# EVALUATION OF STATISTICAL METHODS FOR ESTIMATING LONG-TERM POPULATION CHANGE FROM EXTENSIVE WILDLIFE SURVEYS 

by<br>Len Thomas<br>B.Sc., The University of Sheffield, U.K., 1990<br>M.Sc., The University of York, U.K., 1991<br>A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF<br>Doctor of Philosophy<br>in<br>THE FACULTY OF GRADUATE STUDIES<br>Department of Forestry and<br>Centre for Applied Conservation Biology<br>We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
September 1997
OLen Thomas, 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of FORESTRy
The University of British Columbia
Vancouver, Canada
Date 19 dept 1997


#### Abstract

Monitoring long-term changes in population abundance is an integral part of conservation-oriented research and management. Many extensive monitoring programs are based on annual counts at a large number of permanent survey sites. However, estimating population change from the resulting data is challenging and controversial. At least eight statistical methods have been proposed and used, and in this thesis I evaluated these methods using data from the North American Breeding Bird Survey (BBS).

I began by demonstrating the importance of analysis method for the inferences drawn about species trends. I compared the results when three methods of trend analysis were applied to BBS data for 115 species in British Columbia, and found differences in the direction, magnitude and statistical significance of the estimates.

I then explored the issues surrounding the analysis of these data. I showed that even under ideal survey conditions subjective decisions must be made about whether to estimate annual indices or population trends and, for trend methods, which model to use for trend. In reality the analysis is further complicated by measurement error and missing data. I suggested the use of simulation studies to determine the accuracy of the methods, and provided a simple example using three geographic scenarios of population decline.

I next made a preliminary assessment of the reliability of eight candidate analysis methods by applying them to continental data for four case study species (Cerulean Warbler, Carolina Wren, Chestnut-collared Longspur, and Belted Kingfisher). All methods except one (Mountford moving windows) produced seemingly plausible results in most cases. However each method demonstrated obvious flaws in some cases, and I used these to suggest potential areas of further methodological development.

I concluded that loglinear modeling is currently the best method for estimating annual indices. For trend analysis, I could not determine whether route regression or direct estimation (based on generalized modeling) is generally superior: Large, long-term population changes in well-surveyed species are almost certain to be detected by any of these methods. However, their accuracy and robustness remains uncertain, and further research is required both to develop the methods and to evaluate them through comprehensive simulation studies.


## Table of Contents

Abstract ..... ii
List of Tables ..... vi
List of Figures ..... viii
List of Abbreviations .....  x
Acknowledgements ..... xi
Foreword ..... xii
Chapter 1 General Introduction
Introduction .....  .1
The Breeding Bird Survey ..... 2
Methods of Estimating Long-term Population Change ..... 3
Chain Index Methods ..... 3
Linear Route Regression ..... 5
Nonlinear Route Regression ..... 6
Linear Modeling ..... 7
Overview of Thesis ..... 8
Figures

$\qquad$
Chapter 2 The Importance of Analysis Method for Population Trend EstimatesIntroduction.14
Methods ..... 15
Breeding Bird Survey ..... 15
Data Selection ..... 15
Trend Estimation ..... 16
Comparison of Trend Estimates. ..... 19
Results ..... 20
Discussion ..... 22
Causes of the Differences ..... 22
Implications of the Results ..... 25
Priorities for Further Work ..... 26
Summary ..... 27
Tables ..... 29
Figures
Chapter 3 Issues in the Estimation of Long-term Population Change
Introduction ..... 36
Analysis of Survey Data Under Ideal Conditions ..... 36
Selection of the Unit of Analysis ..... 37
Model of Trend ..... 37
Additional Factors the Complicate the Analysis of Survey Data ..... 39
Incomplete Sampling Frame ..... 39
Measurement Error ..... 40
Missing Data ..... 41
Evaluating the Methods Using Monte Carlo Simulations ..... 42
Summary ..... 46
Tables ..... 47
Figures
Chapter 4 Four Case Studies
Introduction ..... 55
Methods ..... 57
Data Selection ..... 57
Data Analysis ..... 58
Comparison of Estimates ..... 62
Results ..... 63
Cerulean Warbler ..... 63
Carolina Wren ..... 68
Chestnut-collared Longspur ..... 71
Belted Kingfisher ..... 73
Summary of Performance of Methods ..... 75
Discussion ..... 76
Linear Route Regression ..... 76
Nonlinear Route Regression ..... 79
Loglinear Modeling ..... 81
Moving-window Mountford's Method ..... 86
Generality of Findings ..... 88
Summary ..... 88
Tables ..... 90
FiguresChapter 5 General Conclusions
Introduction ..... 119
Evaluation of Analysis Methods ..... 119
Annual Index Methods ..... 119
Trend Analysis Methods ..... 121
Future Research ..... 123
Development of Current Methods ..... 123
Further Evaluation of Methods ..... 124
Alternative Approaches. ..... 126
Implications of Research ..... 128
Tables ..... 130
Figure ..... 133
Literature Cited ..... 135
Appendix 1 Computer Programs For Analysis of Population Change ..... 145
Appendix 2 Breeding Bird Survey Data ..... 147

## List of Tables

2.1 Summary of the three Breeding Bird Survey trend estimation methods compared in Chapter 2 ..... 29
2.2 Pairwise comparisons among three Breeding Bird Survey trend estimation methods of the proportion of 115 species with trends in different directions or with different levels of statistical significance. ..... 31
2.3 Estimates of the effect of sample size and species abundance on the proportion of differences in the direction and statistical significance of species trends between three Breeding Bird Survey analysis methods ..... 32
3.1 Methods that have been used to estimate annual indices from bird count data. ..... 47
3.2 Methods that have been used to estimate population trends from bird count data ..... 48
3.3 Parameters used in the simulation study of the performance of three Breeding Bird Survey trend-estimation methods in the context of overall population declines and under three different geographic scenarios of population change. ..... 49
3.4 Bias and precision of two trend estimators (U.S. National Biological Service and Canadian Wildlife Service linear route-regression) from 100 simulations of Breeding Bird Survey data at each of three geographic scenarios of population decline. ..... 50
4.1 List of analysis methods used in Chapter 4, with their associated data selection criteria. ..... 90
4.2 Summary of the Breeding Bird Survey data used in this chapter ..... 91
4.3 Estimates of overall population trend for the Cerulean Warbler between 1966 and 1995 (data selected using stringent data selection criteria). ..... 92
4.4 Proportion of Breeding Bird Survey routes with estimated declines in abundance for four species using six route-regression trend estimators. ..... 93
4.5 Comparison of loglinear models for the analysis of population trends in Cerulean Warbler using Breeding Bird Survey data from 1966-1995 (data selected using stringent data selection criteria) ..... 94
4.6 Goodness-of-fit tests for Mountford moving windows applied to Breeding Bird Survey data for Cerulean Warbler 1966-1995. ..... 95
4.7 Estimates of overall population trend for the Carolina Wren between 1967 and 1995 (data selected using stringent data selection criteria). ..... 96
4.8 Comparison of loglinear models for the analysis of population trends in the Carolina Wren using Breeding Bird Survey data from 1967-1995 ..... 97
4.9 Estimates of overall population trend for the Chestnut-collared Longspur between 1967 and 1995 (data selected using stringent data selection criteria). ..... 98
4.10 Estimates of overall population trend for the Belted Kingfisher between 1968 and 1995 (data selected using stringent data selection criteria). ..... 99
4.11 Comparison of loglinear models for the analysis of population trends in the Belted Kingfisher using Breeding Bird Survey data from 1968-1995. ..... 100
4.12 Summary of problems noted in case studies. ..... 101
4.13 Effect of incorporating observer differences on trend estimates for four species from three analyses. ..... 102
5.1 Summary evaluation of methods that have been used to estimate annual indices from bird count data. ..... 130
5.2 Summary evaluation of methods that have been used to estimate trends from bird count data. ..... 131
A1.1 List of computer programs used in this thesis ..... 146
A2.1 Number of Breeding Bird Survey routes and surveys (1966-1995) in physiographic strata excluded from the case studies of Chapter 4 ..... 148

## List of Figures

1.1 Map showing the location of all North American Breeding Bird Survey routes that have been surveyed one or more times. ..... 11
1.2 Contour map showing the number of annual surveys on Breeding Bird Survey routes between 1966 and 1995. ..... 12
1.3 Number of Breeding Bird Survey surveys performed each year between 1966 and 1995. ..... 13
2.1 Number of species with increasing and decreasing trends from three analysis methods using the same Breeding Bird Survey data for 115 species from British Columbia. ..... 34
2.2 Plot of the magnitude of trends estimated from U.S. National Biological Service route regression against the magnitude of trends from Canadian Wildlife Service route regression for 115 species ..... 35
3.1 Four models of trend applied to the same hypothetical count data. ..... 52
3.2 Three geographic scenarios of population decline. ..... 53
3.3 The direction and statistical significance of trends estimated by three methods of trend analysis from 100 simulations of Breeding Bird Survey data. ..... 54
4.1 Counts of Cerulean Warblers on each Breeding Bird Survey route between 1966 and 1995. ..... 105
4.2 Relationship between measures of variation in counts on a Breeding Bird Survey route and the mean count on that route for four species. ..... 106
4.3 Population indices for Cerulean Warblers between 1966 and 1995, calculated using eight different estimators. ..... 107
4.4 Population indices for the Cerulean Warbler between 1966 and 1995 calculated using model 3 of program TRIM. ..... 108
4.5 Mountford moving-window estimates of population indices for Cerulean Warblers between 1966 and 1996. ..... 109
4.6 Population indices for the Cerulean Warbler between 1966 and 1995, calculated using eight different estimators, with data selected using tolerant data selection criteria. ..... 110
4.7 Population indices for the Carolina Wren between 1967 and 1995, calculated using eight different estimators, with data selected using stringent data selection criteria. ..... 111
4.8 Mountford moving-window estimates of population indices for Carolina Wren between 1966 and 1996 for "without observers" models at four window sizes. ..... 112
4.9 Population indices for the Chestnut-collared Longspur between 1967 and 1995, calculated using eight different estimators, with data selected using stringent data selection criteria. ..... 113
4.10 Population indices for the Chestnut-collared Longspur between 1967 and 1995, calculated using seven different estimators (program TRIM failed to produce log-linear model estimates), with data selected using tolerant data selection criteria. ..... 114
4.11 Population indices for the Belted Kingfisher between 1968 and 1995, calculated using eight different estimators, with data selected using stringent data selection criteria. ..... 115
4.12 Population indices for the Belted Kingfisher between 1968 and 1995, calculated using eight different estimators, with data selected using tolerant data selection criteria. ..... 116
4.13 Relationship between mean abundance and estimated route trend for the linear route regression estimators using tolerantly selected data for Belted Kingfisher between 1968 and 1995. ..... 117
4.14 Residual analysis from loglinear models. ..... 118
5.1 A flow-chart guide to choosing among the analysis methods. ..... 134

## List of Abbreviations

AIC Akaike's Information Criterion
BBS Breeding Bird Survey
CI Confidence interval
CWS Canadian Wildlife Service
GAM Generalized additive modeling
GLM Generalized linear modeling
NNRR Nonlinear nonparametric route regression
NSRR Nonlinear semiparametric route regression
QAIC Quasi-likelihood Akaike's Information Criterion
RR Route regression
SE Standard error
TRIM Trends and indices for monitoring data (a computer program for loglinear modeling)
USNBS U.S. National Biological Service

## Acknowledgements

I thank my research supervisor, Kathy Martin, for her support and encouragement throughout this thesis work, and the many other side projects in which I became entangled over the past six years. I also gratefully acknowledge the help of my supervisory committee, Val LeMay, Jamie Smith, John Smith and Carl Walters, each of whom passed on valuable nuggets of wisdom at various times. Many others helped to make the University of British Columbia (UBC) a lively and stimulating place to study, especially members of the Centre for Applied Conservation Biology and the Centre for Biodiversity Research.

This work would not have been possible without the vision of Chandler Robbins, who originated the Breeding Bird Survey (BBS), and the efforts of the thousands of skilled volunteer observers who collect BBS data each year. In addition, various members of the two federal agencies that co-ordinate the BBS provided support and advice: from the Canadian Wildlife Service I thank Peter Blancher, Brian Collins, Ricky Dunn, Connie Downes and Bob Elner, and from the Biological Resources Division, U.S. Geological Survey, I thank Sam Droege, Bill Link, Bruce Peterjohn and John Sauer. Many of the computer programs I used to analyze BBS data were originally written by others (see Appendix 1), and I am grateful to these people for sharing their code. Financial support for this work came from the Canadian Commonwealth Scholarship and Fellowship Plan, the Natural Sciences and Engineering Research Council of Canada and the Canadian Wildlife Service (Pacific and Yukon Region and National Wildlife Research Centre, Ottawa).

I am grateful for the helpful suggestions made by those who reviewed earlier versions of Chapters 2 and 3 (see Foreword): Brian Collins, Connie Downes, Erica Dunn, John Hagan and John Sauer for Chapter 2, and Brian Collins, Peter Kareiva, David Ward and two anonymous referees for Chapter 3. My ideas for Chapters 4 and 5 were improved by discussions with Steve Buckland, Paul Gustafson, Nancy Heckman and Bent Jorgensen. Tina Schivatcheva, Susan Shirkoff and Pierre Vernier helped me produce Figures 1.1 and 1.2 from GIS data provided by John Sauer.

Although much of this work was done in various offices and cubbyholes at UBC, some of the best insights came while I was away in the mountains. I am therefore indebted to the good friends who dragged me out on many backcountry adventures over the years - you know who you are! Three very special friends must, however, be named. Rachel Holt, your contribution is too important to discuss without a glass of single malt in hand. Wes Hochachka, hi, and thanks for all the hoochery. Last, and most important, I dedicate this thesis to Katrine Vogt.

## Foreword

Parts of this thesis have previously been published. A version of Chapter 2 was published as

Thomas, L., and K. Martin. 1996. The importance of analysis method for Breeding Bird Survey population trend estimates. Conservation Biology 10: 479-490.
and a version of chapter 3 was published as
Thomas, L. 1996. Monitoring long-term population change: why are there so many analysis methods? Ecology 77: 49-58.

The first paper was co-authored with my research supervisor, Dr. Kathy Martin, who played an important part in the early development of the work and provided substantial editorial input. I initiated the work, performed the analysis and drafted the manuscript.

I agree that the contributions of the candidate to our co-authored paper are as stated.
(Dr. K. Martin)

## Chapter 1

## General Introduction

## Introduction

Population monitoring can play a critical role in conservation biology, by providing the information necessary to identify conservation problems at an early stage and suggest possible solutions (Hellawell 1991; James and McCulloch 1995). However, to be effective, population parameters must be measured with sufficient accuracy to determine whether biologically important changes have occurred. Many regional, national and international population monitoring programs are based on a large number of permanent sampling sites that are surveyed at regular intervals using standardized counting techniques. Unfortunately, estimating long-term population changes from the resulting data is not straightforward. A number of statistical methods have been proposed and used but there is currently no consensus as to their relative merits (Geissler and Noon 1981; Sauer and Droege 1990b; Greenwood et al. 1993; ter Braak et al. 1994; Peterjohn et al. 1995; James et al. 1996).

In this thesis I compare the various methods and evaluate them based on theoretical considerations and on their performance when applied to both real and simulated data. I focus on the estimation of change in abundance over time for a single species within a defined sampling area; I do not consider summary analyses that involve groups of species (James et al. 1996; Link and Sauer 1995, 1996) or spatial statistics (Flather and Sauer 1996; Villard and Maurer 1996). The data that I use come from the North American Breeding Bird Survey (BBS), which I describe in the next section. I then briefly introduce the statistical methods that have been used to estimate population change from these data. Lastly, I outline the contents of the rest of this dissertation.

## The Breeding Bird Survey

The BBS is the most important source of information about the status of landbird populations in North America. The survey is operated cooperatively by U.S. National Biological Service ${ }^{1}$ (USNBS) and the Canadian Wildlife Service (CWS), and currently comprises about 370039.2 km long permanent survey routes located along secondary roads within all continental states in the U.S. and provinces and territories in Canada (Fig. 1.1 ${ }^{2}$ ). Routes were originally selected according to a stratified random sampling design, with a predetermined number of routes located in each degree block of latitude and longitude (approx. 112 km by 80 km ; see Robbins and van Velzen 1970 for details of the survey design). Some routes (approx. 200) have since become unusable and have been replaced by newly selected routes. Additional routes have also been added in some states at various times (B. G. Peterjohn, USNBS Patuxent Environmental Science Center, pers. comm.).

The density of routes is relatively uniform within states and the southern parts of the Canadian provinces, but varies between states and provinces from 8 routes per degree block in some eastern states to 0.5 routes per degree block in most western states and provinces (Fig 1.1). Analyses of BBS data therefore often post-stratify by state and province, and many also poststratify by ecologically similar physiographic strata (Fig 1.1; Bystrak 1981).

Surveys are carried out between late May and early July, depending upon latitude.
Starting half an hour before local sunrise and in reasonable weather, a skilled volunteer observer drives the route once, stopping every 0.8 km and recording all the birds seen within a 400 meter radius or heard at any distance in a three minute period (see Robbins et al. 1986 Appendix B for a detailed description of the survey protocol). For each species, the number of birds recorded is summed over the 50 stops to give the survey count on the route in that year.

Approximately $70 \%$ of the 3700 active routes are surveyed in each year, although survey

[^0]coverage varies both spatially and temporally. Routes in some eastern and central states have been surveyed in almost every year, while those in sparsely populated parts of the western states and northern provinces and territories have rarely been visited (Fig. 1.2). Participation in the survey has increased since its inception in 1966 (Fig 1.3).

While there are a number of limitations inherent in the survey methodology (Chapter 2; Robbins and van Velzen 1970; Bystrak 1981; Robbins et al. 1986; Droege 1990; Peterjohn et al. 1995), the BBS is the only systematic surveillance program that covers the breeding ranges of most North American landbird species. Summaries containing species trends have been published regularly by the USNBS and CWS (e.g., Peterjohn et al. 1994; Downes and Collins 1996), as well as more detailed accounts for some species (e.g., Sauer and Droege 1990a, d; Sauer et al. 1994a; Peterjohn and Sauer 1995) and analyses of multispecies patterns (e.g., Robbins et al. 1989; Sauer and Droege 1992). BBS data and trend analyses have been used by third parties for conservation prioritization schemes (e.g., Hunter et al. 1993), for population status reviews (e.g., Terborgh 1989; Askins et al. 1990; Askins 1993), for tests of hypotheses of causation (e.g., Böhning-Gaese et al. 1993; Johnson and Schwartz 1993; Heckert 1997) and as a benchmark in the validation of other surveillance programs (e.g., Holmes \& Sherry 1988; Hagan et al. 1992; Hussell et al. 1992; Witham \& Hunter 1992). Morton (1992: 585) summed up the general feeling among ornithologists by stating: "Population trend determinations by the BBS have become the accepted measure of North American breeding bird populations."

## Methods of estimating long-term population change

## Chain index methods

Missing data are a ubiquitous feature of long-term volunteer-based surveillance programs like the BBS. Because of this, simple averages of the counts in each year are at best inefficient, and will be biased if there is an association between the missing data and bird abundance (Chapter 3; Geissler and Noon 1981; James et al. 1990; ter Braak et al. 1994). The earliest solution to this problem was the use of the chain index method (Bailey 1967; Ogilvie 1967), also known as the ratio method or proportional base-year adjustment. For each pair of adjacent
years, counts from the routes that were surveyed in both years are summed and divided by one another to yield a ratio that estimates the change in abundance over that 2-year period. Successive ratios are "chained" (multiplied) together to produce indices of abundance relative to some "base-year" (arbitrary starting point). Variations on this method have been widely used (e.g., Erskine 1978; Marchant et al. 1990; O’Connor 1992a) but are no longer recommended for two reasons. Firstly, the chaining of successive indices can lead to an accumulation of errors, which produces very imprecise index estimates (Geissler and Noon 1981; Moss 1985). Secondly, ratio estimates can be positively biased (Barker and Sauer 1992; Link et al. 1994; Underhill and Prys-Jones 1994).

An improvement to the chain index method was suggested by Mountford $(1982,1985)$. This method uses not only ratios of counts in adjacent years, but also those from non-adjacent years to fit the model

$$
\text { count }=\text { route effect * year effect }+ \text { error }
$$

where the error terms are random variables with means of zero, variances proportional to the expected count, and decreasing covariances between successive error terms on a route with increasing separation in time. The model is based on the assumption that yearly population changes are constant across all routes (no interaction effect), and this can be checked using a goodness-of-fit test (Mountford 1982). Unfortunately the method appears to be very sensitive to this assumption, which means that only relatively short time periods can be used in each analysis (6-12 years Peach and Baillie 1994; ter Braak et al. 1994; Peterjohn et al. 1995), making it of little use in the analysis of long-term surveys such as the BBS (30 years).

Peach and Baillie (1994) have proposed a "moving windows Mountford method", in which successive overlapping Mountford models are fit to the data, each of which is of a time period short enough to pass the goodness-of-fit test. These successive models are then in effect chained together to produce indices for the entire time period of interest. This ad hoc solution has yet to be rigorously evaluated.

## Linear route regression

The development of linear route regression was begun in response to the problems with chain indexing (Geissler and Noon 1981). Route regression (RR) methods are two-stage estimators in which trends are first estimated for each route and then the route trends are combined into an overall estimate for the region of interest (this is more generally known as multilevel or heirarchical modeling, Goldstein 1995). Linear RR is motivated by the exponential growth model

$$
\begin{equation*}
\text { expected count = route effect * route trend }{ }^{\text {year }} \tag{1}
\end{equation*}
$$

Most published analyses fit this model using linear least-squares regression on log-transformed data, using the equation

$$
\log (\text { count }+ \text { constant })=\log (\text { observer effect })+\log (\text { route trend }) * \text { year }+ \text { error }
$$

where the error terms are independent random variables with identical Gaussian distributions and zero means (Geissler and Sauer 1990). The inclusion of separate intercepts for each observer on a route allows differences in detection efficiency between observers to be taken into account; in the chain index methods the only way to do this is to treat each observer as a different route. Regional trends are calculated as the weighted mean of the route trends, using weightings for the area represented by the route, the relative abundance of the species on the route, and the expected precision of the trend estimate. The merit of the weightings, and the appropriate method of calculating them (especially the abundance weighting), is controversial (James et al. 1990; Peterjohn et al. 1994). Variations of this method have been used by the USNBS and CWS to analyze BBS data since 1986 (e.g., Peterjohn and Sauer 1993; Downes and Collins 1996).

One important disadvantage of using least-squares regression on log-transformed counts is that a constant must be added to the counts before transformation, because the log of a zero count is undefined. This biases the trend estimates towards zero (i.e., no change, Geissler and Link 1988; Collins 1990; Link and Sauer 1994). Another problem is that trends are estimated on the log scale and back-transformation to the arithmetic scale is problematic (Geissler and Link 1988). To bypass these problems Link and Sauer (1994) recently proposed a method based on estimating equations (Godambe 1991) that directly fits the exponential growth model (equation
(1), above). No transformation is performed, so back-transformation is not a problem, and no constant need be added. In addition the error distribution is not specified, so the method is very robust (assuming that the model is correct). This method of estimating route trends was used by the USNBS in their most recent published report (Peterjohn et al. 1994), but appropriate weightings for combining route trends based on this estimator have yet to be published.

## Nonlinear route regression

A recurring criticism of linear RR methods is that they assume trends stay the same throughout the period analyzed, while in reality we would expect them to change over time. This problem is addressed within the linear RR framework by estimating trends for subsets of the entire time period and/or by calculating "residual" annual indices about the trend line (e.g., Peterjohn et al. 1995). Another solution, first proposed by James et al. (1990; see also Taub 1990) is to fit non-linear smoothers (Hastie and Tibshirani 1990 Chapter 2) to the route counts, and then average the smoothed route indices to produce regional trend estimates. In the James et al. method (nonlinear nonparametric route regression, NNRR), the loess smoother is used to smooth indices at both the route and physiographic stratum levels. A recent extension to the method (nonlinear semiparametric route regression, NSRR, James et al. 1996) allows differences between observers to be taken into account.

One disadvantage of this approach is that it requires the specification of the level of smoothing required, and this is set subjectively (James et al. 1990). Oversmoothing may obscure abrupt but biologically important changes in abundance (Peterjohn et al. 1995), while undersmoothing leaves too much noise in the indices (James et al. 1990). The choice of weightings to use when combining route trends is again somewhat controversial, although there is no need for an abundance weighting with this method.

A second nonlinear RR method, rank-trend analysis, was first applied to multi-site count data by Titus (1990; see also Titus et al. 1990). Counts on each route are ranked, a rank-trend statistic is calculated, and these route statistics are combined to give an overall statistic that indicates whether the counts show a tendency to increase or decline through time. Differences among observers are not accounted for, and no weighting scheme has been developed. The
method makes no assumptions about the distribution of counts (except that the distribution is the same for all counts), but is relatively uninformative because it gives no information about the magnitude of the estimated population change.

## Linear modeling

In contrast to RR methods, which use a two-stage (hierarchical or mixed-model) estimation process, linear modeling can be used to estimate the parameters directly from the count data. Sauer and Geissler (1990) used least-squares regression to fit a complex model containing observer, route-specific trend and year effects to log-transformed counts. As with least-squares linear RR a constant must be added to the counts before log-transformation and this biases the estimates. This problem is avoided by ter Braak et al. (1994), who proposed the use of loglinear modeling (Poisson regression), a type of generalized linear modeling (GLM, McCullagh and Nelder 1989) in which the counts are assumed to be Poisson random variables and the logarithm of the expected count is modeled as a linear function of the parameters of interest. The models are fit using the method of maximum likelihood, which enables the use of a suite of well-established techniques for data-based model selection, model checking and parameter estimation in GLMs (McCullagh and Nelder 1989; Burnham and Anderson 1992).

Fitting linear and loglinear models using standard computer software can be very computer memory intensive, especially for large datasets like the BBS which often require many parameters to be modeled adequately (Sauer and Geissler 1990). Recently, Pannekoek and van Strien (1996) have developed PC software (program TRIM) that enables large models to be fit and has modest memory requirements. This program can also be used to model serial correlation (autocorrelation) between counts on a site and overdispersion in counts relative to the Poisson distribution. The program can fit both annual index models
expected count = route effect * year effect
and exponential trend models

$$
\text { expected count }=\text { route effect } * \text { trend }^{\text {year }}
$$

and the year effect and trend parameters can depend upon categorical covariables (such as region, habitat type, etc.).

Another method that has been used to fit linear and loglinear models is impution of missing counts. Both applications of imputing to BBS data (Moses and Rabinowitz 1990; Böhning-Gaese et al. 1993) used linear models and would be improved by the use of loglinear models (e.g., Underhill and Prys-Jones 1994), so avoiding the need to transform the data and backtransform the estimates. However, imputing is not generally recommended because the estimation algorithm (expectation-maximization) is notoriously slow to converge when there are many missing data and identical results can be obtained using loglinear Poisson regression (ter Braak et al. 1994).

## Overview of thesis

In this chapter I have introduced the North American Breeding Bird Survey and the statistical methods that I will be considering in subsequent chapters. Before going on to examine the methods in detail, however, it is important to determine whether differences in analysis method have an important effect on the inferences drawn about population trends. I do this in Chapter 2, by applying three analysis methods to BBS data from 115 species in British Columbia and quantifying the differences in trend estimates. I find that even relatively small differences in methods of data selection and analysis can produce important discrepancies in the results. I discuss possible causes of the observed differences, and their implications for users of BBS trend information.

In Chapter 3 I explore the issues involved in the estimation of long-term population change from extensive survey data. I consider first how the data could be analyzed in an idealized survey, and then discuss the effect of various practical constraints on the analysis. I then suggest the use of Monte Carlo simulations to empirically test the performance of the methods under plausible scenarios of population change, and provide an example of the approach using the same analysis methods as in the previous chapter.

Case studies are a useful way of examining the performance of the analysis methods in detail. In Chapter 4 I present studies of four species, using eight of the most promising analysis methods applied to continent-wide BBS data. The four study species are chosen to represent
extremes with respect to two important factors that influence the reliability of the analysis methods: abundance (i.e., average count per route) and sample size (number of routes containing sightings). I place the results in the context of previous work, consider what inferences can be made about the reliability of the methods, and discuss possible improvements to the methods.

In the final chapter, I bring this work together by giving a summary evaluation of the methods. I then recommend some future directions for research and suggest three promising new approaches to trend analysis. Lastly, I discuss the implications of my findings for the use of extensive wildlife surveys like the BBS in estimating long-term population change.

## Figure Legends

Figure 1.1. Map showing the location of all North American Breeding Bird Survey (BBS) routes that have been surveyed one or more times ( $n=3924$; approximately 200 of these routes are no longer suitable for use, see text). Thick black lines show state and provincial boundaries and thin green lines show boundaries between the 99 physiographic strata (Bystrak 1981) that are often used in analyses of BBS data.

Figure 1.2. Contour map showing the number of annual surveys on BBS routes between 1966 and 1995. On average, routes in central and eastern states have been surveyed more times than those in western and northern states and provinces. Contours were generated using inverse distancing (Environmental Systems Research Institute Inc. 1996).

Figure 1.3. Number of BBS surveys performed each year between 1966 and 1995.

Figure 1.1


Figure 1.2


Number of surveys per route between 1966 and 1995
$0-10$
$\square$
$10-20$
$20-30$

Figure 1.3


## Chapter 2

# The Importance of Analysis Method for Population Trend Estimates 

## Introduction

There is widespread concern that many North American landbird populations are undergoing long-term declines (Terborgh 1989, 1992; Askins et al. 1990; Askins 1993; Finch and Stangel 1993). For these species, the primary source of population information at the continental scale is the North American Breeding Bird Survey (BBS). Trend determinations from the BBS are widely used by government agencies, conservation organizations and independent researchers to set priorities for conservation research and management (e.g., Hunter et al. 1993; U.S. Fish and Wildlife Service 1995) and to test possible hypotheses of causation (e.g., Sauer et al. 1996; Heckert 1997). Unfortunately, estimating trends from the BBS is a not straightforward; a number of statistical methods have been proposed and applied but there is currently no consensus as to their relative merits (Chapter 1). Here, I focus on the practical consequences of the application of different analysis methods for those that use BBS trend information.

I compare trend estimates produced by different analysis methods applied to the same data, and discuss the implications of any observed differences in estimates for our ability to detect population declines and to prioritize species of conservation concern. I use three analysis methods: USNBS route regression (Geissler and Sauer 1990), CWS route regression (Erskine et al. 1992) and rank-trends analysis (Titus 1990) on a common set of BBS data from British Columbia. The first two methods were used by the two agencies in recent published reports (Erskine et al. 1992; Peterjohn et al. 1995) and are based on a similar approach (route regression), but differ in the way that this approach is implemented. The third method is a nonparametric technique that estimates the direction of trend (i.e., whether the population is increasing or decreasing), but not its magnitude.

## Methods

## Breeding Bird Survey

The Breeding Bird Survey comprises approximately 370039.2 km long survey routes selected at random along secondary roads within all states/provinces in North America. Each year, during late May or June, a volunteer observer drives the survey route, stopping every 0.8 km and recording all the birds seen or heard within a 400 meter radius in a three minute period (Robbins et al. 1986, Appendix B). Approximately $70 \%$ of routes are surveyed each year.

Here, I used all routes in British Columbia that were surveyed between 1968 and 1992, excluding those from physiographic strata 29, 30 and 68 (Closed Boreal Forest, Aspen Parklands and Northern Rocky Mountains respectively, Butcher 1990), since they are rarely surveyed, and are not normally included in quantitative analyses of BBS data (J. R. Sauer, USNBS Patuxent Environmental Science Center, pers. comm.). The data set thus included 82 routes (803 individual surveys) within physiographic strata 63, 64 and 94 (Fraser Plateau, Central Rocky Mountains and Northern Pacific Rainforests).

## Data selection

Population trend analysis from BBS data involves two stages: data selection and trend estimation. Data selection criteria are designed to screen out surveys performed under questionable conditions, and to ensure that there are sufficient data to meet the requirements of the trend estimation procedure. Criteria depend largely upon the judgment of the analyst and vary widely among published analyses. The selection procedures used by the USNBS and CWS are similar, but differ in the criteria for the exclusion of individual surveys and the minimum data requirements of the route regression procedures. Standard data selection protocols for ranktrends analysis have not been developed. Because I was chiefly interested in differences due to trend estimation method, I based my comparisons of the three methods on the common subset of data that passed both current USNBS and CWS criteria, as provided by J. R. Sauer (USNBS, pers. comm.) and B. T. Collins (CWS National Wildlife Research Centre, pers. comm.). Nevertheless, in order to determine the effect of data selection, I also compared the results from
the two route regression methods produced using their respective agency's data selection protocols.

USNBS criteria: (1) Only surveys performed within a regionally specified range of dates, times and weather conditions, and by qualified observers were included in the analysis (i.e., surveys with USNBS survey codes one, two or three). (2) Routes with too few surveys remaining to enable estimation of trend and trend variance were then excluded (i.e., where number of surveys minus number of observers was less than two). (3) Routes on which no birds of the species under analysis had been seen were also discarded. (4) If (after data selection) the species had fewer than 10 routes remaining or had a mean count of less than 1.0 birds per survey, trend estimates were not calculated for that species.

CWS criteria: (1) Surveys that did not meet CWS survey criteria were discarded. These criteria are similar (but not identical) to those used in USNBS criterion 1. (2) Routes with too few surveys remaining to enable estimation of trend (but not trend variance) were then excluded (i.e., where the number of surveys minus number of subroutes was less than one. A subroute is a subset of the surveys done on a single route such that all surveys in a subroute were performed by the same observer within a span of 19 or fewer calendar days across years and under similar weather conditions. Subroute designations were provided by B.T. Collins). Criteria 3 and 4 were identical to those used by the USNBS.

## Trend estimation

The three trend estimation methods are summarized in Table 2.1 and are described in detail below.

USNBS route regression: In route regression, the overall trend for the region and species of interest is calculated using the weighted average of the trend on each route (see below). A route trend is the slope of a log-linear regression of birds counted (dependent variable) against time (independent variable). I performed the analysis described here using programs written in C and C++ (Appendix 1), following the methods of Geissler (1984), Geissler and Sauer (1990) and Peterjohn et al. (1995), with additional details provided by J. R. Sauer (pers. comms.). The programs included some numerical routines written by Press et al. (1992).

For each route, I computed the natural logarithm of the number of birds counted (plus 0.5 to accommodate zero counts; Component 1, Table 2.1). Route trends were calculated on the logarithmic scale as the slope of the regression of log count on year of count (Component 2). To account for differences among observers I treated them as covariables in the regression (Component 3). I then back-transformed the route trends using the method developed by Bradu and Mundlac (1970), which is approximately equal to taking the exponent of the trend estimate minus one-half of its variance (Component 4).

The overall species trend for British Columbia was calculated using the weighted mean of the back-transformed route trends (Component 5), in two stages as follows. I first calculated the weighted mean slope for each physiographic area, weighting back-transformed route trends by the marginal mean count on the route (Searle et al. 1980; Component 6), and a measure of the reliability of the route trend estimate, calculated as the variance of the slope divided by the mean square error for the route-specific regression (Component 7). The marginal mean is an estimate of the count that would have been recorded in the mid-year of the analysis period (i.e., 1980) by an average observer. Geometric mean count was used in place of marginal mean on routes where estimating the marginal mean required extrapolating beyond the years when counts took place. In this context geometric mean was defined as the arithmetic mean of the log-transformed counts, backtransformed onto the multiplicative scale. I then calculated the overall trend as the mean of the trends for each physiographic stratum, weighted by the stratum area (Component 8 ).

I estimated the variance of the overall trend estimate from 400 bootstrapped subsamples of the routes within physiographic strata (Component 10). Bootstrapping is a technique that allows the estimation of the sampling distribution of a non-standard statistic (in this case the overall trend estimate). Many repeat subsamples are taken with replacement from the original data and the distribution of the statistic in the subsamples is taken as the best indication of the true distribution of the statistic (Efron 1982). I also used the mean of the bootstrapped trend estimates as the final species trend since it is in theory less biased than the original estimate. Statistical significance of the trend was assessed using a $z$-test (Component 11). This tested the
alternate hypothesis that routes show consistent log-linear trends with a slope different from zero across physiographic strata within the study area.

CWS route regression: The CWS method also uses route regression to calculate regional trends from a weighted average of the trend on each route. However it differs from the USNBS method in the way the route trends are calculated, the method of averaging route trends and the weightings used to produce the regional trend. I performed this analysis using a FORTRAN program supplied by B. T. Collins (Appendix 1), which implements the methodology of Erskine et al. (1992; see also Collins and Wendt 1989).

As with the USNBS route regression, I calculated route trends as the slope of the regression of the natural logarithm of counts on year of count (Components 1,2 and 3 ). In this case 0.23 (rather than 0.5 ) was added to each count before $\log$ transforming to accommodate zero counts, and subroutes were used as covariables in the regression (rather than observers; see CWS data selection criteria for a definition of subroutes).

I calculated the species trend for British Columbia directly from the weighted mean of the route trends, still on the logarithmic scale (Component 5). The weightings used were the marginal mean count on the route (Searle et al. 1980; Component 6), a precision estimate given by the squared deviations of the counts from the subroute mean (Component 7), and an area weighting (Component 8 ). I constrained the marginal mean so that it could not exceed the maximum number of birds seen on all routes, or 200 , whichever was the least. The area weightings were provided by B.T. Collins and these consider the number of routes and the proportion of land area in each degree-block.

I then back-transformed the overall trend estimate to the arithmetic scale by simple exponentiation (Component 9), and calculated the variance of the estimate by jackknifing (Component 10). Jackknifing is similar to bootstrapping, except that in this case the population distribution of the statistic (overall trend estimate) is approximated using the set of subsamples generated when one data point (route) is removed from the sample in turn (Efron 1982). Unlike the USNBS method, no bias adjustment was performed. The statistical significance of the alternate hypothesis, that there are consistent non-zero log-linear trends across routes in British

Columbia, was calculated using a $t$-test, with the degrees of freedom equal to the number of routes minus one (Component 11).

Nonparametric rank trend analysis: With this method, trend statistics are calculated separately for each route using a nonparametric ranking procedure, and are summed to produce regional trend statistics. This enables calculation of the direction (increasing or decreasing) and statistical significance of the trend, but not its magnitude (the size of increase or decrease). The procedure described in Titus et al. (1990) was implemented fully with a C program (Appendix 1 ), and will be briefly summarized here.

On each route, I arranged counts in ascending order and assigned ranks, giving ties the mean of the tied ranks. I calculated a trend statistic, $D$, for each route as the sum of $\left(R_{i}-i\right)^{2}$, where $R_{i}$ is the rank of the $i$ th yearly count (i.e. $i=1$ in 1968 to 25 in 1992; Component 2). If counts tend to increase over time on the route then D will be small, and if counts tend to decrease over time then $D$ will be large. Under the null hypothesis of random change in count over time, this statistic is normally distributed with an expected mean and variance that can be easily calculated (Lehman 1975). To test the alternate hypothesis, that counts on routes tend to increase or decrease over time throughout British Columbia, I summed the values of $D$ (Component 5), the expected mean of $D$ and the expected variance of $D$ (Component 10 ) over routes and performed a $z$-test (Component 11).

## Comparison of trend estimates

For each pairwise combination of methods, I quantified differences in trend estimates by calculating the proportion of species with trends in different directions (i.e., increasing in one method but decreasing in the other). I also calculated the proportion of differences in statistical significance (i.e., not significant ( $p<=0.05$ ) in one method but significant ( $p>0.05$ ) in the other). In order to determine whether the differences were related to sample size (number of routes) or species abundance (mean count per survey), I performed logistic regressions on the proportion of differences (dependent variable) against sample size and log species abundance (independent variables).

For the two route regression methods, I further examined the pattern of differences by plotting estimates of trend against one another and determining the slope of the principal axis (Model II regression, Sokal and Rohlf 1981). If the two methods produce similar trends, on average, then the principal axis will have a slope of 1.0 and run through the origin. I also calculated the median absolute difference between trends, and determined whether this was related to sample size or log species abundance using linear regression.

## Results

Of the 265 species recorded at least once, 119 passed both USNBS and CWS data selection standards. For this common subset of the data, four species had estimated annual rates of population change of greater than $15 \%$ per year using CWS route regression (Tennessee warbler Vermivora peregrina: $+18.8 \%$; Black tern Chlidonias niger: $+22.5 \%$; Magnolia warbler Dendroica magnolia: -33.6\%; Black swift Cypseloides niger: $+41.8 \%$; see Discussion). These extreme trends are biologically implausible, so I excluded them from further analysis. For the remaining 115 species, the median sample size was 39 routes (range $10-61$ ) and the median abundance was 4.76 birds per survey (range 1.09-62.86).

Using these data, about the same number of species were estimated to be increasing or decreasing in all three methods (Fig. 2.1). However, there was considerable variation among methods in the species making up these totals: the proportion of species assigned different trends ranged between 0.21 and 0.33 , depending upon the pair of methods being compared (Table 2.2). The proportion of differences did not appear to be strongly related to sample size or species abundance (Table 2.3). The estimates had quite large confidence intervals (Table 2.3), so I cannot rule out the possibility that there were too few samples ( $N=115$ species) to detect such an effect had one occurred. However, in many cases, the estimates were in the opposite direction to that predicted (i.e., proportion of differences increased with increasing sample size and abundance).

There were large differences among methods in the overall number of species estimated to be undergoing statistically significant population changes, with USNBS route regression
assigning almost twice as many species significant trends as CWS route regression, and nonparametric rank trends assigning twice as many again as the USNBS method (Fig. 2.1). The proportion of species assigned trends of different significance levels was between 0.18 and 0.43 (Table 2.2), and again did not appear to be strongly related either to species abundance or sample size (Table 2.3). The comments about sample size are relevant here also. In all comparisons (i.e., direction and significance), the two route regression methods were the most similar, while the greatest differences were between CWS route regression and rank-trends analysis (Table 2.2).

For the two route regression methods, there was little overall difference in trend estimates (median difference USNBS trend - CWS trend $=+0.04 \%$ per year; $95 \%$ confidence interval (CI) estimated from 10000 bootstrap subsamples of the 115 species $=-0.60$ to +0.44 ). However, the absolute magnitudes of trends were greater in CWS route regression trends than in the USNBS method: negative trends tended to be more negative and positive trends tended to be more positive (Fig. 2.2., slope of principal axis $=0.58,95 \%$ parametric $\mathrm{CI}=0.45$ to 0.72 ). There was considerable scatter about this overall pattern (Fig. 2.2), with the median absolute difference in magnitude of trends being $1.19 \%$ per year, (range $0.00-13.06$ ).

Contrary to the results for the proportion of differences in the direction and significance of trends, the mean absolute difference in the magnitude of trends decreased with increasing sample size (slope of linear regression: $-0.46 ; 95 \%$ parametric $\mathrm{CI}=-0.87$ to -0.05 ) and with increasing log species abundance (slope: $-0.044 ; 95 \%$ parametric $\mathrm{CI}=-0.067$ to -0.020 ), although in both cases the proportion of variance explained by the relationship was small ( $r^{2}=$ 0.04 and 0.11 respectively). I speculated that this may be due to decreasing variance of the estimates, since variance decreases with increasing sample size, and sample size and log abundance were correlated (Pearson $r=0.44$ ). In order to test this, I corrected for decreasing variance by dividing the difference in trend estimates by the square root of the sum of their variances, to yield $z$-statistics. The absolute magnitude of these $z$-statistics did not appear to vary with sample size or with $\log$ abundance in linear regressions (sample size: slope $=-0.002$; $95 \%$ parametric $\mathrm{CI}=-0.008$ to $+0.005 ; r^{2}=0.00 ; \log$ abundance: slope $=0.05 ; 95 \%$ parametric $\mathrm{CI}=-0.06$ to $+0.16 ; r^{2}=0.02$ ), indicating that the increasing similarity between trend estimates
with sample size and abundance was indeed explained by decreasing variance. I conclude that trend estimates that have low variance when derived using route regression are less sensitive to the particular variant of route regression used.

Finally, I evaluated the effect of data selection criteria. When only USNBS data criteria were applied to the data, the average species dataset contained 3.5 more routes (median sample size $=42.5$ routes, range 11-64), and the median trend estimated using USNBS route regression was $0.45 \%$ per year higher than when both data selection criteria were applied (range 1.20 lower to 6.02 higher). For CWS data selection criteria alone, the average increase in sample size was 2 (median sample size $=41$ routes, range $10-62$ ), and the median trend was $0.07 \%$ per year lower (range 4.36 lower to 2.90 higher). Comparing USNBS route regression and CWS route regression using trend estimates derived from the datasets appropriate to each, these changes caused the USNBS method to produce on average more positive trends than the CWS method (median difference USNBS trend - CWS trend $=+0.73 \%$ per year; $95 \%$ CI estimated from 10000 bootstrap subsamples $=+0.36$ to +1.01$)$. This is dissimilar to the result using the uniform data set, where both methods produced similar trends on average. I therefore conclude that data selection criteria can have an important effect on the results obtained.

## Discussion

The Breeding Bird Survey is the most comprehensive source of data available for monitoring trends in North American landbird populations, and for prioritizing species of conservation concern. I have shown that the method of trend estimation can affect the magnitude, direction and statistical significance of population trends assigned to species. In addition, small changes in the way that data are selected for analysis can affect the overall magnitude of trends. In this section I consider possible causes of the differences, their implications for users of BBS trend information, and suggest some priorities for further work.

## Causes of the differences

This study was not designed to explicitly test which of the many differences between the methods caused the observed differences in trend estimates. Nevertheless, based on the
predicted effect of differences between the methods (Table 2.1) and some preliminary analyses, I can suggest which components of the methods could account for the results. In summary, the results were: (1) large differences in the number of statistically significant trends among the three methods; (2) greater absolute magnitude of trends in CWS route regression compared with the USNBS method; (3) slightly more positive trends in the USNBS method compared with CWS route regression when the respective agencies' data selection criteria were applied separately; and (4) large variation about these patterns in the direction, magnitude and significance of trends. I deal with each of these in turn.

Differences in the number of statistically significant trends were consistent with differences in the way that regional trends and trend variances were calculated by the different methods, although many other factors may have been involved. Nonparametric rank trends analysis produced the greatest proportion of statistically significant results, probably because it did not consider between-route variance when calculating the expected variance of the regional trend (Component 10, Table 2.1). USNBS route regression produced more statistically significant results than CWS route regression. The USNBS method tested for consistency of trends nested within physiographic areas, while CWS route regression tested for consistency of trends over the whole region (Component 5). Hence in the USNBS method, the variance of trends among physiographic areas was removed from the statistical test (analogous to a nested analysis of variance), resulting in an increased chance that the trend was significant.

The greater absolute magnitude of trends in CWS route regression relative to the USNBS method may have been due to differences in the constant added before log-transformation (Component 1). Addition of a constant biases the trend towards zero (no trend), and this effect is stronger for larger values of the constant (Geissler and Link 1988; Collins 1990). The constant was larger in the USNBS method than in the CWS method ( 0.5 vs. 0.23 ), so the potential bias in USNBS route regression was greater. The bias is also stronger for low abundance species (Geissler and Link 1988; Collins 1990). I thus predicted that if differences in the constant were causing the observed differences in absolute magnitude of trend, the effect would be greater for low abundance species than those with high abundance. I tested this
possibility by dividing the 115 species into two equal groups by abundance and calculating the slope of the principal axis separately for each group (one species was excluded from the test to make both groups of equal size). The median abundance in the low abundance group was 2.08 birds per survey (range 1.09-4.59), and the slope of the principal axis was 0.52 ( $95 \%$ parametric $\mathrm{CI}=0.32$ to 0.76 ). For the high abundance group, the median abundance was 8.29 birds per survey (range 4.96-62.86) and the slope of the principal axis was 0.63 ( $95 \%$ parametric $\mathrm{CI}=$ 0.49 to 0.79 ). The slopes vary in the anticipated direction (low abundance group smaller) but are clearly not significantly different. Simulations performed by Collins (1990) showed that the bias due to addition of a constant decreases rapidly with increasing abundance. I thus suspect that the minimum data requirement for a mean of at least 1.0 birds per survey was effective in controlling the difference in bias to below the level that I could detect statistically. Other explanations must account for most of the observed difference in the absolute magnitude of trends between route regression methods.

The systematic change in USNBS trend estimates resulting from changes in data selection criteria currently defies explanation. I cannot imagine any scenario in which the small differences in data selected would cause systematic differences in trend. These results highlight the need for standardization of data selection, since many published analyses use data selection criteria even more divergent than the two I compared here (e.g., Böhning-Gaese et al. 1993; James et al. 1996).

Differences in trend estimates among methods that are not explained by the above patterns are likely due to the additive effect of the many differences among the methods, each acting independently. Of these, I predict that the most important are the weightings used when combining route trends to produce trends for the region (Components 6, 7 and 8). These weightings are designed to correct for geographic variation in the number of routes and the abundance of birds, and to increase the precision of the regional trend estimate. The method of calculating the weightings differs between route regression methods, which can result in differences of up to three orders of magnitude in the weightings assigned individual routes. No system of weightings has been developed for use with nonparametric rank trends analysis. The
importance of weightings is demonstrated by the four species assigned improbably large trend estimates in CWS route regression and excluded from the comparisons. In all four of these species, one route with a large apparent change in the density of birds but few surveys was assigned a very large abundance weighting (due to the extrapolation of the route trend to calculate the least squares mean) and thus dominated the regional trend estimate. The abundance weighting was calculated differently in USNBS route regression (geometric mean was used where extrapolation was necessary), and these routes were assigned small weightings, leading to more credible regional trend estimates for the four species.

## Implications of the Results

To be effective, conservation programs must be based upon accurate information. Partners in Flight, the multi-agency coalition responsible for initiating the Neotropical Migratory Bird Conservation Program (Hagan 1992; Finch and Stangel 1993), has defined an effective monitoring scheme as one that has a $90 \%$ chance of detecting a $50 \%$ decline in species abundance over 25 years (Butcher et al. 1993). I have shown here that for British Columbia BBS data the direction, magnitude and significance of trends attributed to species shows considerable variation due to analysis method alone. For the two route regression methods, which used a similar approach, 15 of 115 species ( $17 \%$ ) showed a difference in the magnitude of the trend due to analysis method of greater than $50 \%$ over the 25 years analyzed. For these species the uncertainty due to analysis method alone was larger than the absolute change in population size that Partners in Flight wishes to detect. Differences in the magnitude of route regression estimates were related to sample size (number of routes in the analysis); hence, I expect these differences to be smaller for estimates based on larger geographic areas or states/provinces with a greater density of routes (e.g., those in eastern North America). I do not know whether trend estimates from methods based on very different assumptions, such as nonlinear nonparametric route regression (James et al. 1992) or Mountford's method (Mountford 1982), will show similar convergence with increasing sample size.

A number of species prioritization schemes have used the statistical significance of the trend estimate as one criterion for setting the conservation priority of a species (e.g., Hunter et al.
1993). The results show large differences among methods in the statistical significance assigned to species (Table 2.2), with no consistent increase in similarity with increasing sample size or abundance (Table 2.3). Of the 31 species estimated to be showing significant declines (i.e., $p<0.05)$ by one or more of the methods, $15(47 \%)$ were estimated to be declining significantly by USNBS route regression, 8 (25\%) by CWS route regression and 29 ( $91 \%$ ) by rank-trends analysis. Hence species prioritization schemes based upon the statistical significance of trends are likely to highlight a very different set of species depending upon which analysis method is used.

The differences reported here indicate that analysis method will also affect the results of geographic analyses of the significance of population trends (e.g., Sauer and Droege 1990a, d; Peterjohn et al. 1995). I do not know whether analysis method will influence the outcome of multispecies comparisons (e.g., Böhning-Gaese et al. 1993; Peterjohn and Sauer 1993; James et al. 1996), since differences in individual species trends may "average out" when species are grouped (Fig 2.1). I thus tentatively support Droege (1990:3) in his statement: "Statistical analyses of these data and their subsequent interpretation should dwell on the patterns of population change rather than on the magnitudes of calculated trends and variances."

## Priorities for Further Work

In this chapter, I have demonstrated that the manner in which BBS data are selected and analyzed can affect the result obtained and inferences drawn. These results raise two important questions: 1) which components of the methods have the largest effect on the resulting estimates, and 2) which method is the most accurate? The first question requires a detailed comparison of many variants of each method applied to the same dataset. This type of study would act as a sensitivity analysis, enabling detailed methodological research (e.g., Sauer et al. 1994b) to focus on the aspects of the methods that are the most sensitive to the approach used. The second question is much harder to resolve. Comparisons of results among methods are useful for highlighting problems with the methods that lead to unfeasible estimates (e.g., the abundance weighting problem in CWS route regression). However, these comparisons cannot definitively tell which method is the most accurate because as there is no benchmark of known trends with
which to compare the estimates. True population trends in British Columbia, or for any other large geographic part of the continent, are unknown for any species monitored. One possibility is to compare results to those from other surveillance programs, such as migration monitoring from bird observatories or checklist programs (e.g., Hussell et al. 1992). However, this approach is flawed: because all programs measure trends with error, neither concordance nor lack of it can be used to infer which method is biased.

Monte Carlo studies offer a possible solution, in which "known trends" are generated $a$ priori using stochastic simulations, and are used to compare analysis methods. Trends could be simulated under a variety of realistic scenarios, such as those suggested by Wilcove and Terborgh (1984). In general, the most appropriate method will be the one that gives the most accurate results over the widest range of conditions. An example of the approach is given in Chapter 3.

The BBS has the potential to provide population trends for the majority of breeding birds in North America over a large part of their ranges. Although trend estimation from BBS data is a formidable statistical challenge, many of the causes of bias are predictable and can be minimized by the choice of a method appropriate to the hypothesis under examination. Given the current high levels of concern about population trends of North American landbirds it is clearly imperative that we determine how best to draw inference from the wealth of information that is collected by the BBS each year.

## Summary

The purpose of this chapter was to determine whether analysis method has an important influence on the inferences drawn about population trends. I compared trend estimates produced by three different analysis methods applied to the same set of BBS data for 115 species in British Columbia. The methods were U.S. National Biological Service route regression, Canadian Wildlife Service route regression, and nonparametric rank trends analysis. Overall, the number of species estimated to be declining was similar among the three methods, but the number of statistically significant declines was not similar (15, 8 and 29 respectively). In addition, there
were many differences among methods in the trend estimates assigned to individual species. Comparing the two route regression methods, Canadian Wildlife Service estimates had a greater absolute magnitude on average than those of the U.S. National Biological Service method. U.S. National Biological Service estimates were on average more positive than Canadian Wildlife Service estimates when the respective agencies data selection criteria were applied separately. These results imply that our ability to detect population declines, and to prioritize species of conservation concern are strongly dependent upon the analysis method used. I discussed the possible causes of the observed differences, and suggested the use of simulation studies to evaluate the performance of the methods.
Table 2.1. Summary of the three Breeding Bird Survey trend estimation methods compared in this chapter.

| Component | Methods |  |  | Predicted effect of differences |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | USNBS ${ }^{1}$ route regression ${ }^{2}$ | CWS ${ }^{1}$ route regression ${ }^{3}$ | Nonparametric rank trends ${ }^{4}$ | Direction | Size |
| 1. Transformation of counts | $\log ($ count +0.5$)$ | $\log ($ count +0.23$)$ | - | Larger constant in USNBS decreases absolute magnitude of trend and variance relative to CWS | Small, but greater at low abundances |
| 2. Calculation of route trend | slope of linear regression of log count | slope of linear regression of log count | rank trend | Unknown | Unknown |
| 3. Covariables | observers | subroutes | none | USNBS / CWS: unknown; no covariables in rank trends may cause positive bias due to increase in observer quality | USNBS / CWS: <br> likely small; rank trends: likely small-medium |
| 4. Backtransform | approx. exp (route trend - variance/2) | performed later | - | Back-transformation at this stage makes USNBS trends more positive than CWS | Small |
| 5. Regional trend | weighted mean of backtransformed route trends within physiographic strata | weighted mean of route trends on log scale | sum of route trends | Variance estimates likely higher for CWS route regression than USNBS method | Unknown |
| 6. Abundance weight | least squares mean / geometric mean | Winsorized least squares mean | none | Unknown | May be large |

Table 2.1. (continued)

| Component | Methods |  |  | Predicted effect of differences |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | USNBS ${ }^{1}$ route regression ${ }^{2}$ | CWS ${ }^{1}$ route regression ${ }^{3}$ | Nonparametric rank trends ${ }^{4}$ | Direction | Size |
| 7. Precision weight | sum of squares of years surveyed | sum of squares of years surveyed | none | Unknown | May be large |
| 8. Area weight | area of physiographic stratum within state | proportion of degree block covered | none | Unknown | May be large |
| 9. Backtransform | performed earlier | $\exp$ (regional trend) | - | Simple exponentiation by CWS method makes it more positive than USNBS | Small |
| 10. Variance estimate | bootstrap | jackknife | parametric estimate | USNBS / CWS: unknown; within-site variance not accounted for in rank trends, so variances smaller | USNBS / CWS: <br> likely small; rank trends unknown |
| 11. Significance test | $z$-test | $t$-test | $z$-test | USNBS / CWS: USNBS will give more statistically significant results | Small, but greater at low abundances |

[^1]Table 2.2. Pairwise comparisons among three Breeding Bird Survey trend estimation methods of the proportion of 115 species with trends in different directions (i.e., increasing in one method but decreasing in the other) or with different levels of statistical .significance (i.e., $p<=0.05$ in one method but $p<0.05$ in the other).

| Trend | USNBS $^{1}$ route regression vs. <br> CWS $^{1}$ route regression | USNBS ${ }^{1}$ route regression vs. <br> nonparametric rank trends | CWS ${ }^{1}$ route regression vs. <br> nonparametric rank trends |
| :--- | :---: | :---: | :---: |
| Direction | $0.21(0.14 \text { to } 0.28)^{2}$ | $0.25(0.18 \text { to } 0.32)^{2}$ | $0.33(0.26 \text { to } 0.40)^{2}$ |
| Significance | $0.18(0.11 \text { to } 0.25)^{2}$ | $0.37(0.28 \text { to } 0.49)^{2}$ | $0.43(0.34 \text { to } 0.52)^{2}$ |

${ }^{2}$ Values given are proportions, with $95 \%$ confidence interval (calculated using the normal approximation to the binomial distribution) in parentheses. If methods produced identical trends then the expected proportion of disagreements would be 0.0 . If there was no association between trends produced by the methods then the expected proportion would be approximately 0.5 .
Table 2.3. Estimates of the effect of sample size (number of routes) and species abundance (log mean count per survey) on the proportion of differences in the direction and statistical significance of species trends between three Breeding Bird Survey analysis
methods. Estimates are from a logistic regression with $\mathrm{N}=115$ species.

| Dependent <br> variable | Independent <br> variable | USNBS ${ }^{1}$ route regression vs. <br> CWS ${ }^{1}$ route regression | USNBS ${ }^{1}$ route regression vs. <br> nonparametric rank trends | CWS ${ }^{1}$ route regression vs. <br> nonparametric rank trends |
| :--- | :--- | :---: | :---: | :---: |
| Direction sample size $0.14(-0.15 \text { to } 0.44)^{2}$ $-0.05(-0.33 \text { to } 0.26)^{2}$ | $0.27(-0.06 \text { to } 0.56)^{2}$ |  |  |  |
| of trend | log abundance | $0.25(-0.11 \text { to } 0.50)^{2}$ | $-0.19(-0.58 \text { to } 0.19)^{2}$ | $0.16(-0.27 \text { to } 0.48)^{2}$ |
| Significance | sample size | $0.06(-0.21 \text { to } 0.35)^{2}$ | $0.16(-0.18 \text { to } 0.47)^{2}$ | $0.10(-0.24 \text { to } 0.43)^{2}$ |
| of trend | log abundance | $0.10(-0.27 \text { to } 0.38)^{2}$ | $0.07(-0.36 \text { to } 0.43)^{2}$ | $-0.18(-0.56 \text { to } 0.25)^{2}$ |

${ }^{1}$ USNBS = U.S. National Biological Service; CWS = Canadian Wildlife Service. lowest value, with $95 \%$ confidence interval in parentheses. If analysis methods produce more similar results with increasing values of the independent variable then the estimated change will be negative. In all cases the confidence limits span zero, indicating that we could not detect an effect of the independent variable.

## Figure Legends

Figure 2.1. Number of species with increasing and decreasing trends from three analysis methods using the same Breeding Bird Survey data for 115 species from British Columbia. USNBS is U.S. National Biological Service and CWS is Canadian Wildlife Service.

Figure 2.2. Plot of the magnitude of trends estimated from U.S. National Biological Service (USNBS) route regression against the magnitude of trends from Canadian Wildlife Service (CWS) route regression for 115 species. If the methods produced identical trends then all the species would be on the dashed line. The solid line is the principal axis.

Figure 2.1


Figure 2.2


## Chapter 3

# Issues in the Estimation of Long-term Population Change 

## Introduction

A number of issues make the analysis of long-term multi-site population monitoring data challenging and controversial. In this chapter I explore these issues, and suggest one possible method for resolving them. In the first part of the chapter I consider how the data could be analyzed in an idealized survey, and then discuss the effect of various practical constraints on the analysis. As a concise reference, I summarize the major features of analysis methods that have been used to date in Tables 3.1 and 3.2. In the second part I suggest the use of Monte Carlo simulations to empirically test the performance of these different methods under plausible scenarios of population change, and provide an example of the approach. Much of the emphasis is on the estimation of long-term population change in North American landbirds from the Breeding Bird Survey (BBS); however these methods and the discussion are applicable to any population where repeated counts are performed at a large number of sites.

## Analysis of survey data under ideal conditions

Since it is not feasible to census the entire population of most bird species, population monitoring is almost always based upon surveys of a sample of the population. In an ideal sample survey, the population of interest is divided into non-overlapping units, a sample of these units is selected according to a pre-defined procedure and a complete census is taken of each unit. In bird surveys, the sample unit is usually all the birds within a specified survey site ("route" in BBS terminology). Because population change (i.e., change in abundance) is the main parameter of interest, the most efficient procedure involves repeated annual censuses of the same sample of sites. Two issues arise in the estimation of population change from the resulting data: (1) selection of the unit of analysis and (2) model of trend.

## Selection of the unit of analysis

Population change is usually portrayed using annual indices of abundance or as population trend (prevailing tendency underlying annual indices). In the ideal situation, annual indices are given by the average of the counts in each year, and population trend is the average of the trends at each sample site. If the sample design involves stratification then the average within each stratum should be weighted by the proportion of units in the stratum, usually calculated as the proportion of the total area covered that is in the stratum (area weighting). In addition, trends should be weighted by the abundance of individuals at the sample site (abundance weighting), since declines at a site containing many individuals represent a greater proportion of the population than declines at a site containing few individuals (see James et al. 1990 for an overview of weightings).

Alternatively, population trends can be calculated from the estimates of annual indices, or annual indices could be calculated as the residuals from a trend analysis. In the ideal situation, these indirect estimators produce the same results as using counts as the unit of analysis.

However, because the analysis methods used in practice are modified to deal with departures from the ideal, the estimates may diverge. Hence, analyses that present both annual indices and population trends (e.g., Greenwood et al. 1994, Peterjohn et al. 1995) tend to use estimates of one parameter to produce the other. The unit of analysis for each of the methods is given in Tables 3.1 and 3.2.

## Model of trend

Repeated counts at a survey site form an ecological time series, and so may be thought of as being composed of four sources of pattern: trend (prevailing tendency), interventions (irregular perturbations such as environmental effects), autocorrelation (due to population processes such as density dependence) and sampling error (Barker and Sauer 1992). Estimating population trend thus involves separating the prevailing tendency from the other components. Because the relative contribution of these components is not known a priori, the model used depends upon the judgment of the analyst (Jassby and Powell 1990). Many of the major differences between trend-estimation methods can be attributed to differences in the way that
trend is modeled.
Although there are a number of methods available for deriving trends from time series (Jassby and Powell 1990), analyses of bird counts have focused on regression-based approaches. Four regression models have been applied to count data (Table 3.2), each of which can produce a different picture of the prevailing tendency (Fig. 3.1), and each of which may be advantageous under certain conditions.
(1) In linear-multiplicative models (e.g. linear route regression), population change is modeled as a multiplicative process in which the expected count in year $y+1$ is a multiple of the expected count in year $y$, leading to $E\left(C_{y}\right)=E\left(C_{0}\right) \beta^{y}$ where $E\left(C_{y}\right)$ is the expected count in year $y$ and $\beta$ is the trend. This model is conceptually simple, involves estimating the minimum number of parameters and is unbiased in the face of random interventions and autocorrelation. However, it will be inaccurate if the prevailing tendency changes over time (i.e., non-linear trend).
(2) Polynomial-multiplicative models (e.g. loglinear modeling) can allow for changes in trend over time, using polynomial parameters to estimate higher order effects, i.e., $E\left(C_{y}\right)=E\left(C_{0}\right) \beta_{1}^{y} \beta_{2}^{y^{2}} \beta_{3}^{y^{3}} \ldots$. The order of the model is usually determined a priori, but could be derived from the data using model selection techniques, such as likelihood ratio tests (Burnham and Anderson 1992). Although more general, it is also more susceptible to incorporating inappropriate patterns such as population cycles.
(3) In additive models (e.g. smoothing route regression), no a piori model is postulated, and non-trend sources of variation in counts are removed through filters such as LOESS. The advantage of this model is that no explicit assumptions are made about the shape of the trend. One disadvantage is the requirement for specification of a smoothing parameter, which determines how much variation is filtered from the data. This must be assigned subjectively. A common criticism of both polynomial-multiplicative and additive models is that it is hard to draw inferences about overall patterns of population change since they estimate many parameters. Both can, however, provide a single parameter estimate if
trend is defined as the difference in expected count between the beginning and the end of the time period of interest (e.g., James et al. 1996).
(4) In rank models (e.g. rank-trends), trend is modeled as the tendency for counts to increase or decrease over time. Counts are ranked, and the distribution of ranks over time is used to estimate whether the population has shown a tendency to increase or decline (Fig. 3.1). Because ranks are used, only the direction of trend is estimated, not its magnitude. This model is only appropriate if the prevailing tendency is uni-directional, but does not require that the trend be linear.

Ultimately, the most appropriate model of trend will depend upon the pattern of population change in the population being modeled. While the true pattern is not known for any species monitored, the robustness of alternative models given expected population changes over space and time can be tested using population simulations (e.g., Anganuzzi 1993). Note also that the accuracy of all models is conditional on the length of the time series and the validity of the starting point (Jassby and Powell 1990, Barker and Sauer 1992). For example, if a species shows population cycles but the time series covers only part of one cycle then the observed changes in abundance will be interpreted as trends. A second example is the increasing trends recorded for some species in BBS analyses which may be due to recoveries from pesticide use or severe winters that occurred near the start of the survey (references in Barker and Sauer 1992).

## Additional factors that complicate the analysis of survey data

Extensive bird surveys often depart from the ideal situation in three ways: (1) the population from which the sample is drawn is a subset of the population about which information is required (incomplete sampling frame), (2) the number of individuals in each sample site is not counted accurately (measurement error), and (3) most sites in the selected sample are not censused every year (missing data).

## Incomplete sampling frame

If the sample frame is incomplete then the degree to which results obtained from the sample can be extrapolated to the population as a whole is limited. For example, in the BBS,
survey sites are allocated along secondary roads; hence limited inferences can be drawn about species in which a large proportion of individuals breed in habitats such as floodplains, mountains or forest interior that are not in close proximity to roads, or in the roadless northern part of Canada. Although this can be an important limitation to the utility of the survey program (Droege 1990; but see Bart 1995), it does not affect the way that estimates are derived from the population sampled, and so is not considered further here.

## Measurement error

Measurement error is a feature of almost all bird counts. The probability that a bird is counted given that it is present at the survey site that year (its detectability) is a function of a large number of factors, such as survey method, species, observer, habitat, time of year, weather, bird density, and breeding phenology (references in Droege 1990). Hence, estimates of population size or absolute abundance are not possible from the count data, and indices of abundance are seldom comparable between species or survey programs. In addition, sources of error that change systematically over time, such as habitat, traffic noise, inter- and intra-observer ability can generate spurious estimates if they are not modeled in the analysis method (Barker and Sauer 1992). In particular, differences in detection rates between observers are often cited as a major potential source of bias (Barker and Sauer 1992, Sauer et al. 1994b, Peterjohn et al. 1995). Sauer et al. (1994b) used several approaches to show that the number of birds counted by observers has increased, independently of trend, for the majority of species in the BBS. This result was corroborated by James et al. (1996), who found that trends were usually more positive if observer differences were not modeled in their nonlinear trend-estimation technique.

However, incorporating observer differences (or other systematic sources of error) into models may result in a drastic reduction in efficiency due to overfitting, and the resulting accuracy of the estimate may be decreased. Currently, some analyses model observer differences, while others do not (Tables 3.1 and 3.2).

The problem of determining whether to include additional explanatory variables in a model has been encountered in other areas of ecology (e.g., mark-recapture studies: Lebreton et al. 1992), and the use of model selection techniques such as likelihood ratio tests and Akaike's

Information Criterion have proven useful in selecting the most parsimonious model (Burnham and Anderson 1992). One possible solution, therefore, is to incorporate these techniques into analyses of count data that use likelihood-based estimation techniques (e.g., loglinear modeling) to determine whether inter-observer differences should be modeled on a case-by-case basis. Another solution is offered by an extension to the estimating equations approach to trend estimation, in which observer differences may be modeled using relatively few parameters, with a corresponding increase in efficiency (Link and Sauer 1994). A third possibility is to apply the BBS field technique in areas where more intensive studies are taking place in an attempt to quantify observer differences and, more generally, to estimate the effect of variation in detectability. Lastly, simulation studies can be used to determine the robustness of the methods of analysis to inter-observer variation and other systematic sources of error (e.g., Geissler and Link 1988).

## Missing data

Missing data occur when sites are not surveyed in a particular year, when sites are discontinued or new ones initiated, or when the location of a site changes. Such occurrences are a ubiquitous feature of extensive volunteer-based surveys: in the BBS approximately $35 \%$ of sites are missed each year. Complete time series are not required to estimate population trends, although missing data does decrease the precision of the estimate. For this reason many published trend analyses do not include sites that are rarely or irregularly surveyed, although criteria used to exclude sites varies widely (e.g., Geissler and Sauer 1990, Böhning-Gaese et al. 1993, James et al. 1996). A complimentary technique used to increase overall precision is to weight site trends by their precision when averaging them to produce regional trends ("precision weighting'). In this way, site trends estimated with low precision have little influence on the overall population trend. The calculation of the weightings should not be based upon the estimated variance of the site trends, because the autocorrelation of counts between years makes this estimate biased (James et al. 1990). Methods of analysis that use precision weightings (Tables 3.1 and 3.2) base the weightings on statistics that are proportional to the expected precision, calculated from the number and distribution of surveys.

For annual indices, imbalance in the data structure caused by missing observations is a potential source of bias and must be accounted for (Geissler and Noon 1981, James et al. 1990, ter Braak et al. 1994). A number of approaches have been taken, most of which involve either implicitly or explicitly replacing missing observations with predicted values (Table 3.1). The chaining method is an exception in that it creates balance by excluding observations. This further decreases precision of chaining indices and is one reason why that method is not recommended (Geissler and Noon 1981, Mountford 1985, Peterjohn et al. 1995).

Missing data and measurement error combine to obscure the true pattern of population change over space and time, and thus make the selection of an appropriate statistical model of counts difficult. A number of methods assume that counts are log-normally distributed (Tables 3.1 and 3.2). This simplifies the calculation of estimates, but has a number of disadvantages, including the implicit assumption that counts are continuous (i.e., non-integer), and the need to add a constant before log-transformation (Collins 1990, Link and Sauer 1994). These problems are avoided by loglinear modeling (Tables 3.1 and 3.2), which assumes that counts are a Poisson random variable. All of these methods may, however, be biased if the model of counts is incorrect (e.g., Geissler and Link 1988). Because of this a number of trend estimation methods use non-parametric estimators, which do not require specification of the distribution of counts (Table 3.2), but are usually less efficient where a parametric model would be justified.

## Evaluating the methods using Monte Carlo simulations

From the above discussion it is clear that there are a number of unresolved issues that prevent a complete evaluation of the relative merits of the methods of analysis. One approach is to use Monte Carlo simulations (Manly 1991) to generate count data containing known population changes, and to use these data to evaluate the methods of analysis. While the true patterns of population change and measurement error over space and time are not known, a number of plausible hypotheses can be used to construct realistic count data. These data can then be used to determine the accuracy of the methods under a variety of models for population change. An example of the approach is given below.

Partners in Flight, the multi-agency coalition responsible for initiating the Neotropical Migratory Bird Conservation Program (Finch and Stangel 1993), has suggested that continental monitoring programs such as the BBS should be able to detect a $50 \%$ decline in the population of a species over 25 years with $90 \%$ power (Butcher et al. 1993). The purpose of the simulation study described here was to evaluate the bias and efficiency of three BBS trend-estimation methods in the context of linear-multiplicative declines of this magnitude and under three different geographic scenarios of population change. The methods evaluated were the same as those compared in Chapter 2: the U.S. National Biological Service (USNBS) implementation of linear route regression (Geissler and Sauer 1990), the Canadian Wildlife Service (CWS) implementation of the same approach (Erskine et al. 1992) and rank-trends analysis (Titus 1990). The USNBS and CWS co-administer the BBS program, and results from their analyses are widely cited in secondary publications. The methods of analyses are very similar: both logtransform the count data, calculate trends for each site (route) as the slope of a linear regression of log count against time (hence "linear route regression"), and take a weighted average of these slopes to produce an overall trend for the population of interest. However the implementations differ in many details, such as the criteria for the exclusion of data, the constant that is added before log-transformation, the treatment of observer differences, the way that site trends are averaged to form the population trend (arithmetic mean in the USNBS implementation, geometric mean in the CWS version), the method of back-transforming the population trend from the $\log$ scale to the arithmetic scale, the method of calculating area, abundance and precision weightings, and the technique used to estimate confidence intervals. Rank-trends analysis is a nonparametric method that estimates the direction and consistency (but not the magnitude) of population change over time at each site, and averages these statistics over sites to produce an estimate of population change. No weightings are used, and no attempt is made to account for observer differences. In Chapter 2, I found that these three methods produced different results when they were applied to the same subset of BBS data; however, because there was no baseline of known trends, I could not tell which method was the most accurate.

Wilcove and Terborgh (1984) have proposed that species undergoing population declines will show one of three broad geographic patterns of population change: (1) declines in abundance at the center of the species geographic range but little contraction in the size of the range, (2) range contractions with little change in abundance at the center, and (3) both declines at the center and range contraction. These scenarios were modeled by assuming that species abundance was proportional to the normal distribution with respect to distance from the center of the range (Brown 1984, Lacy and Bock 1986, Lawton 1993), as shown in Fig. 3.2. Population declines were simulated by reducing the overall abundance by $2.73 \%$ per year (i.e., $50 \%$ over the 25 years of the survey, Fig 3.2). Annual variation in addition to trend was modeled by multiplying the overall abundance by a random log-normal variable with mean (on the log scale) of zero and variance set by a model parameter.

The actual count at a site was calculated as the abundance at that site, given its distance from the center of the species range and the year, multiplied by variables representing random within-site variation and measurement error due to inter-observer differences, and rounded to the nearest integer value. Within-site variation was modeled as a random variable in a manner similar to annual variation. Each observer was assigned an observer error from a random lognormal distribution with a mean that was proportional to the year that the observer began surveying the route and variance set by a model parameter. This allowed the modeling of increases or decreases in inter-observer quality over time (Sauer et al. 1994b). The locations of sample sites were selected at random from the circular sample space centered around the center of the species range and with a radius such that $99.9 \%$ of the initial population was contained in the space. The pattern of surveillance at each site was simulated using three parameters: the probability that a new observer adopts the site, the probability of the observer surveying the site in a given year and the probability of the observer never surveying the site again.

In a full simulation study, data would be generated at a number of levels for each parameter. Here, for the purposes of illustration, results are presented for each geographic scenario of population change at one level of the other parameters (Table 3.3). These parameter values were chosen to represent median values obtained from an analysis of BBS data for British

Columbia. For each scenario, 100 sets of data were generated and analyzed by the three methods. The main features of the results were:
(1) Both route regression methods were efficient in "detecting" the declines (i.e., high proportion of significantly declining trends) in all three scenarios (Fig. 3.3). O'Connor (1992b) suggested that these methods may be less sensitive to declines that are concentrated at the edge of species ranges, due to the abundance weightings used. No evidence for this was found in these simulations.
(2) Rank-trends analysis produced inaccurate results when declines were concentrated at the center of the species range (Fig. 3.3). Observer differences were not accounted for in the rank-trends method, but increasing observer quality was a major component of this simulation, resulting in increasing counts on some routes. Rank-trend estimates for scenarios involving range contraction were "correct" due to the rapid decline in abundance at peripheral sites outweighing the tendency for increases due to observer differences. Since route trends are not weighted by abundance in this method, trends at peripheral sites were given equal weight to those at the center. These results demonstrate the potential importance of accounting for observer differences, and the effect of weightings on the overall estimates.
(3) The CWS route regression method showed a small negative bias in all three scenarios (median $-0.82 \%$ year $^{-1}$ ), while the USNBS method was not significantly biased (Table 3.4). The bias may be due to the use of the geometric mean in estimating overall population trends from route trends rather than the arithmetic mean, although there are many other differences between the methods (Chapter 2).
(4) Both route regression methods underestimated the standard deviation of the trend estimates (Table 3.4). This implies that the frequency of type I errors (significant population trends detected when none are occurring) may be higher than the nominal level of $\alpha=0.05$, and suggests that estimates of the number of species adequately sampled by the BBS based upon power analysis (e.g., Peterjohn et al. 1995) may be overestimates. Simulations under the null hypothesis of no population change are required to evaluate this possibility.

In conclusion, this example has demonstrated the utility of simulation studies in highlighting potential areas of bias in the methods under a small number of "What if..." scenarios. A more complete study, over a wider range of conditions, is required to confirm the preliminary results presented here.

## Summary

In this chapter I considered the issues involved in the estimation of long-term population change from multi-site count data. It is not feasible to census the entire population of most bird species, so monitoring programs are based on sample surveys. In an ideal sample survey the population is divided into sample units (sites), a random sample of these units is selected, and the selected units are censused at regular intervals. When analyzing the resulting data, two subjective decisions are required: (1) the primary unit of analysis (annual indices or population trend), and (2) the assumed model of trend (linear multiplicative, polynomial multiplicative, additive or rank). Real surveys often depart from the ideal in three ways: (1) the population from which the sample is drawn is a subset of the population about which information is required (incomplete sample frame), (2) the number of individuals at each sample site is not counted accurately (measurement error), and (3) most sites in the selected sample are not censused every year (missing data). The analysis methods take different approaches to deal with these problems and it in many cases it is not possible to determine which is the best on theoretical grounds alone. Simulation studies offer a method of evaluating the analysis methods under known conditions. I give an example where three analysis methods are tested using three geographic scenarios of population decline: declines at the center of the range, declines at the periphery and a mixture of the two. One of the methods, rank-trends analysis, produced inaccurate results in one scenario. Another, Canadian Wildlife Service route regression, was slightly negatively biased. Both this method and US National Biological Service route regression underestimated the variance of their estimates of trend. Although the example is simplistic, it demonstrates the potential for the use of simulation studies to test the methods under realistic scenarios of population change.
Table 3.1. Methods that have been used to estimate annual indices from bird count data.

| Method: | Chaining, |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |

Table 3.2. Methods that have been used to estimate population trends from bird count data.

| Method: | Linear route regression |  | Nonlinear route regression |  | Loglinear modeling ${ }^{6}$ | Regression of annual indices ${ }^{7,8}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Least squares ${ }_{2}^{1,}$ | Estimating equations ${ }^{3}$ | Smoothing ${ }^{4}$ | Rank trends ${ }^{5}$ |  |  |
| Unit of analysis | count | count | count | count | count | annual indices |
| Model of trend | linearmultiplicative | linear- / polynomialmultiplicative | additive | ranks | linear- / polynomialmultiplicative | linear-/ polynomialmultiplicative |
| Population trend calculated as | average of site (route) trends | population trend effect from model | average of site (route) trends | average of site (route) trends | population trend effect from model | regression of annual indices |
| Assumed distribution of counts | log-normal | not specified / Poisson | not specified | not specified | (overdispersed) Poisson (with autocorrelation) | $\mathrm{n} / \mathrm{a}$ |
| Transformation of counts (c) | $\begin{aligned} & \log (c+0.5)^{1} / \\ & \log (c+0.23)^{2} \end{aligned}$ | none | $\sqrt{c}$ | none | none | $\mathrm{n} / \mathrm{a}$ |
| Abundance weighting? | yes | yes | implicit | no | not developed | $\mathrm{n} / \mathrm{a}$ |
| Observers incorporated? | yes | yes | no (nonparametric)/ yes (semiparametric) | no | possible | $\mathrm{n} / \mathrm{a}$ |
| Precision weighting? | yes | yes | no | no | not developed | $\mathrm{n} / \mathrm{a}$ |
| Variance estimated by | bootstrap ${ }^{1 /}$ jackknife ${ }^{2}$ | bootstrap | parametric | parametric | parametric | bootstrap ${ }^{7} /$ not estimated ${ }^{8}$ |

[^2]Table 3.3. Parameters used in the simulation study of the performance of three BBS trendestimation methods in the context of overall population declines and under three different geographic scenarios of population change. Where three values are given, the first refers to the scenario declines at center of range, the second to range contraction and the third to both declines at the center and range contraction (see Fig. 2).

| Parameter | Value |
| :--- | :--- |
| Abundance in year 0 at center of range | 50 |
| Variance of range in year $0\left(\mathrm{~km}^{2}\right)$ | 90000 |
| Rate of change of abundance at center of range (\% year $\left.{ }^{-1}\right)$ | $-2.73 / 0 /-1.37$ |
| Rate of change of range variance (\% year $)$ | $0 /-2.73 /-1.37$ |
| Year-to-year fluctuations about trend (variance, log scale) | 0.16 |
| Within-site variation in counts (variance, log scale) | 0.04 |
| Rate of change of inter-observer mean (log scale) | 0.05 |
| Inter-observer variation (variance, log scale) | 0.16 |
| Site density (sites km²) | 0.0005 |
| Probability that a new observer adopts a site | 0.1 |
| Probability of an observer surveying a site in a given year | 0.8 |
| Probability of the observer never surveying the site again | 0.2 |

Table 3.4. Bias and precision of two trend estimators (USNBS ${ }^{1}$ and CWS ${ }^{1}$ linear routeregression) from 100 simulations of Breeding Bird Survey data at each of three geographic scenarios of population decline.

| Scenario | Median bias ${ }^{2,3}$ |  | Median estimated std. dev. ${ }^{2,3}$ (Actual std. dev.) ${ }^{2}$ |  |
| :---: | :---: | :---: | :---: | :---: |
|  | USNBS ${ }^{1}$ | CWS ${ }^{1}$ | USNBS ${ }^{1}$ | CWS ${ }^{1}$ |
| Declines at center | 0.27 n.s. | $-0.64 * * *$ | $\begin{gathered} 0.54 * * * \\ (1.45) \end{gathered}$ | $\begin{gathered} 0.60 * * * \\ (1.28) \end{gathered}$ |
| Range contraction | -0.14 n.s. | $-1.11^{* * *}$ | $\begin{gathered} 0.27 * * * \\ (1.29) \end{gathered}$ | $\begin{gathered} 0.22 * * * \\ (1.44) \end{gathered}$ |
| Both | -0.21 n.s. | -0.86 *** | $\begin{gathered} 0.60^{* * *} \\ (1.36) \end{gathered}$ | $\begin{gathered} 0.68 * * * \\ (1.40) \end{gathered}$ |

[^3]
## Figure legends

Figure 3.1. Four models of trend applied to the same hypothetical count data.
Figure 3.2. Three geographic scenarios of population decline. In each case the total population declines by $50 \%$ over 25 years. Curves show the distribution of abundance across a cross section of the species range at successive 5 year intervals, with the outer top line being year 0 and the inner bottom line being year 25 .

Figure 3.3. The direction and statistical significance of trends estimated by three methods of trend analysis from 100 simulations of BBS data. USNBS $=$ U.S. National Biological Service linear route-regression method; CWS = Canadian Wildlife Service linear routeregression method; RT = rank-trends method.





Figure 3.2


Figure 3.3


## Chapter 4

## Four Case Studies

## Introduction

In Chapter 2, I showed that the way in which Breeding Bird Survey (BBS) data are selected and analyzed can have important consequences for the inferences drawn about regional population trends. It is therefore important to choose the method of analysis carefully, and to evaluate the results cautiously. Here I provide a detailed empirical evaluation of the results when eight candidate analysis methods are applied to continent-wide BBS data for four species. These four case studies allow the performance of many methods to be discussed at a level of detail not possible when all species are being analyzed. Previous case studies (Geissler and Noon 1981; Mountford 1982; Sauer and Droege 1990c; Peach and Baillie 1994) and comparisons of results across methods based on many species (Link and Sauer 1994; Peterjohn, Sauer and Link 1996 Appendix 1; Chapter 2) have included only a small subset of the available methods.

The analysis methods are listed in Table 4.1, and are described in Chapter 1 and Tables 3.1 and 3.2. The six route regression (RR) methods were chosen because they have all been advocated in recent analyses of BBS data (refs. in Table 4.1). The other two approaches, loglinear modeling and the Mountford moving windows method, have been applied to other extensive bird surveys, and have been recommended for general use. A secondary aim of this chapter, therefore, is to investigate the feasibility of applying them to the BBS.

The four bird species were chosen to represent extremes with respect to abundance (average count per route) and sample size (number of routes). Trends for low abundance species may be poorly estimated by methods that log-transform the data before analysis (e.g., least squares RR, Table 4.1), because of the need to add a constant before log-transformation (Geissler and Link 1988; Collins 1990; Link and Sauer 1994; Chapters 1, 2 and 3). High abundance species can show large variation in counts between sites and time periods and may therefore be
poorly modeled by some estimators. Differences among methods may be more apparent in species that are seen on few routes and therefore have a small sample size (Chapter 2), but widespread species that occur on many routes will likely show greater regional variation in trends. I therefore chose to analyze data for the following species:
(1) the Cerulean Warbler (Dendroica cerulea), an eastern-forest breeding Neotropical migrant that is seen on relatively few routes and in low numbers, and is thought to be sensitive to forest fragmentation and be in decline (Ehrlich et. al. 1988; Robbins et al. 1992; but see Villard and Maurer 1996);
(2) the Carolina Wren (Thyrothorus ludovicianus), a widespread and abundant species, resident in the eastern and central US, that appears to have undergone a population crash in the harsh winter of 1976 (Robbins et al. 1986) and may then have recovered (Barker and Sauer 1992);
(3) the Chestnut-collared Longspur (Calcarius ornatus), a central prairie dwelling shortdistance migrant that occurs in relatively large numbers but on few BBS routes and may be declining along with many other grassland species due to overgrazing and modern farming practices (Johnson and Schwartz 1993); and
(4) the Belted Kingfisher (Ceryle alcyon), a pan-continental but low abundance shortdistance migrant that inhabits riparian areas and is not thought to be of conservation concern.

For all of these species we have a vague idea of the broad-scale patterns of population change, but the actual levels of change are unknown. Because the "truth" is unknown, it is not possible to tell which analysis method produces the most accurate results (simulations are . required for this, Chapter 3). Nevertheless, as we shall see, many of the methods give results for some species that are not feasible. Detailed case studies are a useful way to reveal these problems and so to make preliminary judgements about the reliability of the methods.

I present the results of the Cerulean Warbler first and in some detail, and then treat the other species in successively less detail as common patterns emerge. I discuss the performance of the methods and, based on this and the previous chapters, suggest how they may be improved.

## Methods

Data selection
All methods require some amount of data screening to block out poorly performed surveys and to ensure that the data are not too sparse to allow the estimator to work. Selection criteria vary widely among published analyses, from stringent criteria that remove all but the best surveys and exclude all routes except those with a nearly continuous record of surveys (e.g., James et al. 1996) to tolerant criteria that remove only the most questionable surveys and exclude only those routes with too few surveys to allow population change to be estimated (e.g., Link and Sauer 1994). Even small differences in data selection criteria can cause important differences in results (Chapter 2).

In this analysis I used two sets of criteria: (1) stringent selection criteria that excluded many routes and surveys, but allowed the estimates from (almost) all of the methods be based on exactly the same data, and (2) tolerant criteria that excluded few routes and surveys, but resulted in some of the estimators not being able to use all of the data. The tolerant criteria were used to examine the effect of a more representative amount of missing data ( $\sim 35 \%$ missing surveys per route on average, as opposed to only $\sim 10 \%$ after stringent data selection). However, differences among estimates based on tolerantly selected data will be partly caused by differences in the data that each method could use (see Results).

The stringent criteria comprised choosing only data that passed the selection criteria of all analysis methods except Canadian Wildlife Service (CWS) RR and Mountford moving windows (Table 4.1). CWS RR selection criteria were not included because the appropriate information on survey quality (subroute designation, Chapter 2) is not available outside of Canada. Nevertheless, the CWS RR estimator was able to utilize all of the data that passed the stringent selection criteria, so results are directly comparable with those of the other estimators. The Mountford moving windows estimator requires that a route has at least two surveys in a "window" (block of successive years) for that window to be analyzed. With a window size of 4 (the smallest possible), this means that there must be almost no missing data for all windows on a route to be analyzed. Very few routes qualify under this criterion ( $2.8 \%$ of the total BBS
routes). However most of the routes that pass the stringent selection criteria (Table 4.1) have only one or two windows that could not be analyzed. Thus the Mountford estimator with window size 4 used slightly less data than the other estimators (approx. 3\% fewer surveys). With window size 6 and greater, the difference in data used was trivial.

The tolerant criteria were the same as those for the U.S. National Biological Service (USNBS) RR method (Table 4.1). This resulted in a much larger sample size and a more realistic amount of missing data, but also meant that some estimators (particularly nonlinear semiparametric RR (NSRR) and Mountford moving windows with small window sizes) could not use many of the routes in the dataset.

For both criteria the initial dataset consisted of all BBS data collected between 1966 and 1995, excluding routes located in some peripheral state-strata and routes placed non-randomly or off roads (Appendix 2). For three species, data from the first one or two years could not be analyzed using TRIM model 5 because there were no surveys in some physiographic strata in those years (see Loglinear modeling in Data Analysis section). In these cases the dataset was reduced to 1967 or 1968 to 1995 before data selection.

## Data analysis

For each species, I began with a simple exploratory analysis of the data. Many of the estimators assume that the data are log-normally distributed, implying that the standard deviation of counts is proportional to their expected value. Others assume a Poisson distribution, or they square-root transform the data, assuming that the variance of counts is equal to the expected value. As an initial check, I calculated the mean, standard deviation/mean (coefficient of variation) and variance/mean on each route. If the either measure of variation divided by the mean produces nearly constant values, that measure is proportional to the mean. These analyses are preliminary because variation among counts is also caused by covariables (such as time and observers; see Loglinear modeling subsection in Discussion). To get an initial impression of the change in counts over time I calculated the mean count in each year. I also calculated the area weighted mean count, which corrected for geographic variation in sample intensity by weighting
each count by a measure of the area represented by the route (area of the physiographic stratum within the state or province divided by the number of routes).

I then estimated annual indices and population trends using each of the eight analysis methods. These methods are discussed in Chapters 1, 2 and 3, and references are given in Table 4.1. The programs used to implement the methods are listed in Appendix 1. Below I briefly describe how each method was applied to the data in this analysis. Descriptions of the loglinear modeling and Mountford moving window methods are given in more detail because I used slightly different methods from those previously applied to bird count data.

Linear route regression: In linear $R R$, linear-multiplicative trends are calculated for each route, and a weighted average of these route trends is used to estimate the overall trend for the region of interest. For each of the three linear RR methods (Table 5.1) I calculated the overall trend, and tested whether this was statistically different from zero trend at an $\alpha$ level of 0.05 using $z$-tests. (Note that the implementation of CWS RR used here was slightly different from that used in Chapter 2, reflecting recent changes in practice (Downes and Collins 1996). The abundance weight (Component 6, Table 2.1) in this case was the geometric mean of the route counts.) I calculated annual indices of abundance using the residual methods for USNBS RR (Sauer and Geissler 1990) and CWS RR (Collins 1994), but not for USNBS estimating equations RR because appropriate methods of annual index calculation have not been published for this method. For comparison with the non-linear RR estimators I also calculated the proportion of routes on which populations were estimated to be declining, and estimated confidence intervals (CIs) for this proportion using the normal approximation to the binomial distribution (Zar 1996).

Nonlinear route regression: In both smoothing methods (nonlinear nonparametric RR (NNRR) and NSRR, Table 4.1), counts are first smoothed at the route level and then the smoothed counts are aggregated into strata. These stratum averages are smoothed again before being aggregated (with area weightings) to form overall trend estimates. As with the original publications, (James et al. 1990, 1992, 1996) I used smoothing parameters of 0.5 for the route smoothing and 0.25 for the stratum smoothing. To test statistically for overall change in
abundance, I used a $z$-test to compare the mean of the estimated indices for the fifth, sixth and seventh years with the mean estimated indices for the fifth, sixth and seventh from last years in each route (James et al. 1990, 1996; e.g., for 1966-95 data the comparison would be mean estimates from 1970, 71 and 72 with those from 1989, 90, and 91).

For the rank trend method, I calculated a trend statistic D is calculated for each route, and summed these over all routes. I compared the total D with its expected value using a $z$-test, based on an expectation of no overall increase or decrease in counts over time (Chapter 2).

Loglinear modeling: I implemented this approach using program TRIM (TRends and Indices for Monitoring data, Pannekoek and van Strien 1996). TRIM allows five models for population change: (1) no change over time, (2) loglinear trend, (3) annual indices, (4) separate loglinear trend for each level of a categorical covariable, and (5) separate set of annual indices for each level of a categorical covariable. All models allow for differences in expected counts among routes. In addition, there are options to incorporate extra-Poisson variance (overdispersion) and first-order serial correlation (autocorrelation) between counts on a route into the models. Because of the exploratory nature of this analysis, I attempted to fit each of the five models both with and without overdispersion and serial-correlation. Instead of using the estimating equation method for fitting overdispersed models provided in TRIM, I estimated overdispersion for each model as Pearson's chi-squared goodness-of-fit statistic divided by the degrees of freedom (McCullagh and Nelder 1989; Lebreton et al. 1992; Anderson et al. 1994) because this is computationally more efficient. I then incorporated overdispersion into the parameter estimates by multiplying their estimated variance by the overdispersion parameter. For models 4 and 5 I used a nested set of 3 covariables, representing increasingly fine geographic scales: BBS region (Eastern, Mid, Western), physiographic stratum (up to 99 strata, Fig 1.1) and route. (Note that model 5 could not be fit using route as a covariable, because of missing counts in many years on most routes.) In all models, I used area weighting (as described for mean counts, above) to correct for geographic variation in the density of routes.

In addition to model-based indices of abundance, program TRIM also reports "imputed indices", which are similar in spirit to the residual method annual indices of the linear RR
methods. Imputed indices are calculated as the mean count in each year, with missing data replaced by the predicted count estimated from the model. For comparison with the linear RR indices, I calculated imputed indices for TRIM models 2 and 4.

Variation among observers in the number of birds counted can be an important factor in influencing counts at a site (Sauer et al. 1994b; Chapter 2). I therefore repeated the above modeling in program TRIM but assigning a new route number to each observer on each route. In a few cases for rare species some observers on a route saw no birds during their surveys. Routes with all zero counts cannot be analyzed by TRIM, so I set these counts to 1 for this method. This had a negligible effect on the results, while ensuring that the "with observer effects" and "without observer effects" analyses used exactly the same number of observations.

In some cases the numerical algorithm failed to converge, especially for models that included serial correlation (these use a different fitting algorithm based on estimating equations) and models that had a moderate number of estimable parameters relative to the number of observations. In these cases I used the last calculated deviance value for model selection if the algorithm appeared to be moving towards convergence (steadily decreasing difference in deviance over successive iterations) and the last deviance was less than $1 \%$ different from the previous value.

I selected the best fitting model using Akaike's Information Criterion (AIC, Sakamoto et al. 1986) rather than the Wald tests produced by TRIM, because AIC values can be compared for non-nested models (Burnham and Anderson 1992). If the largest-parameter model contained significant overdispersion, as indicated by a low $p$-value in the Pearson goodness-of-fit test, then I selected models using the quasi-likelihood AIC (QAIC: Lebreton et al. 1992; Anderson et al. 1994). To calculate QAIC for each sub-model, I used the estimate of overdispersion from the largest-parameter model (McCullagh and Nelder 1989). In most cases (see Results), the final model that was selected was not dependent on whether AIC, QAIC or Wald tests (where applicable) were used.

Mountford moving-window method: To implement the moving-window approach, I selected the appropriate window size using the goodness-of-fit test described by Mountford
(1982). I started with a window size of 4 (the minimum possible) and increased window size in steps of 2 until one or more windows failed the test, at an alpha-level of 0.05 . (This procedure is identical to that recommended by Peach and Baillie (1994), except that they started with window size 6). To quantify the uncertainty in the yearly indices at each window size, I conducted 1000 bootstrap replications (resampling routes), and calculated pointwise $95 \%$ CIs for each yearly index using the percentile method (Buckland 1984; Peach and Baillie 1994). I also calculated linear-multiplicative trend estimates, using linear regression of the Mountford indices on the log scale, and back-transforming by the method of Bradu and Mundlac (1970; see USNBS RR methods, Chapter 1). I calculated CIs for the trend estimate from the bootstrap replicates. As with the loglinear modeling, I calculated Mountford indices both with and without observer differences (i.e., coding each new observer as a new route).

## Comparison of estimates

Quantitative comparisons among estimators are difficult because they estimate different parameters (annual population index, loglinear trend or smoothed trend) at different scales (logscale or untransformed). In addition, statistical tests of differences among estimates for each species are not appropriate because the estimates are based on the same (or similar) data and so are not independent, but their covariance is unknown. I therefore focus on graphical displays of the estimates to allow qualitative comparisons. I also compare loglinear trend estimates and their corresponding CIs, for those estimators where this is possible. For route-regression estimators I compare the proportion of routes with estimated declines.

To enable comparison of indices among estimators, graphs of population change over time are scaled so that the index for the first year is 1.0 . The only exceptions are annual indices of abundance from the linear RR estimators, which are properly displayed as residuals about the linear trend line. Because population change is thought to be a multiplicative process, results are presented using a log-scaled axis for the indices: this gives equal emphasis to proportional increases as decreases (Moses and Rabinowitz 1990). I used the log base 2 because this makes the axes easy to interpret: a change in index from 1 to 2 indicates an estimated doubling of
populations and a change from 1 to 0.5 indicates a halving; both are equally spaced along a $\log _{2}$ axis.

Here I focus exclusively on estimates of population change at the continental level. Regional estimates are likely to have similar properties to the continental estimates for the two species that were seen on few routes. For comparability I use the same models for each estimator for both the tolerant and strict data selection criteria.

## Results

## Cerulean Warbler

Stringent data selection: Between 1966 and 1995, Cerulean Warblers were counted on 388 routes, but only 41 of these passed the stringent selection criteria (Table 4.2). The distribution of counts among these routes was highly skewed, with most routes having a median count of zero, while two, 82031 (Elk Valley, Tennessee) and 82032 (Smokey Junction, Tennessee), contained $31 \%$ and $11 \%$ of the total counts respectively (Fig. 4.1). The coefficient of variation (standard deviation of counts within routes/mean route count) decreased with increasing mean (Fig 4.2a, left plot), implying that counts are likely not approximately lognormally distributed. The pattern was reminiscent of a simple inverse relationship, but this would imply that the standard deviation was independent of the mean when in fact there was a significant positive relationship (least-squares linear regression slope $=0.49,95 \% \mathrm{CI} 0.52$ to 0.46 ; the relationship was similar if the two high abundance routes were deleted). The variance/mean ratio was greater than 1.0 and increased with increasing mean count (Fig 4.2a, right plot; linear regression slope $=3.72,95 \%$ CI 3.97 to 3.48 ; the slope was still significantly greater than zero if the two outlying high abundance routes were deleted), indicating possible departures from the Poisson distribution as well.

The mean counts remained relatively stable until about 1977, then declined quite sharply until the mid 1980s, with some indication of a recovery in the last few years (Fig 4.3a). The area-weighted mean counts showed a similar pattern but were slightly more negative, with the decline perhaps starting after 1972 (Fig 4.3a). The means are clearly heavily influenced by
counts on routes 82031 and 82032: for example the increase in mean count in 1970 is an artifact of the missing data on route 82031 before that time (Fig. 4.1).

All three linear RR methods estimated that Cerulean Warbler populations declined significantly over the period of the survey (Table 4.3). The magnitude of the estimated change differed among the methods (USNBS estimating equations $<$ USNBS RR $<$ CWS RR, Fig. 4.3b), although CIs for the estimates showed a large overlap (Table 4.3). The estimating equations estimate was much less precise than the other estimates, with a CI three times as wide as the comparable USNBS RR estimate (Table 4.3). Annual indices of abundance showed similar patterns for the USNBS RR and CWS RR methods (Fig 4.3b; annual indices for estimating equations approach have not been developed). In both cases the annual indices fluctuated around their loglinear trend line, which appeared to describe the overall pattern of population change well. The results from this methods imply that Cerulean Warblers have been declining at a reasonably constant rate for the time period of the survey.

Although all three nonlinear RR methods also estimated overall declines, the estimates were not statistically significant for either smoothing RR method (Table 4.3) and there appeared to be some important changes in trend over time which reflect changes seen in the mean counts (compare Figs. 4.3c and d with 4.3a). Estimates that incorporated differences among observers (NSRR, Fig. 4.3c) were slightly more positive than those that did not (NNRR, Fig. 4.3d). In both methods, indices for the last 3 years were not joined smoothly to those from earlier years, indicating a problem with the methods. NSRR estimates were less precise than those from NNRR (compare CI widths in Table 4.3). The rank-trends method estimated highly significant declines, although the proportion of routes with estimated declines was considerably less than the other RR methods and was not significantly different from 0.5 (Table 4.4).

The results of loglinear model fitting and selection using program TRIM (Table 4.5) can be summarized as follows: (1) when observer effects were included, the fit of the model to the data was adequate for all but model 1 ( $p$ for Pearson $\chi^{2}$ non-significant, Table 4.5) with no indication of overdispersion (Pearson $\chi^{2} /$ d.f. close to 1.0 , Table 4.5); (2) TRIM model 3 (separate effects for each year) had the lowest AIC and is therefore the selected model; (3) modeling serial
correlation caused a small reduction in deviance and AIC so could be included in the final model, although the estimated correlation was small for the well-fitting models (Table 4.5).

Index estimates from model 3 including observer effects (Figs. 4.3c and 4.4a) indicated that Cerulean Warblers were relatively stable or perhaps declining slowly until 1977, declined quite rapidly up to the late 1980s and then stabilized, or perhaps showed a small recovery. After 1980, all yearly indices were statistically significantly less than in the first year of the survey (upper $95 \%$ confidence limit less than the first year index of 1.0, Fig. 4.4a). Estimates with or without serial correlation were essentially identical (compare Figs. 4.4a and b), which is not surprising given the small correlation estimate (Table 4.5, upper half). If observer effects were not included the estimated decline stops in the early 1980s (Fig 4.4c), but the "without observers" model did not fit the data very well (Table 4.5). The lack of fit translated into an estimated overdispersion greater than 1.0 (Table 4.5), which resulted in slightly wider CIs when this was incorporated into the estimate of standard error (Figs. 4.4c and d). The estimated serial correlation was also nontrivial in this model (Table 4.5, lower half), which also produced wider CIs, although it had almost no effect on the estimate (compare Figs. 4.4c and d). The estimates from Model 3 "without observers" were very similar to the yearly mean counts and area weighted mean counts (compare Figs. 4.3a and d), but when observer differences were taken into consideration the indices were noticeably more negative (Fig. 4.3c).

Although the data were better modeled in TRIM as annual indices, it is informative to compare linear trend estimates from program TRIM (model 2) with those from the linear RR methods (Table 4.3). When observer effects were included, the estimated magnitude of decline was between that of the USNBS estimating equations estimator and CWS RR, but with a smaller CI than for these methods. Without observer effects, the estimated decline was still statistically significant, but was of much smaller magnitude.

The Mountford moving window method did not perform well with these data. When observer differences were incorporated, all windows passed the goodness-of-fit test for all years with window sizes 4 and 6 (Table 4.6, columns 2 and 3), but not with larger windows. Nevertheless, the estimated population index in 1991 with window size 6 (Fig. 4.5b) was
implausibly high and exceeded the CIs by a large margin. Estimates were outside of the CIs in several other years, indicating that they are probably highly biased. When observers were not included, the goodness-of-fit test gave negative $\chi^{2}$ values for one window with window size 4 and another with window size 6 (Table 4.6). Negative $\chi^{2}$ values are not admissible, and indicate that one or more of the estimated counts in those windows were below zero. In addition, the estimates for window size 6 were clearly spurious for years 1978 and 1979 (Fig. 4.5d).

The overall patterns of population change estimated with the Mountford movingwindows method did not match any of the other estimators (Figs 4.3c and d). The most credible estimates from the Mountford method ("with observers", window size 4: Fig 4.5a) showed population fluctuations but no consistent trend, and the Mountford trend estimates with this window size showed close to no trend with narrow CIs (Table 4.3).

Tolerant data selection: When the tolerant data selection criteria were applied to the data, the number of routes available to the estimators was increased from 41 to 198 (Table 4.2). This was accompanied by an increase in the percentage of missing data, from $11 \%$ to $32 \%$ (Table 4.2). The annual mean counts showed a qualitatively similar pattern to those after stringent data selection but with smaller fluctuation (compare Figs. 4.3a and 4.6a), probably indicating a reduction in variation due to sampling error. Even with almost 200 routes, however, the indices were still noticeably influenced by a few routes with high counts: for example the increases in mean counts after 1991 disappear with the removal of two routes that were surveyed only in these later years ( 66080 Dell, Ohio and 90028 Strange Creek, West Virginia).

Both the USNBS RR and CWS estimates were more positive using tolerantly selected data than with stringent data selection (Fig. 4.6b), and the residual annual indices showed much less variation. The USNBS estimating equations RR estimate, on the other hand, was more negative, and $95 \%$ CIs for the two USNBS estimates no longer overlapped (USNBS RR: estimate $-2.16 \%$ per year, $95 \%$ CI -1.65 to -2.86 ; USNBS estimating equations RR: estimate $5.30 \%$ per year, $95 \%$ CI -3.20 to -7.49 ). No identifiable small group of routes seemed to be responsible for this difference between the methods.

The smoothing RR methods were not able to use all of the data that passed the tolerant selection criteria: NNRR used 177 of 198 routes for the smoothed index estimates and 125 routes for the $t$-tests, NSRR used 121 routes for the indices and 90 for the $t$-tests. Both methods produced clearly spurious estimates of population change between 1966 and 1967 (Fig. 4.6c and d), because there were no suitable routes in three strata in 1966 (see Discussion). This year apart, the indices for both methods no longer showed the estimated stable period in the early years of the survey, but now estimated relatively constant declines over that period (compare Figs. 4.3 c and d with 4.6 c and d ). The overall estimates of population change were still not statistically significant for either method (NNRR: estimate $-2.42 \%$ per year, $95 \% \mathrm{CI}+0.76$ to 4.76; NSRR estimate $-4.60 \%$ per year, $95 \%$ CI +1.26 to -8.07 ).

The results of model fitting and selection in program TRIM differed slightly from the stringently selected dataset: models 3 and 4 (strata) "with observers" had almost the same AIC (4481.99 and 4481.68 respectively; the next closest was model 2 "with observers" 4494.79. Note that AIC values are only comparable across models fitted to the same data so these values should not be compared with those from the stringently selected data). Model 5 (strata) would likely have had an even lower AIC, but unfortunately it could not be fit because there were no surveys in several years for several strata.

Index estimates from model 3 were more positive using the tolerant data selection criteria (Fig. 4.6c and d) than the stringent criteria (Fig. 4.3 c and d), particularly for the years 1979 1992 in the "with observers" model. Once again the "without observers" model estimates are very similar to the yearly mean counts (compare Figs 4.6 a and d), except for the last few years. In addition both model estimates track the smoothed indices from NNRR and NSRR quite well (excluding the first year, Fig. 4.5c and d). Estimated trends using Model 2 were also more positive ( $-2.96 \%$ per year "with observers", $-1.25 \%$ per year "without observers"; compare with Table 4.3). CIs with the larger dataset were narrower, although the estimated overdispersion in the "without observers" models was comparable to those reported in Table 4.5.

The Mountford method was able to make use of some of the extra data (e.g., mean number of routes used per window, "without observers" models: 146.5 of a possible 198 for
window size $4,156.9$ for window size 6 ). The data passed the goodness-of-fit test for window sizes 4 and 6 for both "with observers" and "without observers" models, and the estimates exceeded the $95 \%$ consistency bounds on only two occasions, and then by only a small amount. Nevertheless, the estimated indices showed a similar overall pattern to those from the stringently selected data: seemingly random fluctuations that did not match estimates from any of the other methods.

Summary of estimates: Excluding the Mountford method results, and the few clearly spurious estimates from the smoothing RR methods, the results for Cerulean Warbler showed overall declines of somewhere between $30 \%$ and $90 \%$ between 1966 and 1995 (Table 4.3 and tolerant selection criteria results above). Declines were either relatively continuous over that time period (Linear RR methods: Figs. 4.3b and 4.6b) or occurred largely in the late 1970s and early-to-mid 1980s (other methods: Figs 4.3c, d and 4.6c, d).

## Carolina Wren

Stringent data selection: After stringent selection of Carolina Wren data, 268 routes remained from a possible 1435 (Table 4.2). Counts from 1966 were excluded because there were no counts in some strata in that year (see Data Selection Criteria in Methods section), leaving 29 years of data. Both the mean and median counts were much higher than for the Cerulean Warbler (Table 4.2), but the pattern of decreasing standard deviation/mean with increasing mean was similar (Fig. 4.2b, left plot). On average, the variance was appreciably larger than the mean (mean $($ variance $/$ mean $)=4.33$; median 3.54 ) indicating overdispersion relative to a Poisson distribution. The variance/mean ratio increased with increasing mean (Fig 4.2 b , right plot; least-squares linear regression slope $=0.078,95 \% \mathrm{CI} 0.112$ to 0.044 ), again indicating departure from the Poisson model.

The yearly mean counts clearly demonstrated the main feature of this time series: the catastrophic effect of the cold winter in 1976 (Fig. 4.7a). Both before and after this, it appeared that counts of Carolina Wrens were increasing at a near-exponential rate (i.e., linear on the log scale). This pattern was mirrored in the residual indices from the USNBS and CWS RR methods (Fig. 4.7b). There was also some evidence of a second recent reduction in numbers
(1993 to 1995), although of much smaller magnitude. All three linear RR estimators gave almost identical estimates of trend over the whole time period (Table 4.7), although these are not particularly meaningful given the obviously non-linear changes.

The smoothing RR methods poorly depicted the dynamics of this species, because the smooths obscured the interesting pattern. In both NNRR and NSRR, the decline was estimated to have been of relatively small magnitude, and to have occurred gradually over the period 1974 to 1981 (Fig 4.7c and d). The methods gave similar estimates, except that NNRR estimated a stronger recovery at the end of the time period. The nonparametric rank-trends estimate was also uninformative: it showed an overall statistically significant increase over the whole time period (Table 4.7), but provided no other information.

None of the loglinear models fit the data very well (Pearson $\chi^{2} / \mathrm{df}=2.41$ to 4.32 and goodness-of-fit $p<0.001$ for all models, Table 4.8), indicating either that the counts were overdispersed relative to a Poisson distribution or that the models were incorrectly specified (see Discussion). Assuming for the moment that the lack of fit was due to overdispersion, I selected a model based on the QAIC criterion with the overdispersion parameter equal to 2.41, the estimate from the highest-dimension model for which an estimate was available (model 5 (region) "with observers"; the higher-dimension models 5 (stratum) and 4 (route) failed to converge). The model with the lowest QAIC also had the lowest AIC, so in this case the model selection process was robust to the choice of overdispersion parameter.

Models that included observer effects fit considerably better than those that did not, and models including serial correlation fit slightly worse than those that did not (Table 4.8). Models with separate parameters for each time period fit better than those with one overall trend parameter, and, generally, the finer the spatial scale of the covariable the better the model fit. Model 5 (stratum) with observer effects had the lowest QAIC and AIC (Table 4.8). Unfortunately, parameter estimates were not available for this model because the fitting algorithm did not reach convergence. I therefore selected the next best model, Model 5 (region), as the model to use for parameter estimates (Figs. 4.7 c and d). Population indices from this
method were generally similar to the residual linear $R$ R indices, although the amplitude of the population fluctuations (peaks in 1976 and 1992 and troughs in 1977 and 1994) were less.

The Mountford moving window estimates were again problematic. When observer differences were ignored all windows passed the goodness-of-fit test for window sizes 4,6 and 8 , and one window narrowly failed the test at window size 10. Although population indices for all windows showed the decline from 1976 to 1977, the estimated size of the decline increased with window size, as did the estimated rate of recovery (Fig. 4.8). The recovery rates were implausibly fast for window sizes 8 and 10. Bootstrap CIs were suspiciously narrow for all window sizes, indicating that important sources of variation were likely being ignored. When observer differences were included, one window narrowly failed the goodness-of-fit test at window sizes 4,8 , and 10 , but the index estimates were very similar to the "with observer" indices. In both cases using window size 6 gave the most plausible index estimates, but these indices still showed a relatively flat trend after the mid 1980s when all other methods indicated an increase in abundance (Figs. 4.7c and d).

Tolerant data selection: Population change estimates for all methods were almost identical to those described for the stringent data selection criteria (not shown). The weighted annual means were slightly more positive towards the end of the time series, and this was reflected in more positive overall trend estimates (approx. $0.5 \%$ per year) for all of the estimators except the Mountford method.

Summary of estimates: The well-known crash in Carolina Wren populations between 1976 and 1977 was documented in all of the estimators except the loess smoothing methods, although estimates of the size of the decline varied (approximately $40 \%$ reduction for Mountford method, $50 \%$ for TRIM model 5 (region), and $70 \%$ for linear RR methods, Fig. 4.7). All of the estimators except the Mountford method indicated that Wren populations were increasing at a near-exponential rate both before and after the crash (Fig. 4.7); there was also some suggestion of a recent set-back in numbers between 1992 and 1994 (Fig. 4.7b).

## Chestnut-collared Longspur

Stringent selection criteria: For the Chestnut-collared Longspur, only 21 of a possible 190 routes remained after stringent data selection. There was a very large variation in the number of birds counted per survey, from 0 to 340 , and the average counts were much higher than for the other species analyzed (Table 4.2). Nevertheless, the patterns of declining standard deviation/mean and increasing variance/mean with increasing mean per route appeared to be similar to those of the other species (Fig. 4.2; not investigated quantitatively because of the small sample size).

The annual mean counts fluctuated around the initial value until 1982, after which there was some suggestion of a step to a lower level (Fig. 4.9a). This pattern was reflected to some extent in the residual linear RR indices (Fig. 4.9b), although an increase in indices in the mid 1970s gave the impression of little overall change. Trend estimates were almost identical for the three linear RR methods: all showed a non-significant decline of approximately $0.5 \%$ per year with relatively similar CIs (Table 4.9). The nonlinear RR estimates were more negative, with the NNRR and rank-trend estimates being statistically significant (Table 4.9). These were also the only RR methods where the proportion of routes with estimated declines was significantly different from 0.5 (Table 4.4). The estimated indices for NNRR (the "without observers" model) matched the overall pattern in the mean counts (compare Figs. 4.9 a and d), while the estimates for NSRR were somewhat similar to the pattern in the residual linear RR indices (compare Figs. 4.9b and c).

The loglinear models fit the data very poorly, with an estimated overdispersion for the highest parameter, best fitting model of 7.02 (all models failed the Pearson goodness-of-fit test with $p<0.001$ ). Using AIC as the model selection criterion, the best model was model 5 (stratum) "with observers" (Deviance 2618.95, 139 parameters, AIC 2896.95, QAIC 651.24). However, when overdispersion was taken into account (i.e., using QAIC), the lower-dimension model 4 (route) "with observers" was favored (Deviance 3056.89, 76 parameters, AIC 3208.89, QAIC 587.65). As with previous species, models that included observer effects invariably had lower AIC and QAIC. Including serial correlation in the models worsened the fit in all cases.

The index estimates for model 4 (route) "with observers" showed an initial decline, but then an increasing trend over the period of the survey (Fig 4.9c). This was quite different to the estimates from the linear RR methods, which also fit individual loglinear trends to each route (but used a different method to combine these trends into regional estimates). The corresponding "without observers" model estimates showed an increasingly shallow declining trend, closer to the estimates from the other methods (Fig. 4.9d).

Imputed indices for model 4 (route) may be expected to show similar variation to the residual indices of abundance in the linear RR methods, especially for the "with observers" model. However, the method that I used to incorporate observer differences (coding each new observer as a new site) created many missing cells in the data matrix, and the use of model estimates in place of this "missing data" resulted in imputed indices that were very similar to the model indices (Fig 4.9c). The imputed indices in the "without observers" model (Fig. 4.9d) were similar to the linear $R R$ residual indices and the mean counts in the latter part of the survey (after 1985), but not in the first part of the survey, where there were more missing data.

The Mountford models did not pass the goodness-of-fit test for any window size from 4 to 8 , giving negative $\chi^{2}$ values in many cases. Population indices showed unfeasible estimated population changes in some years (e.g., Fig. 4.9c and d), exceeding the $95 \%$ CIs which were themselves very wide (e.g., for window size 4 the index estimate for 1995 was 1.43 with $95 \%$ CI 0.28 to 28.12).

Tolerant selection criteria: With tolerant data selection, the number of useable routes increased to 128. The annual means showed a somewhat different pattern, with an initial rise in counts followed by a large decline between 1975 and 1985 to about half the initial level (Fig. 4.10a). The CWS RR trend estimate was nearly unchanged, but the USNBS RR and estimating equations estimates were substantially higher (Fig. 4.10b) with much greater variance (see Discussion). Like the CWS RR indices, the NSRR smooth showed no overall trend (Fig 4.10c), while the NNRR smooth was similar to the annual means (Fig. 4.10d). In program TRIM, neither model 4 nor model 5 reached convergence, so parameter estimates were not
available. The Mountford moving window indices showed small fluctuations about the initial value, but they did not appear to correspond with either the annual means or CWS indices.

Summary of estimates: It is difficult to summarize the findings for this species because the methods gave so many different estimates, and the effect of data selection appeared to be strong. Both the annual means and the NNRR smoothed indices indicated that the population declined during the mid 1980s, but neither of these took account of possible differences among observers. The other methods either indicate no overall change, or gave unrealistic estimates. Species like the Chestnut-collared Longspur that show a very wide variation in counts both spatially and temporally seem to be particularly difficult to analyze.

## Belted Kingfisher

Stringent selection criteria: Belted Kingfishers were counted on over 2000 routes throughout the area of survey coverage, but the maximum count on any one route was 15 (far fewer than the other species, Table 4.2). Surveys on the western routes were begun in 1968 so only data from 1968 to 1995 were used in this analysis. After stringent data selection 529 routes remained, with a mean count per survey of 0.58 . The patterns of declining standard deviation/mean and increasing variance/mean with increasing mean were similar to the other species (Fig. 4.2d; least-squares linear regression slope $=1.67,95 \%$ CI 1.78 to 1.56 ).

Compared with the other species, the annual means showed relatively small fluctuations, although (Fig. 4.11a). Overall there was perhaps some evidence of a gradual decline in mean count, although this was largely due to several high count years between 1972 and 1977. The small overall decline was reflected in the linear RR estimates, all of which were statistically significant (Table 4.10). The estimating equations estimate was noticeably lower than the other two, but all three CIs overlapped. The pattern of fluctuations about the trend line mirrored those of the mean counts, but the amplitude was considerably less (compare Figs. 4.11a and b).

The nonlinear smoothing RR methods both showed an increase in counts up to 1973 followed by a consistent decline (Fig. 4.11c and d). This was different from the other methods, but the difference was mainly due to an estimated decline between the first and second years in all other methods. Because the statistical tests for trend in these methods only considered the
period from 1972 to 1991 (see Methods) the initial increase was excluded and both NNRR and NSRR estimated an overall significant decline (Table 4.10). The rank-trends method also estimated an overall significant decline (Table 4.10). The proportion of routes with estimated declines was similar among the RR methods, and all were significantly greater than 0.5 (Table 4.4).

The outcome of the loglinear modeling was different from previous species because models without observer effects had a lower AIC and QAIC than those that accounted for observer differences (Table 4.11). The models fit the data better than for the previous two species, but the large sample size meant that all models failed the Pearson goodness-of-fit test ( p $<0.001$ ) even though the estimated overdispersion was relatively small ( 1.11 to 1.24 ). The final model included parameters for each year, but surprisingly (given the large amount of data and the wide geographic range of the species) did not include region or stratum effects. Note, however, that neither model 4 (route) nor model 5 (stratum) reached convergence and these could conceivably provide a better fit. Models with or without serial correlation were very similar. The estimated indices were very similar to the annual mean counts (Figs. 4.11 c and d ).

Unlike the previous species, all windows passed the Mountford goodness-of-fit tests for window sizes from 4 to 16 for the "with observers" data and 4 to 14 "without observers". I therefore used the window size 14 for both datasets. The annual indices were very similar to those from TRIM model 3 (Fig. 5.11c and d) except that the decline in abundance between 1977 and 1979 was larger and became the major feature of the dataset. The linear trend estimates were well within the range of the other methods (Table 4.10). As with the loglinear model estimates, there were only minor differences between the "with" and "without observers" indices.

Tolerant selection criteria: When data were selected using the tolerant criteria the sample size nearly tripled (Table 4.2), but the estimates remained essentially the same (Fig. 4.12). The only exception was the NSRR method, which estimated strong declines at the end of the time series (Fig 4.12c).

Summary of estimates: There was generally good agreement among the methods for this species, apart from estimates for the first two years in the smoothing RR methods, and
estimates for the period 1977 to 1979 in the Mountford moving window method. There appears to have been a small but significant decline ( $\sim 15-35 \%$ ) in the abundance of Belted Kingfishers that either occurred mainly in the years 1977-79 (Mountford moving window method) or was relatively continuous from the mid-1970s onward (other methods).

## Summary of performance of methods

The methods produced quantitatively similar estimates in many cases, but there were also a number of obvious problems (Table 4.12). The linear RR methods produced seemingly reasonable estimates in all but one case (the tolerantly selected Chestnut-collared Longspur data, Fig 4.10b). There were some quite differences in results among the three implementations (e.g., Fig 4.6b, CIs given in text), although the $95 \%$ CIs often overlapped indicating that high variance of estimates was a contributing factor. Annual indices based on the residual method appeared to be biased toward the trend line in some cases (Figs. 4.3b and 4.10b).

The smoothing RR methods produced seemingly realistic summaries of the patterns in some cases (Fig. 4.9c and d). However, they were plagued by problems at the beginning and end of the time series (Fig. 4.3c and d, Fig. 4.6c and d). The two different implementations, NNRR and NSRR, could produce quite different estimates (Fig. 4.10 c and d, Fig. 4.12 c and d), but the $95 \%$ CIs for the overall trend always overlapped. The rank-trend method apparently suffered no problems, but it does not give enough information in the results for lack of fit to be judged to the same extent as the other methods.

Loglinear modeling gave seemingly reasonable results for three of the four species, although many of the more complex models did not reach convergence. Estimates for Chestnutcollared Longspur were clearly unreliable, and the fit of the models (judged by deviance/d.f) was very poor.

There were significant problems with the Mountford moving-window method, including highly unrealistic estimates (Fig. 4.5), estimates that exceeded the 95\% bootstrap CIs (Fig. 4.5b and d), and difficulties in selecting an appropriate window size (Fig. 4.8). The method gave seemingly reliable results in only one case, the Belted Kingfisher (Figs. 4.11 and 4.12).

## Discussion

Empirical comparisons of estimates cannot determine which estimation method is the most accurate, because the true parameter values are unknown. Nevertheless, case studies are useful for testing expectations about the differences among methods, for exposing situations under which they produce implausible estimates, and for suggesting areas that need further development. Here I discuss each group of estimators in turn. I place the results in the context of previous work, consider what inferences can be made about the reliability of the methods, and suggest some possible improvements to the methods.

## Linear route regression

Linear RR methods are probably the most extensively studied of the population trend estimators. Several studies have investigated the bias caused by the addition of a constant before log-transformation in both CWS and USNBS RR methods (simulations by Geissler and Link 1988; Collins 1990; Link and Sauer 1994). Adding a constant biases the route trend estimates towards zero (i.e., no trend); this effect is greater for larger values of the constant (although the reverse is not always true, Collins 1990) and for routes that have few birds counted. In Chapter 2 I showed that there was a tendency for USNBS RR estimates ( 0.5 added) to be closer to zero than CWS trend estimates ( 0.23 added), and for both estimates to become more similar as mean abundance and sample size increased. The results in this chapter show a similar pattern. The CWS estimate was larger for the low abundance, low sample size species (Cerulean Warbler) under stringent data selection criteria (Table 4.3, Fig. 4.3b), but the two became more similar when the sample size was increased (Fig. 4.6b). The estimates were very similar for the low abundance, high sample size species (Belted Kingfisher, Table 4.10, Figs. 4.11 b and 4.12 b ), but again the CWS estimate was slightly larger. The estimates were also similar for the high abundance species (except for Chestnut-collared Longspur with tolerant data selection, discussed presently). The same pattern is evident when looking at individual route trends within a species: the magnitude of the trend estimate tends to be closer to zero for lowabundance routes, and the effect is slightly stronger for the USNBS method than the CWS method (compare Figs. 4.13a and b).

The estimating equations method was introduced because it does not require the data to be log-transformed and so bypasses the issue of what constant to add and the accompanying bias (e.g., Fig 4.13c). It also makes no assumptions about the underlying distribution of the counts other than the expected value being a multiplicative function of an observer effect and an exponential (loglinear) trend (Link and Sauer 1994, equation 1). Simulations have shown that it is slightly positively biased (Link and Sauer 1994; Barker and Sauer 1992 have suggested that the other RR estimators are also positively biased). However, comparisons with the USNBS RR estimator (using 348 species that occurred on 40 or more BBS routes after tolerant data selection criteria) showed that the estimating equations estimates were slightly more negative, especially for high abundance species (Link and Sauer 1994). Of the four species studied in this chapter, estimating equation estimates were more negative than both USNBS and CWS RR estimates in all cases, including the species with an increasing trend (Carolina Wren). The causes of the difference in behavior between the simulated and real data are unclear.

Another difference between the linear $R R$ methods is the weightings used in combining route trends to produce regional estimates. In Chapter 2 I found that trend estimates for four of the 115 species analyzed were unrealistically large using CWS RR because of the method of calculating the abundance weightings. In all four cases, one route with a large apparent change in density but few surveys was assigned a very large abundance weighting due to the extrapolation of the route trend to calculate the least-squares mean weighting value. In this chapter, I used the geometric mean as the abundance weight in CWS RR to reflect recent changes in practice (Downes and Collins 1996). However, both USNBS RR and the estimating equations estimator use the least-squares abundance estimate, rather than geometric mean, when this does not involve extrapolation (Chapter 2). When there are few surveys close to the midyear this can produce very imprecise estimates and lead to the same effect as that noted for the former CWS weighting. This is exactly what happened for the tolerantly-selected Chestnutcollared Longspur data (Fig 4.10b): route 04008 (Bow City, Alberta) had an abundance weighting of 5939 , which is more than 10 times larger than the highest count for that species. The overall trend estimates were unrealistically high ( $+7.22 \%$ per year for USNBS RR and
$+3.22 \%$ per year for the estimating equations estimator) and the variances were also large (SE 8.54 and 3.72 respectively). When this route is removed, the USNBS RR estimate becomes $-0.32 \%$ per year (SE 0.61) and the estimating equations estimate becomes $-0.33 \%$ per year (SE 0.44 ); both are closer to the CWS estimate of -0.23 . There are reasonable theoretical reasons for using the least-squares mean as an abundance weight; however even in the USNBS estimators it is currently used only part of the time (when there are surveys both before and after the midyear). To prevent the sabotage of the trend estimate by poor estimates on single routes it seems necessary to either constrain (Winsorize) the least-squares mean estimate (for example, to be less than the maximum observed count), or to use the geometric mean count throughout. Another possibility is to incorporate uncertainty in the estimate of abundance weighting into the precision weight, although the use of precision weights is controversial (James et al. 1990).

The calculation of annual indices from route regression estimates has received less attention. The only simulation study (Collins 1994) compared CWS residual indices with two ANOVA-based methods and concluded that an index that incorporated both year effects and route trends (non-parallel ANOVA) showed the least bias, although the results were preliminary. Despite this, residual indices are still the accepted approach for linear RR (e.g., Peterjohn et al. 1995), and are used as informal checks on the regional trend estimate. The CWS and USNBS implementations differ in three ways: (1) CWS indices are calculated entirely on the log-scale while USNBS indices are calculated mostly on the multiplicative scale; (2) CWS indices are calculated relative to the mean survey year (i.e., mean of the years on which surveys took place), while USNBS indices use the mid-year (i.e., mid point of the time period analyzed, regardless of when the surveys took place, e.g., 1980.5 for the time period 1966 to 1995); (3) CWS indices are based on residuals from the route trend estimates, while the USNBS estimator first re-fits the regional trend to each route before calculating residuals. Despite these differences, the estimates were very similar. Both appeared to accurately portray the non-linearity in the Carolina Wren series (Fig. 4.7b), but there was a suggestion that both were biased towards the straight trend line in the Cerulean Warbler series (Fig 4.3b). In addition, the USNBS residual indices did not strongly highlight the problem with route 04008 in the Chestnut-collared Longspur data (Fig
4.10b). There is clearly a need to further investigate the use of residual indices before they can confidently be used as a diagnostic to judge the fit of the linear trend estimates or to draw conclusions about patterns of population change (see also discussion by Collins 1994).

## Nonlinear route regression

At their best, the smoothing-based RR methods (NNRR and NSRR) can produce a plausible picture of the trend underlying noisy BBS data (e.g., Fig. 4.3d, Fig. 4.9d). However, in many cases it appeared that the estimates were unreliable near the beginning and the end of the time series. The problem arises because many routes have missing data in the first or last few years (even after data selection), and the smoothing-based methods do not extrapolate index estimates beyond the limits of the available data. When smoothed route indices are averaged to produce strata means, which are then smoothed again, the strata smooths at the beginning and end of the series will be based on a subset of the routes used in the middle of the series. This leads to obvious discontinuities such as at the end of the stringently selected Cerulean Warbler time series (Fig. 4.3d), and may account for the implausible estimates at the beginning of the Belted Kingfisher series (Figs. 4.11c and d, and 4.12c and d; also the end of 4.12c). When there are no surveys in a particular stratum, jumps may occur, such as at the beginning of the tolerantly selected Cerulean Warbler time series (Fig 4.6c and d). Missing index estimates also cause the statistical tests for overall trend to be based on fewer data points than the smooths, which is clearly undesirable.

These problems are minimized in published smooth-based estimates (James et al. 1990, 1996) by using stringent data selection criteria. They are also less likely to be important with large sample sizes (although stringent data selection criteria lead to a drastic reduction in the amount of available data). Nevertheless, given the results of the case studies, it would be useful to investigate the consequences of extrapolating the smoothed estimates to the beginning and end of the time series. The loess smoother used in NNRR and NSRR is capable of extrapolation, as are other smoothers such as the cubic smoothing spline (which is linear beyond the bounds of the data, Hastie and Tibshirani 1990). Smoothing at the route level will, however, always be datahungry. It will probably be necessary to use stringent data selection criteria to ensure that there
are some surveys near the beginning and end of the time series in each route and so prevent overly long extrapolations.

One aspect of the NSRR estimator not discussed by James et al. (1996) is the failure of the numerical algorithm for more than $30 \%$ of the routes that would otherwise pass the stringent data selection criteria. This failure occurs because the matrix of smoothed observer residuals ( X 'Xres in James et al. 1996 Appendix 1) is close to singular, and is not properly inverted by the LU-decomposition and back-substitution algorithm. Rescaling the initial matrix will sometimes solve such problems (Bard 1974), but in this case the poor condition of the matrix is likely due to the distribution of observers among years and not to large differences in the matrix elements. Alternative numerical methods, such as singular value decomposition (Press et al. 1992) and other direct decomposition methods (McCullagh and Nelder 1989) cope better with illconditioned matrices and are worth investigating.

James et al. (1996) noted that NSRR estimates were, on average more negative than NNRR estimates for 26 species of warblers using stringently selected data from eastern and central North America. This result was consistent with a detailed analysis of the differences in counts among observers in the BBS based on linear RR analyses (Sauer et al. 1994b), which showed that for many species observers in recent years tended to count more birds than those in earlier years. Sauer et al. (1994b) speculated that this may be due to increasingly stringent screening by BBS coordinators of new observers, leading on average to increased observer quality over time. They conclude that ignoring observer differences will tend to positively bias trend estimates, although the effect varied among species. The results of one test were reported on a species-by-species basis, and can be compared with the differences observed for three of the four species analyzed in this chapter (Table 4.13; the fourth species showed non-significant results in Sauer et al. 1994b and so results were not reported). The effect of incorporating observer differences into the analysis is qualitatively similar in only one of the three species. Further comparisons based on more species are required to determine whether the conclusions of Sauer et al. (1994b) about observer differences are robust to the choice of analysis method.

The smoothing estimators were not flexible enough to accurately portray the sudden change in abundance of Carolina Wrens between 1976 and 1977 (Figs. 4.7c and d). The estimated low point using NNRR and NSRR was 1981, 4 years after the actual crash had occurred. As implemented, these methods are clearly unsuitable for detecting large, short term changes in abundance that may be of conservation concern (Peterjohn et al. 1995) and annual index analysis may be more appropriate if this is the goal. Alternatively, a smaller smoothing parameter may be used if sudden changes in trend are suspected, although this would be highly subjective. A better solution would be to allow the smoothing parameter to be dependent on the year-to-year variability in the data (data-driven bandwidth selection, Li and Heckman 1996). Unfortunately, this is not straightforward for time series data because positive correlations between successive counts tends to produce over-smoothing (Hart and Wehrly 1986, Li and Heckman 1996). The problem of choosing an appropriate smoothing parameter deserves more study.

The rank-trends estimates followed the pattern established in Chapter 2 of having higher statistical significance than the other estimates for each species. Titus (1990) suggested that the rank-trends method may not be suitable for species with many zero counts. In this analysis, the trend estimates did not appear to be any more different for the low-abundance species than the high-abundance species, although the proportion of routes with estimated declines was the most divergent from the other methods for Cerulean Warbler (Table 4.3). In general, the rank-trend method is quite unsatisfactory because it provides little information about population trends, and no way of testing for biologically interesting patterns of population change over space or time.

## Loglinear modeling

The loglinear modeling approach appears to have much potential. It is flexible, allowing data-based selection of the appropriate scale for parameters in space and time. It appears to provide plausible estimates of population change and variance of estimates, except when the model is clearly inadequate (e.g., Chestnut-collared Longspur, see below). In addition, it allows for model checking through formal and informal analyses of residuals.

There are, however, a number of practical problems in implementing the method. Firstly, it is computationally very expensive, especially when the number of sites or years is large. Previous attempts to apply linear modeling to BBS data have had problems with the large amount of memory needed to invert the design matrix (Sauer and Geissler 1990). Program TRIM bypasses these problems by taking advantage of the block-diagonal nature of the design matrix (Pannekoek and van Strien 1996), and so allows the analysis of data from a very large number of sites. Nevertheless, the program is still computationally intensive, especially for models that incorporate serial correlation (these use a different fitting algorithm). For example, fitting the model 5 (region) "with observers" to Belted Kingfisher data (1618 parameters) took 26 minutes on a 90 Mhz Pentium PC without inclusion of serial correlation, but 286 minutes when serial correlation was included (see Appendix 1 for computer, operating system and compiler specifications).

A second problem is the failure of the estimates to converge for many of the more complex models, especially those that use the generalized estimating equations estimator (i.e., the serial correlation models). Two types of failure can be distinguished: (1) where the difference in deviance between successive iterations becomes small but some parameter estimates become increasingly large, and (2) where the deviance actually increases in some iterations, and there is no sign that it is reaching a minimum. In the first case the likelihood function for the troublesome parameters is very flat or is asymptotic. The estimated deviance value can be used in model selection, but the parameter estimates will be very inaccurate. The second type of failure is an indication of a complex likelihood function; in this case the deviance from the last iteration may not be close to the maximum-likelihood value so cannot be used. Rescaling may help in both cases, together with an increase in the accuracy of computations (to decrease problems with rounding errors and calculation of the gradient), and the use of alternative fitting algorithms (Bard 1974). It will, however, always be difficult to reliably fit large-dimension or poorly specified models.

For three of the four species, the final model chosen did not adequately fit the data (Pearson $\chi^{2} /$ d.f. from the final model was very close to the expected value of 1.0 for Cerulean

Warbler, but was 2.41 for Carolina Wren, 8.03 for Chestnut-collared Longspur and 1.24 for Belted Kingfisher). There are two likely explanations for this: either the counts are not distributed as a Poisson random variable or there are some important covariables missing from the model (model incorrectly specified). Biological count data can show greater variability than allowed by the Poisson distribution for numerous ecological reasons (Eberhardt 1978; White and Bennetts 1995), such as clumping of birds in space due to habitat heterogeneity or social behaviors. In addition, variation in the probability of detection of individual birds both within and between surveys creates extra variability in the counts. Nevertheless, model fits as poor as those observed for the Chestnut-collared Longspur are unlikely to be due to overdispersion alone (Anderson et al. 1994), and incorrect model specification must be suspected. I do not pursue this possibility further here, although the BBS dataset contains some additional covariables (such as weather during the survey, time of day and date) that could prove useful. Instead, I pursue the issue of overdispersion, which is almost certainly responsible for at least some of the lack of fit.

A common solution to dealing with extra-Poisson variability, and the one implemented by program TRIM, is to estimate the overdispersion parameter, $\sigma$, based on the Pearson $\chi^{2}$ divided by the degrees of freedom, and to use this to inflate the variance estimates (McCullagh and Nelder 1989). However, this assumes that the variance is a linear function of the mean (variance is now assumed to be $\sigma^{2} \mu$, rather than $\mu$ for a Poisson distribution). This assumption can be checked with a plot of the estimated variance of each count divided by the predicted count against predicted count (Cameron and Trivedi 1990; Winkelmann 1994). Estimated variance is calculated as (observed count - predicted count) ${ }^{2}$. Using the most parsimonious model for each species, these plots show a great deal of scatter with little apparent pattern (Figure 4.14).

However, linear regression slopes (Cameron and Trivedi 1990) are significantly greater than zero for all species except Cerulean Warbler (Cerulean Warbler $0.013 \pm 0.024$ (slope $\pm$ SE); Carolina Wren $0.051 \pm 0.004$; Chestnut-collared Longspur $0.031 \pm 0.010$; Belted Kingfisher $0.232 \pm$ 0.057 ), indicating the possibility of a higher-order relationship between the variance and mean.

An alternative distribution often suggested for overdispersed count data is the negative binomial (Winkelmann 1994; White and Bennetts 1996). In the negative binomial distribution,
the variance is a quadratic function of the mean, and is specified by a dispersion parameter, k , such that

$$
\text { variance }=\text { mean }+\mathrm{k}^{*}\left(\text { mean }^{2}\right)
$$

In a preliminary investigation, I compared the fit of the overdispersed Poisson distribution with the negative binomial for the stringently selected Chestnut-collared Longspur data. I chose the Chestnut-collared Longspur because the sample size is relatively small and it is feasible to fit the models using ordinary statistical software. I used SAS PROC GENMOD (SAS Institute Inc. 1993a) to repeat the Poisson regression from program TRIM and a SAS macro by Hilbe (1994) that used PROC GENMOD to fit the negative binomial distribution with a log link-function. For model 4 (routes) "with observers", the negative binomial regression (deviance of fit 497.99 on 76 parameters, AIC 649.99) produced very similar parameter estimates to the Poisson regression, and the regression line of estimated variance/predicted count against count was almost identical (slope $0.032 \pm 0.010$ ). Similar results were obtained when the other TRIM models were used. For these data the negative binomial distribution offered no apparent advantage over the overdispersed Poisson, although this question deserves further investigation. An alternative distribution worthy of study is the Poisson-logistic (Winkelmann 1994), in which a binomial process, such as the detection of individual birds, is superimposed on a Poisson process, such as the occurrence of birds within the counting radius of each survey stop. This line of reasoning could be further extended to produce other biologically-motivated mixture distributions. The residual analysis presented here could be extended to include numerous other formal and informal checks for specification (McCullagh and Nelder 1989; Venables and Ripley 1994). One obvious improvement to program TRIM in this respect would be reporting of the hat (influence) matrix.

Ecological time series are generally assumed to be serially correlated (Jassby and Powell 1990). However, including serial correlation in the models produced only a small improvement in fit for three of the four species, and resulted in a worse fit for the fourth (Carolina Wren). The estimated magnitude of serial correlation was small for all species except Carolina Wren, and so had little effect on the estimated variances or the statistical inferences drawn about the parameter
estimates. Little work has been done to date on serial correlation in annual songbird counts, so it is hard to generalize these findings. Fewster and Buckland (1996) incorporated serial correlation into their additive models of British Common Bird Census data using an autocovariate (the observed count in the previous year), and this had a reasonably strong effect on the estimates.

Model selection using AIC and QAIC does not select the model that is the most faithful representation of the true data generating process (the "true" model); rather it attempts to select the most parsimonious model, i.e., the model with the optimum balance between bias (too few parameters) and imprecision (too many parameters, Burnham and Anderson 1992). One would therefore expect that, other factors being equal, higher sample sizes (number of routes) would lead to more complex (higher parameter number) models being selected. This trend was evident when comparing the selected models for the stringently selected (low sample size) and tolerantly selected (high sample size) datasets within species. It was not evident when comparing the selected models between high and low sample size species: for example the final model for the largest sample size species (Belted Kingfisher) contained only one set of geographic parameters. Thus, the amount of temporal and spatial variation in population change depends on the species being considered. This highlights the advantage of an estimator that has the flexibility to choose the appropriate scale based on the patterns evident in the data. For example, modeling of loglinear trends on a route-by-route basis (as in linear RR estimators) was the selected model in only one of the three species.

The models of temporal change used here could be extended in four ways. Firstly, more complex models that combine trends and annual indices could be used to try and improve the fit of poorly modeled species such as the Chestnut-collared Longspur. For example, the linear model of Sauer and Geissler (1990), which included route-specific linear-multiplicative trends and continental indices, could be extended to the loglinear context. Secondly, models that are intermediate between annual indices (models 3 and 5 and linear trends (models 2 and 4) could be used to try and improve the parsimony of the models for species that are well modeled (e.g., Cerulean Warbler). This could be done by (1) fitting different loglinear trends in different time
periods ("broken stick" regression), e.g., in the Carolina Wren dataset fitting one trend for 19671976 and one for 1977 to 1995; (2) adding polynomial terms to the model; and (3) using Generalized Additive Models to estimate smoothed trends. This last option, was implemented on British Common Bird Census data by Fewster and Buckland (1996; see Chapter 5).

Incorporating observer differences produced a more parsimonious fit for all species except the Belted Kingfisher. As with the loess smoothing RR estimators, this result is different from Sauer et al. (1994b), who found significant observer effects for all species except the Cerulean Warbler, although their analysis procedure was quite different. Interestingly, the effect of adding observer covariables on the overall trend estimate (Model 2) was qualitatively similar to the smoothing RR methods (Table 4.13).

Differences among observers on a route could be caused by variation in observer ability (i.e., in the detection function), or by differences in the way the survey is performed (e.g., in the exact place used as a survey location), or by a combination of both factors. If observer ability is the more important factor, then a more parsimonious model could be fit by using one covariable for each observer by name, rather than treating the same observer on different routes as a new observer as I did in this analysis. A second extension of the observer effects modeling would be to incorporate "start-up" effects, whereby each observer tends to count fewer birds on the first year they survey a route (Kendall et al. 1996). This could be modeled using a binary covariable to indicate the first year by an observer, and tested using AIC or other measures of model fit.

## Moving-window Mountford's method

The Mountford method is based on large-sample theory (Mountford 1982), so it is not surprising that it performed poorly on species observed on few routes. Peterjohn et al. (1995, discussing Mountford's method without moving windows) mention that "there is evidence that technical and practical aspects of model fitting [...] invalidate its use at smaller scales." It was, however, disappointing to find significant problems associated with the method at larger sample sizes as well (for Carolina Wren). Based on this limited sample of four species, it appears that larger window sizes lead to larger estimates of population change and smaller CIs. Larger window sizes give a greater overlap between successive windows, which could result in more
accurate "chaining" of successive models. Peach et al. (1994) suggested choosing the largest window size for which all windows that passed the Mountford goodness-of-fit test, but this strategy can apparently lead to implausible estimates of population change in some cases (e.g., for Carolina Wren, Fig. 4.8) but not others (e.g., for Belted Kingfisher, Fig. 4.11).

Mountford's method was introduced because it is more efficient than the chaining method it replaced, and because it is less subject to the "random walk" that plagues chaining estimates at low sample sizes (Geissler and Noon 1981, Moss 1985). Because species with low sample sizes in this analysis could only be fit using short window sizes, we may expect to see evidence of random walk in their estimates. This could be tested by fitting an auto-regressive model to the indices and testing for any residual trend, but such a test would have low power with 30 time points. Informal inspection of the plotted indices (Figs 4.3 c and d , and 4.9 c and d ) shows that apart from dramatic fluctuations in one poorly fitting model (1970 and 1971 estimates for Chestnut-collared Longspur) the indices tend to be clustered around the no-change line, and that this effect also occurs for large sample size species (e.g., Fig. 4.8a). I suspect that, rather than random walk, some other failures of the model assumptions are occurring.

One possible reason for model failure would be heterogeneous population change between sites, although this is expected to be more prevalent with large sample and window sizes. Plots of counts in successive years provide an informal test of this assumption: the principal axis of the plot of counts in one year against the counts in another should be a straight line running through the origin (Mountford 1982). An indirect piece of evidence for the importance of between-site heterogeneity is that the only species that produced feasible parameter estimates using Mountford's method was the Belted Kingfisher. Despite the large sample size for this species, the most parsimonious model using TRIM was model 3 (no regional variation), indicating that the population changes were likely very similar throughout the species range. By contrast, the other large sample size species, Carolina Wren, was poorly modeled using Mountford's method. This species was best fit with model 5 (strata) in TRIM, indicating significant geographic variation in patterns of population change. Because of its lack of
robustness, it appears that the Mountford method is not useful for estimating population change from BBS data.

## Generality of findings

In this chapter I have estimated population change for four species, chosen from the BBS dataset to represent extremes of abundance and sample size. By using a small number of species I have been able to present and discuss the results in some detail. Nevertheless, this means that my findings are based on a small subset of the entire BBS data. Some conclusions, therefore, must be treated as preliminary.

One reliable conclusion is that all of the estimators require further development to solve the obvious problems encountered (Table 4.12). However, based on these case studies alone I cannot reliably conclude when or how often these problems will occur in the whole dataset, or in data from other similar surveys. Further case studies, or less detailed comparisons using the entire BBS dataset, would therefore be useful. In the short term, however, I believe that a focus on methodological development to solve the obvious problems highlighted by these case studies would be more rewarding.

## Summary

I analyzed continental BBS data for four case study species using eight analysis methods, and used the results to make inferences about the reliability of the methods. The species were Cerulean Warbler (low abundance, low sample size), Carolina Wren (high abundance, high sample size), Chestnut-collared Longspur (high abundance, low sample size), and Belted Kingfisher (low abundance, high sample size). I used three linear route-regression (RR) methods (USNBS and CWS RR and the estimating equations method), two nonlinear RR methods (NNRR, NSRR and rank-trends analysis), loglinear modeling and the moving-windows Mountford Method (see Tables 3.1 and 3.2 for summaries of the methods). I also compared the results under both stringent and tolerant data selection criteria.

The linear route-regression methods produced reasonable estimates in all cases except for one, in which a single route-specific weighting caused poor estimates by the USNBS methods.

Weighting schemes are seemingly the bugbear of route regression methods and will require further research before an appropriate and robust scheme is found. The estimating equations estimator has, in theory at least, solved the problem of estimating loglinear trends at the route level. Annual indices calculated using the residual method appeared to be biased toward the linear trend estimates in some cases so are probably not a useful way of diagnosing nonlinearities in trend.

The loess smoothing methods produced some useful summaries of the data, but also encountered a number of problems: (1) poor performance because of missing data at the beginning and end of the time series; (2) numerical problems with the NSRR estimator; (3) oversmoothing of large, sudden, biologically interesting population changes. Development is needed in all three of these areas. As in Chapter 2, the rank-trends method gave more statistically significant results than the other methods. Although there were no obvious problems with the method, it is hard to recommend because it summarizes the patterns of population change into one statistic which has no biological meaning.

The loglinear modeling approach appeared promising, although there were numerical problems in program TRIM with higher-dimension models. All species except Belted Kingfisher showed significant lack of fit, indicating that the models were inadequate and/or that the counts are generally overdispersed relative to the Poisson distribution. This was particularly true for the Chestnut-collared Longspur, which showed the greatest variation in counts. The models used here could be extended to incorporate more survey- and observer-specific covariables and to give smoothed estimates of trend over time and/or space. Alternative overdispersed count distributions could also be modeled explicitly, rather than through quasilikelihood.

The Mountford moving-window method performed poorly with the low sample-size species, and there were some questions regarding the appropriate window size to choose with one of the high-sample size species (Chestnut-collared Longspur). Belted Kingfisher was the only species that was well fit, and the estimates appeared similar to those from one of the TRIM models. Based on these results, this method cannot be recommended for use with BBS data.
Table 4.1. List of analysis methods used in this chapter, with their associated data selection criteria.

| Method | Reference | Data selection criteria |
| :---: | :---: | :---: |
| Linear RR |  |  |
| Least squares |  |  |
| CWS RR | Downes and Collins 1996 | Not used in this chapter (see Methods). |
| USNBS RR | Geissler and Sauer 1990 | (1) Exclude unacceptable surveys (where USNBS survey code $>3$ ); (2) exclude routes where \#surveys - \#observers < 2; (3) exclude routes where no birds counted. |
| Estimating equations | Link and Sauer 1994 | As above, except add (4) only include routes where there was at least 1 observer with $>1$ non-zero survey (i.e., a survey where 1 or more birds were counted). |
| Nonlinear RR |  |  |
| Smoothing methods |  |  |
| NNRR | James et al. 1990 | (1) Exclude unacceptable surveys (where USNBS survey code $>1$ ); (2) exclude routes with $<9$ surveys, or with no surveys in any 6 year period: 1966-71, 72-77, 78-$83,84-89,90-95^{1}$; (3) exclude physiographic strata with $<5$ non-zero routes (i.e., routes where 1 or more birds were counted), or with $<5 \%$ non-zero surveys. |
| NSRR | James et al. 1996 | As above, except add (2a) exclude routes where any element of the matrix of smoothed observer residuals ( $X^{\prime}$ Xres in Appendix 1 of James et al. 1996) > 2 . ${ }^{2}$ |
| Rank-trends | Titus et al. 1990 | None implemented. |
| Loglinear modeling | Pannekoek and van Strien 1994 | Exclude years where any of the physiographic strata does not contain at least one survey. |
| Mountford moving windows | Peach and Baillie 1994 | Not used in this chapter (see Methods). |
| ${ }^{1}$ This requirement is in the spirit of James et al. (1990, 1996). For example James et al. (1996) analyzed slightly fewer years of from 1966-1992) and excluded routes with < 8 surveys and no surveys in any of the periods 1966-72, 73-75, 76-80, 81-83 and |  |  |
| ${ }^{2}$ This requirement was not documented in James et al. (1996), but is necessary to prevent the NSRR estimation algorithm from p unreliable results (D. A. Wiedenfeld, Museum of Natural Science, Louisiana State University, personal communication; see Dis |  |  |

Table 4.2. Summary of the Breeding Bird Survey data used in this chapter.

| Species | Selection criteria ${ }^{1}$ | Years | No. of routes | No. of surveys | No. of BBS regions | No. of strata | $\begin{gathered} \% \\ \text { missing } \\ \text { surveys } \end{gathered}$ | Total count | Mean count / survey | Median count / survey | Max. <br> count/ <br> survey |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Cerulean warbler | Stringent | 1966-95 | 41 | 1093 | 1 | 4 | 11.14 | 1342 | 1.23 | 0 | 31 |
|  | Tolerant | 1966-95 | 198 | 4020 | 1 | 13 | 32.32 | 4392 | 1.09 | 0 | 46 |
|  | Minimal | 1966-95 | 362 | 7261 | 2 | 23 | 33.14 | 4836 | 0.67 | 0 | 46 |
| Carolina wren | Stringent | 1967-95 | 268 | 7028 | 2 | 20 | 9.57 | 75406 | 10.73 | 6 | 104 |
|  | Tolerant | 1967-95 | 1077 | 19961 | 2 | 26 | 36.09 | 202674 | 10.15 | 5 | 142 |
|  | Minimal | 1966-95 | 1341 | 23785 | 2 | 32 | 40.88 | 214038 | 9.00 | 3 | 142 |
| Chestnut- <br> collared <br> Longspur | Stringent | 1967-95 | 21 | 534 | 2 | 3 | 12.32 | 19559 | 36.63 | 9 | 340 |
|  | Tolerant | 1967-95 | 128 | 1944 | 2 | 6 | 47.63 | 51480 | 26.48 | 7 | 447 |
|  | Minimal | 1966-95 | 187 | 2558 | 2 | 10 | 54.40 | 53922 | 21.08 | 3 | 447 |
| Belted kingfisher | Stringent | 1968-95 | 529 | 13037 | 3 | 33 | 11.98 | 7545 | 0.58 | 0 | 15 |
|  | Tolerant | 1968-95 | 1603 | 31253 | 3 | 55 | 30.37 | 18183 | 0.58 | 0 | 15 |
|  | Minimal | 1966-95 | 2316 | 41888 | 3 | 64 | 39.71 | 20682 | 0.49 | 0 | 15 |

[^4]Table 4.3. Estimates of overall population trend for the Cerulean Warbler between 1966 and 1995 (data selected using stringent data selection criteria).

| Estimation method | Trend (\% per year) | SE | 95\% CI | $z$ | $p$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Linear route regression |  |  |  |  |  |
| USNBS RR | -2.76 | 0.48 | -1.80 to -3.72 | -5.88 | $<0.001$ |
| USNBS est. eqns. RR | -4.85 | 1.51 | -1.83 to -7.87 | -3.21 | 0.001 |
| CWS RR | -3.80 | - ${ }^{1}$ | -1.56 to -5.98 | -3.38 | 0.001 |
| Nonlinear route regression |  |  |  |  |  |
| NSRR | $-3.04{ }^{2}$ | - ${ }^{1}$ | +1.01 to -5.90 | -1.52 | 0.129 |
| NNRR | $-2.58{ }^{2}$ | - ${ }^{1}$ | +0.67 to -4.86 | -1.72 | 0.085 |
| Rank-trends | - | - | - | -5.97 | $<0.001$ |
| TRIM Model 2, incorporating serial correlation |  |  |  |  |  |
| with observers | -4.06 | 0.48 | -3.10 to -5.02 | -8.46 | $<0.001$ |
| without observers ${ }^{3}$ | -2.41 | 0.43 | -1.55 to -3.27 | -5.60 | $<0.001$ |
| Mountford window size 4 |  |  |  |  |  |
| with observers | -0.13 | - 1 | +0.40 to -0.30 | -0.01 | 0.992 |
| without observers | -0.05 | - ${ }^{1}$ | +0.34 to -0.39 | -0.03 | 0.976 |

${ }^{1}$ Standard errors not calculated on the multiplicative scale and so not comparable.
${ }^{2}$ Population change in NNRR and NSRR is calculated as the weighted mean over all routes of the mean estimated value in the years 1970, 71 and 72 minus the mean estimated value in years 1989, 90 and 91 (see appendix 1). Therefore the estimates cover only the middle 22 years of the survey.
${ }^{3}$ Standard error (SE) and confidence intervals (CIs) calculated after adjusting for overdispersion (Table 4.4).
Table 4.4. Proportion of BBS routes with estimated declines in abundance (p. dec.) for four species using six route-regression (RR) trend estimators. Figures in brackets after each species name indicate the number of routes used in the analysis (data selected using stringent criteria, see Methods). The $95 \%$ confidence intervals (CIs) were calculated using the normal approximation to the binomial distribution. If a $95 \%$ CI spans 0.5 then the null hypothesis that the same number of routes show declines as increases cannot be rejected at an $\alpha$-level of 0.05 .

| Estimation method | Cerulean warbler <br> $(41)$ |  | Carolina wren <br> $(286)$ |  | Chestnut-collared <br> Longspur (21) |  | Belted kingfisher <br> ( 529$)$ |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | p. dec. | $95 \% \mathrm{CI}$ | p. dec. | $95 \% \mathrm{CI}$ | p. dec. | $95 \% \mathrm{CI}$ | $\ldots$ | p. dec. | $95 \% \mathrm{CI}$ |
| USNBS RR | 0.78 | 0.90 to 0.65 | 0.36 | 0.42 to 0.30 | 0.52 | 073 to 0.31 | 0.58 | 0.62 to 0.54 |  |
| USNBS est. eqns. RR | 0.76 | 0.89 to 0.63 | 0.34 | 0.39 to 0.29 | 0.62 | 0.83 to 0.41 | 0.56 | 0.60 to 0.52 |  |
| CWS RR | 0.76 | 0.89 to 0.63 | 0.35 | 0.41 to 0.29 | 0.52 | 0.73 to 0.31 | 0.57 | 0.61 to 0.53 |  |
| NNRR | 0.68 | 0.82 to 0.54 | 0.31 | 0.36 to 0.26 | 0.76 | 0.94 to 0.58 | 0.57 | 0.61 to 0.54 |  |
| NSRR | 0.78 | 0.90 to 0.65 | 0.36 | 0.42 to 0.30 | 0.57 | 0.78 to 0.36 | 0.58 | 0.62 to 0.54 |  |
| Rank-trends | 0.58 | 0.73 to 0.43 | 0.25 | 0.30 to 0.20 | 0.76 | 0.94 to 0.58 | 0.55 | 0.59 to 0.51 |  |

Table 4.5. Comparison of loglinear models for the analysis of population trends in Cerulean Warbler using BBS data from 1966-1995 (data selected stringent data selection criteria; sample sizes given in Table 4.2). Calculations were performed using program TRIM. Models are ordered according to the number of parameters estimated; for models 4 and 5 the covariable is given in parentheses. (Because this species only occurred in 1 BBS region, Models 4 (region) and 5 (region) are equivalent to models 2 and 3.) Definitions of the models are given in the text. The selected models (lowest AIC) are shown in boldface.

| Model | No. of params. | Deviance | Pearson $\chi^{2} /$ d.f. | $\begin{gathered} p \text { for } \\ \text { Pearson } \chi^{2} \end{gathered}$ | Serial correlation | $\mathrm{AIC}^{1}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| With observer effects, with serial correlation |  |  |  |  |  |  |
| 5 (stratum) $^{2}$ | 228 | 776.48 | - | - | - | 1232.48 |
| 4 (route) | 153 | 857.61 | 0.98 | 0.627 | 0.01 | 1163.61 |
| 3 | 141 | 868.38 | 1.02 | 0.347 | 0.05 | 1150.38 |
| 4 (stratum) | 116 | 929.99 | 1.02 | 0.330 | 0.07 | 1161.99 |
| 2 | 113 | 930.12 | 1.02 | 0.364 | 0.07 | 1156.12 |
| 1 | 112 | 996.66 | 1.10 | 0.016 | 0.11 | 1220.66 |
| With observer effects, without serial correlation |  |  |  |  |  |  |
| 5 (stratum) | 228 | 776.79 | 0.99 | 0.604 | - | 1232.79 |
| 4 (route) | 153 | 857.75 | 0.98 | 0.627 | - | 1163.75 |
| 3 | 141 | 869.60 | 1.02 | 0.354 | - | 1151.60 |
| 4 (stratum) | 116 | 932.29 | 1.02 | 0.335 | - | 1164.29 |
| 2 | 113 | 932.43 | 1.01 | 0.372 | - | 1158.43 |
| 1 | 112 | 1002.62 | 1.10 | 0.015 | - | 1226.62 |
| Without observers effects, with serial correlation ${ }^{3}$ |  |  |  |  |  |  |
| 5 (stratum) $^{2}$ | 157 | 980.20 | - | - | - | 1294.20 |
| 3 | 70 | 1107.59 | 1.30 | $<0.001$ | 0.15 | 1247.59 |
| 4 (stratum) | 45 | 1168.40 | 1.30 | $<0.001$ | 0.16 | 1258.40 |
| 2 | 42 | 1172.17 | 1.30 | $<0.001$ | 0.17 | 1256.17 |
| 1 | 41 | 1224.62 | 1.32 | $<0.001$ | 0.20 | 1306.62 |
| Without observers effects, without serial correlation |  |  |  |  |  |  |
| 5 (stratum) $^{2}$ | 157 | 980.20 | - | - | - | 1294.20 |
| 4 (route) | 82 | 980.23 | 1.09 | 0.019 | - | 1144.23 |
| 3 | 70 | 1108.52 | 1.30 | $<0.001$ | - | 1248.52 |
| 4 (stratum) | 45 | 1175.15 | 1.31 | $<0.001$ | - | 1265.15 |
| 2 | 42 | 1179.87 | 1.31 | $<0.001$ | - . | 1263.87 |
| 1 | 41 | 1234.97 | 1.33 | <0.001 | - - | 1316.97 |

[^5]Table 4.6. Goodness-of-fit tests for Mountford moving windows applied to BBS data for Cerulean Warbler 1966-1995 ( $\mathrm{n}=41$ routes, 112 observers; stringent data selection criteria) . $\chi^{2}$ values $>\chi_{\text {crit }}^{2}$ are shown in boldface.

| Years | With observers |  | Without observers |  |
| :---: | :---: | :---: | :---: | :---: |
|  | No. "routes" (i.e. observers) in model (max 112) | $\chi^{2}$ | No. routes in model (max 41) | $\chi^{2}$ |
| Window size 4; d.f. $=3 ; \chi_{\text {crit }}^{2}=7.82(\alpha=0.05)$ |  |  |  |  |
| 66-69 | 33 | 5.19 | 34 | -0.59 |
| 68-71 | 38 | 1.31 | 38 | 2.66 |
| 70-73 | 41 | 0.10 | 41 | 2.02 |
| 72-75 | 41 | 1.54 | 41 | 0.67 |
| 74-77 | 43 | 4.83 | 41 | 1.78 |
| 76-79 | 42 | 3.07 | 41 | 2.64 |
| 78-81 | 41 | 4.12 | 40 | 4.91 |
| 80-83 | 38 | 2.99 | 40 | 3.60 |
| 82-85 | 41. | 2.89 . | 40 | 2.24 |
| 84-87 | 41 | 3.26 | 40 | 3.00 |
| 86-89 | 44 | 1.52 | 39 | 1.76 |
| 88-91 | 40 | 1.67 | 39 | 2.19 |
| 90-93 | 38 | 1.78 | 39 | 2.30 |
| 92-95 | 36 | 0.53 | 36 | 0.60 |
| Window size 6; d.f. $=10 ; \chi_{\text {crit }}^{2}=18.30(\alpha=0.05)$ |  |  |  |  |
| 66-71 | 41 | 8.72 | 39 | 5.99 |
| 68-73 | 45 | 3.33 | 41 | 7.15 |
| 70-75 | 44 | 3.68 | 41 | 2.48 |
| 72-77 | 45 | 5.43 | 41 | 1.84 . |
| 74-79 | 46 | 5.31 | 41 | 8.34 |
| 76-81 | 46 | 10.72 | 41 | -18.13 |
| 78-83 | 45 | 0.16 | 41 | 3.79 |
| 80-85 | 43 | 8.11 | 41 | 9.49 |
| 82-87 | 43 | 9.81 | 40 | 2.54 |
| 84-89 | 52 | 7.24 | 41 | 5.31 |
| 86-91 | 48 | 4.10 | 40 | 2.74 |
| 88-93 | 42 | 4.73 | 39 | 0.78 |
| 90-95 | 42 | 4.67 | 40 | 2.59 |

Table 4.7. Estimates of overall population trend for the Carolina Wren between 1967 and 1995 (data selected using stringent data selection criteria).

| Estimation method | Trend <br> (\% per year) | SE | $95 \% \mathrm{CI}$ | $z$ | $p$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Linear route regression |  |  |  |  |  |
| USNBS RR | +1.04 | 0.25 | +1.54 to +0.54 | 4.14 | $<0.001$ |
| USNBS est. eqns. RR | +0.85 | 0.25 | +1.34 to +0.35 | 3.43 | $<0.001$ |
| CWS RR | +1.27 | -1 | +1.89 to +0.67 | 4.16 | $<0.001$ |
| Nonlinear route regression |  |  |  |  |  |
| NSRR | $+1.38^{2}$ | -1 | +2.03 to +0.78 | 4.80 | $<0.001$ |
| NNRR | $+1.62^{2}$ | -1 | +2.09 to +1.17 | 7.70 | $<0.001$ |
| Rank-trends | - | - | - | 19.31 | $<0.001$ |

TRIM Model 2, incorporating serial correlation and overdispersion

| with observers | +1.02 | 0.15 | +1.33 to +0.72 | 6.80 | $<0.001$ |
| :--- | :--- | :--- | :--- | :---: | :--- |
| without observers | +1.95 | 0.14 | +2.08 to +1.68 | 13.93 | $<0.001$ |

Mountford window size 6

| with observers | +0.37 | -1 | +0.47 to +0.26 | 5.69 | $<0.001$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| without observers | +0.40 | $-^{1}$ | +0.50 to +0.30 | 6.19 | $<0.001$ |

${ }^{1}$ Standard errors not calculated on the multiplicative scale and so not comparable.
${ }^{2}$ Population change in NNRR and NSRR is calculated as the weighted mean over all routes of the mean estimated value in the years 1971, 72 and 73 minus the mean estimated value in years 1989, 90 and 91 (see appendix 1). Therefore the estimates cover only the middle 21 years of the survey.

Table 4.8. Comparison of loglinear models for the analysis of population trends in the Carolina Wren using BBS data from 1967-1995 (see Table 4.5 legend for further information). The model with the lowest AIC and QAIC is shown in boldface, but this model was not used for parameter estimation because the estimates were unreliable (see ${ }^{4}$ and text). All models failed the Pearson $\chi^{2}$ goodness-of-fit test ( $\mathrm{p}<0.001$ in all cases).

| Model | No. of params. | Deviance | Pearson $\chi^{2} /$ d.f. | Serial correlation | AIC ${ }^{1}$ | QAIC ${ }^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| With observer effects, with serial correlation ${ }^{3}$ |  |  |  |  |  |  |
| 4 (route) ${ }^{4}$ | 1084 | 17516.25 | - | - | 19684.25 | 9437.66 |
| 4 (stratum) | 836 | 19737.89 | 3.14 | 0.32 | 21409.89 | 9863.69 |
| 4 (region) | 818 | 20045.21 | 3.18 | 0.33 | 21681.21 | 9955.24 |
| 2 | 817 | 20050.71 | 3.18 | 0.33 | 21684.71 | 9955.51 |
| 1 | 816 | 20275.97 | 3.20 | 0.33 | 21907.97 | 10047.01 |
| With observer effects, without serial correlation |  |  |  |  |  |  |
| 5 (stratum) ${ }^{4}$ | 1376 | 11277.19 | - | - | 14029.19 | 7432.30 |
| 4 (route) $^{4}$ | 1084 | 17034.22 | - | - | 19202.22 | 9237.60 |
| 5 (region) | 872 | 15364.78 | 2.41 | - | 17108.89 | 8120.75 |
| 3 | 844 | 15756.89 | 2.47 | - | 17444.89 | 8227.44 |
| 4 (stratum) | 836 | 19301.55 | 3.12 | - | 20973.55 | 9682.60 |
| 4 (region) | 818 | 19635.38 | 3.17 | - | 21271.38 | 9785.15 |
| 2 | 817 | 19641.91 | 3.17 | - | 21275.91 | 9758.86 |
| 1 | 816 | 19882.43 | 3.20 | - | 21514.43 | 9883.68 |
| Without observers effects, with serial correlation ${ }^{3}$ |  |  |  |  |  |  |
| 4 (route) ${ }^{4}$ | 536 | 22512.87 | - | -. | 23584.87 | 10415.37 |
| 4 (stratum) ${ }^{4}$ | 288 | 26116.21 | - | - | 26692.21 | 11414.84 |
| 4 (region) | 270 | 27481.00 | 4.08 | 0.46 | 28021.00 | 11945.26 |
| 2 | 269 | 27515.16 | 4.09 | 0.46 | 28053.16 | 11957.44 |
| 1 | 268 | 28945.45 | 4.32 | 0.48 | 29481.45 | 12549.04 |
| Without observer effects, without serial correlation |  |  |  |  |  |  |
| 5 (stratum) ${ }^{4}$ | 828 | 16879.06 | - | - | 17227.06 | 8661.21 |
| 4 (route) ${ }^{4}$ | 536 | 21830.71 | ${ }^{-}$ | - | 22902.71 | 10132.26 |
| 5 (region) | 324 | 22224.37 | 3.23 | - | 23880.37 | 9871.64 |
| 3 | 296 | 22616.00 | 3.30 | - | 23208.00 | 9978.17 |
| 4 (stratum) | 288 | 26116.21 | 3.96 | - | 26992.21 | 11414.84 |
| 4 (region) | 270 | 26997.42 | 4.06 | - | 27537.42 | 11744.57 |
| 2 | 269 | 27041.64 | 4.07 | - | 27579.64 | 11760.92 |
| 1 | 268 | 28731.30 | 4.32 | - | 29267.30 | 12460.17 |

${ }^{1}$ Akaike information criterion (AIC) $=$ deviance $+2 *$ no. free parameters.
${ }^{2}$ Quasi-likelihood AIC (QAIC) $=$ deviance/dispersion $+2 *$ no. free parameters. The dispersion value used was 2.41 , from model 5 (region) with observers, the best fitting model for which an estimate of dispersion could be obtained.
${ }^{3}$ Models 5 (stratum), 5 (region) and 3 failed to converge.
${ }^{4}$ Models came close to convergence, but failed to find accurate parameter estimates.

Table 4.9. Estimates of overall population trend for the Chestnut-collared Longspur between 1967 and 1995 (data selected using stringent data selection criteria).

| Estimation method | Trend <br> (\% per year) | SE | $95 \% \mathrm{CI}$ | $z$ | $p$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Linear route regression |  |  |  |  |  |
| USNBS RR | -0.41 | 0.62 | +0.83 to -1.65 | -0.66 | 0.51 |
| USNBS est. eqns. RR | -0.48 | 0.58 | +0.68 to -1.64 | -0.83 | 0.41 |
| CWS RR | -0.33 | -1 | +1.02 to -1.82 | -0.45 | 0.65 |
| Nonlinear route regression |  |  |  |  |  |
| NSRR | $-1.05^{2}$ | -1 | +2.28 to -3.44 | -0.71 | 0.48 |
| NNRR | $-2.94^{2}$ | -1 | -0.32 to -4.93 | -2.20 | 0.03 |
| Rank-trends | - | - | - | -5.17 | $<0.001$ |

TRIM Model 2, incorporating serial correlation and overdispersion

| with observers | -0.84 | 0.91 | +0.99 to -2.65 | -0.92 | 0.356 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| without observers | -2.17 | 1.09 | -0.04 to -4.33 | -1.99 | 0.046 |

Mountford window size 4

| with observers | -1.56 | -1 | +2.58 to -1.64 | -0.03 | 0.976 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| without observers | +4.01 | -1 | +1.01 to -1.31 | +0.32 | 0.746 |

${ }^{1}$ Standard errors not calculated on the multiplicative scale and so not comparable.
${ }^{2}$ Population change in NNRR and NSRR is calculated as the weighted mean over all routes of the mean estimated value in the years 1971, 72 and 73 minus the mean estimated value in years 1989, 90 and 91 (see appendix 1). Therefore the estimates cover only the middle 21 years of the survey.

Table 4.10. Estimates of overall population trend for the Belted Kingfisher between 1968 and 1995 (data selected using stringent data selection criteria).

| Estimation method | Trend (\% per year) | SE | 95\% CI | $z$ | $p$ |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Linear route regression |  |  |  |  |  |
| USNBS RR | -0.70 | 0.13 | -0.43 to -0.97 | -5.21 | $<0.001$ |
| USNBS est. eqns. RR | -1.47 | 0.41 | -0.66 to -2.28 | -3.63 | <0.001 |
| CWS RR | -0.81 | - ${ }^{1}$ | -0.35 to -1.26 | -3.50 | $<0.001$ |
| Nonlinear route regression |  |  |  |  |  |
| NSRR | $-2.10^{2}$ | - ${ }^{1}$ | -0.39 to -3.53 | -2.42 | 0.015 |
| NNRR | $-1.41^{2}$ | - 1 | -0.45 to -2.27 | -2.88 | 0.004 |
| Rank-trends | - | - | - | -3.32 | 0.001 |
| TRIM Model 2, incorporating serial correlation and overdispersion |  |  |  |  |  |
| with observers | -1.08 | 0.20 | -0.66 to -1.49 | -5.14 | <0.001 |
| without observers | -0.75 | 0.15 | -0.45 to -1.04 | -5.00 | $<0.001$ |
| Mountford window size 4 |  |  |  |  |  |
| with observers | -0.52 | - 1 | +0.14 to -1.11 | -1.27 | 0.204 |
| without observers | -0.42 | - 1 | +0.15 to -1.03 | -1.13 | 0.258 |

${ }^{1}$ Standard errors not calculated on the multiplicative scale and so not comparable.
${ }^{2}$ Population change in NNRR and NSRR is calculated as the weighted mean over all routes of the mean estimated value in the years 1972, 73 and 74 minus the mean estimated value in years 1989, 90 and 91 (see appendix 1). Therefore the estimates cover only the middle 20 years of the survey.

Table 4.11. Comparison of loglinear models for the analysis of population trends in the Belted Kingfisher using BBS data from 1968-1995 (see Table 4.4 legend for further information). All models failed the Pearson $\chi^{2}$ goodness-of-fit test ( $p<0.001$ in all cases).

| Model | No. of params. | Deviance | Pearson $\chi^{2} /$ d.f. | Serial correlation | AIC ${ }^{1}$ | QAIC ${ }^{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| With observer effects, with serial correlation ${ }^{3}$ |  |  |  |  |  |  |
| 5 (region) ${ }^{4}$ | 1618 | 11553.46 | 1.11 | -0.05 | 14789.46 | 13673.01 |
| 4 (stratum) | 1570 | 11687.10 | 1.11 | -0.04 | 14827.10 | 13697.74 |
| 3 | 1564 | 11646.63 | 1.11 | -0.04 | 14774.63 | 13649.18 |
| 4 (region) | 1540 | 11719.67 | 1.12 | -0.04 | 14799.67 | 13667.16 |
| 2 | 1538 | 11732.80 | 1.12 | -0.04 | 14808.80 | 13675.02 |
| 1 | 1537 | 11760.91 | 1.12 | -0.04 | 14834.91 | 13698.42 |
| With observer effects, without serial correlation ${ }^{3}$ |  |  |  |  |  |  |
| 5 (region) ${ }^{4}$ | 1618 | 11555.99 | 1.11 | - | 14791.99 | 13675.30 |
| 4 (stratum) | 1570 | 11691.19 | 1.11 | - | 14831.19 | 13701.43 |
| 3 | 1564 | 11648.46 | 1.11 | - | 14776.46 | 13650.83 |
| 4 (region) | 1540 | 11723.62 | 1.12 | - | 14803.62 | 13670.73 |
| 2 | 1538 | 11736.63 | 1.11 | - | 14812.63 | 13678.48 |
| 1 | 1537 | 11764.47 | 1.12 | - | 14838.47 | 13701.63 |
| Without observers effects, with serial correlation ${ }^{3}$ |  |  |  |  |  |  |
| 5 (region) ${ }^{4}$ | 610 | 13314.98 | - | - | 14534.98 | 13248.31 |
| 4 (stratum) | 562 | 13445.86 | 1.24 | 0.02 | 14569.86 | 13270.54 |
| 3 | 556 | 13415.91 | 1.24 | 0.02 | 14527.91 | 13231.49 |
| 4 (region) | 532 | 13529.54 | 1.25 | 0.03 | 14647.00 | 13286.14 |
| 2 | 530 | 13532.04 | 1.25 | 0.03 | 14592.04 | 13284.40 |
| 1 | 529 | 13558.41 | 1.25 | 0.03 | 14616.41 | 13306.22 |
| Without observer effects, without serial correlation ${ }^{3}$ |  |  |  |  |  |  |
| 5 (region) ${ }^{4}$ | 610 | 13313.47 | 1.23 | - | 14553.47 | 13246.95 |
| 4 (stratum) | 562 | 13442.69 | 1.24 | - | 14566.69 | 13267.68 |
| 3 | 556 | 13415.59 | 1.24 | - | 14527.59 | 13231.20 |
| 4 (region) | 532 | 13526.12 | 1.24 | - | 14590.12 | 13283.05 |
| 2 | 530 | 13528.61 | 1.24 | - | 14588.61 | 13281.30 |
| 1 | 529 | 13554.61 | 1.24 | - | 14612.61 | 13302.79 |

${ }^{1}$ Akaike information criterion (AIC) $=$ deviance $+2 *$ no. free parameters.
${ }^{2}$ Quasi-likelihood AIC (QAIC) $=$ deviance/dispersion $+2 *$ no. free parameters. The dispersion value used was 1.11 , from model 5 (region) with observer effects and serial correlation, the best fitting model for which an estimate of dispersion could be obtained.
${ }^{3}$ Model 4 (route) and model 5 (stratum) filed to converge
${ }^{4}$ Models came close to convergence, but failed to find accurate parameter estimates.
Table 4.12. Summary of problems noted in case studies. Eight analysis methods were applied to continental BBS data for four species under two different data selection criteria (stringent and tolerant). ${ }^{1}$

| Method | Cerulean Warbler |  | Carolina Wren |  | Chestnut-collared Longspur |  | Belted Kingfisher |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Stringent | Tolerant | Stringent | Tolerant | Stringent | Tolerant | Stringent | Tolerant |
| Linear RR |  |  |  |  |  |  |  |  |
| Least squares |  |  |  |  |  |  |  |  |
| CWS RR | B | A | G | G | A | A | A | A |
| USNBS RR | B | A | G | G | A | B, F | A | A |
| Estimating equations | A | A | G | G | A | A | A | A |
| Nonlinear RR |  |  |  |  |  |  |  |  |
| Smoothing methods |  |  |  |  |  |  |  |  |
| NNRR | C | C | G | G | A | A | A | A |
| NSRR | C | C | G | G | A | A | C | C |
| Rank-trends | A | A | G | G | A | A | A | A |
| Loglinear modeling | D | D | D | D | B, D, F | D | D | D |
| Mountford moving windows | E, F | E, F | H | H | E, F | A | A | A |

[^6]Table 4.13. Effect of incorporating observer differences on trend estimates for four species from three analyses (one prior analysis that used the linear RR method and two analyses in this chapter that used both "with observer" and "without observer" models).

| Species | Sauer et al. 1994b ${ }^{1}$ | Smoothing RR ${ }^{2}$ | Loglinear modeling $^{3}$ |
| :--- | :---: | :---: | :---: |
| Cerulean Warbler | not significant | Decrease | decrease |
| Carolina Wren | decrease | decrease | decrease |
| Chestnut-collared | decrease | increase | increase |
| Warbler | increase | decrease | decrease |
| Belted Kingfisher |  |  |  |

[^7]
## Figure Legends

Figure 4.1. Counts of Cerulean Warblers on each BBS route between 1966 and 1995. Only the 41 routes that passed the stringent data selection criteria are shown. The two routes that together contain 45 percent of the total counts are highlighted with dark lines (upper line is route 82031 , lower line is route 82032 ).

Figure 4.2. Relationship between measures of variation in counts on a BBS route and the mean count on that route for four species. If counts on route follow a log-normal distribution then the standard deviation would be proportional to the mean and the plots of standard deviation (stdev)/mean (i.e., coefficient of variation) against mean would show no trend. If counts were Poisson distributed then the variance would be equal to the mean and the plots of variance/mean against mean would show no trend, and have an average value of 1.0. Values higher than 1.0 in the variance/mean-mean plots indicate overdispersion relative to the Poisson distribution.

Figure 4.3. Population indices for Cerulean Warblers between 1966 and 1995, calculated using eight different estimators. BBS data were screened using stringent selection criteria (see text). As with all subsequent figures, the population indices have been scaled so that their value in the first year is 1.0 , and are plotted on a logarithmic axis with log-base 2.

Figure 4.4. Population indices for the Cerulean Warbler between 1966 and 1995 calculated using model 3 of program TRIM. The dashed lines are $95 \%$ parametric confidence limits for the comparison between the index estimate and the index estimate in first year. In figures c ) and d) the confidence limits have been scaled to account for the estimated overdispersion in these models (Table 4.4).

Figure 4.5. Mountford moving-window estimates of population indices for Cerulean Warblers between 1966 and 1996. The dashed lines are $95 \%$ bootstrap consistency intervals. Estimates outside the range of the consistency intervals indicate failure of assumptions of the Mountford model.

Figure 4.6. Population indices for the Cerulean Warbler between 1966 and 1995, calculated using eight different estimators, with data selected using tolerant data selection criteria.

Figure 4.7. Population indices for the Carolina Wren between 1967 and 1995, calculated using eight different estimators, with data selected using stringent data selection criteria.

Figure 4.8. Mountford moving-window estimates of population indices for Carolina Wren between 1966 and 1996 for "without observers" models at 4 window sizes. The dashed lines are $95 \%$ bootstrap confidence intervals.

Figure 4.9. Population indices for the Chestnut-collared Longspur between 1967 and 1995, calculated using eight different estimators, with data selected using stringent data selection criteria.

Figure 4.10. Population indices for the Chestnut-collared Longspur between 1967 and 1995, calculated using seven different estimators (program TRIM failed to produce log-linear model estimates), with data selected using tolerant data selection criteria.

Figure 4.11. Population indices for the Belted Kingfisher between 1968 and 1995, calculated using eight different estimators, with data selected using stringent data selection criteria.

Figure 4.12. Population indices for the Belted Kingfisher between 1968 and 1995, calculated using eight different estimators, with data selected using tolerant data selection criteria.

Figure 4.13. Relationship between mean abundance and estimated route trend for the linear route regression estimators using tolerantly selected data for Belted Kingfisher between 1968 and 1995 ( $\mathrm{n}=1603$ routes). For both the USNBS and CWS methods, trend estimates for lowest abundance routes are biased towards zero.

Figure 4.14. Residual analysis from loglinear models. These graphs are comparable with those on the right side of Fig. 4.2 (variance/mean against mean), except that these use the residuals and predicted values of each count using the best-fitting loglinear model for each species. If counts were distributed as Poisson random variables the graphs should show no trend and an average variance/predicted value of 1.0. In all cases, a linear regression line shows an increasing trend with increasing predicted value.

Figure 4.1

a) Cerulean Warbler

b) Carolina wren

c) Chestnut-collared longspur

d) Belted kingfisher






Figure 4.3


c) without observers, with serial correlation

d) without observers, without serial correlation


b) with observers, window size 6


d) without observers, window size 6



c) Other methods, with observers

b) Linear route regression

d) Other methods, without observers


Figure 4.8


Figure 4.9





Figure 4.13

b) CWS route regression

c) USNBS est. eqns. route regression



## Chapter 5

## General Conclusions

## Introduction

In this thesis I have investigated the statistical methods used to estimate long-term population change from extensive wildlife surveys like the North American Breeding Bird Survey (BBS). I have reviewed previous work (Chapter 1), demonstrated the importance of analysis method and data selection (Chapter 2), synthesized the issues involved and shown the potential of simulation studies to resolve some of these issues (Chapter 3), and compared the methods empirically using case studies (Chapter 4).

Here, I bring this work together by providing an evaluation of the methods, with recommendations for when they should be used and how they may be further developed. The evaluations are summarized in Tables 5.1 (annual index methods) and 5.2 (trend estimation methods); these are intended to complement the summary descriptions of the methods given in Tables 3.1 and 3.2. I also provide a simplified guide to selecting among the methods (Fig 5.1).

In the second part of the chapter, I discuss some future directions for research and suggest three promising new approaches to trend analysis. Lastly, I discuss the implications of my research for the use of extensive wildlife surveys like the BBS in estimating population change.

## Evaluation of analysis methods

## Annual index methods

Calculating annual indices is appropriate where the focus is on the actual yearly population changes, rather than the underlying trend. This is useful when the goal is to detect sudden changes in abundance (e.g., the 1976-77 population crash in Carolina Wren, Fig. 4.7), or when further modeling is to be undertaken (e.g., the effect of winter temperature on breeding population fluctuations in Song Thrush, Baillie 1990). Annual indices are also used for judging the fit of trend estimates (e.g., Peterjohn et al. 1995; see Collins 1994 for further discussion of
the uses of annual indices). In all of these cases the goal is to produce index estimates that accurately reflect the true population changes.

Of the six annual index methods summarized in table 3.1 , three (chaining, imputing and linear modeling) can be discounted from consideration on the basis of previous work (Chapter 1; Table 5.1) and were not included in the empirical evaluations in this thesis. This leaves three candidate methods: Mountford's method, the residual method, and loglinear modeling.

Mountford method: Even with the moving-window extension, Mountford's method performed poorly in three out of four case study species (Chapter 4). Poor performance was probably caused by violation of the assumption that population change is the same at all sites. This assumption will likely only be met in small datasets or geographically homogeneous species, so the method cannot be recommended for general use.

Residual method: This approach was developed to provide informal indices that complement linear regression trend estimates (Sauer and Geissler 1990). As the original authors noted, the indices are biased when the trend lines are invalid, and this was confirmed in at least one of the case studies (Fig 4.10b; see also Figs. 4.3b, 4.6b, 4.11b and Discussion, Chapter 4). The method is clearly not useful for judging the fit of trend estimates, and cannot be generally recommended without further studies to determine the situations under which it works well.

Loglinear modeling: Loglinear modeling, with separate effects for each year, can be recommended on both theoretical and empirical grounds. Theoretically, it allows a defensible model of the data to be constructed using the well-established tools of generalized linear modeling (GLM, McCullagh and Nelder 1989). Empirically, the method worked well in the case studies, although the final model was not satisfactory for one species (see discussion, Chapter 4). One disadvantage of the method is the practical problems associated with implementation: complex models require a large amount of computer time and sometimes fail to converge, even using the specially written computer program TRIM (Chapter 4).

Conclusion: The best annual index method appears to be loglinear modeling, although there may be problems implementing the method for large datasets and for species that show high variation in counts over time and space.

## Trend analysis methods

Determining population trends is the principal goal of many long-term monitoring programs like the BBS. Unlike annual index analyses the focus is no longer on the actual population changes. Instead it is the underlying tendency in abundance that is of interest, and this is defined subjectively (Chapter 3). Thus, the assumed model of trend is a key factor that differentiates the analysis methods (Table 3.1).

The methods evaluated in this thesis used three trend models: linear multiplicative, additive and rank (Fig 3.1; polynomial multiplicative models have been suggested but not implemented, and can be viewed as a type of additive model (Hastie and Tibshirani 1990)). The linear model requires estimation of fewest parameters (so is useful when there are many missing data), but is only realistic for most populations for relatively short time periods (perhaps 5-15 years for most species monitored by the BBS). Additive trend models are more realistic, but require more parameters and are also susceptible to incorporating changes in abundance that may not be regarded as trend. In rank models only the direction of trend is estimated, not its magnitude. Rank models are therefore more robust, but less informative.

The second major difference between trend analysis methods is in the approach to trend estimation (Table 3.1). Route regression (RR) methods first estimate trends on each site (route) and then combine these to produce regional trend estimates (a form of multilevel or hierarchical modeling, Goldstein 1995). By contrast, loglinear modeling estimates trend directly from the counts at a level that can be determined by the geographic variation in the data. Both approaches are well established in the statistical literature and I have not been able to determine in this thesis whether one is generally superior. Both have practical problems associated with them: route regression methods have problems associated with the imprecision of route trend estimates and the methods for combining route trends to form regional estimates, while loglinear modeling can run into computational problems with large, complex models. I consider these issues in the specific context of each method in the following paragraphs.

Linear route regression: This method has a number of practical advantages including low computational requirement and ability to deal with much missing data. Of the two methods
for estimating route trends (Table 3.1), the estimating equations approach is preferred because it is not biased by low counts (Link and Sauer 1994; Chapter 4). However, a number of other decisions must be made when implementing the method (Table 2.1) and it is often not clear which is best. Different implementations usually produced reasonable (although divergent) estimates in the empirical comparisons, except in some cases where poorly estimated weightings (used in combining route trends) lead to obviously unfeasible results (Chapters 2 and 4). The example simulation study (Chapter 3) indicated that the variance estimates may be too small and that one implementation (CWS RR) may be negatively biased, although these results are preliminary. In conclusion, the method may be recommended where there are a priori reasons to wish to estimate linear trends, and it will be particularly useful in large surveys with many missing counts.

Smoothing route regression: The primary advantage of the smoothing methods (NNRR and NSRR) is their ability to estimate nonlinear (additive) trends. This flexibility comes at a price: the methods cannot deal with much missing data. The case studies (Chapter 4) also showed that there can be significant problems with estimates at the beginning and end of the time series, and with fitting the observer-effects model (NSRR). Assuming that these problems can be resolved (see Discussion, Chapter 4), the smoothing RR methods can be recommended for most situations where there are few missing data, or where the analyst is willing to use stringent data selection to exclude sites with few counts.

Rank trends route regression: As implemented, the rank trends method uses a naive approach to combine route trends into regional estimates and does not take account of differences among observers. The empirical comparisons showed that it tends to give more statistically significant results than the other methods (Chapters 2 and 4), and the simulations showed that it may be biased by geographic variation in trend (Chapter 3). The method may be further developed, but given that rank models are intrinsically uninformative (because they only estimate the direction of trend, not its magnitude) this is not a high priority.

Loglinear modeling: This method provides an appealing alternative to linear RR. The modeling framework is well established and there are fewer arbitrary decisions to be made. The
case studies did not focus on linear trend modeling using this approach because annual indices were found to give a better fit to the data in most cases (using Akaike's Information Criterion, Chapter 4). It would be useful to extend this method to allow fitting of additive trend models (see next section). Practical problems fitting large, complex models mean that this method is best suited to relatively small datasets.

Regression of annual indices: This approach is a natural extension of annual index analyses, but relies on the indices being unbiased (e.g., poor performance of linear regression on Mountford indices, Chapter 4). The best annual index method is loglinear modeling, but I did not evaluate the use of loglinear model indices because the method can also be used to produce trend estimates directly. It is unlikely that this approach would offer any improvements over direct trend estimation.

Conclusion: Unlike annual index estimation, it is not possible to recommend one single method for trend estimation in all situations. The choice of method depends on the assumed model of trend and the approach to trend estimation (Fig. 5.1). The length of time series, size of dataset and amount of missing data are all important considerations when making these decisions. Depending on the situation, one or more of the following may be appropriate: linear $R R$, smoothing RR or loglinear modeling. All of these methods suffer from practical drawbacks and could benefit from further development.

## Future research

## Development of current methods

For linear RR, the biggest issue is the choice of weightings. Three weightings are used: area, abundance and precision weights (Table 2.1). Of these, it is the occasional very high abundance weight that causes the few unreliable trend estimates (Chapters 2, 4). This can be avoided by capping (Winsorizing) the abundance weights, by using another method to estimate the weights, or by including the uncertainty in abundance weight estimates into the precision weights (Chapter 4). Precision weighting is also controversial (James et al. 1990). One option worth investigating is the use of hierarchical Bayes methods (Gustafson 1997) to combine route
trends into regional estimates. This approach would automatically account for the precision of the route trend estimates and could also be used to constrain the estimates to within reasonable limits by choosing a suitable prior distribution for route trend.

The issue of weightings is not so problematic for the smoothing RR methods because they do not require abundance weights (James et al. 1990). There are, however, two other improvements that must be made before the methods can be considered credible: (1) increased reliability of estimates at both ends of the time series, perhaps through extrapolation of the route trend estimates, and (2) solution of the numerical problems associated with fitting the NSRR model (Chapter 4). Another area of potential development is in the selection of the amount of smoothing to perform on the counts, which is currently set subjectively but could be set according to the amount of between-year variability in the data (Chapter 4).

The loglinear models used in Chapter 4 (both annual index and trend) could be improved in several ways. Alternative count distributions should be investigated, such as the negative binomial and Poisson-logistic (Chapter 4). Models that incorporate site-specific linear trends and overall regional annual indices simultaneously (e.g., Sauer and Geissler 1990; Collins 1994) should be tried. More realistic models for trend should be considered, such as additive models (see below). Lastly, if the method is to be useful for large datasets then there must be improvements in the fitting algorithm (Chapter 4).

## Further evaluation of methods

The empirical and theoretical work in this thesis has resulted in a substantial advance in our understanding of the relative merits of the analysis methods. Nevertheless, a number of important questions remain unanswered: (1) what is the relative accuracy of the $R R$ approach compared with loglinear modeling? (2) which weighting schemes produce the most reliable results in $R R$ methods? (3) how robust are the methods to mis-specification of trend? (4) are the variance estimates of the methods biased? I suggest that these questions can be answered by simulation studies similar to the example in Chapter 3.

The model used to simulate data in Chapter 3 could be improved in a number of respects. Firstly it could be made more realistic by incorporating a simple stochastic population model.

For example, assume the actual population size in year $y$ at site $s, n_{y s}$, is a Poisson random variable:

$$
n_{y s} \sim \operatorname{Poisson}\left(N_{y s}\right)
$$

where $\mathrm{N}_{\mathrm{ys}}$ is the expected population size. This is given by a discrete version of the simple logistic growth model:

$$
N_{y s}=n(y-1) s\left(1+R\left(1-n(y-1) s / K_{y s}\right)\right)
$$

where $R$ is the per capita growth rate and $K_{y s}$ is the carrying capacity (Krebs 1985). $R$ could be set at various values between 1 and, say, 1.5 to represent different species. (In a more complex model $R$ could depend upon $y$ and/or $s$.) $K_{y s}$ could be set to change through time using an additional function that depends upon the assumed local trend, regional trend, and distribution of random interventions. Trends could be linear-multiplicative or could change over time. The distribution of $N_{0 s}$ and $K_{0 s}$ among sites could be set by the geographic models proposed in Chapter 3 (Fig. 3.2) or by some other non-geographic model. A similar simulation model was proposed by Greenwood (1991).

A second improvement would be to explicitly model observation error by assuming

$$
c_{y s} \sim \operatorname{binomial}\left(n_{y s}, p_{y s}\right)
$$

where $c_{y s}$ is the count in year $y$ at site $s$, and $p_{y s}$ is the expected proportion of the $n_{y s}$ individuals that are counted. $p_{y s}$ could change over time depending upon observer, and could also vary stochastically (giving $c_{y s}$ an overdispersed binomial distribution).

Simulation studies are a useful tool for investigating the properties of the methods under known conditions, but they can be contentious because the results depend upon the model and the range of parameter values used to generate the simulated data. One solution to this problem would be to invite all interested parties (developers of methods and users of results) to co-operate in planning and executing the simulation trials. In this approach, an initial workshop would be held to agree a general reference model, range of parameters, and evaluation criteria. Many simulated datasets would then be generated by an independent coordinator and sent to the developers of the methods for analysis. The analysts would not know the exact parameters used in the simulations. These would then be returned to the coordinator, analyzed, and the results
discussed at a follow-up workshop. Further simulations that are more specific may be agreed upon to address questions that arise. This approach (on a much larger scale) was successfully used by the International Whaling Commission to resolve the contentious issue of whale stock management procedures (International Whaling Commission 1992).

## Alternative approaches

Generalized additive modeling (GAM): GAM is an extension of generalized linear modeling where the independent variable is nonlinearly related to the covariables using smooth functions (Hastie and Tibshirani 1990). This approach provides an additive model counterpart to loglinear modeling, allowing smooth trends in abundance over time to be modeled. In a preliminary analysis, Fewster and Buckland (1996) investigated the feasibility of using GAMs to model British Common Bird Census Data. They compared GAM with loglinear modeling and found GAMs to be superior on both theoretical and empirical grounds (comparison of trend estimates for 14 species). They suggested the use of auto-regression (in which the predicted count from the previous year is included as a covariable) to model serial dependence of counts, and also provided a method of identifying significant changes in trend (by testing for non-zero approximate second derivative of predicted count).

This approach could, in theory, readily be applied to other datasets. However, no specialist software is available, and the large memory requirements of complex models using standard software (e.g., function gam in S-plus, StatSci 1995) may severely limit the amount of data that can be analyzed. Nevertheless, GAMs often require fewer interaction terms than comparable linear models (Hastie and Tibshirani 1990) so this may not be a problem in practice.

As with loglinear modeling, the approach could be extended to include alternative count distributions, such as negative binomial and Poisson-logistic. Another intriguing possibility is to fit a spatio-temporal GAM for counts, allowing more precise inferences to be made about the geographic patterns of population change.

State-space modeling: The observed variation in counts comes from many sources: population trend, irregular perturbations, density dependent responses and measurement error (Chapter 3). However, all the current estimators assume a form for trend and put the remaining
non-trend variation into a single catch-all error term. This is inefficient and uninformative. State-space models (Harvey 1989; Aoki 1990) attempt to model and estimate these components explicitly. The approach is theoretically equivalent to traditional ARIMA time-series analysis (Box and Jenkins 1970) but offers several practical advantages, such as easier model construction, fitting and interpretation, and better handing of multivariate (i.e., multi-site) data.

In state-space modeling, the system (population in this case) is described in terms of parameters which change through time according to rules described in transition equations. The system is not observed directly but is sampled, so introducing noise (error). This is modeled by an observation equation. For example, a simple model of counts has the observation equation

$$
c_{y s}=n_{y s} * e_{y s}
$$

(where $c_{y s}$ is the observed count at time $y$ and site $s, n_{y s}$ is the actual population size, and $e_{y s}$ is the measurement error) and the transition equations

$$
\begin{aligned}
& n_{y s}=n_{(y-1) s} * b_{y-1} * f_{y} \\
& b_{y}=b_{y-1} * g_{y}
\end{aligned}
$$

(where $b_{y}$ is the regional trend in year $y, f_{y}$ represents stochastic interventions and $g_{y}$ is the change in trend over time. The distribution family of the error terms $e, f$ and $g$ is assumed known). Estimates of $b$ give the trend over time, and estimates of $n$ can be used to calculate annual indices (area weighted mean, $n_{y o}$ ). The model could be refined by allowing $b, f$ and $g$ to vary by geographic stratum and by the addition of other explanatory covariables (such as weather and observer) in the same way as loglinear modeling. One practical advantage over the GLM and GAM approaches is that the estimation procedure (Kahlman filtering) is much less computer-intensive and that updating the estimates to account for a new year of data is simple (Harvey 1989).

Fitting ecological population models: As Solow (1995) has commented, ecologists are fond of constructing theoretical population models, but rarely attempt to fit them to real data. However, ecologically-based (as opposed to purely empirical) modeling has two important advantages. Firstly, the parameters are biologically meaningful. It is far easier to interpret estimates of, say, changes in carrying capacity or per-capita growth rate than a smoothed
population trend curve. Secondly, the models are likely to be a better approximation to reality, and may therefore provide a better (i.e., more parsimonious) fit to the data. An example of a candidate population model is the simulation model proposed in the previous section. Such a model could be fit using maximum likelihood techniques or, with the addition of suitable priors, Bayesian methods. An example of the former approach for intensive bird count data is given in Lebreton (1990).

The models could be extended to incorporate information from other data sources. For example, in North America there are at least six regional or continental surveillance programs that provide information about relative bird abundance and three that give demographic information (Butcher et al. 1993). These could be combined using a simple demographic model in either a likelihood or Bayesian framework (Hillborn and Walters 1992; Rafferty et al. 1995 and subsequent papers). This approach has the potential to provide far more information about the dynamics of wildlife populations, and begin to explicitly address the second major goal of biological monitoring programs: generating preliminary hypotheses of causation.

In reality it is likely that the quality of the data will seriously constrain the inferences that can be drawn from fitting population models, especially for the more ambitious demographic models. The exercise may serve instead to highlight the parameters for which more information is required. In either case, it will be informative and is worth attempting.

## Implications of research

To be effective, population monitoring programs must provide reliable estimates of population change. However, as I have shown in this thesis, estimating population change from extensive survey data is a considerable statistical challenge. A wide diversity of analysis methods are available, but each is problematic in some areas. I have recommended that the use of some methods be discontinued, and have attempted to provide guidelines for choosing among those that remain (Fig 5.1). Nevertheless, there are a number of outstanding issues that make it difficult to select an appropriate method in many situations. I have suggested a consensus-based simulation approach to help resolve these remaining issues, but it seems likely that there will
always be more than one seemingly appropriate method available in some circumstances. In these cases I recommend using all of the candidate methods and comparing the results to see how robust they are to the different approaches. If the different methods lead to substantially different results, the conclusions about population changes should be tempered accordingly.

The field of population change estimation is evolving rapidly. All of the methods recommended here were newly developed or significantly improved during the period of this research, and I have suggested some further developments and alternative approaches. Improvements in current methods will, I hope, resolve the practical problems that constrain their use, while new approaches will allow for increased scope of inferences (e.g., estimating separate components of the time series with state-space modeling). Nevertheless, the accuracy of the results will always be constrained by unquantified biases in the survey methodology, such as an incomplete sampling frame or unknown but systematic changes in measurement error (Chapter 3). New surveys should be carefully designed to minimize these biases (e.g., the new British Breeding Bird Survey, Gregory and Carter 1994), while existing surveys must be validated as much as possible (e.g., Bart et al. 1995).

Confronted with all of the problems outlined in this thesis, the reader may be tempted to ask whether reliable estimates of population change can ever be derived from large-scale monitoring schemes such as the BBS. To some extent they already are: trends from the BBS are positively correlated with those from a wide range of independent monitoring schemes, each of which has different survey biases and uses different methods of analysis (Butcher et al. 1990; Hagan et al. 1992, Hussell et al. 1992, Cyr and Larivée 1993). Long-term, extreme changes in the population status of birds well sampled by the BBS are almost certain to be detected using any of the recommended methods of analysis. However the degree to which these schemes can meet more stringent targets for accuracy, such as those laid out by Partners in Flight (Butcher et al. 1993, Chapters 2 and 3), is unknown. Determining and realizing the full potential of extensive monitoring programs will require strong commitments by survey biologists to investigate and minimize potential sources of bias and by statisticians to evaluate and improve their methods of analysis.
Table 5.1. Summary evaluation of methods that have been used to estimate annual indices from bird count data ${ }^{1}$.

| Method | Discussed in chapters | Advantages | Disadvantages | Extensions | Recommendations for use |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Chaining | 1,3 | - easy to compute and understand | - biased <br> - can be very inaccurate | - Mountford method | Not recommended |
| Mountford | 1,3 | - higher precision than chaining method | - sensitive to assumption that population change Is the same at all sites | - Mountford with moving window <br> - loglinear modeling | Not generally recommended may be satisfactory for small or spatially homogeneous datasets over a short time span. |
| - with moving window | 1, 4 | - allows more years than Mountford method | - sensitive to choice of window size; - questionable performance in case studies (Ch. 4) | - loglinear modeling | Same as above, except can work over a longer time span. |
| Loglinear modeling | 1, 3, 4 | - flexible approach to analysis, allowing covariables to be incorporated as needed - use of GLM tools for model selection, criticism and inference | - fitting complex models can be problematic <br> - computationally intensive <br> - may not adequately model some species (Ch. 4) | - alternative count distributions <br> - include route*linear trend interactions <br> - include other covariables (e.g., weather) | Recommended for general use (with reservations about large datasets and species with large variation in counts between times or routes). |
| Imputing | 1,3 |  | - as implemented, requires log-transform <br> - slow convergence | - loglinear modeling | Not recommended |
| Linear | 1,3 | - flexible approach to analysis possible | - assumption of normal distribution of response var | - loglinear modeling | Not recommended |
| Residual | 3, 4 | - indices match linear route regression trends | - indices may be biased towards trend line |  | Not generally recommended requires further study. |

${ }^{1}$ See Table 3.1 for summary of methods and references to their use. Rows in this table correspond with columns in Table 3.1.
Table 5.2. Summary evaluation of methods that have been used to estimate trends from bird count data ${ }^{1}$.

| Method | Discussed in chapters | Advantages | Disadvantages | Extensions | Recommendations for use |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Linear route regression |  | Route regression: <br> - route-specific variation in trend and covariables easily incorporated <br> - low computation requirement <br> - facilitates geographic modeling of trends <br> Log-linear trend: <br> - simple trend statistic <br> - not biased by other components of time series <br> - works when much missing data | Route regression: <br> - combining route trends into regional estimates is problematic <br> Log-linear trend: <br> - not realistic for longer time periods <br> - test of model (residual indices) questionable |  | Log-linear trend useful for relatively short time periods (approx. 5-10 years in BBS data) or when there are a priori reasons to suspect linearity. Good when many missing counts. |
| - Least squares | 1,2, 3, 4 |  | - biased for routes with small counts | - estimating equations | - not recommended |
| - Estimating equations | 1,3,4 | - not biased by small counts <br> - minimal assumptions about distribtuion of counts |  | - more flexible trend models (additive) | - recommended |
| Nonlinear route regression |  | Route regression: <br> - as linear route regression | Route regression: <br> - combining route trends still problematic (but less so than linear route regression) |  |  |
| - Smoothing methods (NNRR and NSRR) | 1,3,4 | -additive trend model is flexible | - cannot tolerate much missing data <br> - requires subjective specification of smoothing parameter - as implemented, end-effects and fitting can be problematic | - data-based selection of smoothing parameter <br> - alternative fitting <br> algorithms <br> - alternative smoothers <br> - extend estimates beyond end of route time series | Assuming extensions solve implementation problems, recommended for longer time periods where there are few missing counts. |
| - Rank-trends | 2, 3, 4 | -rank trend model makes minimal assumptions | - results not informative <br> - no route-specific covariables <br> - no weightings developed | - inter-observer <br> differences <br> - weighting scheme | Not recommended (except as a last resort) |

Table 5.2. (continued)

| Method | Discussed in chapters | Advantages | Disadvantages | Extensions | Recommendations for use |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Loglinear modeling | 1, 3, 4 | - flexible approach to analysis, allowing covariables to be incorporated as needed - use of GLM tools for model selection, criticism and inference | - fitting complex models can be problematic <br> - computationally intensive - may not adequately model some species (Ch. 4) <br> - log-linear trend not realistic for longer time periods | - alternative count <br> distributions <br> - include other covariables <br> (e.g., weather) <br> - smooth trends <br> (generalized additive <br> modeling, GAM) | Relatively short time series as implemented. With GAM extension, may be generally applicable, although complex models may prove problematic to fit. (Species with large variation in counts between times or routes may also be problematic.) |
| Regression of annual indices | 3 | - natural extension to annual index analyses | - cannot model geographic variation in trend | - use of annual index variance-covariance estimates | Not evaluated, but unlikely to be superior to loglinear modeling. |

${ }^{1}$ See Table 3.2 for summary of methods and references to their use. Rows in this table correspond with columns in Table 3.2.

- 132 -


## Figure Legends

Figure 5.1. A flow-chart guide to choosing among the analysis methods. Only methods that are recommended are included in the chart.

Figure 5.1


## Literature Cited

Anderson, D. R., K. P. Burnham, and G. C. White. 1994. AIC model selection in overdispersed capture-recapture data. Ecology 75: 1780-1793.

Anganuzzi, A. A. 1993. A comparison of tests for detecting trends in abundance indices of dolphins. Fishery Bulletin 91: 183-194.

Aoki, M. 1990. State Space Modeling of Time Series, 2nd Edition. Springer-Verlag, Berlin, Germany.

Askins, R. A. 1993. Population trends in grassland, shrubland, and forest birds in eastern North America. Pages 1-34 in D. M. Power, editor. Current Ornithology. Volume 11. Plenum Press, New York.

Askins, R. A., J. F. Lynch, and R. S. Greenberg. 1990. Population declines in migratory birds in eastern North America. Pages 1-57 in D. M. Power, editor. Current Ornithology. Volume 7. Plenum Press, New York.

Bailey, R. S. 1967. An index of bird population changes on farmland. Bird Study 14: 195-209.
Baillie, S. R. 1990. Integrated population monitoring of breeding birds in Britain and Ireland. Ibis 132: 151-166.

Bard, Y. 1974. Nonlinear Parameter Estimation. Academic Press, New York, New York.
Barker, R. J., and J. R. Sauer. 1992. Modeling population change from time series data. Pages 182-194 in D. R. McCullough and R. Barrett, editors. Wildlife 2001: Populations. Elsevier, New York, New York.

Bart, J., M. Hofschen, and B. G. Peterjohn. 1995. Reliability of the Breeding Bird Survey: effects of restricting surveys to roads. The Auk 112: 758-761.

Böhning-Gaese, K., M. L. Taper, and J. H. Brown. 1993. Are declines in North American insectivorous songbirds due to causes on the breeding range? Conservation Biology 7: 76-86.

Borland International Inc. 1994. Borland C++ for OS/2 User's Guide, Version 2.0. Borland International Inc., Scotts Valley, CA.

Box, G. E. P., and G. M. Jenkins. 1970. Time Series Analysis, Forecasting and Control. Holden-Day, San Fransisco, CA.

Bradu, D., and Y. Mundlac. 1970. Estimation in lognormal linear models. Journal of the American Statistical Association 65: 198-211.

Brown, J. H. 1984. On the relationship between the abundance and distribution of species. The American Naturalist 124: 255-279.

Buckland, S. T. 1984. Monte Carlo confidence intervals. Biometrics 40: 811-817.

Burnham, K. P., and D. R. Anderson. 1992. Data-based selection of an appropriate biological model: the key to modern data analysis. Pages 16-30 in D. R. McCullough and R. Barrett, editors. Wildlife 2001: Populations. Elsevier, New York, New York.

Butcher, G. S. 1990. Audubon Christmas Bird Counts. Pages 5-14 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Butcher, G. S., M. R. Fuller, L. S. McAllister, and P. H. Geissler. 1990. An evaluation of the Christmas Bird Count for monitoring population trends for selected species. Wildlife Society Bulletin 18: 129-139.

Butcher, G. S., B. G. Peterjohn, and C. J. Ralph. 1993. Overview of national bird population monitoring programs and databases. Pages 192-203 in D. M. Finch and P. W. Stangel, editors. Status and Management of Neotropical Migratory Birds; Proceedings of the 1992 Partners in Flight National Training Workshop; September 21-25; Estes Park, Colorado. General Technical Report RM-229. U.S.D.A. For. Serv. Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Bystrak, D. 1981. The North American Breeding Bird Survey. Pages 34-41 in C. J. Ralph and J. M. Scott, editors. Estimating Numbers of Terrestrial Birds. Studies in Avian Biology 6. Cooper Ornithological Society, Lawrence, Kansas.

Cameron, A. C., and P. K. Trivedi. 1990. Regression-based tests for overdispersion in the Poisson model. Journal of Econometrics 46: 347-364.

Collins, B. T. 1990. Using rerandomizing tests in route-regression analysis of avian population trends. Pages 63-70 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Collins, B. T. 1994. Annual population indices from BBS data. Unpublished Canadian Wildlife Service internal document, March 311994.

Collins, B. T., and J. S. Wendt. 1989. The Breeding Bird Survey in Canada, 1966-1983: Analysis of Trends in Breeding Bird Populations. Technical Report 75. Canadian Wildlife Service, Ottawa, Ontario.

Cyr, A., and J. Lariveé. 1993. A checklist approach for monitoring neotropical migrant birds: twenty-year trends in birds of Québec using EPOQ. Pages 229-237 in D. M. Finch and P. W. Stangel, editors. Status and Management of Neotropical Migratory Birds; Proceedings of the 1992 Partners in Flight National Training Workshop; September 21-25; Estes Park, Colorado. General Technical Report RM-229. U.S.D.A. For. Serv. Rocky Mountain Forest and Range Experiment Station, Fort Collins, CO.

Downes, C. M., and B. T. Collins. 1996. The Canadian Breeding Bird Survey, 1966-1994. Progress Note 210. Canadian Wildlife Service, Ottawa, Ontario.

Droege, S. 1990. The North American Breeding Bird Survey. Pages 1-4 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Eberhardt, L. L. 1978. Appraising variability in population studies. Journal of Wildlife Management 42: 207-238.

Efron, B. 1982. The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

Ehrlich, P. R., D. S. Dobkin, and D. Wheye. 1988. The Birder's Handbook. Simon \& Schuster Inc., New York, New York.

Erskine, A. J. 1978. The First Ten Years of the Co-operative Breeding Bird Survey in Canada. Canadian Wildlife Service Report 42. Canadian Wildlife Service, Ottawa, Ontario.

Erskine, A. J., B. T. Collins, E. Hayakawa, and C. M. Downes. 1992. The Cooperative Breeding Bird Survey in Canada, 1989-91. Progress Note 199. Canadian Wildlife Service, Ottawa, Ontario.

Fewster, R. M., and S. T. Buckland. 1996. Quantifying population trends from CBC counts. Unpublished contract report to British Trust for Ornithology, MAFF contract CSA3109.

Finch, D. M., and P. W. Stangel. 1993. Introduction. Pages 1-4 in D. M. Finch and P. W. Stangel, editors. Status and Management of Neotropical Migratory Birds; Proceedings of the 1992 Partners in Flight National Training Workshop; September 21-25; Estes Park, Colorado. General Technical Report RM-229. U.S.D.A. For. Serv. Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

Flather, C. H., and J. R. Sauer. 1996. Using landscape ecology to test hypotheses about largescale abundance patterns in migratory birds. Ecology 77: 28-35.
fPrint UK Ltd. 1996. Virtual Pascal Users Guide, Version 1.0. fPrint UK Ltd., London, UK.
Geissler, P. H. 1984. Estimation of animal population trends and annual indices from a survey of call-counts or other indications. Proceedings of the American Statistical Association, Section on Survey and Research Methods 1984: 472-477.

Geissler, P. H., and W. A. Link. 1988. Bias of animal population trend estimates. Pages 755759 in E. J. Wegman, D. T. Gantz and J. J. Miller, editors. Computing Science and Statistics; Proceedings of the 20th Symposium on the Interface; Fairfax, Virginia; April 1988. American Statistical Society, Alexandria, Louisiana.

Geissler, P. H., and B. R. Noon. 1981. Estimates of avian population trends from the North American Breeding Bird Survey. Pages 42-51 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. Studies in Avian Biology 6. Cooper Ornithological Society, Lawrence, Kansas.

Geissler, P. H., and J. R. Sauer. 1990. Topics in route-regression analysis. Pages 54-57 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Godambe, V. P. 1991. Estimating Functions. Clarendon Press, Oxford, UK.
Goldstein, H. 1995. Multilevel Statistical Models. Edward Arnold, London, UK.

Greenwood, J. J. D. 1991. Notes on simulations to investigate 'random walk' in CBC and WBS. Unpublished British Trust for Ornithology internal document, June 131991.

Greenwood, J. J. D., S. R. Baillie, H. Q. P. Crick, J. H. Marchant, and W. J. Peach. 1993. Integrated population monitoring: detecting the effects of diverse changes. Pages 267-342 in R. W. Furness and J. J. D. Greenwood, editors. Birds as Monitors of Environmental Change. Chapman and Hall, London, U.K.

Greenwood, J. J. D., S. R. Baillie, R. D. Gregory, W. J. Peach, and R. J. Fuller. 1994. Some new approaches to conservation monitoring of British breeding birds. Ibis 137: S16-S28.

Gregory, R. D., and S. Carter. 1994. Results from the Pilot Census Project. BTO News 194: 1213.

Gustafson, P. 1997. Large hierarchical Bayesian analysis of multivariate survival data. Biometrics 53: 230-242.

Hagan, J. M. 1992. Conservation biology when there is no crisis - yet. Conservation Biology 6: 475-476.

Hagan, J. M., T. L. Lloyd-Evans, J. L. Atwood, and D. S. Wood. 1992. Long-term changes in migratory landbirds in the north-eastern United States: evidence from migration capture data. Pages 115-130 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington D.C.

Hart, J. D., and T. E. Wehrly. 1986. Kernel regression estimation using repeated measurements data. Journal of the American Statistical Association 81: 1080-1088.

Harvey, A. C. 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge, UK.

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized Additive Models. Chapman and Hall, London, UK.

Heckert, J. R. 1997. Bobolink Dolichonyx oryzivorus population decline in agricultural landscapes in the midwestern USA. Biological Conservation 80: 107-112.

Hellawell, J. M. 1991. Development of a rationale for monitoring. Pages 1-14 in F. B. Goldsmith, editor. Monitoring for Conservation and Ecology. Chapman and Hall, London, U.K.

Hilbe, J. H. 1994. Log negative binomial regression using the GENMOD procedure SAS/STAT software. Pages 1199-1204 in Proceedings of the Ninteenth Annual SAS Users Groups International Conference. SAS Institute, Inc., Cary, North Carolina.

Hillborn, R., and C. J. Walters. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty. Chapman and Hall, London, UK.

Holmes, R. T., and T. W. Sherry. 1988. Assessing population trends of New Hampshire forest birds: Local vs. regional patterns. The Auk 105: 756-768.

Hunter, W. C., M. F. Carter, D. N. Pashley, and K. Barker. 1993. The Partners in Flight species prioritization scheme. Pages 109-120 in D. M. Finch and P. W. Stangel, editors. Status and Management of Neotropical Migratory Birds; Proceedings of the 1992 Partners in Flight National Training Workshop; September 21-25; Estes Park, Colorado. General Technical Report Rm-229. U.S.D.A. For. Serv. Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

Hussell, D. J. T., M. H. Mather, and P. H. Sinclair. 1992. Trends in numbers of tropical- and temperate-wintering migrant landbirds in migration at Long Point, Ontario, 1961-1988. Pages 101-114 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington, D.C.

International Whaling Commission. 1992. Report of the Sub-Committee on Management Procedures. Pages 87-136 in Forty-Second Report of the International Whaling Commission. International Whaling Commission, Cambridge, UK.

James, F. C., and C. E. McCulloch. 1995. The strength of inferences about causes of trends in bird populations. Pages 40-51 in D. M. Finch and T. E. Martin, editors. Ecology and Management of Neotropical Migrant Birds: a Synthesis and Review of the Critical Issues. Oxford University Press, Oxford, U.K.

James, F. C., C. E. McCulloch, and L. E. Wolfe. 1990. Methodological issues in the estimation of trends in bird populations with an example: the Pine Warbler. Pages 84-98 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

James, F. C., D. A. Wiedenfeld, and C. E. McCulloch. 1992. Trends in breeding populations of warblers: declines in the southern highlands and increases in the lowlands. Pages 43-56 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington D.C.

James, F. C., C. E. McCulloch, and D. A. Wiedenfeld. 1996. New approaches to the analysis of population trends in land birds. Ecology 77(1): 13-27.

Jassby, A. D., and T. M. Powell. 1990. Detecting changes in ecological time series. Ecology 71: 2044-2052.

Johnson, D. H., and M. D. Schwartz. 1993. The Conservation Reserve Program and grassland birds. Conservation Biology 7: 934-937.

Kendall, W. L., B. G. Peterjohn, and J. R. Sauer. 1996. First-time observer effects in the North American Breeding Bird Survey. The Auk 113: 823-829.

Krebs, C. J. 1985. Ecology: The Experimental Analysis of Distribution and Abundance, 3rd Edition. Harper \& Row, New York, New York.

Lacy, R. C., and C. E. Bock. 1986. The correlation between range size and local abundance of some North American birds. Ecology 67: 258-260.

Lawton, J. H. 1993. Range, population abundance and conservation. Trends in Evolution and Ecology 8: 409-413.

Lebreton, Jean-D. 1990. Modelling density dependence, environmental variability, and demographic stochasticity from population counts: an example using Wytham Wood Great Tits. Pages 89-102 in J. Blondel, editor. Population Biology of Passerine Birds. SpringerVerlag, Berlin, Germany.

Lebreton, Jean-D., K. P. Burnham, J. Clobert, and D. R. Anderson. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monographs 62: 67-118.

Lehman, E. L. 1975. Nonparametrics - Statistical Methods Based on Ranks. Holden-Day Inc., San Francisco, California.

Li, X., and N. E. Heckman. 1996. Local linear forecasting. Technical Report 167. University of British Columbia, Department of Statistics, Vancouver, Canada.

Link, W. A., and J. R. Sauer. 1994. Estimating equation estimates of trends. Bird Populations 2: 23-32.

Link, W. A., and J. R. Sauer. 1994. Within-site variability in surveys of wildlife populations. Ecology 75: 1097-1108.

Link, W. A., and J. R. Sauer. 1995. Estimation and confidence intervals for empirical mixing distributions. Biometrics 51: 810-821.

Link, W. A., and J. R. Sauer. 1996. Extremes in ecology: avoiding the misleading effects of sampling variation in summary analyses. Ecology 77: 1633-1640.

Manly, B. F. J. 1991. Randomization and Monte Carlo Methods in Biology. Chapman and Hall, London, U.K.

Marchant, J. 1990. Common Bird Census report 1988-1989. Pages 52-61 in D. Stroud and D. Glue, editors. Britain's Birds in 1989-90: the Conservation and Monitoring Review. British Trust for Ornithologý, Thetford, U.K.

McCullagh, P., and J. A. Nelder. 1989. Generalized Linear Models, 2nd Edition. Chapman and Hall, London, UK.

Morton, E. S. 1992. What do we know about the future of migrant landbirds? Pages 75-84 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington D.C.

Moses, L. E., and D. Rabinowitz. 1990. Estimating (relative) species abundance from route counts of the Breeding Bird Survey. Pages 71-79 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Moss, D. 1985. Some statistical checks on the BTO Common Birds Census index - 20 years on. Pages 175-179 in K. Taylor, R. J. Fuller and P. C. Lack, editors. Bird Census and Atlas Studies, Proceedings of the CIII International Conference on Bird Census and Atlas Work. British Trust for Ornithology, Tring, UK.

Mountford, M. D. 1982. Estimation of population fluctuations with application to the Common Bird Census. Applied Statistics 31: 135-143.

Mountford, M. D. 1985. An index of population change with application to the Common Bird Census. Pages 121-132 in B. J. T. Morgan and P. M. North, editors. Statistics in Ornithology. Lecture Notes in Statistics 29. Springer-Verlag, Berlin, Germany.

O'Connor, R. J. 1992a. Population variation in relation to migrancy status in some North American birds. Pages 64-74 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington D.C.

O'Connor, R. J. 1992b. Trends in populations: introduction. Pages 23-25 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington.

Ogilvie, M.A. 1967. Population changes and mortality of the mute swan in Britain. Wildfowl 18: 64-73.

Pannekoek, J., and A. J. van Strien. 1996. TRIM (TRends \& Indices for Monitoring data). Research paper 9634. Statistics Netherlands, Voorburg, The Netherlands.

Peach, W. J., and S. R. Baillie. 1994. Implementation of the Mountford indexing method for the Common Bird Census. Pages 653-662 in E. J. M. Hagemeijer and T. J. Verstrael, editors. Bird Numbers 1992. Distribution, monitoring and ecological aspects. Proceedings of the 12th International Conference of International Bird Census Committee and European Ornithological Atlas Committee. Statistics Netherlands, Voorburg, The Netherlands.

Peterjohn, B. G., and J. R. Sauer. 1993. North American Breeding Bird Survey annual summary 1990-1991. Bird Populations 1: 1-15.

Peterjohn, B. G., and J. R. Sauer. 1995. Population trends of the loggerhead shrike from the North American Breeding Bird Survey. Pages 117-121 in R. Yosef and F. Lohrer E, editors. Shrikes (Laniidae) of the world: biology and conservation. Proceedings of the First International Shrike Symposium. Proceedings of the Western Foundation of Vertebrate Zoology 6.

Peterjohn, B. G., J. R. Sauer, and W. A. Link. 1994. The 1992-1993 summary of the North American Breeding Bird Survey. Bird Populations 2: 46-61.

Peterjohn, B. G., J. R. Sauer, and W. A. Link. 1996. The 1994 and 1995 summary of the North American Breeding Bird Survey. Bird Populations 3: 48-66.

Peterjohn, B. G., J. R. Sauer, and C. S. Robbins. 1995. Population trends from the North American Breeding Bird Survey. Pages 3-39 in D. M. Finch and T. E. Martin, editors. Ecology and Management of Neotropical Migrant Birds: a Synthesis and Review of the Critical Issues. Oxford University Press, Oxford, U.K.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in C: the Art of Scientific Computing, 2nd Edition. Cambridge University Press, Cambridge, U.K.

Rafferty, A. E., G. H. Givens, and J. E. Zeh. 1995. Inference from a deterministic population dynamics model for Bowhead Whales (with discussion). Journal of the American Statistical Association 90: 402-430.

Robbins, C. S., and W. T. van Velzen. 1970. Progress report on the North American Breeding Bird Survey. Pages 22-30 in S. Svensson, editor. Bird Census Work and Environmental Monitoring. Bulletin from the Ecological Research Committee 9. Swedish Natural Science Research Council, Stockholm, Sweden.

Robbins, C. S., D. Bystrak, and P. H. Geissler. 1986. The Breeding Bird Survey: Its First Fifteen Years, 1965-1979. Resource Publication 157. U.S. Fish and Wildlife Service, Washington D.C.

Robbins, C. S., J. R. Sauer, R. S. Greenberg, and S. Droege. 1989. Population declines in North American birds that migrate to the neotropics. Proceedings of the National Academy of Science, U.S.A. 86: 7658-7662.

Robbins, C. S., J. W. Fitzpatrick, and P. B. Hamel. 1992. A warbler in trouble: Dendroica cerulea. Pages 549-578 in J. M. Hagan and D. W. Johnston, editors. Ecology and conservation of neotropical migrant landbirds. Smithsonian Institution Press, Washington, DC.

Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1986. Akaike Information Criterion Statistics. Mathematics and Its Applications (Japanese Series). KTK Scientific Publishers, Tokyo, Japan.

SAS Institute Inc. 1993a. The GENMOD Procedure, Release 6.09. Technical Report P-243. SAS Institute Inc., Cary, North Carolina.

SAS Institute Inc. 1993b. SAS Companion for the OS/2 Environment, Version 6, 2nd Edition. SAS Institute Inc., Cary, North Carolina.

Sauer, J. R., and S. Droege. 1990a. Recent population trends of the Eastern Bluebird. Wilson Bulletin 102: 239-252.

Sauer, J. R., and S. Droege, editors. 1990b. Survey Designs and Statistical Methods for the Estimation of Animal Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Sauer, J. R., and S. Droege, editors. 1990c. Scissor-tailed Flycatcher analysis. Pages 157-166 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Sauer, J. R., and S. Droege. 1990d. Wood Duck population trends from the North American Breeding Bird Survey. Pages 225-231 in L. H. Fredrickson, G. V. Burger, S. P. Havera, D. A. Graber, R. E. Kirby and T. S. Taylor, editors. Proceedings of the 1988 North American Wood Duck Symposium, St. Louis, Missouri.

Sauer, J. R., and S. Droege. 1992. Geographic patterns in population trends of neotropical migrants in North America. Pages 26-42 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington D.C.

Sauer, J. R., and P. H. Geissler. 1990. Estimation of annual indices from roadside counts. Pages 58-62 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Sauer, J. R., D. D. Dolton, and S. Droege. 1994a. Mourning dove population trend estimates from Call-Count and North American Breeding Bird Surveys. Journal of Wildlife Management 58: 506-515.

Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994b. Observer differences in the North American Breeding Bird Survey. The Auk 111: 50-62.

Sauer, J. R., G. W. Pendleton, and B. G. Peterjohn. 1996. Evaluating causes of popualtion changes in North American insectivorous songbirds. Ecology 10: 465-478.

Searle, S. R., F. M. Speed, and G. A. Milliken. 1980. Population marginal means in the linear model: an alternative to least squares means. The American Statistician 34: 216-221.

Sokal, R. R., and J. F. Rohlf. 1981. Biometry. W H Freeman and Co., New York, New York.
Solow, A. R. 1995. Fitting population models to time series data. Pages 21-27 in T. M. Powell and J. H. Steele, editors. Ecological Time Series. Chapman and Hall, New York, New York.

StatSci. 1995. S-PLUS Guide to Statistical and Mathematician Analysis, Version 3.3. StatSci, a division of Mathsoft, Inc., Seattle, WA.

Taub, S. R. 1990. Smoothed scatterplot analysis of long-term Breeding Bird Census data. Pages 80-83 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Terborgh, J. 1989. Where Have All the Birds Gone?: Essays on the Biology and Conservation of Birds That Migrate to the American Tropics. Princeton University Press, Princeton, New Jersey.

Terborgh, J. 1992. Why American songbirds are vanishing. Scientific American 1992(May): 98-104.
ter Braak, C. J. F., A. J. van Strien, R. Meijer, and T. J. Verstrael. 1994. Analysis of monitoring data with many missing values: which method? Pages 663-673 in E. J. M. Hagemeijer and T. J. Verstrael, editors. Bird Numbers 1992. Distribution, monitoring and ecological aspects. Proceedings of the 12th International Conference of International Bird Census Committee and European Ornithological Atlas Committee. Statistics Netherlands, Vooburg, The Netherlands.

Titus, K. 1990. Trends in counts of scissor-tailed flycatchers based on a nonparametric ranktrend analysis. Pages 164-166 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Titus, K., M. Fuller, and D. Jacobs. 1990. Detecting trends in hawk migration count data. Pages 105-113 in J. R. Sauer and S. Droege, editors. Survey Designs and Statistical Methods for the Estimation of Avian Population Trends. Biological Report 90(1). U.S. Fish and Wildlife Service, Washington D.C.

Underhill, L. G., and R. P. Prys-Jones. 1994. Index numbers for waterbird populations. I: Review and methodology. Journal of Applied Ecology 31: 463-480.
U.S. Fish and Wildlife Service. 1995. Nongame Birds of Management Concern - The 1995 List. U.S. Fish and Wildlife Service, Office of Migratory Bird Management, Washington, D.C. (Available on the internet at http://www.fws.gov/~r9mbmo/reports/speccon/tblconts.html)

Venables, W. N., and B. D. Ripley. 1994. Modern Applied Statistics with S-Plus. Statistics and Computing. Springer-Verlag, New York, New York.

Villard, Marc-A., and B. A. Maurer. 1996. Geostatistics as a tool for examining hypothesized declines in migratory songbirds. Ecology 77: 59-68.

White, G. C., and R. E. Bennetts. 1996. Analysis of frequency count data using the negative binomial distribution. Ecology 77: 2549-2557.

Wilcove, D. S., and J. Terborgh. 1984. Patterns of population decline in birds. American Birds 38: 10-13.

Winkelmann, R. 1994. Count Data Models. Econometric Theory and an Application to Labor Mobility. Lecture Notes in Econometric and Mathematical Systems 410. Springer-Verlag, Berlin, Germany.

Witham, J. W., and M. L. J. Hunter. 1992. Population trends of neotropical migrant landbirds in northern coastal New England. Pages 85-95 in J. M. Hagan and D. W. Johnston, editors. Ecology and Conservation of Neotropical Migrant Landbirds. Smithsonian Institution Press, Washington D.C.

## Appendix 1

## Computer programs for analysis of population change

In this appendix I list the computer programs that I used, and give details of how they may be obtained.

Most of the analyses described in this thesis cannot be performed using standard statistical software without considerable macro programming. There is one exception: loglinear modeling can be performed by any package that allows generalized linear modeling, and I did some of the loglinear analyses in Chapter 4 using SAS (v6.11, SAS Institute Inc. 1993b; see Chapter 4, Discussion). In all other cases, however, I used custom-written computer programs that I either developed myself or adapted from programs supplied by the originators of the methods (Table A1.1). I modified the programs that were sent to me for four reasons: (1) to enable them to work with the continent-wide BBS dataset (programs NNRR, NSRR, TRIM, MOUNTF); (2) to update the analysis methods (MOUNTF); (3) to alter the input, output, and error handling capabilities (NNRR and NSRR); and (4) to enable them to work with my compilers (TRIM, MOUNTF).

All programs were compiled and run on a 90 Mhz Pentium PC with 32MB RAM operating under IBM OS/2 (v2.11 for the analyses in Chapter 2 and $3, v 4$ for the analyses in Chapter 4). The compilers were Borland C++ v2.0 for OS/2 (Borland International Inc. 1994), Virtual Pascal v1.0 (fPrint UK Ltd. 1996), and a port of GNU Fortran (g77-0.5.19) by Michael Holzapfel (unpublished) for the emx runtime environment (version 0.9c fix 2, Eberhard Mattes, unpublished; available from ftp://ftp.hobbes.nmsu.edu/pub/os2/dev/emx/).

I will supply the source code or complied versions of the programs that I wrote upon request; however I cannot distribute source code for any programs that are based on the work of others without their prior consent. The names and addresses of the original authors are given in Table A1.1.
Table A1.1. List of computer programs used in this thesis.

| Program Name | Analysis method(s) | Language ${ }^{\text {1 }}$ | Authors of original code | Address |
| :---: | :---: | :---: | :---: | :---: |
| USNBS | USNBS RR, est. eqn. RR, residual method | C | L. Thomas ${ }^{2}$ | c/o Dr. K. Martin, Faculty of Forestry, University of British Columbia, \#270-2357 Main Mall, Vancouver, BC V6T $1 Z 4$. Canada. |
| BBSRR | CWS RR, residual method | F77 | B. T. Collins | Canadian Wildlife Service, National Wildlife Research Centre, Ottawa, Ontario K1A 0H3. Canada. |
| NNRR | NNRR | F77 | D. A. Wiedefeld | Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana 70803. USA. |
| NSRR | NSRR | F77 | D. A. Wiedenfeld |  |
| NRT | rank-trends analysis | C | L. Thomas | see earlier |
| TRIM | loglinear modeling | Pascal | J. Pannekoek and <br> A. van Strien | Statistics Netherlands, PO Box 4000, 2270 JM Voorburg. The Netherlands. |
| MOUNTF | Mountford movingwindow | F77 | W. J. Peach, P.A. Whittington and S.R. Baillie | British Trust for Ornithology, The Nunnery, Thetford, Norfolk IP24 2PU. UK. |
| SETUP | data handling ${ }^{3}$ | C++ | L. Thomas | see earlier |
| SETUPSP | data handling ${ }^{3}$ | C++ | L. Thomas | see earlier |
| BBSSIM | simulation study, Ch. 3 | C | L. Thomas | see earlier |

## ${ }^{1} \mathrm{C}=$ ANSI standard $\mathrm{C} ; \mathrm{C}++=$ Borland $\mathrm{C}++;$ F77 = FORTRAN 77; Pascal = Borland Turbo Pascal.

${ }^{2}$ The USNBS program includes some numerical routines written by Press et al. (1992)
${ }^{3}$ Program SETUP extracts survey information from the raw data files and performs data selection, storing the selected surveys to intermediate files; program SETUPSP extracts observations for one species from the raw data files, merges the count data with the selected survey data, and outputs files in the formats that are required by the analysis programs.

## Appendix 2

## Breeding Bird Survey data

The Breeding Bird Survey data are in the public domain, and are available upon request to the Canadian Wildlife Service (Canadian data) or the Biological Resources Division, U.S. Geological Survey (North American data). Addresses are given below.

The data used in this thesis were taken from the 1966-1995 dataset, supplied on CDROM by the Biological Resources Division, U.S. Geological Survey. However, not all of the 4191 routes in this dataset should be analyzed (B. G. Peterjohn, pers. comm.). A small number (219) have been established in non-random locations (such as within national parks) or use nonstandard methods (such as surveying away from roads). These are recognizable in the dataset because they are coded with route numbers of 800 or greater. In addition 48 routes were established in northern Mexico in 1994 and have not yet built up enough data to be useful for trend analysis. These routes were excluded from the dataset before any analysis was performed, leaving 3924 routes (Fig. 1.1).

For the analyses in Chapter 4, I excluded a further 260 routes that are located in states or provinces that contain physiographic strata where the density of routes is so low that a meaningful analysis of population change is not possible (Table A2.1). Another 3 routes are the only ones in a physiographic stratum within a state, which makes estimation of between-trend variance impossible using linear route regression methods (Table A2.1). This left 3664 routes that were submitted to the data selection criteria outlined in Table 4.1 (see also Table 4.2).

## Addresses for obtaining data

Canadian Wildlife Service, Migratory Bird Surveys Division, National Wildlife Research Centre, Ottawa, Ontario K1A 0H3. Canada.

Biological Resources Division, U.S. Geological Survey, Breeding Bird Survey, Patuxent Environmental Science Center, 12100 Beech Forest Road, Laured MD 20708. USA.

Table A2.1. Number of Breeding Bird Survey routes and surveys (1966-1995) in physiographic strata excluded from the case studies of Chapter 4.

| Physiographic stratum | State(s) / Province(s) | n. routes | n. surveys |
| :--- | :---: | :---: | :---: |
| 3 - Coastal Flatwoods | Tennessee | 1 | 3 |
| 9 - Glaciated Coastal Plain | Rhode Island | 1 | 8 |
| 14 - Upper Coastal Plain | Virginia | 1 | 11 |
| 25 - Open Boreal Forest | All | 44 | 170 |
| 29 - Closed Boreal Forest | All | 107 | 769 |
| 68 - Northern Rockies | All | 49 | 236 |
| 94 - N. Pacific Rainforest | Alaska | 18 | 94 |
| 96 - S. Alaska Coast | Alaska | 20 | 117 |
| 99 - Tundra | All | 19 | 53 |
| Total excluded |  | 260 | 1461 |
| Total remaining in BBS dataset |  | 3664 | 54027 |
| after exclusion of these data |  |  |  |


[^0]:    ${ }^{1}$ The U.S. federal agency that administers the BBS has changed name (and sometimes mandate) several times since the survey began (Bureau of Sport Fisheries and Wildlife 1966-1974; Fish and Wildlife Service 1974-1993; National Biological Survey 1993-1995; National Biological Service 1995-1996; Biological Resources Division, Geological Survey 1996-present). Throughout this dissertation I will use the name by which it is currently most commonly known: the U.S. National Biological Service (USNBS).
    ${ }^{2}$ All figures and tables appear at the end of the chapter in which they are referenced.

[^1]:    ${ }^{1}$ USNBS $=$ U.S. National Biological Service; CWS $=$ Canadian Wildlife Service.
    ${ }^{2}$ See Geissler (1984), Geissler \& Sauer (1990) and Peteriohn et al. (1995).
    ${ }^{3}$ See Collins and Wendt (1989) and Erskine et al. (1992).

[^2]:    References: ${ }^{1}$ Geissler and Sauer (1990), ${ }^{2}$ Erskine et al. (1992), ${ }^{3}$ Link and Sauer (1994), ${ }^{4}$ James et al. (1996), ${ }^{5}$ Titus et al. (1990), ${ }^{6}$ Pannekoek and van Strien (1996), ${ }^{7}$ Greenwood et al. (1994), ${ }^{8}$ Böhning-Gaese et al. (1993)

[^3]:    ${ }^{1}$ USNBS = U.S. National Biological Service; CWS = Canadian Wildlife Service
    ${ }^{2}$ Units are \% rate of annual change
    ${ }^{3}$ Asterisks indicate the statistical significance of Wilcoxon signed-rank sum tests of the hypothesis that the median bias is 0 (Median bias column) or that the median estimated standard deviation is equal to the actual standard deviation calculated from the 100 estimates (Median estimated std. dev. column). n.s. $=\mathrm{p}>0.05 ;{ }^{* * *}=\mathrm{p}<0.001$.

[^4]:    ${ }^{1}$ The stringent and tolerant data selection criteria are defined in the text. Minimal criteria exclude only routes with non-standard survey procedures and those located in some peripheral state-strata (Appendix 2), and routes where no birds of the species were counted. The minimal criterion was not used in the analysis, but is included for comparison.

[^5]:    ${ }^{1}$ Akaike information criterion (AIC) $=$ deviance $+2 *$ no. free parameters.
    ${ }^{2}$ Models came close to convergence, but failed to find accurate parameter estimates.
    ${ }^{3}$ Model 4 (route) failed to converge.

[^6]:    ${ }^{1} \mathrm{Key}: \mathrm{A}=$ no problems noted; $\mathrm{B}=$ residual method annual indices appeared biased toward trend line; $\mathrm{C}=$ problems with smoothed indices at the beginning/end of the time series; $\mathrm{D}=$ some models failed to converge; $\mathrm{E}=$ index estimates outside the $95 \%$ confidence intervals; $\mathrm{F}=$ unrealistic estimates; $\mathrm{G}=$ trends model cannot represent the underlying dynamics of the species; $\mathrm{H}=$ problems determining a suitable window size.

[^7]:    ${ }^{1}$ Analysis of the correlation between difference between observers (calculated from linear RR analysis) and time. Pearson correlations were reported for the three species with statistically significant results: $r=0.15,0.23$ and -0.05 respectively.
    ${ }^{2}$ Difference in overall trend estimates between NSRR and NNRR methods, from Tables 4.3, 4.7, 4.9 and 4.10. In all cases, $95 \%$ confidence intervals for NSRR and NNRR estimates overlap by a large margin, indicating that the differences are small in relation to the variability in trend estimates.
    ${ }^{3}$ Difference in overall trend estimates between model 2 "with observers" and "without observers", from Tables 4.3, 4.7, 4.9, and 4.10. Introducing observer effects into the models resulted in a lower AIC in all cases except Belted Kingfisher, but in all cases the $95 \%$ confidence intervals for the estimates overlap.

