potentially useful in the analysis of live-recapture data from a wide range of studies on a variety of animal taxa.

The research presented in this thesis provides wide scope for future work, such as the inclusion of environmental variables, incorporating demographic stochasticity into the population model, allowing for overdispersion in the annual count data, the potential to use CES data alone to obtain information on juvenile survival, and model selection and the assessment of goodness-of-fit in integrated modelling. We comment upon these in turn. We also discuss the possibility of modelling the CES data at the level of the individual within-year visit using Robust design methodology.

### *Environmental Covariates*

Environmental changes, often resulting from human activity, can potentially alter resource availability and thereby affect the demographic parameters (survival, productivity and emigration/immigration) which in turn determine population size (Freeman et al., 2007b). If a population is in decline, conservationists typically wish to identify the important demographic mechanisms driving this decline, and furthermore, identify the factors affecting these demographic parameters so that appropriate remedial action can be instigated. For example, the massive decline in the number of UK farmland birds in the 1970s and 80s is well documented. These declines have been linked to the intensification of agricultural practices that resulted in a reduction in the availability of food (invertebrates and seeds) and suitable breeding habitats (Newton, 2004). In turn, this has had a detrimental effect on the over-winter survival of seed-eating birds and the productivity of ground-nesting birds (Gregory et al., 2004). In light of this knowledge, to help conserve farmland wildlife the UK government-funded Environmental Stewardship was initiated across England, a scheme that rewards farmers for managing the land in a more wildlife friendly way (DEFRA, 2005).

The inclusion of environmental covariates into the analyses presented in this thesis (independent or integrated), and the investigation of their relationship with

the underlying demographic rates, could be of interest. Extending the current methodology to accommodate environmental variables is relatively straightforward. For example, Peach et al. (1991) found that the annual variation in Sedge Warbler adult survival could be explained by annual changes in the amount of wet-season rainfall in the Sahel, their west African wintering quarters - the more rainfall, the higher the probability of survival. In this study Peach et al. (1991) introduced an annual index of Sahel rainfall (denoted by  $R_t$ ) as a covariate for adult survival ( $S_{a,t}$ ), modelling:

$$\text{logit}(S_{a,t}) = \alpha + \beta R_t.$$

Such an expression is readily incorporated into the independent models for adult survival and into the integrated population models presented in this thesis. Note, this covariate has also been used to explain the variation in survival of White Storks (*Ciconia ciconia*), another trans-Saharan migrant bird species (Gimenez et al., 2009). Similarly, productivity or juvenile survival maybe related to the weather conditions during the British summer. Using the appropriate models, these relationships can also be incorporated into the independent and integrated analyses. The inclusion of environmental covariates into an integrated analysis is not new, for example Besbeas et al. (2002), Brooks et al. (2004) and King et al. (2008b) use the number of frost days, a proxy for the severity of the winter, as a covariate for juvenile and adult survival rates in the integrated analysis of Lapwing (*Vanellus vanellus*) data.

Covariate analyses are restricted by the availability of data. Freely available, online from the British Atmospheric Data Centre, are monthly temperature data for central England (Parker et al., 1992; BADC Website, 2009). Monthly England and Wales rainfall data are also freely available online from the Hadley Centre of the UK Met Office (Met Office Website, 2009). Exploratory analyses using these data suggest some promise for summer (May-August) rainfall and temperatures as covariates for adult and juvenile survival, and productivity (for example, see Table 6.1). In particular, minimum summer temperature maybe useful in explaining some of the annual variation in the survival of juvenile Sedge Warblers and the productivity of Reed Warblers, with a cold snap appearing to decrease Sedge Warbler juvenile survival and Reed Warbler productivity.

The relationship between the underlying demographic rates and the environmental variables driving them is almost certainly complex, with potentially several variables acting upon the demographic rates at different times during the year. For example, Sedge and Reed Warbler survival is expected to be in-

fluenced by both the weather conditions in Britain, during the summer months, and in west Africa, during the winter months, and indeed conditions on the arduous migration itself. A detailed understanding of the ecology of the species of interest, along with careful consideration of the appropriate covariates to use and stage of the life cycle they affect, is required when including covariates into the models for the demographic parameters.

**Table 6.1** Exploratory environmental covariate analysis - correlations between the demographic parameters (adult survival  $\phi_{a,t}$ , juvenile survival  $\phi_{j,t}$ , and productivity  $P_t$ ) with the year  $t$  total summer (May-August) rainfall, and the mean, minimum and maximum summer (May-August) temperatures. Posterior means of the demographic parameters, obtained from the fully integrated analysis of the incomplete coverage CES data, allowing for permanent emigration, as in Chapter 5, Section 5.2, are used as estimates.

	<i>Summer Rainfall</i> , mm	<i>Summer Temperature</i> , °C		
		Mean	Minimum	Maximum
<b>Sedge Warbler</b>				
$\text{logit}(\phi_{a,t})$	-0.19	-0.11	0.09	-0.19
$\text{logit}(\phi_{j,t})$	0.19	0.05	0.28	-0.07
$P_t$	-0.09	-0.18	-0.23	-0.14
<b>Reed Warbler</b>				
$\text{logit}(\phi_{a,t})$	-0.06	-0.27	-0.22	-0.27
$\text{logit}(\phi_{j,t})$	-0.10	0.15	0.11	0.12
$P_t$	0.05	0.22	0.40	0.14

### *Demographic Stochasticity*

Recall that, to formulate an integrated population model for CES data the demographic parameters ( $A_t$ ,  $J_t$ ,  $P_t$ ,  $\phi_{a,t}$ ,  $\phi_{j,t}$ ) of the joint likelihood were linked using a deterministic population model, described by Equation (3.1). This model, representing the underlying biological system, expresses the adult abundance in year  $t$  as a function of the adult and juvenile abundances in year  $(t-1)$  and the intermediate survival rates. That is:

$$\begin{aligned} A_t &= A_{t-1}\phi_{a,t-1} + kJ_{t-1}\phi_{j,t-1} \\ &= A_{t-1}\phi_{a,t-1} + kP_{t-1}A_{t-1}\phi_{j,t-1}, \end{aligned} \quad (\text{Equation (3.1)})$$

where  $A_1$ ,  $J_1$ , and  $P_1$  are constrained to 1.

Integrated approaches, which specify deterministic population evolution for birds, have also been employed by Peach et al. (1999); Siriwardena et al. (2001); Freeman and Crick (2003); Besbeas and Freeman (2006); and Freeman et al. (2007b) for example. However, stochastic differences between individual birds

in their survival and reproductive rates will mean the actual number of juveniles and adults in year  $(t - 1)$  that survive to join the adult population in year  $t$  are random variables, giving rise to “demographic stochasticity” (Caswell, 2001). Permitting such demographic stochasticity, for instance by using a state-space model, has been achieved both classically (for example Besbeas et al., 2002, 2003; Véran and Lebreton, 2008; Borysiewicz et al., 2009) and using Bayesian methods (for example Brooks et al., 2004; Schaub et al., 2007; King et al., 2008b; Reynolds et al., 2009).

To introduce demographic stochasticity into the integrated population model presented in this thesis we propose the following extension. Let:

$$\begin{aligned} A_t &= A_{t-1}\phi_{a,t-1} + kJ_{t-1}\phi_{j,t-1} + \epsilon_t \\ &= A_{t-1}\phi_{a,t-1} + kP_{t-1}A_{t-1}\phi_{j,t-1} + \epsilon_t, \end{aligned} \tag{6.1}$$

where the  $\epsilon_t$  terms are independent, identically distributed, Normal random variables with mean 0 and common variance  $\sigma^2$ , i.e.

$$\epsilon_t \sim \text{Normal}(0, \sigma^2).$$

These  $\epsilon_t$  terms thus allow for the additional “process error” associated with the stochastic variability in population size. We might specify a vague inverse-Gamma prior for the hyper-parameter  $\sigma^2$ , which, being conjugate, enables the posterior conditional distribution of  $\sigma^2$  to be readily updated using the Gibbs sampler (see Section 1.5).

The importance of demographic stochasticity is inversely proportional to population size (Buckland et al., 2004a). Thus for the abundant UK Sedge and Reed Warbler populations, with approximately 300 thousand and 60-120 thousand breeding pairs respectively (Robinson, 2005), the consequences of demographic stochasticity are not expected to be great. Moreover, it is likely that the process error is small in comparison to the uncertainty in the parameter estimates themselves (c.f. Besbeas and Freeman, 2006).

#### *Overdispersion in Poisson Count Data*

The models for abundance presented in this thesis assume a Poisson distribution for the CES count data. Recall that the Poisson model for adult abundance is given by:

$$n_{gt}^a \sim \text{Poisson}(\lambda_{gt}^a),$$

where  $\ln(\lambda_{gt}^a) = \beta^a + s_g^a + y_t^a$ , (Equation (2.1))

with an analogous model formed for juvenile abundance.

Analyses by Peach et al. (1998) showed that for most species, CES count data were generally well fitted by such a Poisson model. However, overdispersion is likely in count data from flocking birds that tend to be caught in groups. If required, the models for abundance can readily be adapted to accommodate overdispersion. A simple approach would be to replace the Poisson distribution with the Negative Binomial distribution, allowing the variance to exceed the mean. For example, for adult count data we would assume that

$$n_{gt}^a \sim \text{NegBin}(\lambda_{gt}^a, d^a),$$

where  $d^a \in [0, \infty)$  is the overdispersion parameter. As with the Poisson model,  $\ln(\lambda_{gt}^a) = \beta^a + s_g^a + y_t^a$ . Such a model is readily fitted using Bayesian techniques, as described in Section 2.2, with the Metropolis-Hastings algorithm including an extra step for updating the overdispersion parameter,  $d^a$ . Note, this is equivalent to the Poisson-Gamma hierarchical model (see King et al., 2008b).

Alternatively, overdispersion can be accommodated using an overdispersed Poisson regression model, following the approach of Link and Sauer (2002). Here an extra term representing the overdispersion effects (denoted by  $\epsilon_{gt}$ ) is included in the models for abundance. For example, for adult counts it would be assumed that

$$n_{gt}^a \sim \text{Poisson}(\lambda_{gt}^a),$$

where  $\ln(\lambda_{gt}^a) = \beta^a + s_g^a + y_t^a + \epsilon_{gt}^a$ .

The overdispersion effects are treated as identically distributed Normal random variables, with mean 0, and common variance  $\sigma_{\epsilon_a}^2$ . Under a Bayesian analysis, a prior is specified for the hyper-parameter  $\sigma_{\epsilon_a}^2$ , for example the conjugate inverse-Gamma prior which enables the Gibbs sampler to be employed in the updating step of the MCMC algorithm for this parameter.

### *Estimating Juvenile Survival from CES Data*

On account of the high dispersal of juvenile birds, and consequently the low number of juveniles that are ever recaptured in years subsequent to ringing (see Table 1.3), currently the analysis of CES live-recapture data is restricted to capture histories from adult birds. Further, as ring-recovery data for small

songbirds are typically sparse (Baillie et al., 1986), it is only through integration that we have obtained estimates of juvenile survival,  $\phi_{j,t}$ , with a useful level of precision (see, for example, Figure 3.6).

In Section 3.3.2 the importance of ring-recovery data in the integrated analysis for separating  $\phi_{j,t}$  from the productivity constant  $k$  was highlighted. Given the paucity of direct information on  $\phi_{j,t}$  from ring-recovery alone, it is appealing to utilise the limited information pertaining to juvenile survival from the few recaptures of birds ringed as juveniles, building an age-stratified model for the full CES live-recapture data from both adult and juvenile birds. Modelling the portion of data from the newly-marked juvenile birds is a relatively straightforward extension of the current methodology, although allowances must be made for age-specific effects.

The temporary presence of juvenile and adult transient birds will need to be accounted for by explicit probabilities within the age-specific analysis. Juvenile and adult birds can be caught in transit near the end of the mist-netting season, corresponding to the time they depart CES sites for their pre-migratory feeding grounds. In addition, adult birds can also be caught in transit at the start of the mist-netting season when they return from west Africa to their breeding sites. Consequently, the probability a newly-marked bird is a transient,  $\tau$ , is expected to differ according to whether it was ringed as an adult or as a juvenile. Furthermore, as juvenile birds are yet to have fledged by the start of the CES season, the evasion probability,  $\varepsilon$ , the probability a bird is only caught once within a year given that it is caught, is also expected to differ between adult and juvenile birds in their inaugural ringing year.

Estimates of juvenile survival from CES data must be considered those of apparent survival,  $S_{j,t}$ , due to the low site-fidelity of juvenile birds to their natal grounds. To integrate, the discrepancy between true and apparent juvenile survival must be accounted for. This can be achieved following the approach of Section 5.2, using an age-specific permanent emigration rate, i.e. in the likelihood for the live-recapture data we set  $S_{j,t} = \gamma_j \phi_{j,t}$  for juveniles, and if appropriate  $S_{a,t} = \gamma_a \phi_{a,t}$  for adults.

When integrating the usual non-independence issue between the likelihoods for count data and live-recapture data arises, with now the likelihoods for both the adult and juvenile count data ( $L_a$  and  $L_j$ , respectively) being non-independent of the age-specific likelihood for the full live-recapture data. Split data analyses can be used to accommodate this.

Finally we note that the potential for incorporating recaptures of birds ringed as juveniles into the analysis will depend on the species considered, as some songbirds exhibit greater levels of natal site-fidelity than others. For example, the juvenile dispersal of Reed Warbler is less than that of Sedge Warbler on account of its more specialised habitat, and hence using CES data to estimate  $S_{j,t}$  is likely more feasible for Reed Warbler (see Table 1.3).

#### *Model Selection and Goodness-of-Fit Assessment in Integrated Modelling*

No formal framework for model selection with integrated data has been presented in this thesis. However, in the Bayesian paradigm tools for model selection are well established. For example, DIC, the Bayesian analogue of AIC, is a commonly used statistic for comparing competing models (Gelman et al., 2004). An alternative tool is the Reversible-jump MCMC algorithm (Green, 1995); a type of MCMC sampler that allows for dimensional changes in the probability distribution being simulated, enabling model selection and parameter estimation to be performed simultaneously. Using these tools the integrated analysis presented in this thesis can be extended to include a model selection step. For example, to choose between the set of integrated models that include or exclude time and/or site invariance in the demographic parameters. We note, however, care should be taken to ensure that a competing model is not parameter redundant.

At present goodness-of-fit assessment for integrated models, if considered at all, is typically *ad hoc*. Common approaches include informal graphical checks in which fitted values, such as the number of ring-recoveries, are compared to their observed values (for example, Besbeas et al., 2002). Such an approach should identify any serious lack of fit, although it does not constitute a formal, or rigorous, test. An alternative approach is to use standard tests (such as Bayesian p-values) to assess the goodness-of-fit for each component model in the integrated analysis (for example, Brooks et al., 2004). However, this approach involves performing a number of separate tests. The development of a formal, single, test of goodness-of-fit for integrated models thus provides great scope for future work.

---

*Potential of Robust Design Models for CES Data*

The CES data satisfy the criteria of the “Robust” design capture-recapture models (Pollock, 1982), with the “primary sampling periods”, years, containing 12 “secondary sampling sessions”, visits (see Section 2.4.7). Adopting CES data at the individual within-year visit level, Robust design models provide a means of jointly estimating the probabilities of apparent survival and recapture, the number of animals in the population (abundance) and the number of new recruits. Furthermore, Robust design methodology can be employed for populations stratified according to age, providing estimates of age-specific parameters (Pollock et al., 1990). The appropriate Robust design model may therefore provide a viable alternative to the integrated population model presented here.

CES data arise from an “open population”, in which the population size at a site changes between years as a consequence of births, deaths and migration (Pollock et al., 1990; Amstrup et al., 2005). Furthermore, as birds typically arrive to, and depart from, their breeding site at different times, the population is also open over the secondary periods (the mist-netting season). CES data thus require analysis using the “Open Robust” design model of Schwarz and Stobo (1997), a model that permits this form of non-closure over the secondary sampling session.

Several important advantages of employing Robust design methodology over our current integrated approach are identified:

1. The issue of non-independence between the likelihood for the count data and the likelihood for the live-recapture data is overcome.
2. All available CES data, from both between and within-year captures, are fully utilised.
3. Data from sites with missed visits pose no problems, as here the probabilities of recapture at such visits are simply constrained to zero.
4. A population model is not specified.
5. Estimates of the number of birds recruited into the population each year (at a given site) are available which maybe of interest in their own right.
6. Temporary emigration is permitted, although in the context of CES data, songbirds being short-lived, this is probably not required (see Section 5.3).

However, before implementing a Robust design approach to modelling CES data, several important issues would need to be addressed. In particular, monitoring adult songbird abundance, and to a slightly lesser extent productivity

and survival, is the driving motivation behind the CES scheme. Age-stratified Robust design models provide estimates of the total number of adult/juvenile birds alive at each site at the start of each year (denoted by  $N_{gt}^a$  and  $N_{gt}^j$  respectively). This raises the issue of how to combine these site-specific estimates to form an index of abundance and productivity ( $N_{gt}^j/N_{gt}^a$ ) for the wider UK population. This task is problematical as a different set and number of sites are operational in any given year.

The presence of transient birds in the CES data must be adequately accounted for or negative bias in the estimates of survival will be introduced. Furthermore, Clavel et al. (2008) note that as transients do not belong to the population of interest in the sample of captured animals (i.e. the resident birds at site  $g$ ), their presence in the data positively biases population size estimates. Following the approach of Hines et al. (2003) for Robust design models with closed secondary sampling periods, the Schwarz and Stobo (1997) model can (probably) be adjusted to account for transients, and using an approach similar to Clavel et al. (2008) estimates of population size modified accordingly.

To integrate the CES live-recapture data and the ring-recovery data, the discrepancy between true survival  $\phi_t$  (from the ring-recovery data) and apparent survival  $S_t$  (from the CES data) still needs to be addressed, as is done so in our integrated model of Section 5.2. If the sample of recaptured birds and recovered birds were from the same set of marked birds the approach of Lindberg et al. (2001) can be used to estimate the probability of permanent emigration (in our notation  $(1 - \gamma)$ ). However, the ring-recovery data and the CES live-recapture data arise from two entirely separate studies and correspond to mostly different birds, thus an alternative approach must be adopted. Recall that the likelihoods for ring-recovery data and live-recapture data ( $L_{dead}$  and  $L_{live}$ , respectively) are assumed to be independent, with their joint likelihood given by  $L_{dead} \times L_{live}$ . By setting  $S_t = \gamma\phi_t$  in the likelihood for live-recapture data, it should be possible to disentangle  $\gamma$  from  $\phi_t$  in  $L_{live}$  using the additional information provided by the ring-recovery data.

For juveniles, on account of high natal dispersal, estimates of their survival rates from CES data are likely to lack precision. Precision will be improved slightly by integration with the limited ring-recovery data, although not substantially. In our integrated model ancillary information provided by the deterministic population model dramatically improves precision. As this will not be the case for analysis using the Robust design approach, usefully precise estimates of

juvenile survival may not be obtainable.

In general, we do not believe that the above issues are intractable, however formulating an appropriate Robust design model for CES data will require careful consideration, and an assessment of which parameters are separately identifiable (for example see Schwarz and Stobo, 1997). We note that such a model would require the estimation of many extra nuisance parameters, for example parameters for the recapture, apparent survival, and transient probabilities at each secondary sampling period, some of which will be site-specific. Hines et al. (2003) doubt the usefulness of such Robust design models, adjusted for transients, for estimating survival using CES-style bird ringing data due to the loss of precision incurred from the decrease in parsimony. The same is likely also true for the other biologically interesting parameters associated with abundance and productivity. Further we note the extra computation expense incurred by the estimation of these additional parameters. Nonetheless, the potential for using Robust design methodology to analyse CES data is promising, and is an interesting avenue for future research.

*The UK government is committed to protecting populations of wild birds in Britain, but for this to be achieved it is essential to accurately monitor their populations through time. The CES scheme is a particularly important long-term monitoring scheme with respect to populations of Britain's common songbirds located in scrub and wetland habitats. The models for CES data presented in this thesis provide precise estimates of the key demographic variables (abundance, productivity and survival), and are therefore valuable with regards to accurately assessing the state of these songbird populations.*



# Appendix A

## Notation

### Indices

$a$	adult
$j$	juvenile
$g \in [1, G]$	site
$t \in [1, T]$	year
$k \in [1, K]$	visit
$h$	MCMC iteration

## Data Summary Statistics

### CES Count Data

*For the summary statistics that follow superscripts “a” and “j” are used to denote the adult and juvenile subsets of the CES count data respectively.*

$n_{gt}$  yearly-site count: the number of unique birds caught at site  $g$  in year  $t$  given all  $K$  visits are made. Superscripts “*obs*” and “*cen*” are used to denote, respectively, a known (if all  $K$  visits were made) and an unknown (if 1 or more visits were missed) yearly-site count.

$\ell_{gt}$  the observed number of unique birds caught at site  $g$  during the visits made in year  $t$ . The superscript “*obs*” is used to denote data where 1 or more visits were missed.

$E_{gt}$  yearly-site count at site  $g$  in year  $t$  corrected for missed visits under the method of Peach et al. (1998).

- $N_g$  the total number of unique birds caught at site  $g$  across all years of complete coverage (i.e. for years in which all  $K$  visits were made).
- $N_{gt}$  the number of birds that belong to  $N_g$  that are caught at visits corresponding to those made to site  $g$  in year  $t$ .
- $\eta_{gt}$  degree of censoring: the amount the yearly-site count for site  $g$  in year  $t$  is reduced on account of missed visits (i.e.  $\eta_{gt} = n_{gt}^{cen} - \ell_{gt}^{obs}$ ).
- $n_{gtk}$  visit-level count: the number of unique birds caught at visit  $k$  to site  $g$  in year  $t$ .

### CES Live-Recapture Data

For site  $g$  data corresponding to *newly-marked* birds:

- $F_{gi\bullet}$  the number of birds ringed and first released in year  $i$ .
- $F_{git}$  the number of birds ringed and first released in year  $i$  that were next recaptured in year  $t$ .
- $X_{gi}$  the number of birds ringed and first released in year  $i$  that were never seen again.

For site  $g$  data corresponding to *previously-marked* birds (or assumed residents):

- $R_{gi\bullet}$  the number of birds captured and released in year  $i$ .
- $R_{git}$  the number of birds captured and released in year  $i$  that were next recaptured in year in  $t$ .
- $Z_{gi}$  the number of birds captured and released in year  $i$  that were never seen again.

For site  $g$  data:

- $M'_{gt}$  the number of ringed resident birds recaptured in year  $t$  exactly once.
- $M''_{gt}$  the number of ringed resident birds caught repeatedly in year  $t$ .

### Ring-Recovery Data

- $D_{it}^a$  the number of birds ringed as adults in year  $i$  recovered dead in year  $t$ .
- $D_{it}^j$  the number of birds ringed as juveniles in year  $i$  recovered dead in year  $t$ .

## Models

### Abundance Model

For the notation that follow superscripts “a” and “j” are used to denote the adult and juvenile models respectively.

- $\lambda_{gt}$  parameter of the Poisson distribution for the yearly-site counts from site  $g$ , in year  $t$ .
- $\lambda_{gtk}$  parameter of the Poisson distribution for the visit-level counts from site  $g$ , in year  $t$ , at visit  $k$ .
- $\beta$  intercept term.
- $s_g$   $g^{th}$  site effect.
- $y_t$   $t^{th}$  year effect.
- $v_k$   $k^{th}$  visit effect.
- $A_t$  index of adult abundance in year  $t$ .
- $J_t$  index of juvenile abundance in year  $t$ .
- $L_a$  likelihood under the adult abundance model.
- $L_j$  likelihood under the juvenile abundance model.

### Productivity Model

- $P_t$  index of productivity in year  $t$ .
- $P_t^s$  true seasonal productivity in year  $t$ , defined as the ratio of juvenile to adult birds in the population.
- $k$  productivity constant: positive scaling factor that accounts for the difference between the indices and the true abundances.

### Traditional Productivity Model

- $\theta_{gt}$  parameter of the Binomial distribution: the expected proportion of unique juvenile birds in the annual unique catch for site  $g$  in year  $t$ .
- $\beta^p$  intercept term.
- $s_g^p$   $g^{th}$  site effect.
- $y_t^p$   $t^{th}$  year effect.

**CES Adult Survival Models**

- $S_{a,t}$  apparent adult survival: probability a resident bird alive in year  $t$  survives to year  $(t + 1)$ .
- $\phi_{a,t}$  true adult survival: probability a resident bird alive and present at its site of ringing in year  $t$  survives and is present in year  $(t + 1)$ .
- $\gamma$  probability a ringed resident bird does not permanently emigrate in any given year.
- $S'_a$  probability a ringed resident bird survives the post-ringing period.
- $c_g$  recapture probability: probability a surviving resident bird, ringed at site  $g$ , is recaptured at this site in any given year.
- $\varepsilon_g$  evasion probability: probability a resident bird, caught at site  $g$  in any given year, is caught at exactly one visit within that year.
- $c'_g$  probability a ringed resident bird at site  $g$  is recaptured in the post-ringing period.
- $\tau$  transience probability: probability a bird is a transient at the time of ringing.
- $\varphi$  residence probability: probability a bird is a resident breeder at the site of ringing, and survives the post-ringing period.

For *newly-marked* birds from site  $g$ :

- $f_{git}$  probability a bird ringed and released in year  $i$  is first recaptured in year  $t$ .
- $x_{gi}$  probability a bird ringed and released in year  $i$  is never seen again.

For *previously-marked* birds (or assumed residents) from site  $g$ :

- $p_{git}$  probability a bird caught and released in year  $i$  is next recaptured in year  $t$ .
- $\chi_{gi}$  probability a bird captured and released in year  $i$  is never seen again.

$L_{live}$  likelihood under the model for the CES live-recapture data.

*The following abbreviations are used to denote the different models for the CES live-recapture data.*

- CJS Cormack-Jolly-Seber model.
- AH Ad-hoc method that uses that between-year recaptures.
- EAH Ad-hoc method that uses within-year and between-year recaptures.

P	Pradel et al. (1997) model.
EP	Extended Pradel model.
CD	Freeman (2008) c-dash method.
MV	Cormack-Jolly-Seber model adjusted for missed visits.
EPMV	Extended Pradel model adjusted for missed visits.

### Model for the Visit (within-year) Recapture Probabilities

$v_{gk}$	visit-level recapture probability: probability a surviving resident bird, caught at site $g$ , is recaptured at visit $k$ (to this site) in any given year.
$\kappa$	standardised visit number.
$s_g^r$	$g^{th}$ site effect.
$\beta_o^r$	coefficient for the $o^{th}$ order $\kappa$ effect.
$I_{gtk}$	function indicating whether or not visit $k$ to site $g$ in year $t$ was made.
$\xi_g$	encounter effect for site $g$ .
$c_{gt}$	probability a surviving resident bird, ringed at site $g$ , is recaptured at this site in year $t$ .
$\varepsilon_{gt}$	probability a resident bird, caught at site $g$ in year $t$ , is caught at exactly one visit within that year.

### Ring-Recovery Model

$\phi_{a,t}$	true adult survival: probability an adult bird alive in year $t$ survives to year $(t + 1)$ .
$\phi_{j,t}$	true juvenile survival: probability a juvenile bird alive in year $t$ survives to year $(t + 1)$ .
$\lambda_t^a$	probability a ringed bird, that has died in year $t$ aged as an adult, is reported dead.
$\lambda_t^j$	probability a ringed bird, that has died in year $t$ aged as a juvenile, is reported dead.
$L_{dead}$	likelihood of the ring-recovery data.

## General Notation used in the Bayesian Approach

- $\mathbf{x}$  data.  
 $\boldsymbol{\theta}$  model parameters.  
 $\boldsymbol{\theta}^*$  proposed update for  $\boldsymbol{\theta}$ .  
 $\boldsymbol{\theta}^h$  state of the MCMC chain for  $\boldsymbol{\theta}$  at the start of iteration  $h$ .

*For parameters  $\boldsymbol{\theta}$  and the data  $\mathbf{x}$ :*

- $p(\boldsymbol{\theta})$  prior distribution.  
 $f(\mathbf{x}|\boldsymbol{\theta})$  “likelihood”.  
 $\pi(\boldsymbol{\theta}|\mathbf{x})$  posterior distribution.  
 $\pi(\boldsymbol{\theta}^*|\boldsymbol{\theta})$  proposal distribution.  
 $\alpha(\boldsymbol{\theta}^h, \boldsymbol{\theta}^*)$  Metropolis-Hastings acceptance probability.  
 $\sigma_{\boldsymbol{\theta}}^2$  proposal variance.  
 $\delta_{\boldsymbol{\theta}}$  proposal step length.

## Acronyms and Abbreviations

- BBS Breeding Bird Survey.  
BTO British Trust for Ornithology.  
CBC Common Bird Census.  
CES Constant Effort Sites scheme.  
MCMC Markov chain Monte Carlo.  
NRS Nest Record Scheme.  
NRR National Ring-Recovery programme.  
RSPB Royal Society for the Protection of Birds.

---

## References

- Alker, P. J. and Redfern, C. P. F. (1996). Double brooding and polygyny in Sedge Warblers *Acrocephalus schoenobaenus* breeding in north-west England. *Bird Study*, 43(3):356–363.
- Amstrup, S. C., McDonald, T. L., and Manly, B. F. J., editors (2005). *Handbook of Capture-Recapture Analysis*. Princeton University Press, New Jersey, U.S.A.
- Atkinson, P. W., Austin, G. E., Rehfisch, M. M., Baker, H., Cranswick, P., Kershaw, M., Robinson, J., Langston, R. H. W., Stroud, D. A., Van Turnhout, C., and Maclean, I. M. D. (2006). Identifying declines in waterbirds: the effects of missing data, population variability and count period on the interpretation of long-term survey data. *Biological Conservation*, 130(4):549–559.
- BADC Website (2009). The British Atmospheric Data Centre, Chilton, United Kingdom. <http://www.badc.nerc.ac.uk>.
- Baillie, S. R. (1990). Integrated population monitoring of breeding birds in Britain and Ireland. *Ibis*, 132(2):151–166.
- Baillie, S. R., Brooks, S. P., King, R., and Thomas, L. (2009a). Using state-space model of the British Song Thrush *Turdus philomelos* population to diagnose the causes of a population decline. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling Demographic Processes in Marked Populations*, volume 3 of *Environmental and Ecological Statistics*, pages 541–561. Springer, New York, U.S.A.
- Baillie, S. R. and Green, R. E. (1987). The importance of variation in recovery rates when estimating survival rates from ring recoveries. *Acta Ornithologica*, 23(1):41–60.
- Baillie, S. R., Green, R. E., Boddy, M., and Buckland, S. T. (1986). An evaluation of the Constant Effort Sites scheme. British Trust for Ornithology, Thetford, United Kingdom. Research report 21.
- Baillie, S. R., Marchant, J. H., Crick, H. Q. P., Noble, D. G., Balmer, D. E., Barimore, C., Coombes, R. H., Downie, I. S., Freeman, S. N., Joys, A. C., Leech, D. I., Raven, M. J., Robinson, R., and Thewlis, R. M. (2007). Breeding birds in the wider countryside: their conservation status 2007. Technical

- report, British Trust for Ornithology, Thetford, United Kingdom. Research report 487. <http://www.bto.org/birdtrends2007>.
- Baillie, S. R., Marchant, J. H., Leech, D. I., Joys, A. C., Noble, D. G., Barimore, C., Grantham, M. J., Risely, K., and Robinson, R. A. (2009b). Breeding birds in the wider countryside: their conservation status 2008. Technical report, British Trust for Ornithology, Thetford, United Kingdom. Research report 516. <http://www.bto.org/birdtrends>.
- Bensch, S. and Nielsen, B. (1999). Autumn migration speed of juvenile Reed and Sedge Warblers in relation to date and fat loads. *The Condor*, 101(1):153–156.
- Besbeas, P., Borysiewicz, R. S., and Morgan, B. J. T. (2009). Completing the ecological jigsaw. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling Demographic Processes in Marked Populations*, volume 3 of *Environmental and Ecological Statistics*, pages 513–539. Springer, New York, U.S.A.
- Besbeas, P. and Freeman, S. N. (2006). Methods for joint inference from panel survey and demographic data. *Ecology*, 87(5):1138–1145.
- Besbeas, P., Freeman, S. N., and Morgan, B. J. T. (2005). The potential of integrated population modelling. *Australian & New Zealand Journal of Statistics*, 47(1):35–48.
- Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002). Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58(3):540–547.
- Besbeas, P., Lebreton, J. D., and Morgan, B. J. T. (2003). The efficient integration of abundance and demographic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 52(1):95–102.
- Bibby, C. J. (1978). Some breeding statistics of Reed and Sedge Warblers. *Bird Study*, 25:207–222.
- Bibby, C. J. and Green, R. E. (1981). Autumn migration strategies of Reed and Sedge Warblers. *Ornis Scandinavica*, 12(1):1–12.
- Borysiewicz, R. S. (2008). *The combined analysis of multi-site mark-recapture-recovery and census data*. PhD thesis, Institute of Mathematics, Statistics and Actuarial Science, University of Kent, United Kingdom.

- Borysiewicz, R. S., Morgan, B. J. T., Hénaux, V., Bregnballe, T., and Lebreton, J.-D. (2009). An integrated analysis of multisite recruitment, mark-recapture-recovery and multisite census data. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling Demographic Processes in Marked Populations*, volume 3 of *Environmental and Ecological Statistics*, pages 579–591. Springer, New York, U.S.A.
- Brooks, S. P. (1998). Markov chain Monte Carlo methods and its application. *The Statistician*, 47(1):69–100.
- Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2000a). Bayesian animal survival estimation. *Statistical Science*, 15(4):357–376.
- Brooks, S. P., Catchpole, E. A., Morgan, B. J. T., and Barry, S. C. (2000b). On the Bayesian analysis of ring-recovery data. *Biometrics*, 56(3):951–956.
- Brooks, S. P., Freeman, S. N., Greenwood, J. J. D., King, R., and Mazzetta, C. (2008). Quantifying conservation concern - Bayesian statistics, birds and the red lists. *Biological Conservation*, 141(5):1436–1441.
- Brooks, S. P. and Gelman, A. (1998). Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.
- Brooks, S. P., King, R., and Morgan, B. J. T. (2004). A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation*, 27(1):515–529.
- Brownie, C., Anderson, D. R., Burnham, K. P., and Robson, D. R. (1985). *Statistical Inference from Band Recovery Data: A Handbook*. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C., U.S.A., 2nd edition.
- Brownie, C. and Robson, D. S. (1983). Estimation of time-specific survival rates from tag-resighting samples: a generalization of the Jolly-Seber model. *Biometrics*, 39(2):437–453.
- BTO Website (2009). British Trust for Ornithology, Thetford, United Kingdom. <http://www.bto.org>.
- Bub, H. (1991). *Bird trapping and bird banding: a handbook for trapping methods all over the world*. Cornell University Press, Ithaca, New York, U.S.A. Translated by Hamerstrom, F. and Wuertz-Schaefer, K.

- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2004a). *Advanced Distance Sampling: estimating abundance of biological populations*. Oxford University Press, United Kingdom.
- Buckland, S. T., Newman, K. B., Thomas, L., and Koesters, N. B. (2004b). State-space models for the dynamics of wild animal populations. *Ecological Modelling*, 171(1-2):157–175.
- Burnham, K. P. (1990). Survival analysis of recovery data from birds ringed as young: efficiency of analysis when numbers ringed are not known. *The Ring*, 13(1-2):115–132.
- Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., and Pollock, K. H. (1987). *Design and analysis methods for fish survival experiments based on release-recapture*. American Fisheries Society, Monograph 5, Bethesda, Maryland, U.S.A.
- Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, U.S.A., Texts in Statistical Science, 2nd edition.
- Carothers, A. D. (1979). Quantifying unequal catchability and its effect on survival estimates in an actual population. *Journal of Animal Ecology*, 48(3):863–869.
- Caswell, H. (2001). *Matrix population models: construction, analysis, and interpretation*. Sunderland, MA: Sinauer, 2nd edition.
- Catchpole, C. K. (1974). Habitat selection and breeding success in the Reed Warbler (*Acrocephalus scirpaceus*). *Journal of Animal Ecology*, 43(2):363–380.
- Catchpole, E. A., Freeman, S. N., and Morgan, B. J. T. (1995). Modelling age variation in survival and reporting rates for recovery models. *Journal of Applied Statistics*, 22(5-6):597–610.
- Cave, V. M., Freeman, S. N., Brooks, S. P., King, R., and Balmer, D. E. (2009a). On adjusting for missed visits in the indexing of abundance from constant effort ringing. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling Demographic Processes in Marked Populations*, volume 3 of *Environmental and Ecological Statistics*, pages 949–963. Springer, New York, U.S.A.

- Cave, V. M., King, R., and Freeman, S. N. (2009b). An integrated population model from constant effort bird-ringing data. *Journal of Agricultural, Biological and Environmental Statistics, in press(-)*:–.
- Chamberlain, D. E., Fuller, R. J., Bunce, R. G. H., Duckworth, J. C., and Shrubbs, M. (2000). Changes in the abundance of farmland birds in relation to the timing of agricultural intensification in England and Wales. *Journal of Applied Ecology*, 37(5):771–788.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335.
- Cilimburg, A. B., Lindberg, M. S., Tewksbury, J. J., and Hejl, S. J. (2002). Effects of dispersal on survival probability of adult Yellow Warblers (*Dendroica petechia*). *The Auk*, 119(3):778–789.
- Clark, J. A., Robinson, R. A., Adams, S. Y., Grantham, M. J., Risely, K., Balmer, D. E., Blackburn, J. R., Griffin, B. M., Marchant, J. H., and Kimmel, V. (2007). Bird ringing in Britain and Ireland in 2005. *Ringling & Migration*, 23(3):156–192.
- Clavel, J., Robert, A., Devictor, V., and Julliard, R. (2008). Abundance estimation with a transient model under the Robust design. *Journal of Wildlife Management*, 72(5):1203–1210.
- Coiffait, L., Clark, J. A., Robinson, R. A., Blackburn, J. R., Grantham, M. J., Marchant, J. H., Barber, L., De Palacio, D., Griffin, B. M., and Moss, D. (2008). Bird ringing in Britain and Ireland in 2007. *Ringling & Migration*, 24(1):104–144.
- Cormack, R. M. (1964). Estimates of survival from the sightings of marked animals. *Biometrika*, 51(3-4):429–438.
- Couzens, D. (2006). *Secret lives of British birds*. A & C Black, United Kingdom.
- Cowles, M. K. and Carlin, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434):883–904.
- Cramp, S., editor (1992). *Handbook of the birds of Europe the Middle East and North Africa: The birds of the Western Palearctic, volume VI*. Oxford University Press, United Kingdom.

- Crick, H. Q. P. and Baillie, S. R. (1996). A review of the BTO's Nest Record Scheme. Its value to the Joint Nature Conservation Committee and Country Agencies, and its methodology. British Trust for Ornithology, Thetford, United Kingdom. Research report 159.
- Crick, H. Q. P., Baillie, S. R., and Leech, D. I. (2003). The UK Nest Record Scheme: its value for science and conservation. *Bird Study*, 50(3):254–270.
- Crick, H. Q. P., Dudley, C., Glue, D. E., and Thomson, D. L. (1997). UK birds are laying eggs earlier. *Nature*, 388(6642):526.
- Crick, H. Q. P. and Sparks, T. H. (1999). Climate change related to egg-laying trends. *Nature*, 399(6735):423–424.
- DEFRA (2005). Environmental stewardship: look after your land and be rewarded. Published on behalf of the Department for Environment, Food and Rural Affairs. Rural Development Service, United Kingdom.
- DeSante, D. F., O'Grady, D. R., and Pyle, P. (1999). Measures of productivity and survival derived from standardized mist-netting are consistent with observed population changes. *Bird Study*, 46(suppl.):178–188.
- Dobson, A. (1990). Survival rates and their relationship to life-history traits in some common British birds. In Power, D. M., editor, *Current Ornithology*, volume 7, pages 115–146. Plenum Press, New York, U.S.A.
- Donald, P. F., Green, R. E., and Heath, M. F. (2001). Agricultural intensification and the collapse of Europe's farmland bird populations. *Proceedings of the Royal Society B: Biological Sciences*, 268(1462):25–29.
- Du Feu, C. and McMeeking, J. (1991). Does constant effort netting estimate juvenile abundance? *Ringing & Migration*, 12(1):118–123.
- Eaton, M. A., Balmer, D., Burton, N., Grice, P. V., Musgrove, A. J., Hearn, R., Hilton, G., Leech, D., Noble, D. G., Ratcliffe, N., Rehfisch, M. M., Whitehead, S., and Wotton, S. (2008). The state of the UK's birds 2007. RSPB, BTO, WWT, CCW, EHS, NE and SNH, Sandy, Bedfordshire, United Kingdom.
- Eaton, M. A., Brown, A. F., Nobel, D. G., Musgrove, A. J., Hearn, R., Aebischer, N. J., Gibbons, D. W., Evans, A., and Gregory, R. D. (2009). Birds of conservation concern 3: the population status of birds in the United Kingdom, Channel Islands and the Isle of Man. *British Birds*, 102:296–341.

- EURING Website (2008). European Union for Bird Ringing. <http://www.euring.org>.
- Evans, A. D. and Smith, K. W. (1994). Habitat selection of Cirl Buntings *Emberiza cirrus* wintering in Britain. *Bird Study*, 41(2):81–87.
- Fewster, R. M., Buckland, S. T., Siriwardena, G. M., Baillie, S. R., and Wilson, J. D. (2000). Analysis of population trends for farmland birds using Generalized Additive models. *Ecology*, 81(7):1970–1984.
- Freeman, S. N. (2008). Towards a means of estimating trends in survival for common passerines based on Constant Effort Sites (CES) data. Unpublished report, British Trust for Ornithology, Thetford, United Kingdom.
- Freeman, S. N., Balmer, D. E., and Robinson, R. A. (2005). Survival rates from CES ringing. *BTO News 256*, British Trust for Ornithology, Thetford, United Kingdom.
- Freeman, S. N. and Crick, H. Q. P. (2003). The decline of the Spotted Flycatcher *Muscicapa striata* in the UK: an integrated model. *Ibis*, 145(3):400–412.
- Freeman, S. N. and Morgan, B. J. T. (1992). A modelling strategy for recovery data from birds ringed as nestlings. *Biometrics*, 48(1):217–235.
- Freeman, S. N., Noble, D. G., Newson, S. E., and Baillie, S. R. (2007a). Modelling population changes using data from different surveys: the Common Birds Census and the Breeding Bird Survey. *Bird Study*, 54(1):61–72.
- Freeman, S. N., Robinson, R. A., Clark, J. A., Griffin, B. M., and Adams, S. Y. (2007b). Changing demography and population decline in the common Starling *Sturnus vulgaris*: a multi-site approach to Integrated Population monitoring. *Ibis*, 149(3):587–596.
- Freeman, S. N., Wernham, C. V., and Balmer, D. E. (2001). Long-term changes in the productivity of common songbirds in Britain and Ireland from constant effort ringing. Unpublished report, British Trust for Ornithology, Thetford, United Kingdom.
- Fuller, R. J., Gregory, R. D., Gibbons, D. W., Marchant, J. H., Wilson, J. D., Baillie, S. R., and Carter, N. (1995). Population declines and range contractions among lowland farmland birds in Britain. *Conservation Biology*, 9(6):1425–1441.

- Fuller, R. J., Marchant, J. H., and Morgan, R. A. (1985). How representative of agricultural practice in Britain are Common Birds Census farmland plots? *Bird Study*, 32(1):56–70.
- Furness, R. W. and Greenwood, J. J. D. (1993). *Birds as monitors of environmental change*. Chapman & Hall: London, United Kingdom.
- Gauthier, G., Besbeas, P., Lebreton, J.-D., and Morgan, B. J. T. (2007). Population growth in Snow Geese: a modeling approach integrating demographic and survey information. *Ecology*, 88(6):1420–1429.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2004). *Bayesian Data Analysis*. Chapman & Hall, New York, U.S.A., 2nd edition.
- Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian Statistics*, volume 5, pages 599–607. Oxford University Press, United Kingdom.
- Ghosh, J., Delampady, M., and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*. Springer Texts in Statistics, New York, U.S.A.
- Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC Interdisciplinary Statistics Series, London, United Kingdom.
- Gimenez, O., Bonner, S. J., King, R., Parker, R. A., Brooks, S. P., Jamieson, L. E., Grosbois, V., Morgan, B. J. T., and Thomas, L. (2009). WinBUGS for population ecologists: Bayesian modeling using Markov chain Monte Carlo methods. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling Demographic Processes in Marked Populations*, volume 3 of *Environmental and Ecological Statistics*, pages 883–915. Springer, New York, U.S.A.
- Grantham, M. and Robinson, R. A. (2008). *CES News 21*. British Trust for Ornithology, Thetford, United Kingdom.
- Graveland, J. (1999). Effects of reed cutting on density and breeding success of Reed Warbler *Acrocephalus scirpaceus* and Sedge Warbler *A. schoenobaenus*. *Journal of Avian Biology*, 30(4):469–482.

- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.
- Greenwood, J. D. D. (2003). The monitoring of British breeding birds: a success story for conservation science? *The Science of the Total Environment*, 310(1-3):221–230.
- Gregory, R. D., Noble, D. G., and Custance, J. (2004). The state of play of farmland birds: population trends and conservation status of lowland farmland birds in the United Kingdom. *Ibis*, 146(Suppl. 2):1–13.
- Harrison, N. M., Whitehouse, M. J., Prince, P. A., and Huin, N. (2000). What problems do local habitat change represent for Constant Effort Site ringing scheme? *Ringling & Migration*, 20(1):1–8.
- Hilton, Jr., B. and Miller, M. W. (2003). Annual survival and recruitment in a Ruby-throated Hummingbird population, excluding the effect of transient individuals. *The Condor*, 105(1):54–62.
- Hines, J. E., Kendall, W. L., and Nichols, J. D. (2003). On the use of the Robust design with transient capture-recapture models. *The Auk*, 120(4):1151–1158.
- Hume, R. (2003). *RSPB Complete Birds of Britain and Europe*. Dorling Kindersley, London, United Kingdom.
- Jamieson, L. E. and Brooks, S. P. (2004). Density dependence in north American ducks. *Animal Biodiversity and Conservation*, 27(1):113–128.
- Julliard, R., Jiguet, F., and Couvet, D. (2004). Evidence for the impact of global warming on the long-term population dynamics of common birds. *Proceedings of the Royal Society B: Biological Sciences*, 271(Suppl. 6):490–492.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.
- Kendall, W. L. (2001). The Robust design for capture-recapture analysis using Program MARK. In Field, R., Warren, R. J., Okarma, H., and Sievert, P. R., editors, *Wildlife, land, and people: priorities for the 21st century. Proceedings of the Second International Wildlife Management Congress*, pages 357–360. The Wildlife Society, Bethesda, Maryland, U.S.A.
- Kendall, W. L. and Bjorkland, R. (2001). Using open Robust design models to estimate temporary emigration from capture-recapture data. *Biometrics*, 57(4):1113–1122.

- Kendall, W. L., Nichols, J. D., and Hines, J. E. (1997). Estimating temporary emigration using capture-recapture data with Pollock's Robust design. *Ecology*, 78(2):563–578.
- Kendall, W. L., Pollock, K. H., and Brownie, C. (1995). A likelihood-based approach to capture-recapture estimation of demographic parameters under the Robust design. *Biometrics*, 51(1):293–308.
- King, R. and Brooks, S. P. (2004). Bayesian analysis of the Hector's dolphin data. *Animal Biodiversity and Conservation*, 27(1):343–354.
- King, R. and Brooks, S. P. (2008). On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics*, 64(3):816–824.
- King, R., Brooks, S. P., and Coulson, T. (2008a). Analysing complex capture-recapture data: incorporating time-varying covariate information. *Biometrics*, 64(4):1187–1195.
- King, R., Brooks, S. P., Mazzetta, C., Freeman, S. N., and Morgan, B. J. T. (2008b). Identifying and diagnosing population declines: a Bayesian assessment of Lapwings in the UK. *Journal of Royal Statistical Society C: Applied Statistics*, 57(5):609–632.
- King, R., Morgan, B. J. T., Gimenez, O., and Brooks, S. P. (2009). *Bayesian Analysis for Population Ecology*. Chapman & Hall/CRC Interdisciplinary Statistics Series, New York, U.S.A.
- Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, 62(1):67–118.
- Lindberg, M. S., Kendall, W. L., Hines, J. E., and Anderson, M. G. (2001). Combining band recovery data and Pollock's Robust design to model temporary and permanent emigration. *Biometrics*, 57(1):273–281.
- Link, W. A. and Sauer, J. R. (2002). A hierarchical analysis of population change with application to Cerulean Warblers. *Ecology*, 83(10):2832–2840.
- Loery, G., Nichols, J. D., and Hines, J. E. (1997). Capture-recapture analysis of a wintering Black-capped chickadee population in Connecticut, 1958-1993. *The Auk*, 114(3):431–442.

- Mazzetta, C. (2009). Age-specificity in conditional ring-recovery models. *Under Revision for Journal of Agricultural, Biological and Environmental Statistics*.
- McCarthy, M. A. (2007). *Bayesian Methods for Ecology*. Cambridge University Press, United Kingdom.
- Met Office Website (2009). Met Office, Exeter, United Kingdom. <http://www.metoffice.gov.uk>.
- Miles, W. (2005). The effect of missed Constant Effort Sites mist-netting visits on measures of productivity. Master's thesis, Anglia Polytechnic University, United Kingdom.
- Miles, W., Freeman, S. N., Harrison, N. M., and Balmer, D. E. (2007). Measuring passerine productivity using constant effort sites: the effect of missed visits. *Ringing & Migration*, 23(4):231–237.
- Newton, I. (2004). The recent declines of farmland bird populations in Britain: an appraisal of causal factors and conservation actions. *Ibis*, 146(4):576–600.
- North, P. M. (1990). Analysis of avian ring-recovery and live recapture data: where have we come from Wageningen to Sempach? *The Ring*, 13(1-2):11–22.
- North, P. M. and Morgan, B. J. T. (1979). Modelling Heron survival using weather data. *Biometrics*, 35(3):667–681.
- O' Hagan, A. (1998). Eliciting expert beliefs in substantial practical applications. *The Statistician*, 47(1):21–35.
- Parker, D. E., Legg, T. P., and Folland, C. K. (1992). A new daily central England temperature series, 1772-1991. *International Journal of Climatology*, 12(4):217–342.
- Peach, W. J. (1993). Combining mark-recapture data sets for small passerines. In Lebreton, J.-D. and North, P. M., editors, *Marked Individuals in the Study of Bird Population*, pages 107–121. Birkhäuser, Verlag, Basel, Switzerland.
- Peach, W. J. (2002). Sedge Warbler. In Wernham, C. V., Toms, M. P., Marchant, J. H., Clark, J. A., Siriwardena, G. M., and Baillie, S. R., editors, *The Migration Atlas: movements of the birds of Britain and Ireland*, pages 544–547. T. & A.D. Poyser, London, United Kingdom.

- Peach, W. J., Baillie, S. R., and Balmer, D. E. (1998). Long-term changes in the abundance of passerines in Britain and Ireland as measured by constant effort mist-netting. *Bird Study*, 45(3):257–275.
- Peach, W. J., Baillie, S. R., and Underhill, L. G. (1991). Survival of British Sedge Warblers *Acrocephalus schoenobaenus* in relation to west African rainfall. *Ibis*, 133(3):300–305.
- Peach, W. J., Buckland, S. T., and Baillie, S. R. (1990). Estimating survival rates using mark-recapture data from multiple sites. *Ring*, 13(1-2):87–102.
- Peach, W. J., Buckland, S. T., and Baillie, S. R. (1996). The use of constant effort mist-netting to measure between-year changes in the abundance and the productivity of common passerines. *Bird Study*, 43(2):142–156.
- Peach, W. J., Crick, H. Q. P., and Marchant, J. H. (1995a). The demography of the decline in British Willow Warbler population. *Journal of Applied Statistics*, 22(5-6):905–922.
- Peach, W. J., Du Feu, C., and McMeeking, J. (1995b). Site tenacity and survival rates of Wrens *Troglodytes troglodytes* and Treecreepers *Certhia familiaris* in a Nottinghamshire wood. *Ibis*, 137(4):497–507.
- Peach, W. J., Siriwardena, G. M., and Gregory, R. D. (1999). Long-term changes in over-winter survival rates explain the decline of Reed Buntings *Emberiza schoeniclus* in Britain. *Journal of Applied Ecology*, 36(5):798–811.
- Pledger, S., Pollock, K. H., and Norris, J. L. (2003). Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, 59(4):786–794.
- Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, 46(3):752–757.
- Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Society Monographs*, 107.
- Poole, D. (2002). Bayesian estimation of survival from mark-recapture data. *Journal of Agricultural, Biological, and Environmental Statistics*, 7(2):264–276.

- Pradel, R., Hines, J. E., Lebreton, J.-D., and Nichols, J. D. (1997). Capture-recapture survival models taking account of transients. *Biometrics*, 55(1):60–72.
- Pratt, A. and Peach, W. (1991). Site tenacity and annual survival of a Willow Warbler *Phylloscopus trochilus* population in southern England. *Ringing & Migration*, 12(3):128–134.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.
- Redfern, C. and Alker, P. (2002). Reed Warbler. In Wernham, C. V., Toms, M. P., Marchant, J. H., Clark, J. A., Siriwardena, G. M., and Baillie, S. R., editors, *The Migration Atlas: movements of the birds of Britain and Ireland*, pages 548–551. T. & A.D. Poyser, London, United Kingdom.
- Reynolds, T. J., King, R., Harwood, J., Frederiksen, M., Harris, M. P., and Wanless, S. (2009). Integrated data analyses in the presence of emigration and tag-loss. *Journal of Agricultural, Biological and Environmental Statistics*, in press(-):-.
- Robinson, R. A. (2005). BirdFacts: profiles of birds occurring in Britain & Ireland (v1.22, Oct 2008). British Trust for Ornithology, Thetford, United Kingdom. Research report 407. <http://www.bto.org/birdfacts>.
- Robinson, R. A., Freeman, S. N., Balmer, D. E., and Grantham, M. J. (2007). Cetti's Warbler *Cettia cetti*: analysis of an expanding population. *Bird Study*, 54(2):230–235.
- Robinson, R. A., Green, R. E., Baillie, S. R., Peach, W. J., and Thomson, D. L. (2004). Demographic mechanism of the population decline of the Song Thrush *Turdus philomelos* in Britain. *Journal of Animal Ecology*, 73(4):670–682.
- RSPB Website (2009). The Royal Society for the Protection of Birds, Sandy, United Kingdom. <http://www.rspb.org.uk>.
- Sandercock, B. and Jaramillo, A. (2002). Annual survival rates of Wintering Sparrows: assessing demographic consequences of migration. *The Auk*, 119(1):149–165.

- Saracco, J. F., DeSante, D. F., and Kaschube, D. R. (2008). Assessing landbird monitoring programs and demographic causes of population trends. *Journal of Wildlife Management*, 72(8):1665–1673.
- SAS Institute Inc. (2002-2008). *SAS OnlineDoc<sup>®</sup>9.1.3*. SAS 9.1.3 Help and Documentation, Cary, NC, U.S.A. <http://support.sas.com/onlinedoc/913>.
- Sauer, J. R., Hines, J. E., Gough, G., Thomas, I., and Peterjohn, B. G. (1997). The North American Breeding Bird Survey results and analysis. Version 96.4. Patuxent Wildlife Research Center, Laurel, U.S.A.
- Schaub, M., Gimenez, O., Sierro, A., and Artellaz, R. (2007). Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, 21(4):945–955.
- Schwarz, C. J. and Stobo, W. T. (1997). Estimating temporary migration using the Robust design. *Biometrics*, 53(1):178–194.
- Seber, G. A. F. (1971). Estimating age-specific survival rates from bird-band returns when the reporting rate is constant. *Biometrika*, 58(3):491–497.
- Silkey, M., Nur, N., and Geupel, G. R. (1999). The use of mist-net capture rate to monitor annual variation in abundance: a validation study. *The Condor*, 101(2):288–298.
- Simms, E. (1985). *British Warblers*. Collins, London, United Kingdom.
- Siriwardena, G. M., Baillie, S. R., Buckland, S. T., Fewster, R. M., Marchant, J. H., and Wilson, J. D. (1998a). Trends in the abundance of farmland birds: a quantitative comparison of smoothed Common Birds Census indices. *Journal of Applied Ecology*, 35(1):24–43.
- Siriwardena, G. M., Baillie, S. R., and Wilson, J. D. (1998b). Variation in the survival rates of some British passerines with respect to their population trends on farmland. *Bird Study*, 45(3):276–292.
- Siriwardena, G. M., Baillie, S. R., and Wilson, J. D. (1999). Temporal variation in the annual survival rates of six granivorous birds with contrasting population trends. *Ibis*, 141(4):621–636.
- Siriwardena, G. M., Freeman, S. N., and Crick, H. Q. P. (2001). The decline of the Bullfinch *Pyrrhula pyrrhula* in Britain: is the mechanism known? *Acta Ornithologica*, 36(2):143–152.

- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). *WinBUGS User Manual, Version 1.4*. MRC Biostatistics Unit, Cambridge, United Kingdom.
- Svensson, L. (1992). Identification guide to European passerines. British Trust for Ornithology, Thetford, United Kingdom.
- ter Braak, C. J. F., Van Strien, A. J., Meijer, R., and Verstrael, T. J. (1994). Analysis of monitoring data with many missing values: which method? In Hagemeyer, E. J. M. and Verstrael, T. J., editors, *Distribution, Monitoring and Ecological Aspects., Bird Numbers 1992*, pages 663–673. Proceedings of the 12th International Conference, International Bird Census Council and European Ornithological Atlas Committee. Beek-Ubbergen, The Netherlands, Sovon.
- Thaxter, C. B., Redfern, C. P. F., and Bevan, R. M. (2006). Survival rates of adult Reed Warblers *Acrocephalus scirpaceus* at a northern and southern site in England. *Ringing & Migration*, 23(2):65–79.
- Thomson, D. L., Baillie, S. R., and Peach, W. J. (1999). A method for studying post-fledging survival rates using data from ringing recoveries. *Bird Study*, 46(Suppl.):104–111.
- UK Biodiversity Partnership Standing Committee (2006). Future direction of UK biodiversity. Joint Nature Conservation Committee on the UK Biodiversity Action Plan and Country Biodiversity Strategies, United Kingdom. Report UKBPSC 06-03.
- Underhill, L. G. and Prÿs-Jones, R. P. (1994). Index numbers for waterbird population. I. Review and methodology. *Journal of Applied Ecology*, 31(3):463–480.
- Van Strien, A. J., Pannekoek, J., and Gibbons, D. W. (2001). Indexing European bird population trends using results of national monitoring schemes: a trial of a new method. *Bird Study*, 48(2):200–213.
- Véran, S. and Lebreton, J.-D. (2008). The potential of integrated modelling in conservation biology: a case study of the black-footed albatross (*Phoebastria nigripes*). *The Canadian Journal of Statistics*, 36(1):85–98.
- Vounatsou, P. and Smith, A. F. M. (1995). Bayesian analysis of ring-recovery data via Markov Chain Monte Carlo simulation. *Biometrics*, 51(2):687–708.

- Wernham, C. V., Toms, M. P., Marchant, J. H., Clark, J. A., Siriwardena, G. M., and Baillie, S. R., editors (2002). *The Migration Atlas: movements of the birds of Britain and Ireland*. T. & A.D. Poyser, London, United Kingdom.
- White, G. C. and Burnham, K. P. (1999). Program MARK: survival estimation from populations of marked animals. *Bird Study*, 46(Suppl.):120–139.
- Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002). *Analysis and management of animal populations*. San Diego: Academic Press, U.S.A.