

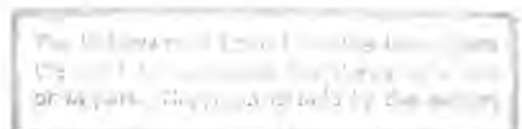
**A COMPARATIVE STUDY OF STOCHASTIC  
MODELS IN BIOLOGY**

by

**ANABELA de GUSMÃO BRANDÃO**

Thesis Presented for the Degree of  
DOCTOR OF PHILOSOPHY  
in the Department of Statistical Sciences  
UNIVERSITY OF CAPE TOWN

October 1997



The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

DST 510 BRAN

98/11127

## ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Walter Zucchini who supervised this thesis before emigrating to Germany and to Professor Les Underhill who took over the supervision. My thanks goes for them having brought the various research topics to my attention, as well as for their encouragement and guidance throughout this study.

Thanks are due to the following people for their cooperation in providing the data sets used in this thesis:

Redshanks <i>Tringa totanus</i>	R.W. Summers	Royal Society for the Protection of Birds, Inverness, Scotland
	L.G. Underhill	Avian Demography Unit, University of Cape Town Cape Town, South Africa
Tristan Buntings <i>Nesospiza acunhae</i>	P.G. Ryan	Percy FitzPatrick Institute, University of Cape Town Cape Town, South Africa
Dunlin <i>Calidris alpina</i>	P.E. Jönsson	Department of Ecology University of Lund Lund, Sweden
Springbok <i>Antidorcas marsupialis</i>	H.J. van Hensbergen	Dept. of Nature Conservation University of Stellenbosch Stellenbosch, South Africa
<i>Solidago altissima</i> (tall goldenrod)	B. Schmid	Botanical Institute University of Basel Basel, Switzerland
Willow Warblers <i>Phylloscopus trochilus</i>	G. Noskov and N. Lapshin	Department of Zoology St. Petersburg State University St. Petersburg, Russia.

The financial assistance of a bursary from the Foundation for Research Development and a Gordon Sprigg Scholarship are acknowledged.

Finally I wish to thank my family whose continual support and encouragement spurred me on to complete this thesis. My thanks also go to all my friends who have prayed and encouraged me throughout this study.



## ABSTRACT

In many instances, problems that arise in biology do not fall under any category for which standard statistical techniques are available to be able to analyse them. Under these situations, specific methods have to be developed to solve and answer questions put forward by biologists. In this thesis four different problems occurring in biology are investigated. A stochastic model is built in each case which describes the problem at hand. These models are not only effective as a description tool but also afford strategies consistent with conventional model selection processes to deal with the standard statistical hypothesis testing situations. The abstracts of the papers resulting from these problems are presented below.

### **Correlation under uncertainty of classification, with application to non-random mating**

A prerequisite for any study of non-random mating appears to be the ability to be able to classify individuals with certainty as male or female. In this study a model for assortative mating is developed when only an imperfect, or probabilistic, classification is available. The Redshank *Tringa totanus* is a species for which the sexes of the breeding pairs of observations could not be determined solely from biometric measurements. The Redshank data of Summers and Underhill (*Bird Study* (1991) 35: 169–180) are used to illustrate the model for assortative mating. The Dunlin *Calidris alpina* data of Jönsson (*Ornis Scandinavica* (1987) 18: 257–264) and the Tristan Bunting *Nesospiza acunhae* data of Ryan (*PhD Thesis* (1992)), for which the sexes are identifiable, are used to test the proposed model and to provide some guidance on the required sample sizes of both sexed birds to give information on the means and the standard deviations for males and females, and of the number of unsexed breeding pairs. Standard statistical hypothesis testing methods are developed to test the assumption of non-assortative mating. Two scenarios are considered: firstly, we regard the parameters of males and females as unknown and needing to be estimated, and secondly, we regard these parameters as known. The proposed model is demonstrated to be able

to determine the presence or absence of assortative mating in bird species, even when there is uncertainty about the sexes of the individuals of a pair.

## **Estimation of the demographic ratios of herding species: a case study of sex ratios of the Springbok *Antidorcas marsupialis***

Estimates of sex ratios are usually based on the assumption that individual animals constitute the natural sampling unit. However all sexually reproductive animals have social structures in order to bring the opposite sexes together and the appropriate sampling unit is therefore the “social structure” rather than the individual. Confidence intervals of sex ratios that ignore this sampling unit can give misleading results. In this study, several populations of Springbok *Antidorcas marsupialis*, which form three different types of herd (breeding herds, bachelor herds and solitary herds) are used to illustrate various confidence intervals constructed for the sex ratio. Both parametric and non-parametric confidence intervals for the sex ratio are considered. Binomial confidence intervals are too conservative. Confidence intervals obtained with standard errors of the ratio computed from the Taylor expansion of the ratio and of the logged ratio are less conservative. The Taylor expansion of the logged ratio produces better standard errors for the ratio. Bootstrap confidence intervals were similar to those obtained by the Taylor expansion of the logged ratio.

## **A model for the relationship between reproductive mass and vegetative mass in *Solidago altissima***

Plant reproductive mass is an example of data censored from the left; it is never negative and becomes positive only once a certain vegetative mass threshold has been reached. These observations follow the pattern of primary moult indices, except that there is no censoring on the right. In this study we adapt the Underhill-Zucchini moult model (*Ibis* (1988) 130: 358–372 and (1990) 132: 118–123) (effectively a two-sided censored regression model)

to handle observations that are only censored from the left. The data of Schmid *et al.* (*The American Naturalist* (1994) 143: 494–507) for *Solidago altissima* L. (tall goldenrod) are used to illustrate the method of maximum likelihood to describe the relationship between reproductive mass and vegetative mass. The approach taken by Underhill and Zucchini adapted to formulate a one-sided censored regression model for the relationship between vegetative mass and reproductive mass problem is shown to be equivalent to the approach taken by Schmid *et al.* . The model parameters from the one approach are transformations of the parameters from the other. Advantages and disadvantages of both algorithms are considered.

### **The analysis of avian primary moult III: inter-annual differences in timing of moult of Willow Warblers *Phylloscopus trochilus* in western Russia**

The Underhill-Zucchini moult model (*Ibis* (1988) 130: 358–372 and (1990) 132: 118–123) is extended to enable the parameters of moult in a number of data sets to be estimated simultaneously in such a way that one of standard statistical hypothesis testing methods are developed and the Akaike Information Criterion is used to facilitate parsimonious model selection procedures. In the application considered, the mean starting date of moult of Willow Warblers *Phylloscopus trochilus* in the St Petersburg region of western Russia is estimated for each of 11 years, while assuming the remaining parameters to be common for all data sets. In eight of the 11 years, the mean starting date, for both sexes combined, lay between 9 and 14 July. With exceptions, the mean starting dates for males and females ran closely in parallel, with males starting moult, on average, about 14 days before females.

### **The analysis of avian primary moult IV: flexible model fitting using the EM algorithm**

Various extensions of the Underhill-Zucchini moult model (*Ibis* (1988) 130: 358–372 and (1990) 132: 118–123) are developed to allow a flexible and

statistically coherent approach to the analysis of variation in the timing and duration of moult in different data sets. The moult indices are constrained to lie between 0 and 1 and as such they are an example of censored observations from both the left and the right. The techniques of censored regression and the expectation-maximization (EM) algorithm (Schmid *et al.* , *The American Naturalist* (1994) 143: 494–507) are used to facilitate the extensions to the moult model. Standard statistical hypothesis testing methods are developed to test the goodness of fit of each model. In the application considered, four models are fitted to the primary moult of Willow Warblers *Phylloscopus trochilus* in the St Petersburg region of western Russia. Each model makes different assumptions about the parameters of moult by allowing none, some or all the parameters to vary between years or localities. Under the assumption of a common standard deviation, the appropriate model for moult when the sex of the birds is ignored, assumes a common duration of moult but variable annual mean starting dates of moult. For males and females, the appropriate moult model assumes variable annual mean starting dates and duration of primary moult. With exceptions, the mean annual starting dates for males and females run closely in parallel, with males starting moult, on average, about 14 days before females. With a few exceptions, annual duration of moult is shorter in males than in females (about 12 days shorter).

# CONTENTS

	PAGE
ACKNOWLEDGEMENTS	iii
ABSTRACT	v
TABLE OF CONTENTS	ix
1 INTRODUCTION	1
2 CORRELATION UNDER UNCERTAINTY OF CLASSIFICATION, WITH APPLICATION TO NON-RANDOM MATING	5
3 ESTIMATION OF THE DEMOGRAPHIC RATIOS OF HERDING SPECIES: A CASE STUDY OF SEX RATIOS OF THE SPRINGBOK <i>Antidorcas marsupialis</i>	43
4 A MODEL FOR THE RELATIONSHIP BETWEEN REPRODUCTIVE MASS AND VEGETATIVE MASS IN <i>Solidago altissima</i>	69
5 THE ANALYSIS OF AVIAN PRIMARY MOULT III: INTER-ANNUAL DIFFERENCES IN TIMING OF MOULT OF WILLOW WARBLERS <i>Phylloscopus trochilus</i> IN WESTERN RUSSIA	81
6 THE ANALYSIS OF AVIAN PRIMARY MOULT IV: FLEXIBLE MODEL FITTING USING THE EM ALGORITHM	109



## CHAPTER 1

# Introduction

Many data analysis problems in biology do not fit into standard statistical methods and require custom-made solutions. This thesis tackles several such problems.

The main objective of this thesis is to take various problems that occur in biology for which no standard statistical procedures are available, to answer the questions posed by the biologist, and to demonstrate that stochastic modelling of such problems can effectively describe these problems. Such models not only provide a concise description of the phenomenon, but also afford the ability of being able to test hypotheses about the system under investigation.

Four different problems have been analysed in this study. One, the primary moult of birds, was analysed using two different approaches and each has been written as a separate chapter. Each problem analysed in this thesis forms a separate entity in that most problems considered here are not related to one another. The thesis has therefore been structured so that each chapter consists of the analysis of a particular problem; each chapter can be extracted from the body of the thesis and is able to stand on its own. To achieve this, each chapter contains an abstract, an introduction, the methods, the conclusions, references and appendices. This introduction to the thesis therefore serves more as a coalition of the thesis and a more in-depth introduction to each problem is given at the beginning of each chapter.

The first problem considered in this thesis is the study of non-random mating, the occurrence of mating combinations that cannot be predicted solely from the marginal distributions of one or more characters in males and females, in species of birds when the criteria for classification as males or females is unreliable or non-existent. Non-random mating is an important topic in evolutionary biology. For many bird species, individual birds cannot easily be categorized as male or female with certainty; yet for many species, especially wader species, sexing criteria are unavailable. A model is proposed for assortative mating which requires only imperfect, or probabilistic, classification. Data on Redshank *Tringa totanus* are used to illustrate the model; data on Dunlin *Calidris alpina* and Tristan Buntings *Nesospiza acunhae*, for which the sexes of the birds were known, are used to validate the procedure and to give guidelines on required sample sizes, both of sexed birds

to provide details on the means and standard deviations for males and females, and of the number of unsexed breeding pairs. This study appears in Chapter 2, entitled “Correlation under uncertainty of classification, with application to non-random mating”.

Secondly, the problem of estimating sex ratios, important demographic parameters for research on, and management of, animal populations, is studied. Usually, estimates of sex ratios are based on the assumption that the sampling unit is the individual. However, most species, such as many species of large and medium sized antelope, form herds as a social structure. Thus, if sampling is done visually while moving through a free population, the appropriate sampling unit is the herd rather than the individual. Data on Springbok *Antidorcas marsupialis*, which form three types of herd; breeding herds, bachelor herds and solitary males, are used to illustrate the construction of sex ratios and confidence intervals for sex ratios when the natural sampling unit is the herd. Chapter 3 describes this study and is entitled “Estimation of the demographic ratios of herding species: a case study of sex ratios of the Springbok *Antidorcas marsupialis*”.

Schmid *et al.* (1994) reviewed applications of censored regression models in biology. The examples they considered were all of one-sided censoring. In particular they considered an example from plant reproduction biology in which both vegetative mass and reproductive mass of plants was measured, and in which reproductive mass is a censored variable in the sense that it is zero until vegetative mass reaches a threshold. The primary moult problem analysed by Underhill and Zucchini (1988) and Underhill *et al.* (1990) is an example of two-sided censoring. Schmid *et al.* (1994), Underhill and Zucchini (1988), and Underhill *et al.* (1990) approached the problems, although similar in structure, in different ways. Schmid *et al.* (1994) formulated the problem as a censored regression model and made use of the EM (expectation-maximization) algorithm to maximize the likelihood function, while Underhill and Zucchini (1988) and Underhill *et al.* (1990) used the maximum likelihood algorithm in their model for censored data and used the Newton-Raphson algorithm to maximize the likelihood function. This study formulates the relationship between reproductive mass and vegetative mass in a similar way to that of the primary moult model of Underhill and Zucchini (1988) and compares the results obtained by the EM and maximum likelihood algorithms. Data on *Solidago altissima* (tall goldenrod) are used to demonstrate the procedure. This study is reported in Chapter 4, entitled “A model for the relationship between reproductive mass and vegetative mass in *Solidago altissima*”.

In Chapter 5 an extension to the model of primary moult proposed by Underhill and

Zucchini (1988) and Underhill *et al.* (1990) is considered. The parameters of interest of primary moult are the mean starting date of moult, the duration of moult and the standard deviation of the starting date of moult. The model of Underhill and Zucchini (1988) and Underhill *et al.* (1990) assumed that all birds sampled have a common set of parameters, regardless of year or place. In this study, the primary moult model is extended using the Newton-Raphson algorithm, to allow a variable mean starting date of moult for each data set. Strategies for hypothesis testing and guidelines for parsimonious selection of model parameters using Akaike's (1973) Information Criterion are also developed. Data collected annually of Willow Warblers *Phylloscopus trochilus* in western Russia are used to illustrate this extension. Chapter 5 contains this study and is entitled: "The analysis of avian primary moult III: inter-annual differences in timing of moult of Willow Warblers *Phylloscopus trochilus* in western Russia".

Further extensions to the primary moult model are possible by allowing some or all the parameters of interest in primary moult to vary with time or place. Allowing the number of parameters to increase in the primary model implies that the use of the Newton-Raphson algorithm to maximize the likelihood function rapidly becomes more complex and intricate to implement because both the first partial derivatives and the matrix of second partial derivatives are needed. In chapter 4 the approaches taken by Schmid *et al.* (1994) and by Underhill and Zucchini (1988) to model censored data were compared. In this study the EM (expectation-maximization) algorithm used by Schmid *et al.* (1994) is extended to model censored data from both the left and from the right as is the case of moult index data. This approach is used because the EM algorithm does not necessitate the computation of either first or second partial derivatives and therefore facilitates further extensions to the primary moult model. The moult model of Underhill and Zucchini (1988) assumed one set of parameters. The extension to the moult model described in chapter 5 allowed the mean starting dates of moult to vary from data set to data set. In this chapter, two further model extensions are considered. These are, the extension in which both the mean starting date of moult and the duration of moult are allowed to vary over the data sets, and the extension in which all parameters of moult are allowed to vary, that is, the mean starting date of moult, the duration of moult and the standard deviation of starting date of moult vary from data set to data set. The annual data of Willow Warblers *Phylloscopus trochilus* in western Russia are used to illustrate these extensions. Strategies for hypothesis testing are also developed to verify assumptions made about the parameters of the moult model. This

study is reported in chapter 6, entitled: “The analysis of avian primary moult IV: flexible model fitting using the EM algorithm”.

The unifying theme of this thesis is the development of statistical models for data analysis problems which do not fit into any standard statistical methods. It is important to do this to prevent the erroneous practice of attempting to fit all data sets into the strait-jacket of existing analyses. There are many problems which fall under this category, the ones investigated in this thesis are just a few examples that illustrate how stochastic modelling can be used as a tool to solve them.

## CHAPTER 2

# Correlation under uncertainty of classification, with application to non-random mating

### ABSTRACT

A prerequisite for any study of non-random mating appears to be the ability to be able to classify individuals with certainty as male or female. In this study a model for assortative mating is developed when only an imperfect, or probabilistic, classification is available. The Redshank *Tringa totanus* is a species for which the sexes of the breeding pairs of observations could not be determined solely from biometric measurements. The Redshank data of Summers and Underhill (*Bird Study* (1991) 35: 169–180) are used to illustrate the model for assortative mating. The Dunlin *Calidris alpina* data of Jönsson (*Ornis Scandinavica* (1987) 18: 257–264) and the Tristan Bunting *Nesospiza acunhae* data of Ryan (*Phd Thesis* (1992)), for which the sexes are identifiable, are used to test the proposed model and to provide some guidance on the required sample sizes of both sexed birds to give information on the means and the standard deviations for males and females, and of the number of unsexed breeding pairs. Standard statistical hypothesis testing methods are developed to test the assumption of non-assortative mating. Two scenarios are considered: firstly, we regard the parameters of males and females as unknown and needing to be estimated, and secondly, we regard these parameters as known. The proposed model is demonstrated to be able to determine the presence or absence of assortative mating in bird species, even when there is uncertainty about the sexes of the individuals of a pair.

### INTRODUCTION

Non-random mating, the occurrence of mating combinations that cannot be predicted solely from the marginal distributions of one or more characters in males and females, is an important topic in evolutionary biology (reviewed by Findley 1987). In most studies of non-random mating, the individuals can be sexed with certainty, the criteria being external genitalia, or secondary sexual characteristics. At face value, the need to classify individuals as male or female with absolute certainty, appears to be a prerequisite for any study of

non-random mating. In this paper we show that this prerequisite can be relaxed, and that results can be obtained even when imperfect, or probabilistic, classification is all that can be attained.

For birds, which have internal gonads, sex classification is based on characteristics such as plumage, behaviour and size, or on the direct observation of copulation. For many species, these criteria are unreliable or even non-existent, and copulation tends to be observed so rarely that sample sizes are inadequate for statistical analyses. However, for a substantial number of bird species, it is known that males and females show significant differences on one or more mensural features (wing-, bill-length, etc), but that the sex of an individual bird cannot be determined from its measurements with certainty. For example, for waders (Charadrii), Engelmoer *et al.* (1987) evaluated measurements that are useful in the determination of sex. Using museum specimens (i.e. birds killed and dissected to determine sex accurately), they found, for example, that discriminant analyses of measurements could classify 96% of Knots *Calidris canutus* correctly, but the success rate for Turnstone *Arenaria interpres* was only 63%.

In the waders, there is considerable interest in non-random mating, particularly because many species in this group of birds display the unusual trait of reversed sexual size dimorphism, with females generally larger than males (Jehl and Murray 1986). Examples of studies of wader species that could be sexed with near certainty include Jehl (1970) and Jönsson (1987). Jönsson (1987), working on Dunlin *Calidris alpina* stated, "Sex was determined from plumage characteristics and measurements, in combination with subsequent behavioural observations," indicating that a range of criteria had to be used to sex the individuals in each pair. For many wader species, sexing criteria remain imperfect, and it is desirable to see what progress can be made in the analysis of non-random mating in the presence of this seemingly critical barrier.

Some progress in this area was made by Summers and Underhill (1991). In their data set, the wing-length measurements for each bird of pairs of Redshanks *Tringa totanus* were available, but the sexes could not be established. In this species it is known that females are, on average, larger than males, but that, even within individual pairs, the larger bird is not necessarily the female (Summers and Underhill 1991). Summers and Underhill (1991) approached the problem using a computer simulation method. In this paper, we develop a maximum likelihood approach to the same problem. We use their data set to illustrate the method, the Dunlin data of Jönsson (1987) supplemented by further observations and the

Tristan Bunting *Nesospiza acunhae* data of Ryan (1992). We consider two scenarios. In the first, we regard the parameters of males and females as unknown and to be estimated from the museum samples of birds of known sex, which may be small in size. In the second, we regard these parameters as known.

We also provide guidance on required sample sizes, both of sexed birds to provide information on the means and standard deviations for males and females, and of the number of unsexed breeding pairs.

## MATERIAL AND METHODS

Wing-, tarsus- and bill-length measurements were made for 204 breeding pairs of Dunlin in Scania in southern Sweden between 1981 and 1991 (see Jönsson 1987). In this study, the sexes were identifiable, therefore this data set can be used to test our methods. Wing-, tarsus-, head- and bill-length, as well as bill-depth and bill-width measurements of 60 breeding pairs of Tristan Buntings were measured on Inaccessible Island in the Tristan da Cunha group of Islands, South Atlantic Ocean (near 31°S, 15°W) over the period October to November 1988 and October 1989 to March 1990 (Ryan 1992). There are morphological differences in this species of Tristan Buntings depending on the region of the island they inhabit. Those found on the upper region (upland) have larger bodies, small bills and yellow colouration (27 breeding pairs), while those on the lower region (lowland) have small bodies, large bills and green/grey colouration (31 breeding pairs). Those birds found on the border of these regions have a mixture of these characteristics (2 breeding pairs). The sexes of these birds were also identifiable and therefore this data set is used to test our methods and to demonstrate problems encountered when the data set is relatively small. The presence or absence of assortative mating is evaluated within Tristan Bunting populations as well as between the populations. Wing-lengths of 27 breeding pairs of Redshanks were measured in Iceland between 6 and 17 June 1981, with the sexes of the birds unknown (Summers and Underhill 1991).

To evaluate the presence or absence of assortative mating, the distributions of wing-length of male and female Redshanks are needed, as are the distributions of wing-, tarsus- and bill-length of male and female Dunlins and the distributions of wing-, tarsus-, head- and bill-length, bill-depth and bill-width of male and female Tristan Buntings. We have taken these distributions to be Normal distributions with means and standard deviations

as set out in Table 1.

The museum-type data available for Dunlins is not appropriate as the measurements given in the handbooks are for a combination of sites and are not restricted to the area in which the field data were collected. It was thus decided to randomly select a portion of the field data and regard it as museum data. No data for Tristan Buntings are available apart from the data acquired by Ryan (1992) and so a portion was also randomly selected from the field data and assumed to be museum data. The means for the museum sample were taken to be those of Summers *et al.* (1988). The standard deviations and sample sizes were taken to be those reported by Cramp and Simmons (1983) for the reasons given by Summers and Underhill (1991).

### A model for assortative mating

The model described in this section is appropriate for any mensural feature (wing-, bill-length, etc). However, to simplify this section, this mensural feature will be referred to as wing-length measurements. Suppose that the wing-lengths  $(X, Y)$  of the two individuals in a pair have a bivariate normal (BN) distribution with parameters  $\mu_M, \mu_F, \sigma_M, \sigma_F,$  and  $\rho$ , that is,  $(X, Y) \sim \text{BN}(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho)$  (Johnson and Kotz 1972), where within the pair,  $x$  is the wing-length of the male, with mean  $\mu_M$  and standard deviation  $\sigma_M$ ,  $y$  is the wing-length of the female with mean  $\mu_F$  and standard deviation  $\sigma_F$ , and  $\rho$  is the parameter describing the correlation between male and female wing-lengths within pairs. Therefore, the hypothesis of non-assortative mating (i.e. that there is no preference for the size of a mate) is equivalent to testing the null hypothesis  $H_0 : \rho = 0$  against the alternative hypothesis  $H_1 : \rho \neq 0$ .

Let  $z_1, z_2, \dots, z_{n_M}$  and  $w_1, w_2, \dots, w_{n_F}$  be the measurements of wing-lengths of the museum sample of  $n_M$  sexed males and  $n_F$  sexed females, respectively. Then  $z_i \sim N(\mu_M, \sigma_M^2)$  and, independently,  $w_i \sim N(\mu_F, \sigma_F^2)$ .

Let  $(u_1, v_1), (u_2, v_2), \dots, (u_r, v_r)$  be the measurements of wing-lengths for the  $r$  unsexed pairs, where  $u_i \leq v_i, \quad i = 1, 2, \dots, r$ .

The density function  $f_{UV}(u, v)$  is shown in Appendix 1 to be

$$f_{UV}(u, v) = \begin{cases} f_{XY}(u, v) + f_{XY}(v, u) & \text{for } -\infty \leq u \leq v \leq \infty \\ 0 & \text{for } u > v \end{cases}$$

where both  $f_{XY}(u, v) \sim \text{BN}(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho)$  and  $f_{XY}(v, u) \sim \text{BN}(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho)$ .

**Table 1.** Museum data for the Dunlin, Tristan Bunting and Redshank.

	Sample size		Mean		Std Dev	
	Male	Female	Male	Female	Male	Female
<b>Dunlin<sup>1</sup></b>						
wing-length	10	10	112.8	116.4	2.49	2.50
tarsus-length	10	10	25.6	26.7	0.84	1.06
bill-length	10	10	27.9	31.7	1.07	1.12
<b>Bunting<sup>2</sup> (Lowlands)</b>						
wing-length	10	10	86.2	84.5	1.03	1.08
tarsus-length	10	10	25.1	24.5	0.52	0.76
head-length	10	10	35.7	35.1	0.78	0.70
bill-length	10	10	12.3	12.2	0.45	0.53
bill-depth	10	10	9.0	8.8	0.31	0.24
bill-width	10	10	7.8	7.7	0.23	0.24
<b>Bunting<sup>2</sup> (Uplands)</b>						
wing-length	10	10	87.7	85.4	1.25	1.43
tarsus-length	10	10	25.2	25.0	0.49	0.56
head-length	10	10	35.4	35.1	0.38	0.31
bill-length	10	10	12.0	11.9	0.26	0.21
bill-depth	10	10	8.8	8.6	0.18	0.23
bill-width	10	10	7.7	7.6	0.18	0.23
<b>Bunting<sup>2</sup> (All Tristan)</b>						
wing-length	10	10	86.7	84.2	1.38	1.48
tarsus-length	10	10	24.9	24.8	0.88	0.65
head-length	10	10	35.5	35.2	0.65	0.62
bill-length	10	10	12.3	12.1	0.43	0.37
bill-depth	10	10	8.9	8.9	0.27	0.25
bill-width	10	10	7.8	7.7	0.19	0.20
<b>Redshank<sup>3</sup></b>						
wing-length	13	16	170.9	174.0	3.9	3.1

1 PE Jönsson 1981–1991.

2 PG Ryan 1992. Unpublished PhD Thesis.

3 Summers and Underhill 1991.

### Estimation of $\rho$

The likelihood function (Cox and Hinkley 1974) under the alternative hypothesis of assortative mating (i.e.  $\rho \neq 0$ ) consists of two parts. The first component of the likelihood function describes the museum observations for which the sex of each bird is known but for which we have independent individuals (i.e. the observations are not of pairs of birds). The second component consists of the likelihood function that describes the field observations for which the sex of the birds is not known, but for which we have paired data. This likelihood function is given by

$$L(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho; z_i, w_i, (u_i, v_i)) = \prod_{i=1}^{n_M} f_Z(z_i) \prod_{i=1}^{n_F} f_W(w_i) \prod_{i=1}^r (f_{XY}(u_i, v_i) + f_{XY}(v_i, u_i)). \quad (1)$$

We do not, in fact, need to have individual museum observations  $z_i$  and  $w_i$ . It is sufficient to have the summary statistics  $\bar{z}_M$ ,  $s_M$ ,  $\bar{w}_F$  and  $s_F$ , the mean and standard deviation of wing-lengths for males and females respectively. This is because the density functions of  $z_i$  and  $w_i$  can be rewritten in terms of  $\bar{z}_M$ ,  $s_M$ ,  $\bar{w}_F$  and  $s_F$  (see Appendix 2). Let  $\ell_1 = \log L(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho; z_i, w_i, (u_i, v_i))$ , then the log-likelihood function under the alternative hypothesis is given by

$$\begin{aligned} \ell_1 = & -n_M \log(\sqrt{2\pi} \sigma_M) - n_F \log(\sqrt{2\pi} \sigma_F) - r \log(2\pi \sigma_M \sigma_F) \\ & - \frac{r}{2} \log(1 - \rho^2) - \frac{1}{2} \left( \frac{(n_M - 1)s_M^2 + n_M(\bar{z}_M - \mu_M)^2}{\sigma_M^2} \right) - \frac{1}{2} \left( \frac{(n_F - 1)s_F^2 + n_F(\bar{w}_F - \mu_F)^2}{\sigma_F^2} \right) \\ & + \sum_{i=1}^r \log \left\{ \exp \left( -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{u_i - \mu_M}{\sigma_M} \right)^2 - 2\rho \left( \frac{u_i - \mu_M}{\sigma_M} \right) \left( \frac{v_i - \mu_F}{\sigma_F} \right) + \left( \frac{v_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right. \\ & \left. + \exp \left( -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{v_i - \mu_M}{\sigma_M} \right)^2 - 2\rho \left( \frac{v_i - \mu_M}{\sigma_M} \right) \left( \frac{u_i - \mu_F}{\sigma_F} \right) + \left( \frac{u_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right\}. \quad (2) \end{aligned}$$

Maximum likelihood estimates are obtained by maximising  $\ell_1$  and this is achieved by setting its partial derivatives with respect to the parameters equal to zero (Appendix 3). The equations obtained by setting the first partial derivatives equal to zero cannot be solved explicitly and therefore the Newton-Raphson iterative algorithm (Ortega and Rheinboldt 1970) (Appendix 4) is used to solve them. For this, the second partial derivatives of  $\ell_1$  are required and these are also given in Appendix 3. The variances of the parameter estimates, and the covariances between them, are given by the elements of the inverse of the  $p \times p$  matrix with entries  $-\partial^2 \ell / \partial \theta_a \partial \theta_b$ , where  $p$  denotes the number of parameters estimated, and  $\ell$  represents the log-likelihood function.

### Estimation of parameters assuming that $\rho = 0$

Under the null hypothesis that  $\rho = 0$ , the likelihood function reduces to

$$L(\mu_M, \mu_F, \sigma_M, \sigma_F; z_i, w_i, (u_i, v_i)) = \prod_{i=1}^{n_M} f_Z(z_i) \prod_{i=1}^{n_F} f_W(w_i) \prod_{i=1}^r (f_X(u_i) f_Y(v_i) + f_X(v_i) f_Y(u_i)), \quad (3)$$

where  $f_X(\cdot) \sim N(\mu_M, \sigma_M^2)$  and  $f_Y(\cdot) \sim N(\mu_F, \sigma_F^2)$ .

Parameter estimates can be obtained by maximizing the log-likelihood function, i.e., the first partial derivatives with respect to the parameters are set to zero. The first partial derivatives with respect to the parameters are the same as those for  $\ell_1$ , but with  $\rho = 0$ . As before, the Newton-Raphson algorithm is used to maximize the log-likelihood function. The required second partial derivatives are given by setting  $\rho = 0$  in the second partial derivatives for  $\ell_1$  (Appendix 3).

### Testing the hypothesis of non-assortative mating

To determine if there is no preference in the choice of the size of a mate, we test the null hypothesis that

$$H_0 : \rho = 0$$

against the alternative hypothesis

$$H_1 : \rho \neq 0.$$

There are two approaches to testing this hypothesis. The likelihood ratio test (Wilks 1962) is defined as follows. Let

$$\lambda = \frac{L(\mu_M, \mu_F, \sigma_M, \sigma_F; z_i, w_i, (u_i, v_i))}{L(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho; z_i, w_i, (u_i, v_i))}, \quad (4)$$

where the numerator is the likelihood function given by equation 3, evaluated at the maximum likelihood estimators of the parameters, i.e. assuming the null hypothesis of non-assortative mating. The denominator is the likelihood function given by equation 1, i.e. assuming the alternative hypothesis of assortative mating.

The test rejects the null hypothesis for small values of  $\lambda$ , or equivalently for large values of  $-2 \log \lambda$ . Asymptotically,  $-2 \log \lambda$  has the chi-squared distribution with  $p - q$  degrees of freedom (Dobson 1983), where  $p$  represents the number of estimated parameters for the model under the alternative hypothesis, and  $q$  represents the number of parameters estimated under the null hypothesis. In this example  $p = 5$  and  $q = 4$ , so  $\lambda$  has the chi-square distribution with one degree of freedom.

A direct approach is also possible. For  $Z \sim N(\mu_M, \sigma_M^2)$  and  $W \sim N(\mu_F, \sigma_F^2)$ , then the distributions for the respective sample variances are given by

$$\frac{(n_M - 1)s_M^2}{\sigma_M^2} \sim \chi_{(n_M-1)}^2$$

and

$$\frac{(n_F - 1)s_F^2}{\sigma_F^2} \sim \chi_{(n_F-1)}^2.$$

Because  $s_F^2$  and  $s_M^2$  are independent

$$\frac{(n_M - 1)s_M^2}{\sigma_M^2} + \frac{(n_F - 1)s_F^2}{\sigma_F^2} \sim \chi_{(n_M+n_F-2)}^2.$$

Let  $U + V$  be the sum of the two measurements values from a single pair. Under the null hypothesis that  $\rho = 0$ , then  $U + V$  has  $N(\mu_M + \mu_F, \sigma_M^2 + \sigma_F^2)$  distribution. Therefore, we have

$$\frac{(r - 1)s_{U+V}^2}{\sigma_M^2 + \sigma_F^2} \sim \chi_{r-1}^2,$$

where  $s_{U+V}^2$  is the sample variance of  $U + V$ . If we now assume that  $\sigma_M^2 = \sigma_F^2 = \sigma^2$  then

$$F = \frac{(n_M + n_F - 2)s_{U+V}^2}{2((n_M - 1)s_M^2 + (n_F - 1)s_F^2)} \sim F_{r-1, n_M+n_F-2}. \quad (5)$$

This test can be used against any of the three alternative hypothesis:

- a)  $H_1 : \rho > 0$  (i.e. reject if  $F$  is too large),
- b)  $H_2 : \rho < 0$  (i.e. reject if  $F$  is too small),
- c)  $H_3 : \rho \neq 0$  (i.e. reject if  $F$  is too large or too small).

### Estimation of $\rho$ assuming that $\mu_M, \mu_F, \sigma_M$ and $\sigma_F$ are known

If the means and standard deviations  $\mu_M, \mu_F, \sigma_M$  and  $\sigma_F$  are known, or if sample sizes of the available museum data set are so large that the means and standard deviations may be assumed to be known, then only the parameter  $\rho$  needs to be estimated. Denote the log-likelihood function under the alternative hypothesis by  $\ell_2 = \log L(\rho; z_i, w_i, (u_i, v_i))$ , which is now given by

$$\begin{aligned} \ell_2 = & -n_M \log(\sqrt{2\pi} \sigma_M) - n_F \log(\sqrt{2\pi} \sigma_F) - \frac{n_M + n_F - 2}{2} - r \log(2\pi \sigma_M \sigma_F) - \frac{r}{2} \log(1 - \rho^2) \\ & + \sum_{i=1}^r \log \left\{ \exp \left( -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{u_i - \mu_M}{\sigma_M} \right)^2 - \frac{2\rho(u_i - \mu_M)(v_i - \mu_F)}{\sigma_M \sigma_F} + \left( \frac{v_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right. \\ & \left. + \exp \left( -\frac{1}{2(1 - \rho^2)} \left[ \left( \frac{v_i - \mu_M}{\sigma_M} \right)^2 - \frac{2\rho(v_i - \mu_M)(u_i - \mu_F)}{\sigma_M \sigma_F} + \left( \frac{u_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right\}. \quad (6) \end{aligned}$$

The first and second partial derivatives with respect to  $\rho$  required for the Newton-Raphson algorithm are given in Appendix 5.

Denote the log-likelihood function under the null hypothesis that  $H_0 : \rho = 0$  by  $\ell_3 = L(z_i, w_i, (u_i, v_i))$ . Then the log-likelihood function simplifies to

$$\begin{aligned} \ell_3 = & -n_M \log(\sqrt{2\pi} \sigma_M) - n_F \log(\sqrt{2\pi} \sigma_F) - \frac{n_M + n_F - 2}{2} - r \log(2\pi\sigma_M\sigma_F) \\ & + \sum_{i=1}^r \log \left\{ \exp \left( -\frac{1}{2} \left[ \left( \frac{u_i - \mu_M}{\sigma_M} \right)^2 + \left( \frac{v_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right. \\ & \left. + \exp \left( -\frac{1}{2} \left[ \left( \frac{v_i - \mu_M}{\sigma_M} \right)^2 + \left( \frac{u_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right\}, \end{aligned}$$

and the likelihood ratio test for the null hypothesis of no assortative mating is

$$\lambda = \frac{L(z_i, w_i, (u_i, v_i))}{L(\rho; z_i, w_i, (u_i, v_i))}, \quad (7)$$

and, as before,  $-2 \log \lambda$  has the  $\chi^2$  distribution with one degree of freedom.

Assuming  $\mu_M$ ,  $\mu_F$ ,  $\sigma_M$  and  $\sigma_F$  are known, then the direct approach reduces to

$$\chi^2 = \sum_{i=1}^r \frac{(u_i + v_i - \mu_M - \mu_F)^2}{\sigma_M^2 + \sigma_F^2} \sim \chi_r^2, \quad (8)$$

where  $r$  is the number of pairs of birds measured.

## RESULTS AND DISCUSSION

We will first discuss the results for the two sets of data for which the sex of the birds are known, that is, the Dunlins and the Tristan Buntings. Lastly, we will discuss the results for the Redshanks for which the sex of the birds is not known.

The parameter estimates for the three different Dunlin measurements, together with the standard errors of the estimates of the parameters for the model assuming that there is assortative mating are shown in Table 2. The parameter estimates with the standard errors under the assumption of no assortative mating are given in Table 3. The estimates for the mean and standard deviation of wing-, tarsus- and bill-length relate closely to the mean and standard deviation obtained from the museum specimens, whether under the assumption of assortative mating or no assortative mating (Tables 1, 2 and 3). From the likelihood ratio results it can be seen that for bill- and wing-length there is no evidence of assortative mating being present, although there is a strong evidence of assortative mating on tarsus-length ( $-2 \log \lambda = 11.50$ ,  $P = 0.0007$ ; Table 4). Looking at the  $F$ -statistic (to test

the null hypothesis directly under the assumption of equal variances between the sexes), we obtain no significant evidence of the presence of assortative mating for bill- and wing-length as well as for tarsus-length ( $F = 1.29, P = 0.54$ ). The differences in conclusions between the likelihood ratio statistic and the direct statistic can be attributed to the fact that the power of the exact test is poor.

Because the sexes of the Dunlins are known, the true correlation coefficient between the sexes for the various measurements can be compared with those obtained by the proposed model. The parameter estimates of the model closely resemble those obtained directly from the data suggesting that the model behaves well (Tables 2, 5 and 6). Testing for significant correlation, we get the same results as those obtained by the model (Table 5), namely that there is no significant evidence of assortative mating in bill- and in wing-length but that there is in tarsus-length.

If one assumes that the means and the standard deviations of the distribution are known then the same conclusions as the previous model are obtained (Table 7). The  $\chi^2$ -statistic gives a direct approach to test the null hypothesis of no assortative mating under the assumption that the means and standard deviations of each sex is known. In this case, the exact test results agrees with those from the likelihood ratio statistic.

In the case of the Tristan Buntings in the upland region, parameter estimates for bill-width were not obtained (Table 2) because the Newton-Raphson iterative algorithm to obtain maximum likelihood estimates did not converge in this instance. One of the disadvantages of the Newton-Raphson algorithm is that to obtain convergence, one needs to give initial parameter estimates that are close to the final estimates. When the number of observations available are few, this problem is intensified. In the case of the Tristan Buntings of the upland region, we only have 17 breeding pairs once ten have been randomly removed to be considered as museum specimens. In the Tristan Buntings, the males are generally larger than the females as seen in the museum specimens (Table 1), while for the other species, the females are larger than the males in the mensural features. As for the Dunlins, the estimates of the means and standard deviations under the assumption of assortative mating resembles those obtained under the assumption of non-assortative mating (Tables 2 and 3).

From Table 4 we see that the likelihood ratio test gives strong evidence that there is assortative mating on wing-length for Tristan Buntings in the lowland region. From the  $F$ -statistic we obtain no significant evidence of assortative mating for any of the mensural

**Table 2.** Parameter estimates (standard errors) under the assumption of assortative mating ( $\rho \neq 0$ ) for the Dunlin, Tristan Bunting and Redshank.

	Parameter estimates				
	$\hat{\mu}_M$	$\hat{\mu}_F$	$\hat{\sigma}_M$	$\hat{\sigma}_F$	$\hat{\rho}$
<b>Dunlin</b>					
wing-length	113.5 (0.21)	116.9 (0.21)	2.36 (0.15)	2.46 (0.16)	-0.002 (0.10)
tarsus-length	25.4 (0.07)	26.5 (0.07)	0.86 (0.05)	0.96 (0.05)	0.41 (0.08)
bill-length	27.7 (0.07)	31.8 (0.09)	1.05 (0.05)	1.30 (0.06)	-0.04 (0.07)
<b>Bunting (Lowlands)</b>					
wing-length	85.2 (0.38)	84.0 (0.36)	2.00 (0.16)	1.62 (0.09)	-0.84 (0.03)
tarsus-length	25.0 (0.09)	24.4 (0.14)	0.42 (0.06)	0.77 (0.11)	-0.04 (0.29)
head-length	35.7 (0.15)	35.1 (0.13)	0.69 (0.10)	0.59 (0.09)	-0.12 (0.28)
bill-length	12.3 (0.09)	12.2 (0.09)	0.39 (0.06)	0.43 (0.06)	-0.24 (0.24)
bill-depth	9.0 (0.07)	8.7 (0.07)	0.28 (0.05)	0.29 (0.04)	-0.15 (0.25)
bill-width	7.8 (0.05)	7.7 (0.06)	0.22 (0.04)	0.24 (0.04)	-0.40 (0.17)
<b>Bunting (Uplands)</b>					
wing-length	88.1 (0.23)	85.0 (0.22)	1.15 (0.17)	1.07 (0.15)	0.13 (0.45)
tarsus-length	25.3 (0.13)	24.9 (0.12)	0.58 (0.08)	0.47 (0.09)	-0.34 (0.22)
head-length	35.4 (0.10)	35.0 (0.10)	0.38 (0.06)	0.41 (0.07)	0.12 (0.30)
bill-length	12.0 (0.07)	11.8 (0.07)	0.25 (0.05)	0.29 (0.04)	0.07 (0.29)
bill-depth	8.8 (0.04)	8.6 (0.04)	0.16 (0.02)	0.20 (0.03)	-0.22 (0.27)
bill-width	-	-	-	-	-
<b>Bunting (All Tristan)</b>					
wing-length	87.2 (0.18)	84.1 (0.21)	1.38 (0.14)	1.58 (0.16)	0.23 (0.16)
tarsus-length	24.8 (0.12)	25.0 (0.07)	0.83 (0.08)	0.46 (0.06)	-0.40 (0.10)
head-length	35.4 (0.11)	35.1 (0.10)	0.63 (0.08)	0.52 (0.07)	-0.24 (0.14)
bill-length	12.2 (0.09)	12.0 (0.08)	0.39 (0.05)	0.36 (0.04)	-0.06 (0.17)
bill-depth	8.8 (0.06)	8.8 (0.05)	0.30 (0.03)	0.24 (0.03)	-0.25 (0.13)
bill-width	7.7 (0.04)	7.6 (0.04)	0.23 (0.02)	0.19 (0.03)	-0.06 (0.15)
<b>Redshank</b>					
wing-length	170.8 (0.71)	173.9 (0.57)	3.89 (0.50)	3.08 (0.38)	0.18 (0.23)

**Table 3.** Parameter estimates (standard errors) under the assumption that  $\rho = 0$ , for the Dunlin, Tristan Bunting and Redshank.

	Parameter estimates			
	$\hat{\mu}_M$	$\hat{\mu}_F$	$\hat{\sigma}_M$	$\hat{\sigma}_F$
<b>Dunlin</b>				
wing-length	113.5 (0.19)	116.9 (0.19)	2.36 (0.14)	2.46 (0.14)
tarsus-length	25.6 (0.13)	26.2 (0.14)	0.92 (0.09)	1.13 (0.10)
bill-length	27.7 (0.07)	31.8 (0.09)	1.05 (0.05)	1.30 (0.06)
<b>Bunting (Lowlands)</b>				
wing-length	84.0 (0.44)	85.6 (0.35)	2.11 (0.28)	1.45 (0.23)
tarsus-length	25.0 (0.08)	24.4 (0.14)	0.42 (0.06)	0.76 (0.10)
head-length	35.7 (0.14)	35.1 (0.12)	0.68 (0.10)	0.58 (0.09)
bill-length	12.3 (0.10)	12.2 (0.10)	0.38 (0.06)	0.43 (0.06)
bill-depth	9.0 (0.05)	8.7 (0.06)	0.27 (0.04)	0.28 (0.04)
bill-width	7.8 (0.05)	7.6 (0.05)	0.21 (0.03)	0.23 (0.03)
<b>Bunting (Uplands)</b>				
wing-length	88.1 (0.22)	85.0 (0.21)	1.15 (0.16)	1.07 (0.15)
tarsus-length	25.3 (0.12)	24.8 (0.10)	0.56 (0.08)	0.44 (0.07)
head-length	35.4 (0.09)	35.0 (0.09)	0.39 (0.07)	0.42 (0.07)
bill-length	12.0 (0.06)	11.8 (0.07)	0.25 (0.05)	0.29 (0.04)
bill-depth	8.8 (0.03)	8.6 (0.04)	0.16 (0.02)	0.20 (0.03)
bill-width	7.7 (0.03)	7.6 (0.04)	0.16 (0.03)	0.20 (0.03)
<b>Bunting (All Tristan)</b>				
wing-length	87.1 (0.19)	84.1 (0.21)	1.40 (0.14)	1.62 (0.16)
tarsus-length	24.8 (0.12)	25.0 (0.09)	0.80 (0.08)	0.49 (0.06)
head-length	35.5 (0.10)	35.1 (0.09)	0.59 (0.07)	0.52 (0.06)
bill-length	12.2 (0.08)	12.0 (0.07)	0.39 (0.05)	0.36 (0.04)
bill-depth	8.7 (0.07)	8.8 (0.08)	0.25 (0.03)	0.28 (0.04)
bill-width	7.8 (0.04)	7.6 (0.03)	0.23 (0.02)	0.19 (0.02)
<b>Redshank</b>				
wing-length	171.0 (0.69)	173.7 (0.55)	3.97 (0.49)	3.12 (0.40)

**Table 4.** Likelihood ratio test and F-test derived from the direct approach for assortative mating in the Dunlin, Tristan Bunting and Redshank (means and standard deviations unknown).

	$-2 \log \lambda$ ( <i>P</i> -value)	<i>F</i> -statistic ( <i>P</i> -value)
<b>Dunlin</b>		
wing-length	0.0004(0.98)	1.07 (0.94)
tarsus-length	11.5 (0.0007)	1.29 (0.54)
bill-length	0.31 (0.58)	1.14 (0.78)
<b>Bunting (Lowlands)</b>		
wing-length	31.89 (0.00)	0.61 (0.28)
tarsus-length	0.02 (0.89)	0.89 (0.80)
head-length	0.15 (0.70)	0.64 (0.33)
bill-length	0.81 (0.37)	0.49 (0.12)
bill-depth	0.25 (0.62)	0.96 (0.92)
bill-width	2.86 (0.09)	0.61 (0.29)
<b>Bunting (Uplands)</b>		
wing-length	0.09 (0.76)	0.57 (0.27)
tarsus-length	1.28 (0.26)	0.77 (0.61)
head-length	0.20 (0.65)	1.84 (0.22)
bill-length	0.07 (0.79)	1.74 (0.26)
bill-depth	0.46 (0.50)	0.66 (0.37)
bill-width	—	1.05 (0.92)
<b>Bunting (All Tristan)</b>		
wing-length	2.31 (0.13)	1.39 (0.45)
tarsus-length	6.44 (0.01)	0.51 (0.06)
head-length	1.96 (0.16)	0.64 (0.21)
bill-length	0.13 (0.72)	0.83 (0.59)
bill-depth	3.45 (0.06)	0.83 (0.58)
bill-width	0.15 (0.70)	1.15 (0.77)
<b>Redshank</b>		
wing-length	0.62 (0.43)	1.29 (0.52)

**Table 5.** Parameters obtained directly from the data for Dunlin and Tristan Bunting. The  $P$ -value is for the test of how significantly the correlation coefficient differs from zero.

	Parameter estimates					$P$ -value
	$\mu_M$	$\mu_F$	$\sigma_M$	$\sigma_F$	$\rho$	
<b>Dunlin</b>						
wing-length	113.5	116.9	2.38	2.45	-0.002	> 0.8
tarsus-length	25.4	26.5	0.85	0.98	0.384	< 0.001
bill-length	27.7	31.8	1.07	1.32	-0.054	> 0.6
<b>Bunting (Lowlands)</b>						
wing-length	86.4	82.6	0.93	1.24	-0.46	0.04
tarsus-length	24.9	24.4	0.54	0.80	-0.20	0.38
head-length	35.7	35.0	0.62	0.56	0.01	0.97
bill-length	12.4	12.0	0.33	0.37	-0.04	0.87
bill-depth	9.0	8.7	0.24	0.31	-0.04	0.87
bill-width	7.8	7.6	0.18	0.25	-0.31	0.17
<b>Bunting (Uplands)</b>						
wing-length	88.4	84.8	1.06	0.81	0.16	0.53
tarsus-length	25.4	24.8	0.65	0.39	-0.29	0.26
head-length	35.4	35.0	0.42	0.49	0.06	0.83
bill-length	12.0	11.8	0.30	0.33	-0.01	0.97
bill-depth	8.8	8.6	0.17	0.19	-0.15	0.58
bill-width	7.7	7.6	0.17	0.17	0.55	0.02
<b>Bunting (All Tristan)</b>						
wing-length	87.3	84.0	1.37	1.60	0.29	0.04
tarsus-length	25.2	24.6	0.49	0.70	-0.18	0.21
head-length	35.5	35.1	0.59	0.50	-0.14	0.32
bill-length	12.1	12.0	0.36	0.39	-0.11	0.45
bill-depth	8.9	8.6	0.24	0.23	-0.01	0.96
bill-width	7.8	7.6	0.20	0.22	-0.04	0.80

**Table 6.** *P*-values for testing how significantly the true correlation coefficient differs from the estimated correlation coefficient obtained by the model for assortative mating. Both the model for assortative mating under the assumption that all model parameters are unknown and the model assuming that means and variances are known are considered for the Dunlin and Tristan Bunting.

	Assuming all parameters unknown	Assuming means and variances known
	<i>Z</i> -statistic ( <i>P</i> -value)	<i>Z</i> -statistic ( <i>P</i> -value)
<b>Dunlin</b>		
wing-length	-0.06 (0.96)	-1.14 (0.26)
tarsus-length	-0.39 (0.69)	0.54 (0.59)
bill-length	-0.19 (0.85)	0.08 (0.94)
<b>Bunting (Lowlands)</b>		
wing-length	3.07 (0.002)	-1.12 (0.26)
tarsus-length	-69 (0.49)	-0.90 (0.37)
head-length	0.55 (0.58)	1.17 (0.24)
bill-length	0.87 (0.39)	1.43 (0.15)
bill-depth	0.47 (0.64)	0.47 (0.64)
bill-width	0.44 (0.66)	0.49 (0.63)
<b>Bunting (Uplands)</b>		
wing-length	0.11 (0.91)	2.19 (0.03)
tarsus-length	0.21 (0.84)	0.29 (0.77)
head-length	-0.23 (0.82)	-0.08 (0.94)
bill-length	-0.30 (0.76)	0.00 (1.00)
bill-depth	0.27 (0.79)	0.63 (0.53)
bill-width	—	-1.24 (0.21)
<b>Bunting (All Tristan)</b>		
wing-length	0.44 (0.66)	1.15 (0.25)
tarsus-length	1.66 (0.10)	1.91 (0.06)
head-length	0.71 (0.48)	0.93 (0.35)
bill-length	-0.35 (0.73)	-1.31 (0.19)
bill-depth	1.68 (0.09)	-0.27 (0.78)
bill-width	0.14 (0.89)	-0.48 (0.63)

**Table 7.** Correlation coefficient estimates (standard errors) when means and variances are assumed known, together with the likelihood ratio test and the  $\chi^2$  test derived from the direct approach for assortative mating in the Dunlin, Tristan Bunting and Redshank.

	Parameter estimate		
	$\hat{\rho}$	$-2 \log \lambda$ ( $P$ -value)	$\chi^2$ -statistic ( $P$ -value)
<b>Dunlin</b>			
wing-length	0.08 (0.09)	0.77 (0.38)	203.2 (0.31)
tarsus-length	0.35 (0.08)	17.14 (0.00004)	264.9 (0.001)
bill-length	-0.06 (0.06)	0.97 (0.32)	219.5 (0.10)
<b>Bunting (Lowlands)</b>			
wing-length	-0.23 (0.10)	4.30 (0.04)	38.5 (0.01)
tarsus-length	0.01 (0.28)	0.001 (0.97)	20.3 (0.50)
head-length	-0.26 (0.28)	0.58 (0.45)	13.0 (0.91)
bill-length	-0.36 (0.26)	1.01 (0.31)	10.4 (0.97)
bill-depth	-0.15 (0.18)	0.64 (0.42)	21.8 (0.37)
bill-width	-0.41 (0.17)	3.63 (0.06)	12.6 (0.92)
<b>Bunting (Uplands)</b>			
wing-length	-0.40 (0.21)	1.90 (0.17)	9.47 (0.92)
tarsus-length	-0.36 (0.17)	3.30 (0.07)	12.6 (0.76)
head-length	0.08 (0.17)	0.20 (0.65)	31.4 (0.02)
bill-length	-0.01 (0.15)	0.003 (0.96)	30.8 (0.02)
bill-depth	-0.31 (0.27)	0.85 (0.36)	10.4 (0.86)
bill-width	0.74 (0.13)	6.39 (0.01)	17.4 (0.38)
<b>Bunting (All Tristan)</b>			
wing-length	0.13 (0.13)	1.08 (0.30)	58.2 (0.20)
tarsus-length	-0.43 (0.13)	5.96 (0.01)	25.4 (0.999)
head-length	-0.27 (0.15)	2.53 (0.11)	32.4 (0.97)
bill-length	0.08 (0.15)	0.27 (0.60)	54.2 (0.32)
bill-depth	0.03 (0.11)	0.06 (0.81)	72.2 (0.02)
bill-width	0.03 (0.11)	0.08 (0.78)	68.3 (0.04)
<b>Redshank</b>			
wing-length	0.19 (0.21)	0.78 (0.38)	32.8 (0.20)

features. Comparing the correlation coefficient between the sexes for the various measurements obtained from the proposed model and the true correlation coefficients (Tables 2 and 5), we see that they differ from each other, but only for wing-length is the difference significant (Table 6). Testing for significant correlations supports the significant evidence of assortative mating on wing-length for Tristan Buntings from the lowland region (Table 5). Assuming that the means and the standard deviations of the distribution are known then the same conclusions as the previous model are obtained (Table 7). The exact  $\chi^2$ -test results only agree with those from the likelihood ratio statistic for wing-length.

For Tristan Buntings in the upland region there is no evidence of assortative mating for the mensural features for which the model could be fitted (Table 4). The  $F$ -statistic again shows no evidence of assortative mating. The true correlation coefficients do not differ significantly from those obtained by the model (Tables 2, 5 and 6). Testing for significance in the correlation coefficients shows evidence of assortative mating on bill-width. Unfortunately this was a variable for which convergence of the Newton-Raphson algorithm could not be achieved and so no parameter estimates were obtained. Under the assumption of known means and standard deviations of the distribution, the likelihood ratio test agrees with the results obtained directly from the data that there is significant evidence of assortative mating on bill-width (Tables 5 and 7), but the model also shows evidence of assortative mating on tarsus-length at the 7% level of significance. From the exact  $\chi^2$ -statistic we obtain no significant evidence of assortative mating on bill-width, but there is significant evidence of assortative mating on head- and bill-length. Comparing the true correlation coefficient with the estimated correlation coefficient we see that they differ significantly for wing-length ( $P = 0.03$ ; Table 6).

When differences in the Tristan Buntings from the two regions of Inaccessible Island are ignored and the two populations combined, including hybrid birds, we see that the likelihood ratio test shows significant evidence of assortative mating on tarsus-length ( $P = 0.01$ ; Table 4). This is supported by the  $F$ -statistic at the 6% level of significance. The likelihood ratio test also shows significant assortative mating on bill-depth with  $P = 0.06$ . Comparing the correlation coefficient between the sexes for the various measurements obtained from the model and those obtained directly from the data we see that they generally differ from each other (Tables 2 and 5), but not significantly (Table 6). Also, only the correlation coefficient for wing-length is significantly different from zero ( $P = 0.04$ ; Table 5). Looking at the results under the assumption of known means and standard deviations of the distribution,

the likelihood ratio test agrees with the results obtained from the previous model of unknown means and standard deviations, except that now, there is no evidence of assortative mating on bill-depth (Tables 4 and 7). The exact test shows significant evidence of assortative mating on bill-depth and bill-width. This result is not supported by either the likelihood ratio test or the results obtained directly from the data. The true correlation coefficient differs significantly from the estimated correlation coefficient at the 6% level of significance (Tables 5, 6 and 7).

The results obtained from the data set on the Tristan Buntings demonstrate the difficulties encountered when the sample size of the "field" data is relatively small. It is not only difficult, but sometimes even impossible, to obtain parameter estimates for the model as seen with the variables wing-length and bill-width for the Tristan Buntings in the upland region, and the estimate for the correlation coefficient between the sexes for the various measurements can vary quite considerably from the actual correlation coefficient obtained directly from the data, leading sometimes to opposite conclusions on the significance of the correlation coefficients. Although these results are disappointing, they do not come as a surprise as the work done on the Dunlin data set demonstrated that sample sizes of at least 100 are needed to have the estimate of  $\rho$  stabilize (see Appendix 6).

Tables 2 and 3 show the parameter estimates for Redshanks under the assumption of assortative mating and non-assortative mating respectively. The estimates for the means and standard deviations are similar. The likelihood ratio statistic for Redshanks shows that there is no significant evidence of assortative mating ( $-2 \log \lambda = 0.62, P = 0.43$ ; Table 4). Assuming that the variances are equal, the  $F$ -statistic tests the hypothesis of assortative mating. The conclusion is that there is no significant evidence of assortative mating ( $F = 1.29, P = 0.52$ ; Table 4). Summers and Underhill (1991) using a computer simulation method obtained a median correlation coefficient between male and female wing-length of 0.16, which compares closely to the correlation coefficient estimated by our model of 0.18. They also concluded that there was no significant evidence of assortative mating on wing-length ( $P = 0.26$ ).

Table 7 gives the parameter estimate for the correlation coefficient between the wing-lengths of males and females when we assume that the means and standard deviations are known. From the log-likelihood ratio it can be seen that there is no significant evidence of assortative mating. The  $\chi^2$ -statistic gives a direct approach to test the null hypothesis. The same conclusion is reached in this case.

## CONCLUSION

The model proposed here to determine the presence or absence of assortative mating in bird species, in the presence of uncertainty about the sexes of the individuals of a pair, has been shown to behave satisfactorily. That is, using a data set in which the sexes of the birds are known, but by pretending that this is unknown we have verified that the results obtained from the model closely resemble those obtained by making use of the knowledge of the sexes. We have furthermore shown that the quality of the “museum” data can be quite poor and the model still behaves reasonably. This fact is quite important and practical as we have established that to be able to determine the correlation between sexes one does not need large numbers of dissected birds to determine the sexes of birds. Priority should be given to increasing the sample size of the field observations of pairs of birds. With field sample sizes of approximately 100, consistent estimates of the correlation coefficient between the sexes are attained even with “museum” samples as small as 10 observations.

The test statistic derived from the direct approach (equations 5 and 8), though simple to compute, cannot be recommended because it is a far less powerful test than the likelihood ratio test (equations 4 and 7). In other words, the probability of rejecting the null hypothesis for a given  $\rho$  is smaller with the test statistic derived from the direct approach than the likelihood ratio test.

## REFERENCES

- CRAMP, S. AND SIMMONS, K.E.L. (Eds). 1983. The birds of the Western Palearctic, Vol. 3. Oxford University Press, Oxford.
- COX, D.R. AND HINKLEY, D.V. 1974. Theoretical statistics. Chapman Hall, London.
- DOBSON, A.J. 1983. An introduction to statistical modelling. Chapman and Hall, London.
- ENGELMOER, M., ROSELAAR, C.S., NIEBOER, E. AND BOERE, G.C. 1987. Biometrics in waders. Wader Study Group Bulletin 51: 44-47.
- FINDLAY, C.S. 1987. Non-random mating: a theoretical and empirical overview with special reference to birds. *In* Cooke, F. and Buckley, P.A. (Eds). Avian Genetics: 289-319. Academic Press, London.
- JEHL, J.R. 1970. Sexual selection for size differences in two species of sandpipers. *Evolution* 24: 311-319.
- JEHL, J.R. AND MURRAY, B.G. 1986. The evolution of normal and reverse sexual size dimorphism in shorebirds and other birds. *In* Johnston, R.F. (Ed). Current Ornithology 3: 1-86. Plenum Press, New York.
- JOHNSON, N.L. AND KOTZ, S. 1972. Distributions in statistics: continuous multivariate distributions. Wiley, New York.
- JÖNSSON, P.E. 1987. Sexual size dimorphism and disassortative mating in the Dunlin *Calidris alpina schinzii* in southern Sweden. *Ornis Scandinavica* 18: 257-264.
- MOOD, A.M., GRAYBILL, F.A. AND BOES, D.C. 1963. Introduction to the theory of statistics. McGraw-Hill, New York.
- ORTEGA, J.M. AND RHEINBOLDT, W.C. 1970. Iterative solution of nonlinear equations in several variables. Academic Press, New York.
- RYAN, P.G. 1992. The ecology and evolution of *Nesospiza* Buntings. Unpublished PhD thesis, University of Cape Town.
- SUMMERS, R.W., NICOLL, M., UNDERHILL, L.G. AND PETERSEN, A. 1988. Methods for estimating the proportions of Icelandic and British Redshanks *Tringa totanus* in mixed populations wintering on British coasts. *Bird Study* 35: 169-180.

SUMMERS, R.W. AND UNDERHILL, L.G. 1991. The relationship between body size and time of breeding in Icelandic Redshanks *Tringa t. robusta*. *Ibis* 133: 134–139.

WILKS, S.S. 1962. *Mathematical statistics*. Wiley, New York.

## APPENDIX 1

The density function  $f_{UV}(u, v)$ .

We are given wing-length measurements for unsexed pairs  $(u_1, v_1), (u_2, v_2), \dots, (u_r, v_r)$ , where  $u_i \leq v_i, i = 1, 2, \dots, r$ . To obtain the density function  $f_{UV}(u, v)$  (the density function of the unsexed paired wing-length measurements) we note that the ordered pair of observations  $(U, V)$  is a transformation of the pair of observations  $(X, Y)$  for which the sex is known. That is, we are making the transformation

$$\begin{aligned} u &= x & \text{if } x \leq y \\ v &= y \end{aligned}$$

or

$$\begin{aligned} u &= y & \text{if } x > y. \\ v &= x \end{aligned}$$

Denote the set of possible values for the pair  $(x, y)$  by  $\mathcal{X}$ . That is,

$$\mathcal{X} = \{(x, y) : -\infty \leq x \leq \infty, -\infty \leq y \leq \infty\}.$$

This set can be decomposed into two sets  $\mathcal{X}_1$  and  $\mathcal{X}_2$  such that

$$\mathcal{X}_1 = \{(x, y) : -\infty \leq x \leq y \leq \infty\}.$$

and

$$\mathcal{X}_2 = \{(x, y) : -\infty \leq y < x \leq \infty\}.$$

Now the transformation  $u = x$  and  $v = y$  is a one-to-one transformation from  $\mathcal{X}_1$  onto set  $\mathcal{N}$ , where

$$\mathcal{N} = \{(u, v) : -\infty \leq u \leq v \leq \infty\}.$$

The Jacobian of this transformation is given by

$$J_1 = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1.$$

Similarly, the transformation  $u = y$  and  $v = x$  is a one-to-one transformation from  $\mathcal{X}_2$  onto  $\mathcal{N}$ , and its Jacobian is now given by

$$J_2 = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = 1.$$

The density function  $f_{UV}(u, v)$  (Mood *et al.* 1963) is then given by

$$f_{UV}(u, v) = \begin{cases} f_{XY}(u, v) + f_{XY}(v, u) & \text{for } -\infty \leq u \leq v \leq \infty \\ 0 & \text{for } u > v \end{cases},$$

where  $f_{XY}(x, y) \sim \text{BN}(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho)$ . That is, the joint density function is given by

$$f_{UV}(u_i, v_i) = \frac{1}{2\pi\sigma_M\sigma_F\sqrt{1-\rho^2}} \left\{ \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{u_i - \mu_M}{\sigma_M}\right)^2 - 2\rho \left(\frac{u_i - \mu_M}{\sigma_M}\right) \left(\frac{v_i - \mu_F}{\sigma_F}\right) + \left(\frac{v_i - \mu_F}{\sigma_F}\right)^2 \right]\right) + \exp\left(-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{v_i - \mu_M}{\sigma_M}\right)^2 - 2\rho \left(\frac{v_i - \mu_M}{\sigma_M}\right) \left(\frac{u_i - \mu_F}{\sigma_F}\right) + \left(\frac{u_i - \mu_F}{\sigma_F}\right)^2 \right]\right) \right\}.$$

## APPENDIX 2

Density functions of  $z_i$  and  $w_i$  rewritten in terms of  $\bar{z}_M$ ,  $s_M$ ,  $\bar{w}_F$  and  $s_F$

The likelihood function under the alternative hypothesis of assortative mating ( $\rho \neq 0$ ) is given by

$$\begin{aligned}
L(\mu_M, \mu_F, \sigma_M, \sigma_F, \rho; z_i, w_i, (u_i, v_i)) &= \prod_{i=1}^{n_M} f_Z(z_i) \prod_{i=1}^{n_F} f_W(w_i) \prod_{i=1}^r (f_{XY}(u_i, v_i) + f_{XY}(v_i, u_i)) \\
&= \left( \frac{1}{\sqrt{2\pi}\sigma_M} \right)^{n_M} \left( \frac{1}{\sqrt{2\pi}\sigma_M} \right)^{n_F} \left( \frac{1}{2\pi\sigma_M\sigma_F\sqrt{1-\rho^2}} \right)^r \\
&\exp \left\{ -\frac{1}{2} \sum_{i=1}^{n_M} \left( \frac{z_i - \mu_M}{\sigma_M} \right)^2 \right\} \exp \left\{ -\frac{1}{2} \sum_{i=1}^{n_F} \left( \frac{w_i - \mu_F}{\sigma_F} \right)^2 \right\} \\
&\prod_{i=1}^r \left\{ \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{u_i - \mu_M}{\sigma_M} \right)^2 - 2\rho \left( \frac{u_i - \mu_M}{\sigma_M} \right) \left( \frac{v_i - \mu_F}{\sigma_F} \right) + \left( \frac{v_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right. \\
&\left. + \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{v_i - \mu_M}{\sigma_M} \right)^2 - 2\rho \left( \frac{v_i - \mu_M}{\sigma_M} \right) \left( \frac{u_i - \mu_F}{\sigma_F} \right) + \left( \frac{u_i - \mu_F}{\sigma_F} \right)^2 \right] \right) \right\}.
\end{aligned}$$

With some algebraic manipulation, the expressions  $\sum_{i=1}^{n_M} \left( \frac{z_i - \mu_M}{\sigma_M} \right)^2$  and  $\sum_{i=1}^{n_F} \left( \frac{w_i - \mu_F}{\sigma_F} \right)^2$  in the above equation can be rewritten in terms of  $\bar{z}_M$ ,  $s_M$ ,  $\bar{w}_F$  and  $s_F$ . This gives

$$\sum_{i=1}^{n_M} \left( \frac{z_i - \mu_M}{\sigma_M} \right)^2 = \frac{1}{\sigma_M^2} \left( (n_M - 1)s_M^2 + n_M(\bar{z} - \mu_M)^2 \right),$$

with an analogous term for  $\sum_{i=1}^{n_F} \left( \frac{w_i - \mu_F}{\sigma_F} \right)^2$ . The log-likelihood function under the alternative hypothesis is then given by equation 2.

## APPENDIX 3

## First and second partial derivatives of equation 2.

## 1. Notation

In order to shorten the expressions for the first and second partial derivatives of the log-likelihood function, we make use of the following notation

$$\begin{aligned}
\theta_1 &= \left( \frac{u_i - \mu_M}{\sigma_M} \right)^2 - 2\rho \left( \frac{u_i - \mu_M}{\sigma_M} \right) \left( \frac{v_i - \mu_F}{\sigma_F} \right) + \left( \frac{v_i - \mu_F}{\sigma_F} \right)^2 \\
\alpha &= \exp \left( -\frac{\theta_1}{2(1-\rho^2)} \right) \\
\theta_2 &= \left( \frac{v_i - \mu_M}{\sigma_M} \right)^2 - 2\rho \left( \frac{v_i - \mu_M}{\sigma_M} \right) \left( \frac{u_i - \mu_F}{\sigma_F} \right) + \left( \frac{u_i - \mu_F}{\sigma_F} \right)^2 \\
\beta &= \exp \left( -\frac{\theta_2}{2(1-\rho^2)} \right) \\
\gamma &= (\alpha + \beta)^{-1} \\
\alpha_{\mu_M} &= \frac{1}{1-\rho^2} \left[ \frac{(u_i - \mu_M)}{\sigma_M^2} - \frac{\rho(v_i - \mu_F)}{\sigma_M \sigma_F} \right] \\
\beta_{\mu_M} &= \frac{1}{1-\rho^2} \left[ \frac{(v_i - \mu_M)}{\sigma_M^2} - \frac{\rho(u_i - \mu_F)}{\sigma_M \sigma_F} \right] \\
\gamma_{\mu_M} &= -(\alpha \alpha_{\mu_M} + \beta \beta_{\mu_M})(\alpha + \beta)^{-2} \\
\alpha_{\mu_F} &= \frac{1}{1-\rho^2} \left[ \frac{(v_i - \mu_F)}{\sigma_F^2} - \frac{\rho(u_i - \mu_M)}{\sigma_M \sigma_F} \right] \\
\beta_{\mu_F} &= \frac{1}{1-\rho^2} \left[ \frac{(u_i - \mu_F)}{\sigma_F^2} - \frac{\rho(v_i - \mu_M)}{\sigma_M \sigma_F} \right] \\
\gamma_{\mu_F} &= -(\alpha \alpha_{\mu_F} + \beta \beta_{\mu_F})(\alpha + \beta)^{-2} \\
\alpha_{\sigma_M} &= \frac{1}{1-\rho^2} \left[ \frac{(u_i - \mu_M)^2}{\sigma_M^3} - \frac{\rho(v_i - \mu_F)(u_i - \mu_M)}{\sigma_M^2 \sigma_F} \right] \\
\beta_{\sigma_M} &= \frac{1}{1-\rho^2} \left[ \frac{(v_i - \mu_M)^2}{\sigma_M^3} - \frac{\rho(u_i - \mu_F)(v_i - \mu_M)}{\sigma_M^2 \sigma_F} \right] \\
\gamma_{\sigma_M} &= -(\alpha \alpha_{\sigma_M} + \beta \beta_{\sigma_M})(\alpha + \beta)^{-2} \\
\alpha_{\sigma_F} &= \frac{1}{1-\rho^2} \left[ \frac{(v_i - \mu_F)^2}{\sigma_F^3} - \frac{\rho(v_i - \mu_F)(u_i - \mu_M)}{\sigma_F^2 \sigma_M} \right] \\
\beta_{\sigma_F} &= \frac{1}{1-\rho^2} \left[ \frac{(u_i - \mu_F)^2}{\sigma_F^3} - \frac{\rho(u_i - \mu_F)(v_i - \mu_M)}{\sigma_F^2 \sigma_M} \right] \\
\gamma_{\sigma_F} &= -(\alpha \alpha_{\sigma_F} + \beta \beta_{\sigma_F})(\alpha + \beta)^{-2} \\
\alpha_\rho &= \frac{(u_i - \mu_M)(v_i - \mu_F)}{(1-\rho^2)\sigma_M \sigma_F} - \frac{\rho \theta_1}{(1-\rho^2)^2} \\
\beta_\rho &= \frac{(v_i - \mu_M)(u_i - \mu_F)}{(1-\rho^2)\sigma_M \sigma_F} - \frac{\rho \theta_2}{(1-\rho^2)^2} \\
\gamma_\rho &= -(\alpha \alpha_\rho + \beta \beta_\rho)(\alpha + \beta)^{-2}.
\end{aligned}$$

## 2. Parameter estimates

The first partial derivatives of the log-likelihood function  $\ell_1$  (equation 2) with respect to the parameters are given by

$$\begin{aligned}\frac{\partial \ell_1}{\partial \mu_M} &= \frac{n_M(\bar{z}_M - \mu_M)}{\sigma_M^2} + \sum_{i=1}^r \gamma [\alpha \alpha_{\mu_M} + \beta \beta_{\mu_M}] \\ \frac{\partial \ell_1}{\partial \mu_F} &= \frac{n_F(\bar{w}_F - \mu_F)}{\sigma_F^2} + \sum_{i=1}^r \gamma [\alpha \alpha_{\mu_F} + \beta \beta_{\mu_F}] \\ \frac{\partial \ell_1}{\partial \sigma_M} &= -\frac{n_M}{\sigma_M} - \frac{r}{\sigma_M} + \frac{((n_M - 1)s_M^2 + n_M(\bar{z}_M - \mu_M)^2)}{\sigma_M^3} \\ &\quad + \sum_{i=1}^r \gamma [\alpha \alpha_{\sigma_M} + \beta \beta_{\sigma_M}] \\ \frac{\partial \ell_1}{\partial \sigma_F} &= -\frac{n_F}{\sigma_F} - \frac{r}{\sigma_F} + \frac{((n_F - 1)s_F^2 + n_F(\bar{w}_F - \mu_F)^2)}{\sigma_F^3} \\ &\quad + \sum_{i=1}^r \gamma [\alpha \alpha_{\sigma_F} + \beta \beta_{\sigma_F}] \\ \frac{\partial \ell_1}{\partial \rho} &= \frac{r\rho}{(1 - \rho^2)} + \sum_{i=1}^r \gamma [\alpha \alpha_\rho + \beta \beta_\rho].\end{aligned}$$

The maximum likelihood estimators are found by setting these five partial derivatives to zero and solving for the five unknown parameters. These equations cannot be solved explicitly and therefore the Newton-Raphson iterative algorithm (Appendix 4) is used to solve them. For this, the second partial derivatives of  $\ell_1$  with respect to the parameters are required and these are given by

$$\begin{aligned}\frac{\partial^2 \ell_1}{\partial \mu_M \partial \mu_M} &= -\frac{n_M}{\sigma_M^2} + \sum_{i=1}^r \left\{ \gamma_{\mu_M} [\alpha \alpha_{\mu_M} + \beta \beta_{\mu_M}] + \gamma [\alpha \alpha_{\mu_M}^2 + \beta \beta_{\mu_M}^2] - \frac{1}{(1 - \rho^2)\sigma_M^2} \right\} \\ \frac{\partial^2 \ell_1}{\partial \mu_M \partial \mu_F} &= \sum_{i=1}^r \left\{ \gamma_{\mu_F} [\alpha \alpha_{\mu_M} + \beta \beta_{\mu_M}] + \gamma [\alpha \alpha_{\mu_M} \alpha_{\mu_F} + \beta \beta_{\mu_M} \beta_{\mu_F}] + \frac{\rho}{(1 - \rho^2)\sigma_M \sigma_F} \right\} \\ \frac{\partial^2 \ell_1}{\partial \mu_M \partial \sigma_M} &= -\frac{2n_M(\bar{z}_M - \mu_M)}{\sigma_M^3} + \sum_{i=1}^r \left\{ \gamma_{\sigma_M} [\alpha \alpha_{\mu_M} + \beta \beta_{\mu_M}] \right. \\ &\quad + \gamma \left\{ \alpha \alpha_{\mu_M} \alpha_{\sigma_M} + \beta \beta_{\mu_M} \beta_{\sigma_M} + \alpha \left( \frac{1}{1 - \rho^2} \left[ \frac{\rho(v_i - \mu_F)}{\sigma_M^2 \sigma_F} - \frac{2(u_i - \mu_M)}{\sigma_M^3} \right] \right) \right. \\ &\quad \left. \left. + \beta \left( \frac{1}{1 - \rho^2} \left[ \frac{\rho(u_i - \mu_F)}{\sigma_M^2 \sigma_F} - \frac{2(v_i - \mu_M)}{\sigma_M^3} \right] \right) \right\} \right\} \\ \frac{\partial^2 \ell_1}{\partial \mu_M \partial \sigma_F} &= \sum_{i=1}^r \left\{ \gamma_{\sigma_F} [\alpha \alpha_{\mu_M} + \beta \beta_{\mu_M}] + \gamma [\alpha \alpha_{\mu_M} \alpha_{\sigma_F} + \beta \beta_{\mu_M} \beta_{\sigma_F}] \right. \\ &\quad \left. + \frac{\rho}{(1 - \rho^2)\sigma_M \sigma_F^2} [\alpha(v_i - \mu_F) + \beta(u_i - \mu_M)] \right\} \\ \frac{\partial^2 \ell_1}{\partial \mu_M \partial \rho} &= \sum_{i=1}^r \left\{ \gamma_\rho [\alpha \alpha_{\mu_M} + \beta \beta_{\mu_M}] + \gamma \left\{ \alpha \alpha_{\mu_M} \alpha_\rho + \beta \beta_{\mu_M} \beta_\rho \right. \right.\end{aligned}$$

$$\begin{aligned}
& -\alpha \left( \frac{v_i - \mu_F}{(1 - \rho^2)\sigma_M\sigma_F} + \frac{\rho}{(1 - \rho^2)^2} \left[ -\frac{2(u_i - \mu_M)}{\sigma_M^2} + \frac{2\rho(v_i - \mu_F)}{\sigma_M\sigma_F} \right] \right) \\
& -\beta \left( \frac{u_i - \mu_F}{(1 - \rho^2)\sigma_M\sigma_F} + \frac{\rho}{(1 - \rho^2)^2} \left[ -\frac{2(v_i - \mu_M)}{\sigma_M^2} + \frac{2\rho(u_i - \mu_F)}{\sigma_M\sigma_F} \right] \right) \Bigg\} \\
\frac{\partial^2 \ell_1}{\partial \mu_F \partial \mu_F} &= -\frac{n_F}{\sigma_F^2} + \sum_{i=1}^r \left\{ \gamma_{\mu_F} [\alpha \alpha_{\mu_F} + \beta \beta_{\mu_F}] + \gamma [\alpha \alpha_{\mu_F}^2 + \beta \beta_{\mu_F}^2] - \frac{1}{(1 - \rho^2)\sigma_F^2} \right\} \\
\frac{\partial^2 \ell_1}{\partial \mu_F \partial \sigma_M} &= \sum_{i=1}^r \left\{ \gamma_{\sigma_M} [\alpha \alpha_{\mu_F} + \beta \beta_{\mu_F}] + \gamma [\alpha \alpha_{\mu_F} \alpha_{\sigma_M} + \beta \beta_{\mu_F} \beta_{\sigma_M}] \right. \\
& \quad \left. + \frac{\rho}{(1 - \rho^2)\sigma_F\sigma_M^2} [\alpha(u_i - \mu_M) + \beta(v_i - \mu_M)] \right\} \\
\frac{\partial^2 \ell_1}{\partial \mu_F \partial \sigma_F} &= -\frac{2n_F(\bar{w}_F - \mu_F)}{\sigma_F^3} + \sum_{i=1}^r \left\{ \gamma_{\sigma_F} [\alpha \alpha_{\mu_F} + \beta \beta_{\mu_F}] + \gamma \left\{ \alpha \alpha_{\mu_F} \alpha_{\sigma_F} + \beta \beta_{\mu_F} \beta_{\sigma_F} \right. \right. \\
& \quad \left. \left. + \alpha \left( \frac{1}{1 - \rho^2} \left[ \frac{\rho(u_i - \mu_M)}{\sigma_F^2\sigma_M} - \frac{2(v_i - \mu_F)}{\sigma_F^3} \right] \right) \right. \right. \\
& \quad \left. \left. + \beta \left( \frac{1}{1 - \rho^2} \left[ \frac{\rho(v_i - \mu_M)}{\sigma_F^2\sigma_M} - \frac{2(u_i - \mu_F)}{\sigma_F^3} \right] \right) \right\} \right\} \\
\frac{\partial^2 \ell_1}{\partial \mu_F \partial \rho} &= \sum_{i=1}^r \left\{ \gamma_{\rho} [\alpha \alpha_{\mu_F} + \beta \beta_{\mu_F}] + \gamma \left\{ \alpha \alpha_{\mu_F} \alpha_{\rho} + \beta \beta_{\mu_F} \beta_{\rho} \right. \right. \\
& \quad \left. \left. - \alpha \left( \frac{u_i - \mu_M}{(1 - \rho^2)\sigma_M\sigma_F} + \frac{\rho}{(1 - \rho^2)^2} \left[ \frac{2\rho(u_i - \mu_M)}{\sigma_M\sigma_F} - \frac{2(v_i - \mu_F)}{\sigma_F^2} \right] \right) \right. \right. \\
& \quad \left. \left. - \beta \left( \frac{v_i - \mu_M}{(1 - \rho^2)\sigma_M\sigma_F} + \frac{\rho}{(1 - \rho^2)^2} \left[ \frac{2\rho(v_i - \mu_M)}{\sigma_M\sigma_F} - \frac{2(u_i - \mu_F)}{\sigma_F^2} \right] \right) \right\} \right\} \\
\frac{\partial^2 \ell_1}{\partial \sigma_M \partial \sigma_M} &= \frac{n_M}{\sigma_M^2} + \frac{r}{\sigma_M^2} - \frac{3((n_M - 1)s_M^2 + n_M(\bar{z}_M - \mu_M)^2)}{\sigma_M^4} \\
& \quad + \sum_{i=1}^r \left\{ \gamma_{\sigma_M} [\alpha \alpha_{\sigma_M} + \beta \beta_{\sigma_M}] + \gamma \left\{ \alpha \alpha_{\sigma_M}^2 + \beta \beta_{\sigma_M}^2 \right. \right. \\
& \quad \left. \left. - \alpha \left( \frac{1}{1 - \rho^2} \left[ \frac{3(u_i - \mu_M)^2}{\sigma_M^4} - \frac{2\rho(u_i - \mu_M)(v_i - \mu_F)}{\sigma_M^3\sigma_F} \right] \right) \right. \right. \\
& \quad \left. \left. - \beta \left( \frac{1}{1 - \rho^2} \left[ \frac{3(v_i - \mu_M)^2}{\sigma_M^4} - \frac{2\rho(v_i - \mu_M)(u_i - \mu_F)}{\sigma_M^3\sigma_F} \right] \right) \right\} \right\} \\
\frac{\partial^2 \ell_1}{\partial \sigma_M \partial \sigma_F} &= \sum_{i=1}^r \left\{ \gamma_{\sigma_F} [\alpha \alpha_{\sigma_M} + \beta \beta_{\sigma_M}] + \gamma [\alpha \alpha_{\sigma_M} \alpha_{\sigma_F} + \beta \beta_{\sigma_M} \beta_{\sigma_F}] \right. \\
& \quad \left. + \frac{\alpha \rho (u_i - \mu_M)(v_i - \mu_F)}{(1 - \rho^2)\sigma_M^2\sigma_F^2} + \frac{\beta \rho (u_i - \mu_F)(v_i - \mu_M)}{(1 - \rho^2)\sigma_M^2\sigma_F^2} \right\} \\
\frac{\partial^2 \ell_1}{\partial \sigma_M \partial \rho} &= \sum_{i=1}^r \left\{ \gamma_{\rho} [\alpha \alpha_{\sigma_M} + \beta \beta_{\sigma_M}] + \gamma \left\{ \alpha \alpha_{\sigma_M} \alpha_{\rho} + \beta \beta_{\sigma_M} \beta_{\rho} \right. \right. \\
& \quad \left. \left. - \alpha \left( \frac{(u_i - \mu_M)(v_i - \mu_F)}{(1 - \rho^2)\sigma_M^2\sigma_F} + \frac{\rho}{(1 - \rho^2)^2} \left[ -\frac{2(u_i - \mu_M)^2}{\sigma_M^3} \right. \right. \right.
\end{aligned}$$

$$\begin{aligned}
& + \frac{2\rho(u_i - \mu_M)(v_i - \mu_F)}{\sigma_M^2 \sigma_F} \Big] \Big) - \beta \left( \frac{(v_i - \mu_M)(u_i - \mu_F)}{(1 - \rho^2)\sigma_M^2 \sigma_F} \right. \\
& \left. + \frac{\rho}{(1 - \rho^2)^2} \left[ -\frac{2(v_i - \mu_M)^2}{\sigma_M^3} + \frac{2\rho(v_i - \mu_M)(u_i - \mu_F)}{\sigma_M^2 \sigma_F} \right] \right) \Big\} \\
\frac{\partial^2 \ell_1}{\partial \sigma_F \partial \sigma_F} &= \frac{n_F}{\sigma_F^2} + \frac{r}{\sigma_F^2} - \frac{3((n_F - 1)s_F^2 + n_F(\bar{w}_F - \mu_F)^2)}{\sigma_F^4} \\
& + \sum_{i=1}^r \left\{ \gamma_{\sigma_F} [\alpha \alpha_{\sigma_F} + \beta \beta_{\sigma_F}] + \gamma \left\{ \alpha \alpha_{\sigma_F}^2 + \beta \beta_{\sigma_F}^2 \right. \right. \\
& - \alpha \left( \frac{1}{1 - \rho^2} \left[ \frac{3(v_i - \mu_F)^2}{\sigma_F^4} - \frac{2\rho(u_i - \mu_M)(v_i - \mu_F)}{\sigma_F^3 \sigma_M} \right] \right) \\
& \left. \left. - \beta \left( \frac{1}{1 - \rho^2} \left[ \frac{3(u_i - \mu_F)^2}{\sigma_F^4} - \frac{2\rho(v_i - \mu_M)(u_i - \mu_F)}{\sigma_F^3 \sigma_M} \right] \right) \right\} \right\} \\
\frac{\partial^2 \ell_1}{\partial \sigma_F \partial \rho} &= \sum_{i=1}^r \left\{ \gamma_{\rho} [\alpha \alpha_{\sigma_F} + \beta \beta_{\sigma_F}] + \gamma \left\{ \alpha \alpha_{\sigma_F} \alpha_{\rho} + \beta \beta_{\sigma_F} \beta_{\rho} \right. \right. \\
& - \alpha \left( \frac{(u_i - \mu_M)(v_i - \mu_F)}{(1 - \rho^2)\sigma_F^2 \sigma_M} + \frac{\rho}{(1 - \rho^2)^2} \left[ \frac{2\rho(u_i - \mu_M)(v_i - \mu_F)}{\sigma_F^2 \sigma_M} \right. \right. \\
& \left. \left. - \frac{2(v_i - \mu_F)^2}{\sigma_F^3} \right] \right) - \beta \left( \frac{(v_i - \mu_M)(u_i - \mu_F)}{(1 - \rho^2)\sigma_F^2 \sigma_M} \right. \\
& \left. \left. + \frac{\rho}{(1 - \rho^2)^2} \left[ \frac{2\rho(v_i - \mu_M)(u_i - \mu_F)}{\sigma_F^2 \sigma_M} - \frac{2(u_i - \mu_F)^2}{\sigma_F^3} \right] \right) \right\} \\
\frac{\partial^2 \ell_1}{\partial \rho \partial \rho} &= \frac{r(1 - \rho^2) + 2r\rho^2}{(1 - \rho^2)^2} + \sum_{i=1}^r \left\{ \gamma_{\rho} [\alpha \alpha_{\rho} + \beta \beta_{\rho}] + \gamma \left\{ \alpha \alpha_{\rho}^2 + \beta \beta_{\rho}^2 \right. \right. \\
& + \alpha \left( \frac{4\rho(u_i - \mu_M)(v_i - \mu_F)}{(1 - \rho^2)^2 \sigma_F \sigma_M} - \frac{\theta_1(4\rho^2 + (1 - \rho^2))}{(1 - \rho^2)^3} \right) \\
& \left. \left. + \beta \left( \frac{4\rho(v_i - \mu_M)(u_i - \mu_F)}{(1 - \rho^2)^2 \sigma_F \sigma_M} - \frac{\theta_2(4\rho^2 - (1 - \rho^2))}{(1 - \rho^2)^3} \right) \right\} \right\}.
\end{aligned}$$

## APPENDIX 4

**Newton-Raphson iterative algorithm to estimate model parameters.**

The Newton-Raphson algorithm is a technique of iteratively solving nonlinear equations of several variables. The algorithm consists of the following steps:

- Step 1:** Calculate initial estimates  $\widehat{\mu}_M, \widehat{\mu}_F, \widehat{\sigma}_M, \widehat{\sigma}_F, \widehat{\rho}$ .
- Step 2:** Compute  $f^{(k)}$  and  $F^{(k)}$ , where  $f^{(k)}$  is the vector of first partial derivatives and  $F^{(k)}$  is the matrix of second partial derivatives, computed at the  $k$ th iteration.
- Step 3:** Compute the vector  $\delta^{(k)}$  which is the solution to the system of  $p$  linear equations

$$F^{(k)}\delta^{(k)} = f^{(k)},$$

where  $p$  represents the number of parameters.

- Step 4:** Set  $\beta^{(k+1)} = \beta^{(k)} - \delta^{(k)}$ , where  $\beta^{(k)}$  contains the parameter estimates at the  $k$ th iteration.
- Step 5:** Test for convergence, for example, if the elements of  $f^{(k)}$  are sufficiently close to zero. If the convergence criterion is met then stop, otherwise increase  $k$  by 1 and return to step 2.

## APPENDIX 5

## First and second partial derivatives for equation 6.

Assuming that the parameters  $\mu_M$ ,  $\mu_F$ ,  $\sigma_M$  and  $\sigma_F$  are known, then the log-likelihood function under the alternative hypothesis is given by equation 6. To estimate the parameter  $\rho$  the first and second partial derivatives of  $\ell_2$  with respect to  $\rho$  are required.

## 1. Notation

As before, to shorten expressions, we make use of the following notation

$$\begin{aligned}\theta_1 &= \left(\frac{u_i - \mu_M}{\sigma_M}\right)^2 - 2\rho \left(\frac{u_i - \mu_M}{\sigma_M}\right) \left(\frac{v_i - \mu_F}{\sigma_F}\right) + \left(\frac{v_i - \mu_F}{\sigma_F}\right)^2 \\ \alpha &= \exp\left(-\frac{\theta_1}{2(1-\rho^2)}\right) \\ \theta_2 &= \left(\frac{v_i - \mu_M}{\sigma_M}\right)^2 - 2\rho \left(\frac{v_i - \mu_M}{\sigma_M}\right) \left(\frac{u_i - \mu_F}{\sigma_F}\right) + \left(\frac{u_i - \mu_F}{\sigma_F}\right)^2 \\ \beta &= \exp\left(-\frac{\theta_2}{2(1-\rho^2)}\right) \\ \gamma &= (\alpha + \beta)^{-1} \\ \alpha_\rho &= \frac{(u_i - \mu_M)(v_i - \mu_F)}{(1-\rho^2)\sigma_M\sigma_F} - \frac{\rho\theta_1}{(1-\rho^2)^2} \\ \beta_\rho &= \frac{(v_i - \mu_M)(u_i - \mu_F)}{(1-\rho^2)\sigma_M\sigma_F} - \frac{\rho\theta_2}{(1-\rho^2)^2} \\ \gamma_\rho &= -(\alpha\alpha_\rho + \beta\beta_\rho)(\alpha + \beta)^{-2}.\end{aligned}$$

## 2. Parameter estimates

The first partial derivative of  $\ell_2$ , as given in equation 6, is

$$\frac{\partial \ell_2}{\partial \rho} = \frac{r\rho}{(1-\rho^2)} + \sum_{i=1}^r \gamma [\alpha\alpha_\rho + \beta\beta_\rho].$$

The second partial derivative of  $\ell_2$  is given by

$$\begin{aligned}\frac{\partial^2 \ell_2}{\partial \rho \partial \rho} &= \frac{r(1-\rho^2) + 2r\rho^2}{(1-\rho^2)^2} + \sum_{i=1}^r \left\{ \gamma_\rho [\alpha\alpha_\rho + \beta\beta_\rho] + \gamma \left\{ \alpha\alpha_\rho^2 + \beta\beta_\rho^2 \right. \right. \\ &\quad + \alpha \left( \frac{4\rho(u_i - \mu_M)(v_i - \mu_F)}{(1-\rho^2)^2\sigma_F\sigma_M} - \frac{\theta_1(4\rho^2 + (1-\rho^2))}{(1-\rho^2)^3} \right) \\ &\quad \left. \left. + \beta \left( \frac{4\rho(v_i - \mu_M)(u_i - \mu_F)}{(1-\rho^2)^2\sigma_F\sigma_M} - \frac{\theta_2(4\rho^2 - (1-\rho^2))}{(1-\rho^2)^3} \right) \right\} \right\}.\end{aligned}$$

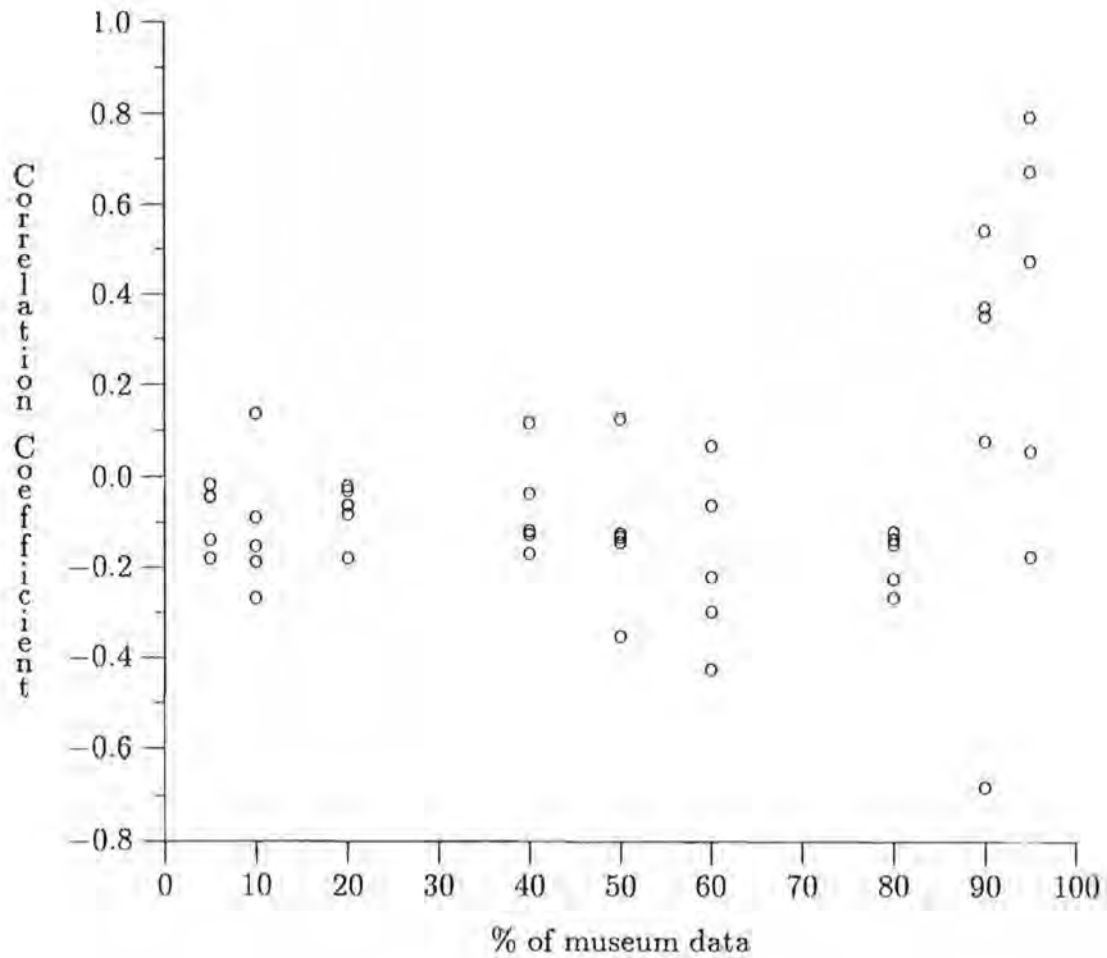
## APPENDIX 6

**Decision on the sample size to randomly select as "museum" data.**

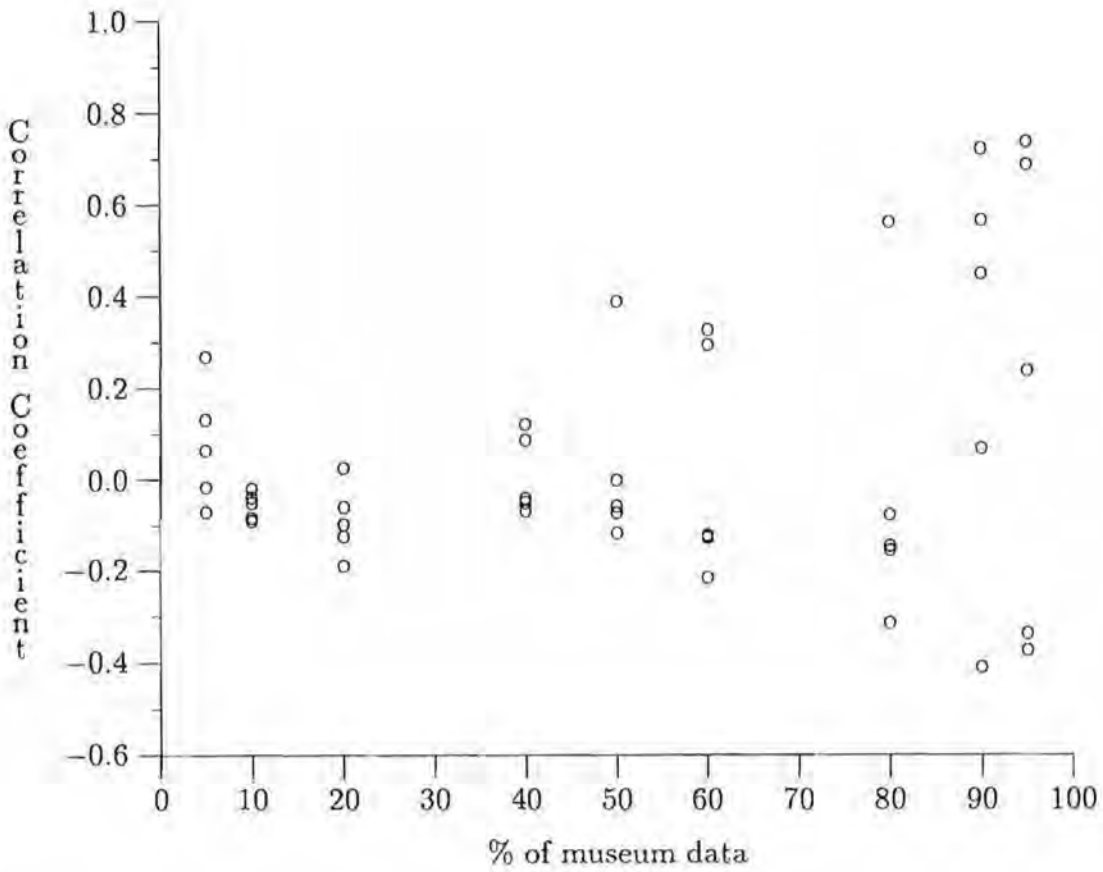
The decision of extracting a sample size of 10 observations to form the "museum data" was made on the basis that several samples of different sizes were chosen from bootstrap samples of the Dunlin data, namely we considered taking out a portion of the data to be treated as "museum data", starting by randomly taking out 10% and then increasing this amount in steps of 10. The extremes of 5% and 95% were also considered. For each of these amounts five repetitions were performed by randomly selecting a portion of the data. The model parameters were then estimated for each group and each repetition. The plot of the estimated correlation coefficients for bill-, wing-, and tarsus-length against % of "museum data" shows a "funnel" effect. This funnel pattern demonstrates that the fewer "field" observations are available, the more erratic is the estimate of the correlation coefficient. Conversely, as the amount of "field" data increases, the estimate of the correlation coefficient stabilises, even though at this stage the "museum" data is fairly small. In fact, at the point when we only have 10 observations for "museum" data, the model parameters are fairly accurately estimated. This indicates that the model is more sensitive to small field samples and fairly insensitive to small historical data sets (Figs 1-3).

In practice the size of the museum sample is frequently small and the option of increasing it is not an attractive one because birds would have to be killed to be able to ascertain the sex of the bird. It would therefore be of some interest to know how the proposed model for estimating assortative mating behaves when the museum sample is rather small. The exercise undertaken was to limit the museum sample size to be very small and to vary the field sample size from small (10 observations) to quite large samples (190 observations). Figs 4-7 show box-and-whisker plots (of 100 repetitions) for the estimated correlation coefficient for tarsus- and bill-lengths for museum sample sizes of 10 and 20 observations. From the plots it can be seen that as the field sample size increases, the estimate of the correlation coefficient stabilises. In fact, when the field sample size reaches approximately 100 observations, the estimate is consistent even when only 10 museum observations are available.

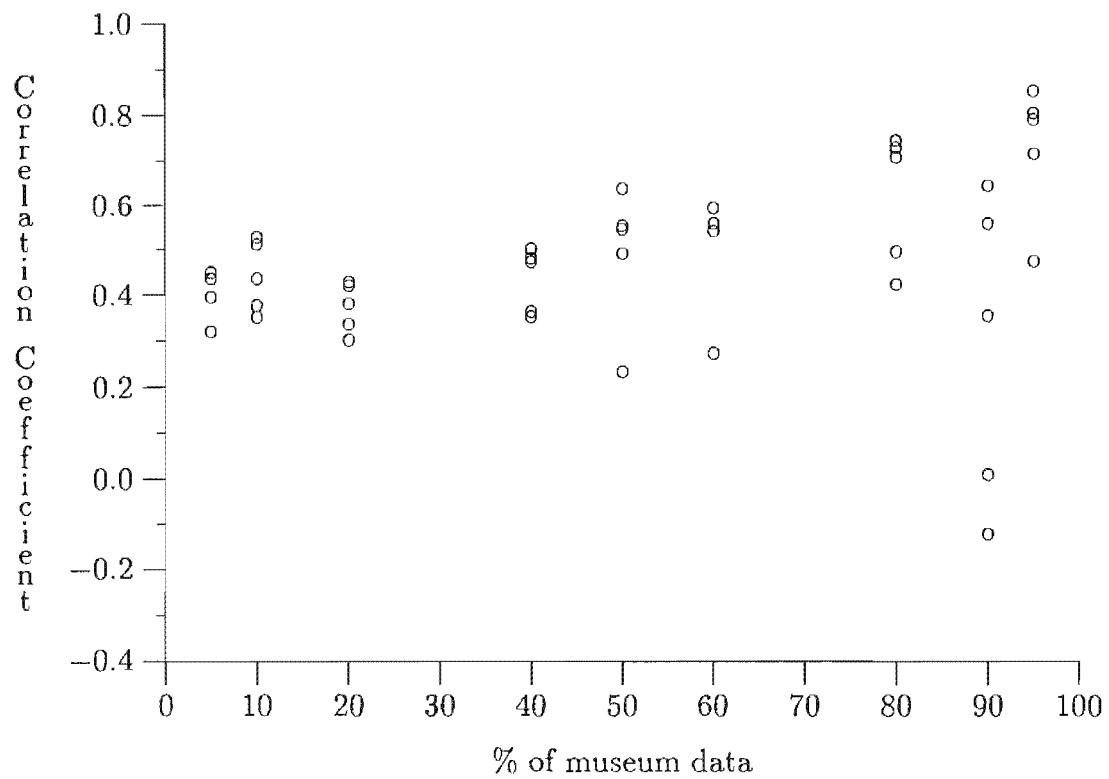
The conclusion we draw is that the sample size of museum observations of birds of known sex need only be small, and that priority should be given to increasing the sample size of the field observations of pairs of birds.



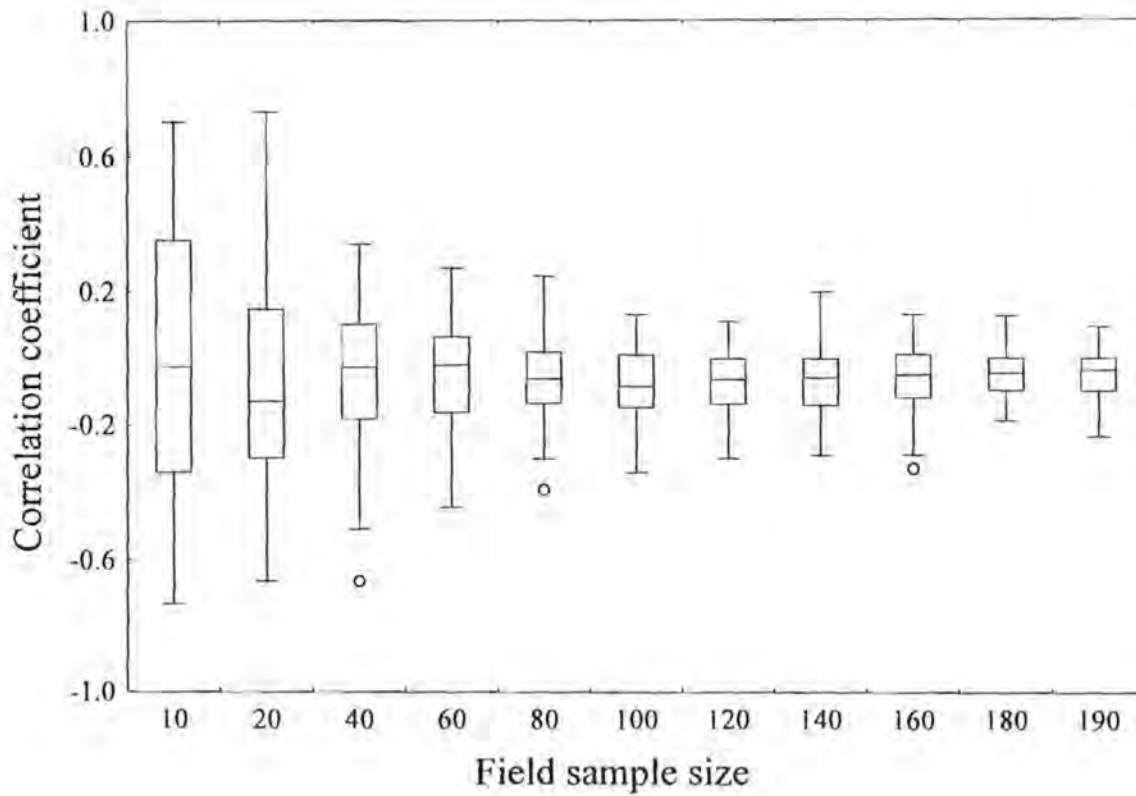
**Figure 1.** Estimated correlation coefficients for bill-length when various percentages of the data were randomly extracted as "museum" data, while the remainder of the observations were considered to be "field" data. The process of randomly selecting each amount as "museum" data was repeated five times.



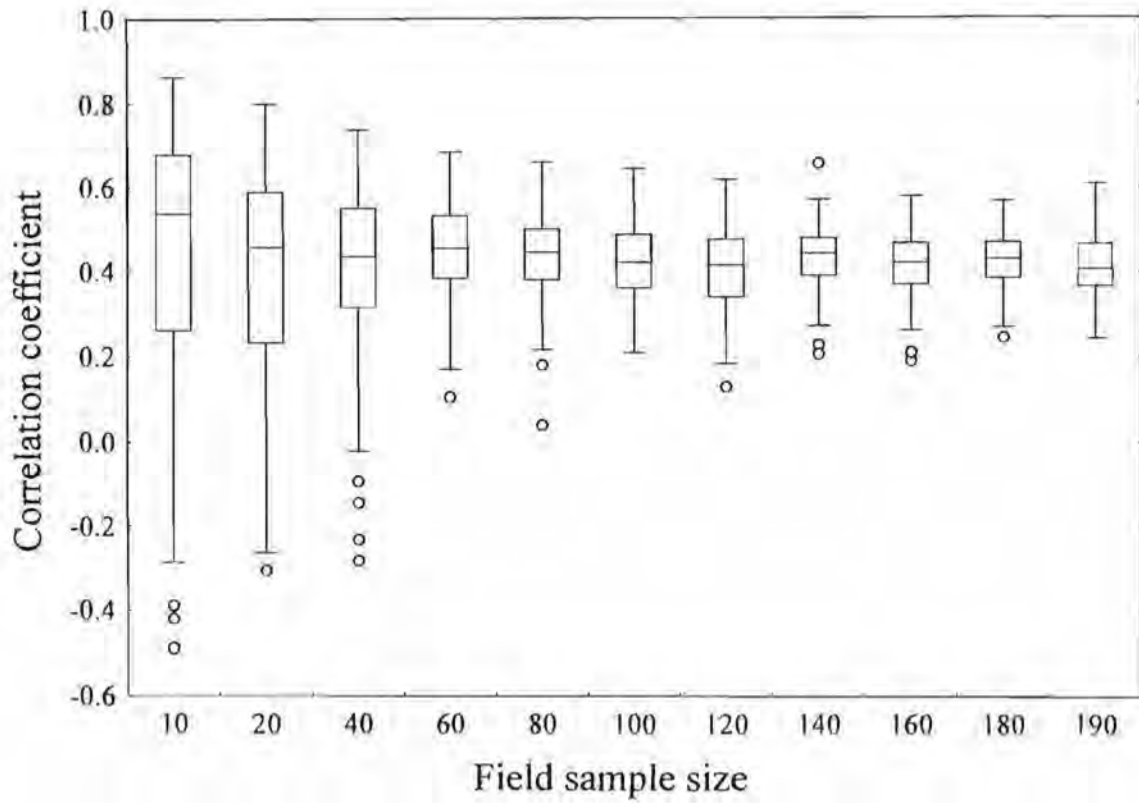
**Figure 2.** Estimated correlation coefficients for wing-length when various percentages of the data were randomly extracted as “museum” data, while the remainder of the observations were considered to be “field” data. The process of randomly selecting each amount as “museum” data was repeated five times.



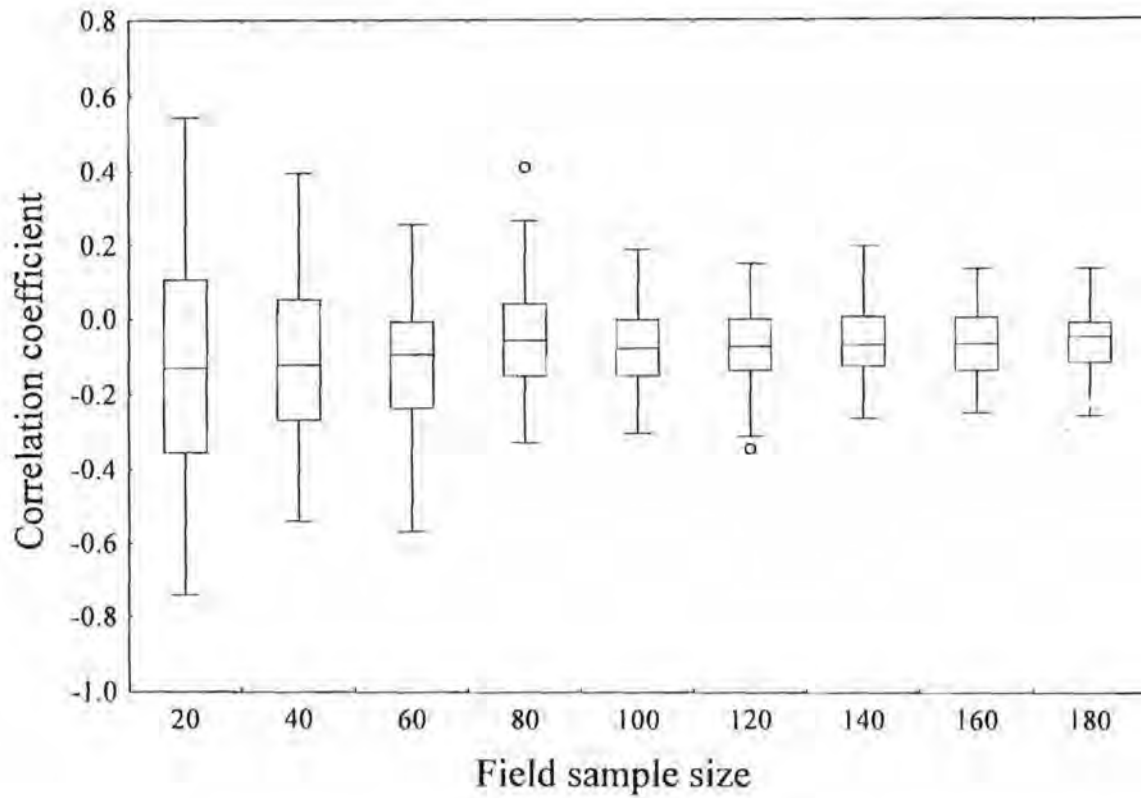
**Figure 3.** Estimated correlation coefficients for tarsus-length when various percentages of the data were randomly extracted as “museum” data, while the remainder of the observations were considered to be “field” data. The process of randomly selecting each amount as “museum” data was repeated five times.



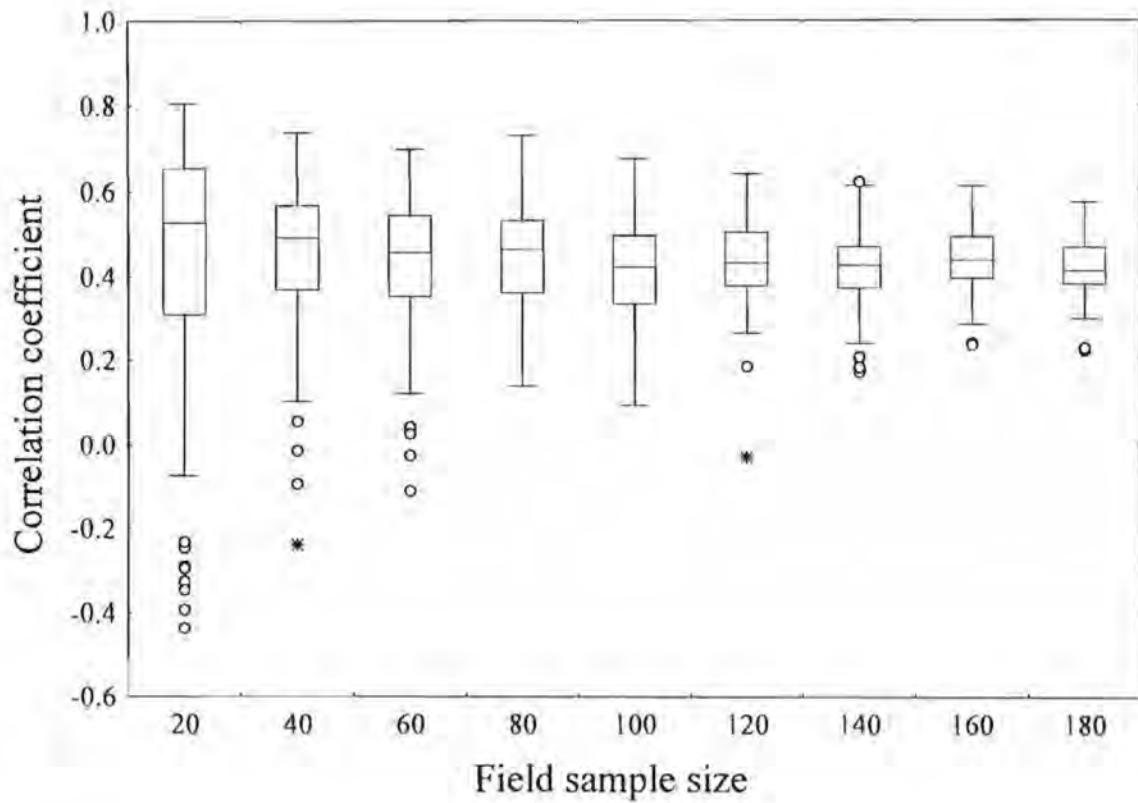
**Figure 4.** Box-and-whisker plots for the estimated correlation coefficient for bill-length. The amount of data randomly selected as “museum” data was held fixed at 10 observations, while the amount of data randomly selected as “field” data was varied. Each selection at given amounts was repeated 100 times.



**Figure 5.** Box-and-whisker plots for the estimated correlation coefficient for tarsus-length. The amount of data randomly selected as “museum” data was held fixed at 10 observations, while the amount of data randomly selected as “field” data was varied. Each selection at given amounts was repeated 100 times.



**Figure 6.** Box-and-whisker plots for the estimated correlation coefficient for bill-length. The amount of data randomly selected as “museum” data was held fixed at 20 observations, while the amount of data randomly selected as “field” data was varied. Each selection at given amounts was repeated 100 times.



**Figure 7.** Box-and-whisker plots for the estimated correlation coefficient for tarsus-length. The amount of data randomly selected as “museum” data was held fixed at 20 observations, while the amount of data randomly selected as “field” data was varied. Each selection at given amounts was repeated 100 times.

**Estimation of the demographic ratios of herding  
species: a case study of sex ratios of the  
Springbok *Antidorcas marsupialis***

ABSTRACT

Estimates of sex ratios are usually based on the assumption that individual animals constitute the natural sampling unit. However all sexually reproductive animals have social structures in order to bring the opposite sexes together and the appropriate sampling unit is therefore the "social structure" rather than the individual. Confidence intervals of sex ratios that ignore this sampling unit can give misleading results. In this study, several populations of Springbok *Antidorcas marsupialis*, which form three different types of herd (breeding herds, bachelor herds and solitary herds) are used to illustrate various confidence intervals constructed for the sex ratio. Both parametric and nonparametric confidence intervals for the sex ratio are considered. Binomial confidence intervals are too narrow in this case. Confidence intervals obtained with standard errors of the ratio computed from the Taylor expansion of the ratio and of the logged ratio are less conservative. The Taylor expansion of the logged ratio produces better standard errors for the ratio. Bootstrap confidence intervals were similar to those obtained by the Taylor expansion of the logged ratio.

INTRODUCTION

Sex ratios and age composition ratios are important demographic parameters, useful in research on, and management of, animal populations. These ratios are often reported in the literature and sex ratios in particular are of considerable theoretical interest for ecologists (Fisher 1930, Trivers and Willard 1973, Maynard Smith 1978, Clutton-Brock *et al.* 1984). Clutton-Brock and Iason (1986) conducted an extensive review of papers reporting sex ratios for mammals. Only those estimates based on a sample size greater than 50 were used. 115 sets of sex ratio data were reported, derived from 70 papers, of these, 50 were considered

to depart significantly from a 1:1 ratio. Mentis (1972) in his review of life history features of African large herbivores reported 169 sex ratios.

The present work was stimulated by problems which arose during intensive management for production of several populations of Springbok *Antidorcas marsupialis* near Kimberley in the Cape Province of South Africa. These populations are counted annually and the sex and adult/infant ratios are estimated from a sample of approximately 20% of the population. Animals are also removed from the population during a commercial cull and the sex of each animal removed is recorded. It proved impossible to reconcile sex ratios estimated from samples taken before and after the cull with the known numbers of each sex removed from the population.

Estimates of these key demographic ratios are usually based on the assumption that the individual animals constitute the natural sampling unit. However all sexually reproductive animals have a social structure. For many species this social organisation may be long-lasting or even permanent. For example, many species of large and medium sized antelope form herds throughout the year.

By definition a herd comprises a group of individuals associated with each other in space and time. Therefore if sampling is done visually while moving through a free population then the natural sampling unit is the herd rather than the individual (where the term "herd" is taken to include solitary individuals).

Often these herds are not formed at random from the population because they are formed for specific purposes. The composition of a herd varies according to its social purpose. For example, Springbok form three different types of herd; breeding herds, bachelor herds and solitary males. Breeding herds consist of adult and subadult females, infants of both sexes and one or more dominant (territorial) rams. Bachelor herds are made up of juvenile and subdominant rams sometimes associated with a few subadult females. Solitary rams occupy territories from which they attempt to exclude other rams but associate with any females which enter the territory.

It is evident that each of these three herd types cannot be regarded as a random sample from the population; they differ in sex ratio and in adult/infant ratio composition. The distribution of the sizes of herds is also different for each type of herd.

This paper considers the estimation of sex ratios for Springbok as well as the construction of confidence intervals for the estimate of sex ratios. The same method could also be applied to the estimation of adult/infant ratio.

## MATERIAL AND METHODS

## Data description

The data for this study consist of counts of the numbers of females and males constituting each herd of Springbok observed in three areas near Kimberley in the Northern Cape Province, South Africa. These areas are the game farms Benfontein (13000ha), Susanna (11000ha) and Uitzicht (2500ha). Three counts were made at approximately six month intervals from the beginning of 1985 to beginning of 1986, except for Uitzicht at which only two population counts were made; at the beginning of 1985 and at the beginning of 1986 (Table 1).

**Table 1.** Descriptive statistics of Springbok.

Area	Date	No. of herds	No. of females	No. of males	Average/herd		std dev/herd		corre- <sup>†</sup> lation
					females	males	females	males	
Benfontein	02/85	19	290	87	15.3	4.6	9.2	4.1	0.45
	09/85	27	257	62	9.5	2.3	11.4	3.5	0.30
	02/86	20	235	111	11.8	5.6	13.6	6.0	0.02
Susanna	01/85	8	135	55	16.9	6.9	18.8	6.9	0.29
	10/85	8	110	18	13.8	2.3	10.0	1.3	0.65
	02/86	13	269	124	20.7	9.5	18.8	13.2	0.34
Uitzicht	01/85	4	102	37	25.5	9.3	17.2	10.6	0.78
	02/86	5	40	16	8.0	3.2	6.1	5.0	-0.66

† Correlation of males and females/herd

The habitat of the farms varies from an open savanna grassland to a lightly wooded savanna with scattered *Acacia erioloba* trees. Grass height rarely exceeds 50cm. Samples were made by driving through the farm in a four wheel drive vehicle. When a herd of Springbok was seen the vehicle was stopped and the individuals were identified by sex and age using either a pair of 10 × 40 binoculars or a 20 × 100 spotting telescope. The route travelled through the farm was chosen to minimize the chance of reencountering a group which had been previously recorded during the sampling day.

### Estimating confidence intervals of sex ratios

A number of complications arise if one wishes to estimate confidence intervals for the sex ratio in a population of animals that form herds of individuals:

1. The animals live in herds and consequently the natural sampling unit is the herd rather than the individual. It is therefore incorrect to base interval estimation on the Bernoulli distribution as this will result in estimates which are too narrow because they are based on larger sample sizes than they ought to be. For example, Springbok live in one of three types of “herds”: single males, bachelor herds (all or mostly males) and breeding herds (mostly females). Thus the addition of a single herd to the sample can alter the estimates of the sex-ratio substantially.
2. Quantitative information relating to herd-size distributions is generally not available. In any case these distributions may be different in different seasons and under different environmental conditions. It is also doubtful whether the sample information which one usually has available is sufficient to estimate models for these distributions. The number of parameters in even simple models is quite large (in the order of 10). Finally it would be necessary to be able to classify each herd observed according to type which it is sometimes difficult to do. For example, it is sometimes not possible to distinguish a “bachelor herd” of Springbok from a “breeding herd” with certainty.
3. The population is finite and generally sampling is without replacement. This needs to be taken into account especially if the sampling proportion is substantial, which it often is (e.g. 0.25). Some estimators are a function of the sampling proportion and thus require an estimate of the population size.

In this study we consider various confidence intervals for the sex ratio of Springbok in which the appropriate sampling unit is the herd and in which sampling is done without replacement. We look at both parametric and nonparametric constructions of confidence intervals for the sex ratio of Springbok.

#### Notation

In what follows we will make use of the following notation:

$N$	number of herds in the population
$n$	number of herds in the sample

$F_i, f_i$	number of females in the $i$ th population and sample herd respectively
$M_i, m_i$	number of males in the $i$ th population and sample herd respectively
$(F_1, M_1), (F_2, M_2), \dots, (F_N, M_N)$	composition of population herds
$(f_1, m_1), (f_2, m_2), \dots, (f_n, m_n)$	composition of sample herds
$f' = n/N$	sampling fraction
$1 - f' = (N - n)/N$	finite population correction
$F = \sum_{i=1}^N F_i$	number of females in the population
$M = \sum_{i=1}^N M_i$	number of males in the population
$f = \sum_{i=1}^n f_i$	number of females in the sample
$m = \sum_{i=1}^n m_i$	number of males in the sample
$n' = f + m$	total number of Springbok in the sample
$\bar{F} = \frac{1}{N} \sum_{i=1}^N F_i$	average number of females per herd in the population
$\bar{M} = \frac{1}{N} \sum_{i=1}^N M_i$	average number of males per herd in the population
$\bar{f} = \frac{1}{n} \sum_{i=1}^n f_i$	average number of females per herd in the sample
$\bar{m} = \frac{1}{n} \sum_{i=1}^n m_i$	average number of males per herd in the sample
$\rho = \frac{F}{M} = \frac{\bar{F}}{\bar{M}}$	sex ratio in the population
$\pi = \frac{F}{(F + M)}$	proportion of females in the population
$\hat{F} = N\bar{f}$	estimate of the number of females in the population
$\hat{M} = N\bar{m}$	estimate of the number of males in the population
$\hat{\rho} = \frac{\hat{F}}{\hat{M}} = \frac{N\bar{f}}{N\bar{m}} = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n m_i}$	sex ratio in the sample

$\hat{\pi} = \frac{\hat{F}}{\hat{F} + \hat{M}} = \frac{\sum_{i=1}^n f_i}{\sum_{i=1}^n f_i + \sum_{i=1}^n m_i} = \frac{f}{n'}$	proportion of females in the sample
$S_{F_i}^2 = \frac{1}{N-1} \sum_{i=1}^N (F_i - \bar{F})^2$	variance of females/herd in the population
$S_{M_i}^2 = \frac{1}{N-1} \sum_{i=1}^N (M_i - \bar{M})^2$	variance of males/herd in the population
$S_{F_i M_i} = \frac{1}{N-1} \sum_{i=1}^N (M_i - \bar{M})(F_i - \bar{F})$	covariance of males and females/herd in the population
$s_{f_i}^2 = \frac{1}{n-1} \sum_{i=1}^n (f_i - \bar{f})^2$	variance of females/herd in the sample
$s_{m_i}^2 = \frac{1}{n-1} \sum_{i=1}^n (m_i - \bar{m})^2$	variance of males/herd in the sample
$s_{f_i m_i} = \frac{1}{n} \sum_{i=1}^n (m_i - \bar{m})(f_i - \bar{f})$	covariance of males and females/herd in the sample

### Confidence limits based on the binomial distribution

If we neglect for the moment that sampling is in fact taken without replacement (this complication will be discussed later) and furthermore assume that the sampling unit is the individual (“herds” are of size 1) then the distribution of the number of females in the sample given the total number of Springbok  $n'$  is the binomial distribution, with the probability  $\pi$  that any given Springbok is a female, that is

$$f \sim B(n', \pi).$$

From this we can compute the distribution of  $f$ , the number of females or of  $\hat{\pi} = \frac{f}{n'}$ , the estimate of the proportion of females in the population.

Confidence limits for the binomial distribution were discussed by Blyth (1986). Of the various approximations which have been suggested Blyth recommended one due to Pratt (1968) (see Blyth 1986, p855). The upper  $100(1 - 2\alpha)\%$  confidence limit is given by

$$\left[ 1 + \left( \frac{f+1}{n'-f} \right)^2 \left\{ \frac{(81(f+1)(n'-f) - 9n' - 8 - 3z^\alpha \sqrt{9(f+1)(n'-f)(9n'+5 - (z^\alpha)^2) + n'+1})}{81(f+1)^2 - 9(f+1)(2 + (z^\alpha)^2) + 1} \right\}^3 \right]^{-1},$$

where  $z^\alpha$  is the  $100\alpha$ th percentile of the standard normal distribution. To obtain the lower  $100(1 - 2\alpha)\%$  confidence limit replace  $f$  by  $f - 1$  and  $z^\alpha$  by  $-z^\alpha = z^{1-\alpha}$ . Thus if 90% confidence limits are required, we set  $\alpha = 0.05$  and  $z^\alpha = 1.64485$ .

Once we have computed confidence limits for  $\pi$  those for the sex ratio  $\rho$  are obtained by direct transformation because  $\rho$  is a monotonic function in  $\pi$ , i.e. if  $\hat{\pi}_L$  and  $\hat{\pi}_U$  are the relevant limits for  $\pi$  in that  $Pr\{\hat{\pi}_L \leq \pi \leq \hat{\pi}_U\} = 0.90$  then the lower and upper limits for  $\rho$  are

$$\hat{\rho}_L = \frac{\hat{\pi}_L}{(1 - \hat{\pi}_L)} \quad \text{and} \quad \hat{\rho}_U = \frac{\hat{\pi}_U}{(1 - \hat{\pi}_U)}$$

respectively, and a 90% confidence interval for  $\rho$  is given by  $(\hat{\rho}_L, \hat{\rho}_U)$ .

For large sample sizes we can use the normal approximation to the binomial distribution, i.e. for  $f \sim B(n', \pi)$ , then  $f \sim N(n'\pi, n'\pi(1 - \pi))$ . Approximate confidence intervals for the proportion of females can be constructed by observing that  $\hat{\pi} \sim N(\pi, \pi(1 - \pi)/n')$  so that  $100(1 - 2\alpha)\%$  confidence intervals for  $\pi$  are given by

$$\hat{\pi} \pm z^\alpha \sqrt{\hat{\pi}(1 - \hat{\pi})/n'}.$$

These confidence limits omit the finite population correction and are also uncorrected for continuity (Cochran 1977).

### Confidence limits based on the hypergeometric distribution

We look now at the situation where we have simple random sampling without replacement. Suppose that  $N' = F + M$  is known and that a sample of size  $n'$  is taken from the population. Again “herds” are of size 1, as in the binomial case. The appropriate distribution for  $f$ , the total number of females in the sample, is the hypergeometric distribution, i.e.  $f \sim H(N', F, n')$ . An unbiased estimator of  $\pi$  is given by

$$\hat{\pi} = \frac{f}{n'}$$

and its variance is

$$\text{Var}(\hat{\pi}) = \frac{\pi(1 - \pi)}{n'} \left( \frac{N' - n'}{N' - 1} \right)$$

(Cochran 1977).

Because  $f \sim H(N', F, n')$ , confidence intervals for  $f$  (and hence  $\pi$  and  $\rho$ ) can be computed. However because  $\text{Var}(\hat{\pi})$  in the hypergeometric distribution is smaller than this variance in the binomial case (i.e.  $\frac{\pi(1 - \pi)}{n'} \left( \frac{N' - n'}{N' - 1} \right) < \frac{\pi(1 - \pi)}{n'}$  for  $n' > 1$ ) we would expect the confidence interval for  $\rho$  to be narrower than is the (already unrealistically narrow) binomial case.

### Confidence intervals for ratios based on the normal distribution

For large samples the normal approximation to the ratio applies and confidence intervals for the ratio  $\rho$  are given by

$$\hat{\rho} \pm z^\alpha \sqrt{\text{Var}(\hat{\rho})}.$$

Under certain conditions (Cochran 1977) the normal approximation is a good approximation. When these conditions are not satisfied  $\text{Var}(\hat{\rho})$  is underestimated and the skewness of  $\hat{\rho}$  can become pronounced.

Interval estimates for a ratio of means were discussed by James *et al.* (1974). The results are applicable to ratios of means in the case of a bivariate normal distribution and are therefore not applicable in the strict sense of the word. Furthermore it is assumed that the covariance matrix is known which is entirely unrealistic in our application.

### Confidence intervals for ratios using Taylor expansions of the ratio

Bowden *et al.* (1984) considered estimates for age and sex ratios of mule deer as well as standard errors for these ratios. Instead of considering the ratio  $\rho = F/M$  (as in our study), they looked at the proportion of females  $\pi = F/(F + M)$ . Once estimates for  $\pi$  and its standard error are obtained they can easily be converted to estimates for  $\rho$  and its standard error because  $\rho = \pi/(1 - \pi)$ . They considered the usual standard error for the estimate of a proportion from a binomial distribution (see section above on this topic) and they used the standard error for a ratio as a robust estimate for the standard error of a proportion to deal with the situation (the present one) when the units of sampling are not individual deers but the unit of sampling is the herd of deer (see also Cochran 1977).

In this section we consider an estimate for the standard error of the estimated sex ratio without placing any distribution function on the sex ratio and by looking at the sampling units of the Springbok as herds and not individuals.

The standard approximate  $100(1-2\alpha)\%$  confidence interval for the sex ratio of Springbok  $\rho$  is given by

$$\rho \in \left( \hat{\rho} + \widehat{\sigma}_{\hat{\rho}} z^\alpha, \hat{\rho} + \widehat{\sigma}_{\hat{\rho}} z^{1-\alpha} \right),$$

where  $\hat{\rho}$  is an estimate of  $\rho$ ,  $\widehat{\sigma}_{\hat{\rho}}$  is an estimate of its standard error and  $z^\alpha$  is the  $100\alpha$ th percentile point of the standard normal distribution. In this section we look at a way to obtain an approximate estimate of the standard error of the estimate  $\hat{\rho}$ ,  $\widehat{\sigma}_{\hat{\rho}}$ .

The estimate of the sex ratio of Springbok  $\hat{\rho}$  is a function of two random variables,  $f$  and  $m$ , the total number of females and males in the sample respectively. Denote this function by  $g$ , that is

$$\hat{\rho} = g(f, m) = \frac{f}{m}.$$

If the first two moments of the random variables  $f$  and  $m$  are known, that is, we know  $E(f) = \mu_f$ ,  $\text{Var}(f) = \sigma_f^2$ ,  $E(m) = \mu_m$ ,  $\text{Var}(m) = \sigma_m^2$ , then the first two moments of  $\hat{\rho} = g(f, m)$  ( $E(\hat{\rho}) = \mu_{\hat{\rho}}$ ,  $\text{Var}(\hat{\rho}) = \sigma_{\hat{\rho}}^2$ ) can be found in terms of  $\mu_f$ ,  $\mu_m$ ,  $\sigma_f^2$  and  $\sigma_m^2$  if the function  $g$  is linear. When the function  $g$  is nonlinear, we can approximate it by a linear function and then approximate moments of  $\hat{\rho}$  can be obtained. The delta method is a procedure by which nonlinear functions can be approximated by linear ones and this is achieved by a Taylor expansion of  $g$  about the point  $(\mu_f, \mu_m)$  (Rice 1988).

The Taylor expansion of  $\hat{\rho}$  to the first order is given by

$$\hat{\rho} = g(f, m) \approx g(\mu_f, \mu_m) + (f - \mu_f) \frac{\partial g(\mu_f, \mu_m)}{\partial f} + (m - \mu_m) \frac{\partial g(\mu_f, \mu_m)}{\partial m}.$$

Approximate first moments of  $\hat{\rho}$  are then given by

$$E(\hat{\rho}) \approx g(\mu_f, \mu_m) = \frac{\mu_f}{\mu_m}$$

and

$$\text{Var}(\hat{\rho}) \approx \sigma_f^2 \left( \frac{\partial g(\mu_f, \mu_m)}{\partial f} \right)^2 + \sigma_m^2 \left( \frac{\partial g(\mu_f, \mu_m)}{\partial m} \right)^2 + 2\sigma_{fm} \left( \frac{\partial g(\mu_f, \mu_m)}{\partial f} \right) \left( \frac{\partial g(\mu_f, \mu_m)}{\partial m} \right),$$

where  $\sigma_{fm} = \text{Cov}(f, m)$ .

The required partial derivatives are given by

$$\begin{aligned} \frac{\partial g(\mu_f, \mu_m)}{\partial f} &= \left. \frac{\partial g(f, m)}{\partial f} \right|_{(\mu_f, \mu_m)} = \frac{1}{\mu_m} \\ \frac{\partial g(\mu_f, \mu_m)}{\partial m} &= \left. \frac{\partial g(f, m)}{\partial m} \right|_{(\mu_f, \mu_m)} = -\frac{\mu_f}{\mu_m^2}. \end{aligned}$$

Therefore

$$\text{Var}(\hat{\rho}) = \frac{\sigma_f^2}{\mu_m^2} + \frac{\sigma_m^2 \mu_f^2}{\mu_m^4} - 2 \frac{\sigma_{fm} \mu_f}{\mu_m^3}.$$

This is the same result as given in Cochran (1977, p155), except that the finite population correction factor  $(1 - f')$  has been omitted. That is, we make the assumption that the sampling fraction  $f'$  is so small that the finite population correction is close to unity and

that therefore the population size has minimal effect on the standard error of the estimate of the sex ratio.

Note that  $E[f_1] = E[f_2] = \dots = E[f_i] = \dots = E[f_n]$ , and we use  $E[f_1]$  to denote this quantity. Similarly for  $E[m_1]$ ,  $\text{Var}[f_1]$ ,  $\text{Var}[m_1]$ , and  $\text{Cov}[f_1, m_1]$ . We estimate  $\text{Var}(\hat{\rho})$  by replacing the population moments by the sample moments and observing that because samples between herds are independent

$$\begin{aligned} E(f) &= E\left(\sum_{i=1}^n f_i\right) = nE(f_1) \\ E(m) &= E\left(\sum_{i=1}^n m_i\right) = nE(m_1) \\ \text{Var}(f) &= \text{Var}\left(\sum_{i=1}^n f_i\right) = n\text{Var}(f_1) \\ \text{Var}(m) &= \text{Var}\left(\sum_{i=1}^n m_i\right) = n\text{Var}(m_1) \\ \text{Cov}(f, m) &= \text{Cov}\left(\sum_{i=1}^n f_i, \sum_{i=1}^n m_i\right) = n\text{Cov}(f_1, m_1), \end{aligned}$$

where  $n$  is the number of herds sampled.

The approximate estimate for the variance of  $\hat{\rho}$  is thus given by

$$\widehat{\text{Var}}(\hat{\rho}) = \widehat{\sigma}_{\hat{\rho}}^2 \approx \frac{\sum_{i=1}^n (f_i - \bar{f})^2}{\left(\sum_{i=1}^n m_i\right)^2} + \frac{\sum_{i=1}^n (m_i - \bar{m})^2 \left(\sum_{i=1}^n f_i\right)^2}{\left(\sum_{i=1}^n m_i\right)^4} - 2 \frac{\sum_{i=1}^n (f_i - \bar{f}) \sum_{i=1}^n (m_i - \bar{m}) \sum_{i=1}^n f_i}{\left(\sum_{i=1}^n m_i\right)^3}.$$

For large samples, the limiting distribution of the sex ratio is the normal distribution with mean  $E(\hat{\rho})$  and variance  $\text{Var}(\hat{\rho})$ , assuming an infinite population (Cochran 1977). Approximate confidence intervals for  $\rho$  are constructed using this result even though in our situation we have a finite population. The  $100(1 - 2\alpha)\%$  approximate confidence interval for  $\rho$  is then given by

$$\rho \in (\hat{\rho} \pm \widehat{\sigma}_{\hat{\rho}} z^\alpha).$$

Extending the Taylor expansion of  $\hat{\rho}$  to the second order gives an improved estimate of the expected value of  $\hat{\rho}$  and from this we can obtain a measure for the bias in our estimate

of sex ratios. The Taylor expansion of  $\hat{\rho}$  to the second order is

$$\begin{aligned}\hat{\rho} = g(f, m) &\approx g(\mu_f, \mu_m) + (f - \mu_f) \frac{\partial g(\mu_f, \mu_m)}{\partial f} + (m - \mu_m) \frac{\partial g(\mu_f, \mu_m)}{\partial m} \\ &+ \frac{1}{2}(f - \mu_f)^2 \frac{\partial^2 g(\mu_f, \mu_m)}{\partial f^2} + \frac{1}{2}(m - \mu_m)^2 \frac{\partial^2 g(\mu_f, \mu_m)}{\partial m^2} \\ &+ (f - \mu_f)(m - \mu_m) \frac{\partial^2 g(\mu_f, \mu_m)}{\partial f \partial m}.\end{aligned}$$

Taking expectations on both sides gives an approximate estimate for  $E(\hat{\rho})$ ,

$$E(\hat{\rho}) \approx g(\mu_f, \mu_m) + \frac{1}{2}\sigma_f^2 \frac{\partial^2 g(\mu_f, \mu_m)}{\partial f^2} + \frac{1}{2}\sigma_m^2 \frac{\partial^2 g(\mu_f, \mu_m)}{\partial m^2} + \sigma_{fm} \frac{\partial^2 g(\mu_f, \mu_m)}{\partial f \partial m}$$

that is

$$E(\hat{\rho}) \approx \frac{\mu_f}{\mu_m} + \frac{1}{2}\sigma_f^2 \left. \frac{\partial^2 \hat{\rho}}{\partial f^2} \right|_{(\mu_f, \mu_m)} + \frac{1}{2}\sigma_m^2 \left. \frac{\partial^2 \hat{\rho}}{\partial m^2} \right|_{(\mu_f, \mu_m)} + \sigma_{fm} \left. \frac{\partial^2 \hat{\rho}}{\partial f \partial m} \right|_{(\mu_f, \mu_m)}.$$

The required second partial derivatives are given by

$$\begin{aligned}\left. \frac{\partial^2 \hat{\rho}}{\partial f^2} \right|_{(\mu_f, \mu_m)} &= 0 \\ \left. \frac{\partial^2 \hat{\rho}}{\partial m^2} \right|_{(\mu_f, \mu_m)} &= \frac{2\mu_f}{\mu_m^3} \\ \left. \frac{\partial^2 \hat{\rho}}{\partial f \partial m} \right|_{(\mu_f, \mu_m)} &= \frac{-1}{\mu_m^2}.\end{aligned}$$

Therefore

$$E(\hat{\rho}) \approx \frac{\mu_f}{\mu_m} + \frac{\sigma_m^2 \mu_f}{\mu_m^3} - \frac{\sigma_{fm}}{\mu_m^2}.$$

The term

$$\frac{\sigma_m^2 \mu_f}{\mu_m^3} - \frac{\sigma_{fm}}{\mu_m^2}$$

is the bias of the estimate  $\hat{\rho}$  and a bias corrected ratio  $\hat{\rho}_a$  can be obtained by

$$\hat{\rho}_a \approx \frac{E(\hat{\rho}) + \frac{\sigma_{fm}}{\mu_m^2}}{1 + \frac{\sigma_m^2}{\mu_m^2}}.$$

The bias term can be expressed as

$$\frac{1}{n} \left( \frac{\sigma_{m_i}^2 \mu_{f_i}}{\mu_{m_i}^3} - \frac{\sigma_{f_i m_i}}{\mu_{m_i}^2} \right),$$

where  $E(f_i) = \mu_{f_i}$ ,  $\text{Var}(f_i) = \sigma_{f_i}^2$ ,  $E(m_i) = \mu_{m_i}$ ,  $\text{Var}(m_i) = \sigma_{m_i}^2$  and  $\text{Cov}(f, m) = \sigma_{f, m_i}$ , and so the bias is of order  $1/n$ . An approximate estimate of the adjusted ratio can be obtained by replacing moments with sample moments to obtain

$$\hat{\rho}_a \approx \frac{\frac{f}{m} + \frac{\sum_{i=1}^n (f - \bar{f})(m - \bar{m})}{\left(\sum_{i=1}^n m_i\right)^2}}{1 + \frac{\sum_{i=1}^n (m - \bar{m})^2}{\left(\sum_{i=1}^n m_i\right)^2}}.$$

When the sample size is small, the distribution of  $\hat{\rho}$  is skewed to the right and usually  $\hat{\rho}$  is a biased estimate of  $\rho$ . As the sample size increases, the distribution of  $\hat{\rho}$  approximates the normal distribution and the bias becomes negligible (Cochran 1977 and Rice 1988). When the sample size is small and the distribution of  $\hat{\rho}$  is skewed to the right, improved confidence intervals for  $\rho$  can be obtained by taking the logarithm of the sex ratio, whose distribution approximates the normal distribution even for small samples. That is we consider the random variable  $\gamma = \log(\rho) = \log(F) - \log(M)$ . An estimate for the logarithm of the sex ratio is given by

$$\hat{\gamma} = \log(\hat{\rho}) = \log(f) - \log(m).$$

Denote  $\hat{\gamma} = \ell(f, m) = \log(f/m)$ , then the Taylor expansion of  $\hat{\gamma}$  to the first order is given by

$$\hat{\gamma} = \ell(f, m) \approx \ell(\mu_f, \mu_m) + (f - \mu_f) \frac{\partial \ell(\mu_f, \mu_m)}{\partial f} + (m - \mu_m) \frac{\partial \ell(\mu_f, \mu_m)}{\partial m}.$$

Approximate first moments of  $\hat{\gamma}$  are then given by

$$E(\hat{\gamma}) \approx \ell(\mu_f, \mu_m) = \log\left(\frac{\mu_f}{\mu_m}\right)$$

and

$$\text{Var}(\hat{\gamma}) \approx \sigma_f^2 \left(\frac{\partial \ell(\mu_f, \mu_m)}{\partial f}\right)^2 + \sigma_m^2 \left(\frac{\partial \ell(\mu_f, \mu_m)}{\partial m}\right)^2 + 2\sigma_{f, m} \left(\frac{\partial \ell(\mu_f, \mu_m)}{\partial f}\right) \left(\frac{\partial \ell(\mu_f, \mu_m)}{\partial m}\right).$$

The required partial derivatives are given by

$$\begin{aligned} \frac{\partial \ell(\mu_f, \mu_m)}{\partial f} &= \left. \frac{\partial \ell(f, m)}{\partial f} \right|_{(\mu_f, \mu_m)} = \frac{1}{\mu_f} \\ \frac{\partial \ell(\mu_f, \mu_m)}{\partial m} &= \left. \frac{\partial \ell(f, m)}{\partial m} \right|_{(\mu_f, \mu_m)} = -\frac{1}{\mu_m}. \end{aligned}$$

Therefore

$$\text{Var}(\hat{\gamma}) = \frac{\sigma_f^2}{\mu_f^2} + \frac{\sigma_m^2}{\mu_m^2} - 2 \frac{\sigma_{fm}}{\mu_f \mu_m}.$$

We estimate  $\text{Var}(\hat{\gamma})$  by replacing the population moments by the sample moments so that  $\widehat{\sigma}_{\hat{\gamma}}^2$  is given by

$$\widehat{\text{Var}}(\hat{\gamma}) = \widehat{\sigma}_{\hat{\gamma}}^2 \approx \frac{\sum_{i=1}^n (f_i - \bar{f})^2}{\left(\sum_{i=1}^n f_i\right)^2} + \frac{\sum_{i=1}^n (m_i - \bar{m})^2}{\left(\sum_{i=1}^n m_i\right)^2} - 2 \frac{\sum_{i=1}^n (f_i - \bar{f})(m_i - \bar{m})}{\left(\sum_{i=1}^n f_i\right) \left(\sum_{i=1}^n m_i\right)},$$

where the summation is taken over the number of herds,  $n$ .

The approximate  $100(1 - 2\alpha)\%$  confidence interval for  $\gamma$  is then given by

$$\gamma \in (\hat{\gamma} \pm \widehat{\sigma}_{\hat{\gamma}} z^\alpha),$$

which can be converted to an approximate confidence interval for  $\rho$  by the transformation  $\rho = \exp(\gamma)$ . Thus the  $100(1 - 2\alpha)\%$  confidence interval for  $\rho$  is

$$\rho \in (\exp(\hat{\gamma} \pm \widehat{\sigma}_{\hat{\gamma}} z^\alpha)).$$

The Taylor expansion of  $\hat{\gamma}$  could again be extended to the second order to obtain an improved estimate of the expected value of  $\hat{\gamma}$  and so get a measure of the bias in our estimate of sex ratios. In this study the bias of the estimate  $\hat{\rho}$  was found to be negligible and therefore the bias was not considered in this case.

### Bootstrap estimates for standard error and bias for sex ratios

The use of the bootstrap to find nonparametric estimates of standard error has been discussed in papers by Efron (1979a, 1979b, 1981a, 1981b, 1982), and Efron and Gong (1983). Efron (1981b, 1982, 1985, 1987), Efron and Gong (1983), Hall (1988), and Diccio and Romano (1988) also discuss bootstrap confidence intervals. Efron and Tibshirani (1993) discuss the whole topic of the bootstrap. In this study we use the bootstrap algorithm to find nonparametric estimates of the bias and standard error of the estimate of the sex ratio  $\hat{\rho}$ . We also develop nonparametric bootstrap confidence intervals for the estimate of sex ratios. Monte Carlo simulation is used to find nonparametric estimates of the standard error and bias of  $\hat{\rho}$  and to construct confidence intervals.

The bootstrap of Efron (1979a, 1981b) to estimate the standard error and the bias of the sex ratio  $\hat{\rho}$  can be described as follows:

1. Suppose that the number of females and males in a given herd  $(f_i, m_i)$ ,  $i = 1, 2, \dots, n$  are independent observations from some bivariate distribution  $F$ . We do not know  $F$ , but we can estimate it by  $\hat{F}$ , where  $\hat{F}$  is the empirical distribution of the  $n$  observed data points  $(f_i, m_i)$ ,  $i = 1, 2, \dots, n$ . That is,  $\hat{F}$  is the probability distribution that assigns the mass  $1/n$  at each observed point  $(f_i, m_i)$ .
2. Obtain the bootstrap sample  $(f_i^*, m_i^*)$  by drawing a random sample of size  $n$  from  $\hat{F}$ , that is, using a random number generator to draw  $n$  new points independently and with replacement from  $\hat{F}$  so that each new point is an independent random selection of one of the  $n$  original data observations.
3. Calculate the bootstrap replication  $\hat{\rho}^* = \hat{\rho}((f_1^*, m_1^*), \dots, (f_n^*, m_n^*))$ , the sex ratio of the bootstrap sample.
4. Repeat steps 2 and 3  $B$  times (where  $B$  is in the range of 25–200 for the estimation of the standard error (Efron 1987, Efron and Tibshirani 1993)), obtaining  $B$  independent realizations of  $\hat{\rho}^*$ , say  $\hat{\rho}^*(b)$ ,  $b = 1, 2, \dots, B$ .
5. Denote the bootstrap estimate of the true standard error of the sex ratio  $\hat{\rho}$  by  $\widehat{\sigma}_B(\hat{\rho})$ , where

$$\widehat{\sigma}_B(\hat{\rho}) = \left[ \text{Var}_*(\hat{\rho}^*) \right]^{1/2}, \quad (1)$$

where  $\text{Var}_*$  indicates variance under the sampling scheme. That is,

$$\widehat{\sigma}_B(\hat{\rho}) \approx \left[ \frac{\sum_{b=1}^B (\hat{\rho}^*(b) - \hat{\rho}^*(\cdot))^2}{B-1} \right]^{1/2}, \quad (2)$$

where  $\hat{\rho}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\rho}^*(b)}{B}$ . As  $B \rightarrow \infty$ , (2) approaches the original definition (1).

6. Denote the bias of  $\hat{\rho}$  by  $\beta = E(\hat{\rho} - \rho)$ . Then the bootstrap estimate of bias is given by

$$\widehat{\beta}_B = E_*(\hat{\rho}^* - \hat{\rho}),$$

where  $E_*$  indicates expectation with respect to the bootstrap sampling scheme. That is,  $\widehat{\beta}_B$  is approximated by

$$\frac{1}{B} \sum_{b=1}^B (\hat{\rho}^*(b) - \hat{\rho}).$$

**Bootstrap nonparametric confidence intervals for ratios**

Efron (1981b, 1982) proposed approximate confidence intervals based on the bootstrap algorithm, mainly the percentile and the bias-corrected (BC) percentile bootstrap confidence intervals. Schenker (1985) commented on the shortcomings of these bootstrap confidence intervals under certain conditions. Efron (1987) proposed an improved bootstrap confidence interval called the  $BC_a$  confidence interval.

**The percentile method interval**

The problem considered here is the case where we would like to obtain some measure of the reliability of some estimator ( $\hat{\rho}$  in this study) which is usually given in the form of a confidence interval. Bootstrap percentile confidence intervals can be described by the following algorithm:

1. We do not know the distribution of  $(f_i, m_i)$ ,  $i = 1, \dots, n$ ,  $F$  but we approximate it by  $\hat{F}$  as before, from which we can obtain the bootstrap sample  $(f_i^*, m_i^*)$ .
2. The bootstrap replication  $\hat{\rho}^*$  is computed from the bootstrap sample and the previous step and this one are repeated  $B$  times (where  $B$  is in the range of 1000 for the construction of confidence intervals (Efron 1987, Efron and Tibshirani 1993)) to obtain  $B$  independent realizations  $\hat{\rho}^*(b)$ ,  $b = 1, 2, \dots, B$ .
3. The bootstrap cumulative distribution function of  $\hat{\rho}^*$  is given by

$$\hat{G}(\hat{\rho}) \equiv \text{Pr}_* \{ \hat{\rho}^*(b) < \hat{\rho} \},$$

where  $\text{Pr}_*$  indicates probabilities calculated according to the bootstrap sampling scheme. That is

$$\hat{G}(\hat{\rho}) = \frac{\# \{ \hat{\rho}^*(b) < \hat{\rho} \}}{B},$$

where  $\#$  denotes the term "number of".

4. Calculate the  $\alpha$  and  $1 - \alpha$  percentiles of the bootstrap distribution, i.e.  $\hat{G}^{-1}(\alpha)$  and  $\hat{G}^{-1}(1 - \alpha)$ .
5. The approximate central  $100(1 - 2\alpha)\%$  confidence interval for  $\rho$  is given by

$$\rho \in [\hat{G}^{-1}(\alpha), \hat{G}^{-1}(1 - \alpha)].$$

The percentile confidence interval will be close to the standard confidence interval if the distribution of  $\widehat{\rho}^*$  is close to the normal distribution or if the sample size is large. For small sample sizes, the distribution of  $\widehat{\rho}^*$  might be far from normal in which case the percentile interval and the standard intervals will differ. The percentile confidence intervals, although undercovering the tails of the distribution of  $\widehat{\rho}^*$ , achieve a better balance of both tails than the standard intervals (Efron and Tibshirani 1993).

### The bias-corrected (BC) percentile method interval

The BC interval introduced by Efron (1981b, 1982) is an approach to improve on the coverage of the percentile interval by adjusting for bias. The assumption is made that there exists a monotone increasing function  $g$  so that

$$\frac{g(\widehat{\rho}) - g(\rho)}{\tau} \sim N(-z_0, 1)$$

and

$$\frac{g(\widehat{\rho}^*) - g(\widehat{\rho})}{\tau} \sim N(-z_0, 1)$$

under the bootstrap sampling distribution, where  $\tau$  is the constant standard error of  $g(\widehat{\rho})$  and  $z_0$  is the bias constant which measures the median bias of  $\widehat{\rho}^*$ . Taking the inverse transformation, confidence intervals can be obtained for  $\rho$ . The advantage of the BC method is that  $g$  need not be known.

The algorithm to obtain nonparametric bootstrap confidence intervals by the bias-corrected method is:

1. The first two steps of the percentile method interval remain the same.
2. Calculate the bias-correction  $z_0$ , and the quantities  $\alpha_L$  and  $\alpha_U$ , where

$$\begin{aligned} z_0 &= \Phi^{-1}(\widehat{G}(\widehat{\rho})) = \Phi^{-1}\left(\frac{\#\{\widehat{\rho}^*(b) < \widehat{\rho}\}}{B}\right) \\ \alpha_L &= \Phi(2z_0 + z^{(\alpha)}) \\ \alpha_U &= \Phi(2z_0 + z^{(1-\alpha)}), \end{aligned}$$

where  $\Phi(x)$  is the standard normal cumulative function and  $\Phi(z^\alpha) \equiv \alpha$ .

3. The central  $100(1 - 2\alpha)\%$  confidence interval for  $\rho$  is given by

$$\rho \in [\widehat{G}^{-1}(\alpha_L), \widehat{G}^{-1}(\alpha_U)].$$

When  $\widehat{G}(\hat{\rho}) = 0.5$ ,  $z_0 = 0$  and the bias-corrected method reduces to the percentile method interval. Both the percentile and the BC confidence intervals differ from the exact interval by order  $O(n^{-1})$ .

### The accelerated bias-corrected percentile bootstrap ( $BC_a$ ) confidence interval

Schenker (1985) demonstrated that the assumptions under the bias-corrected method interval are inapplicable in certain instances and an improvement to the construction of nonparametric confidence intervals called the  $BC_a$  method interval was proposed by Efron (1987). The previous bootstrap confidence intervals assume that there exist a function  $g$  that normalizes the distribution of  $\hat{\rho}$  and that the standard error of  $\hat{\rho}$  is constant for all  $\rho$ , i.e.  $g$  is variance stabilizing. The assumption of a constant standard error of  $\hat{\rho}$  is not always applicable and the  $BC_a$  confidence interval corrects for this by the acceleration constant which measures the rate of change of the standard error of  $\hat{\rho}$  with respect to  $\rho$  (Efron 1987, Efron and Tibshirani 1993).

The assumption now is that for some bias constant  $z_0$  and some acceleration constant  $a$ , there exists a function  $g$  such that

$$\frac{g(\hat{\rho}) - g(\rho)}{\tau} \sim N(-z_0\omega, \sigma^2),$$

where

$$\omega = 1 + ag(\rho).$$

Applying the inverse transformation, we can obtain confidence intervals for  $\rho$ .

When the acceleration constant  $a$  is equal to zero, the  $BC_a$  method interval reduces to the BC method interval. When both the bias ( $z_0$ ) and the acceleration ( $a$ ) constants are zero, the  $BC_a$  method intervals reduce to the percentile method intervals. One advantage of both the BC and the  $BC_a$  intervals is that no knowledge of the function  $g$  is necessary because it is automatically incorporated into the algorithm.

To calculate  $BC_a$  method confidence intervals the following algorithm is used:

1. Perform the same first step as for a BC method confidence interval.
2. Calculate the bias-correction  $z_0$ , and the quantities  $\alpha_L$  and  $\alpha_U$ , where

$$z_0 = \Phi^{-1}(\widehat{G}(\hat{\rho})) = \Phi^{-1}\left(\frac{\#\{\widehat{\rho}^{\wedge}(b) < \hat{\rho}\}}{B}\right)$$

$$\alpha_L = \Phi\left(z_0 + \frac{z_0 + z^{(\alpha)}}{1 - a(z_0 + z^{(\alpha)})}\right)$$

$$\alpha_U = \Phi \left( z_0 + \frac{z_0 + z^{(1-\alpha)}}{1 - a(z_0 + z^{(1-\alpha)})} \right),$$

where  $\Phi(x)$  is the standard normal cumulative function,  $\Phi(z^\alpha) \equiv \alpha$  and the acceleration constant  $a$  is approximated by

$$a \simeq \frac{1}{6} \left[ \frac{\sum_{i=1}^n U_i^3}{\left( \sum_{i=1}^n U_i^2 \right)^{1.5}} \right],$$

where  $U_i$  is the derivative of the estimate  $\hat{\rho}$  with respect to the mass on point  $(f_i, m_i)$ .  $U_i$  is called the empirical influence function of  $\hat{\rho} = t(\hat{F}) = \bar{f}/\bar{m}$ . So to obtain  $U_i$  we need to perturb  $\hat{F}$  by some amount say  $\delta$  at the point  $(f_i, m_i)$ . Denote the perturbed probability distribution function by  $\widehat{F}_\delta$ . To do this we add  $\delta$  to  $\hat{f}$  at the point  $(f_i, m_i)$ , where  $\hat{f}$  is the probability density function corresponding to  $\hat{F}$ . To ensure that the perturbed probability density function still adds to one, it is necessary to scale down the original values of  $\hat{f}$  by a factor of  $(1 - \delta)$ . That is, the perturbed probability density function,  $\widehat{f}_\delta$ , at point  $(f_i, m_i)$  is given by

$$\widehat{f}_\delta = \begin{cases} (1 - \delta)1/n & \text{for points } (f_j, m_j), \quad i \neq j \\ (1 - \delta)1/n + \delta & \text{for point } (f_i, m_i) \\ 0 & \text{otherwise} \end{cases}.$$

Therefore the function  $t(\widehat{F}_\delta)$  at point  $(f_i, m_i)$  is given by

$$\begin{aligned} t(\widehat{F}_\delta) &= \frac{(1 - \delta)1/n \sum_{i=1}^n f_i + \delta f_i}{(1 - \delta)1/n \sum_{i=1}^n m_i + \delta m_i} \\ &= \frac{(1 - \delta)\bar{f} + \delta f_i}{(1 - \delta)\bar{m} + \delta m_i}. \end{aligned}$$

Therefore  $U_i$  is given by

$$\begin{aligned} U_i &= \lim_{\delta \rightarrow 0} \frac{t(\widehat{F}_\delta) - t(\hat{F})}{\delta}, \quad i = 1, 2, \dots, n. \\ &= \lim_{\delta \rightarrow 0} \left( \frac{\frac{(1 - \delta)\bar{f} + \delta f(i)}{(1 - \delta)\bar{m} + \delta m(i)} - \hat{\rho}}{\delta} \right), \quad i = 1, 2, \dots, n. \end{aligned}$$

3. The central  $100(1 - 2\alpha)\%$  confidence interval for  $\rho$  is given by

$$\rho \in \left[ \hat{G}^{-1}(\alpha_L), \hat{G}^{-1}(\alpha_U) \right].$$

The accelerated bias-corrected confidence intervals are accurate to the second order although the coverage can still be erratic when dealing with small sample sizes (Efron and Tibshirani 1993).

## RESULTS AND DISCUSSION

Estimates for the proportion of females and for the ratio of females to males for Springbok are shown in Table 2. For the areas Benfontein and Susanna, the lowest ratio of females to males occurred on the last sampling date, i.e. in 1986, while the highest ratio of females to males occurred on the second sampling, i.e. end of 1985. For Benfontein this represented nearly double the ratio of females to males in September of 1985 to that of February of 1986. For Susanna it represented nearly three times the ratio of females to males in October 1985 to that of February 1986. For Uitzicht there was no sampling for the end of 1985. The sex ratios for January 1985 and February 1986 are similar as are the ones for Benfontein and Susanna for the corresponding dates.

**Table 2.** Estimates for the proportion of females ( $\pi$ ) and the sex ratio ( $\rho$ ) of Springbok together with 90% binomial confidence intervals ( $\rho_N$  indicates that the normal approximation to the binomial distribution has been used to construct the CI, while  $\rho_P$  indicates that Pratt's (1968) approximation to the binomial distribution has been used).

Area	Date	$\hat{\pi}$	$\hat{\rho}$	Binomial CI		
				$\pi$	$\rho_N$	$\rho_P$
Benfontein	02/85	0.77	3.3	(0.73, 0.80)	(2.8, 4.1)	(2.7, 4.3)
	09/85	0.81	4.1	(0.77, 0.84)	(3.3, 5.3)	(3.3, 5.3)
	02/86	0.68	2.1	(0.64, 0.72)	(1.8, 2.6)	(1.7, 2.6)
Susanna	01/85	0.71	2.5	(0.66, 0.76)	(1.9, 3.2)	(1.9, 3.3)
	10/85	0.86	6.1	(0.81, 0.91)	(4.2, 10.1)	(3.9, 9.9)
	02/86	0.68	2.2	(0.65, 0.72)	(1.8, 2.6)	(1.8, 2.6)
Uitzicht	01/85	0.73	2.8	(0.67, 0.80)	(2.1, 3.9)	(2.0, 3.9)
	02/86	0.71	2.5	(0.62, 0.81)	(1.6, 4.4)	(1.5, 4.4)

Binomial 90% confidence intervals are given for the proportion of females and for the ratio of females to males (Table 2). The limits are shown for the normal approximation to the binomial distribution as well as Pratt's (1968) approximation. In this study, the

confidence intervals obtained by the usual normal approximation and that of Pratt differ only slightly, if at all. The binomial confidence intervals for the sex ratio of Springbok can be seen to be very narrow, overestimating the lower limits and underestimating the upper limits. An exception is found for the sample from Susanna made towards the end of 1985, in which case, the binomial confidence interval is able to estimate the lower bound but the upper bound is overestimated in comparison with the other confidence limits.

Confidence intervals for the sex ratio of Springbok using the Taylor expansion of the ratio to estimate the standard error of the ratio estimate are less optimistic than the confidence intervals obtained using the binomial distribution (Table 3). In this instance we are assuming that the distribution of the sex ratio is approximately normal. For small samples this approximation might be poor and the positive skewness of the ratio estimate can become prominent. In this study, because we are now considering the unit of sampling to be the herd and not the individuals, the sample sizes are small in all of the samples. Therefore the distribution of the sex ratio of Springbok could have a prominent positive skewness. Confidence intervals obtained by the Taylor expansion procedure do not take this into account as they are symmetrical (Table 3). Because the distribution of the sex ratio is skewed to the right, the distribution of the logarithm of the sex ratio matches that of the normal distribution more closely. The confidence intervals obtained for the sex ratio of Springbok from the Taylor expansion of the logged ratio are better at reflecting the positive skewness of the sex ratio of Springbok than those obtained from the Taylor expansion of the ratio in that they are positively skewed (Table 3).

One of the advantages of the Taylor expansion approach to constructing confidence intervals for sex ratios is that a measure of bias can be easily obtained. Bias adjusted sex ratio estimates for Springbok are shown in Table 3. The bias adjusted ratio estimates closely resemble the unadjusted ones (there is a slight positive bias) except for the sample taken from Uitzicht in February 1986. The results suggest that the sex ratio for this sample has been slightly overestimated.

Nonparametric bootstrap confidence intervals for the sex ratio of Springbok have been computed for the percentile, the bias-corrected and the accelerated bias-corrected methods (Table 4). Monte Carlo simulations have been performed to obtain the cumulative distribution function of the bootstrap distribution of the ratio. According to Efron (1987, Efron and Tibshirani 1993) about 1000 Monte Carlo replications are needed to construct bootstrap confidence intervals and this is the amount of replications used in this study to construct

**Table 3.** Confidence intervals (90%) for the sex ratio ( $\rho$ ) and for the log of the sex ratio ( $\log(\rho)$ ) of Springbok based on the Taylor expansions. The estimated bias adjusted ratio  $\widehat{\rho}_a$  is also given.

Area	Date	$\hat{\rho}$	$\widehat{\rho}_a$	Taylor series CI	
				$\rho$	$\log(\rho)$
Benfontein	02/85	3.3	3.2	(2.3, 4.3)	(2.5, 4.5)
	09/85	4.1	3.9	(2.0, 6.3)	(2.5, 6.9)
	02/86	2.1	2.0	(0.9, 3.3)	(1.2, 3.7)
Susanna	01/85	2.5	2.3	(0.8, 4.1)	(1.2, 4.9)
	10/85	6.1	6.1	(4.2, 8.0)	(4.5, 8.3)
	02/86	2.2	2.0	(0.9, 3.5)	(1.2, 4.0)
Uitzicht	01/85	2.8	2.5	(1.3, 4.2)	(1.6, 4.7)
	02/86	2.5	1.6	(-1.0, 6.0)	(0.6, 10.1)

90% bootstrap confidence intervals for the sex ratio of Springbok. The bias-corrected and the accelerated bias-corrected methods produced similar confidence intervals with the exception of the sample from Uitzicht in February 1986 in which the bias-corrected method gave a confidence interval with a larger upper limit. The percentile method gives lower limits that are mostly similar to the BC and  $BC_a$  methods. In the cases that it differs, it tends to have a slightly smaller lower limit as compared to the other methods except for the sample from Susanna in October 1985. In this case, the lower limit is slightly biased to the left. The upper limits of the percentile method are also similar to that of the other methods with a slight tendency to overestimate the upper bound. An exception occurs for the sample from Susanna obtained in October 1985, where the direction of the difference is the reverse from that of other samples and the upper limit in this case is slightly underestimated.

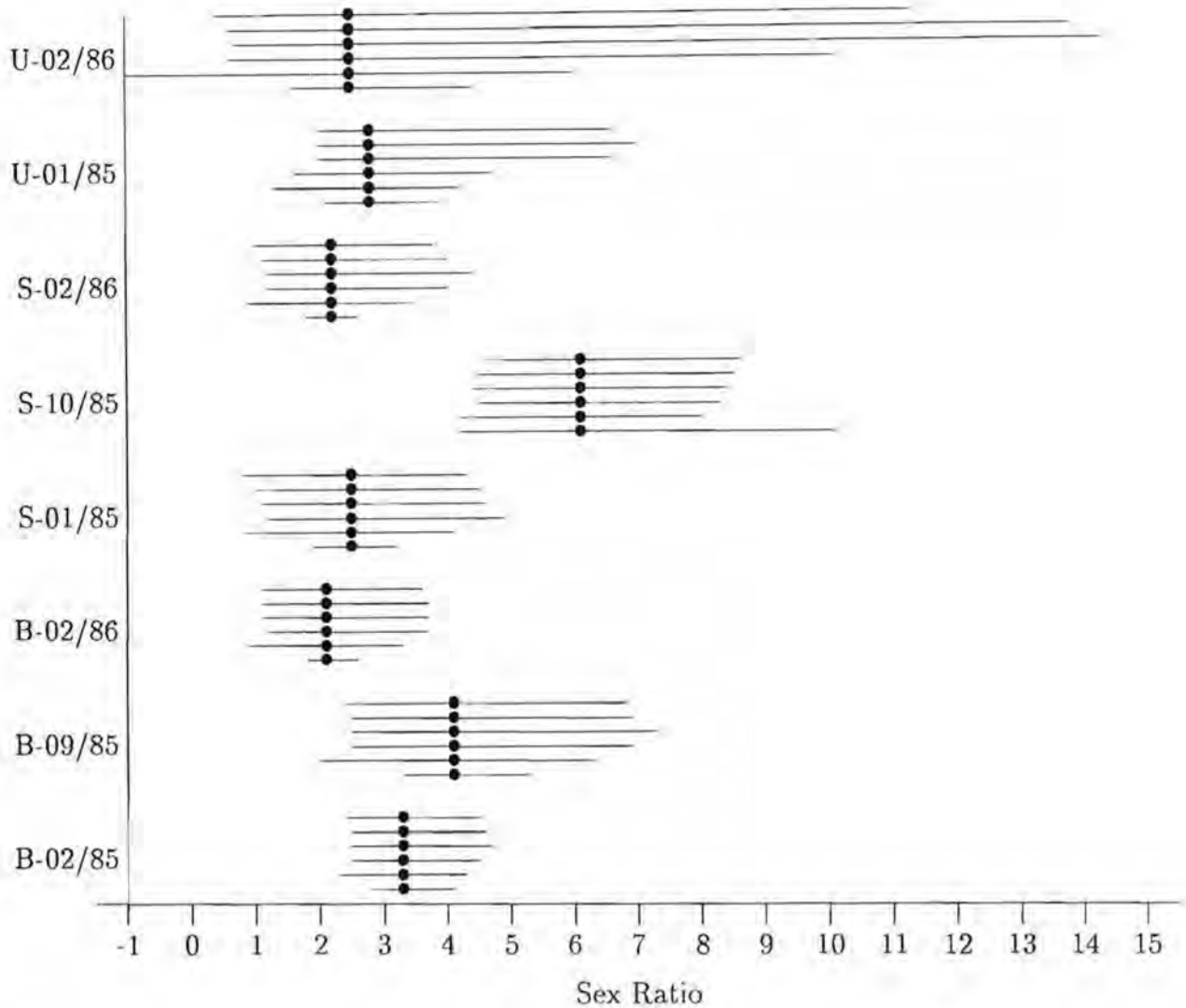
Comparing the confidence intervals obtained from the Taylor expansion of the logarithm of the ratio to those obtained from the bootstrap BC and  $BC_a$  methods, we can see that these intervals are similar to each other (Tables 3 and 4 and Fig. 1). Uitzicht is an exception

in that the upper limits of the bootstrap confidence intervals are slightly overestimated as compared to the Taylor expansion of the logged ratio.

**Table 4.** Bootstrap estimates for the standard error ( $\sigma_B$ ), bias ( $\beta_B$ ) and bootstrap 90% confidence intervals (percentile, BC and  $BC_a$  methods) for the sex ratio ( $\rho$ ) of Springbok. The bias correction ( $z_o$ ) and acceleration correction ( $a$ ) values are also shown.

Area	Date	$\widehat{\sigma}_B$	$\widehat{\beta}_B$	$\hat{\rho}$	Bootstrap CI		
					percentile	BC	$BC_a$
Benfontein	02/85	0.66	0.13	3.3	(2.5, 4.7)	(2.5, 4.6)	(2.4, 4.5)
						$z_o = -0.021$	$a = -0.038$
	09/85	1.57	0.36	4.1	(2.5, 7.3)	(2.5, 6.9)	(2.4, 6.8)
						$z_o = -0.059$	$a = -0.021$
	02/86	0.83	0.14	2.1	(1.1, 3.7)	(1.1, 3.7)	(1.1, 3.6)
						$z_o = -0.004$	$a = -0.023$
Susanna	01/85	1.12	0.14	2.5	(1.1, 4.6)	(1.0, 4.5)	(0.8, 4.3)
						$z_o = -0.051$	$a = -0.076$
	10/85	1.18	0.04	6.1	(4.4, 8.4)	(4.5, 8.5)	(4.6, 8.6)
						$z_o = 0.049$	$a = 0.043$
	02/86	1.02	0.25	2.2	(1.2, 4.4)	(1.1, 4.0)	(1.0, 3.8)
						$z_o = -0.101$	$a = -0.074$
Uitzicht	01/85	1.48	0.42	2.8	(2.0, 6.6)	(2.0, 7.0)	(2.0, 6.6)
						$z_o = 0.149$	$a = -0.030$
	02/86	4.69	2.49	2.5	(0.7, 14.3)	(0.6, 13.8)	(0.4, 11.3)
						$z_o = -0.061$	$a = -0.102$

The bootstrap procedure can also be used to obtain an estimate of the bias of the ratio estimate (Table 4). The bootstrap estimates for the bias of  $\hat{\rho}$  also show a slight upward bias in the sex ratio estimate. For the sample from Uitzicht in February 1986, the bias is very large, as large as the estimate of the ratio itself. We need to keep in mind that there were only four herds observed for this site. Therefore the large bias and other discrepancies



**Figure 1.** Comparison of the 90% confidence intervals for the sex ratio of Springbok. The confidence intervals are stacked for each site and each date that observations were made in the following order from bottom to top: Binomial CI (normal approximation), Taylor expansion of ratio CI, Taylor expansion of the logged ratio CI, bootstrap percentile CI, bootstrap BC CI and  $BC_{\alpha}$  bootstrap CI. The following abbreviations for the site names have been used in the figure: B — Benfontein, S — Susanna and U — Uitzicht.

observed for this sample might be due to the small sample size rather than to the techniques themselves.

## CONCLUSION

Several confidence intervals for the sex ratio of Springbok have been constructed in this study. They vary from confidence intervals that consider the unit of sampling to be the individual Springbok observed to other procedures that take into account that the sampling unit is the herd, where the composition of males and females in each herd differs, as does the size of a herd.

The binomial confidence intervals for the sex ratio of Springbok are in general too narrow, which suggests that we cannot ignore the fact that the sampling unit in this type of problem is a herd and not individuals.

Confidence intervals obtained by Taylor expansion of the ratio are less optimistic, and in this study, it is better to take logarithms of the sex ratios to overcome the positive skewness of the distribution of the ratio.

Bootstrap confidence intervals for the BC and  $BC_a$  methods were similar and they were also similar to the confidence intervals obtained from the Taylor expansion of the logged ratio. The results for the samples from Uitzicht deviated slightly from those of the samples from other areas. The  $BC_a$  method underestimated the upper bounds compared to both the BC and the percentile method. Also the bootstrap confidence intervals were different from the confidence intervals constructed from the Taylor expansion of the logged ratio. For both samples obtained from the game farm Uitzicht, the confidence intervals from the expansion of the logged ratio underestimated the upper bound as compared to the bootstrap confidence intervals. The number of herds observed at this site on both occasions was small (four and five herds respectively) but the total numbers counted differed by a factor of two. These might account for the differences observed between the different methods of constructing confidence intervals.

## REFERENCES

- BLYTH, C.R. 1986. Approximate binomial confidence limits. *Journal of the American Statistical Association* 81: 843–855.
- BOWDEN, D.C., ANDERSON, A.E. AND MEDIN, D.E. 1984. Sampling plans for mule deer sex and age ratios. *Journal of Wildlife Management* 48: 500–509.
- CLUTTON-BROCK, T.H., ALBON, S.D. AND GUINNESS, F.E. 1984. Maternal dominance, breeding success, and birth sex ratios in red deer. *Nature* 308: 358–360.
- CLUTTON-BROCK T.H. AND IASON, G.R. 1986. Sex ratio variation in mammals. *The Quarterly Review of Biology* 61: 339–374.
- COCHRAN, W.G. 1977. *Sampling techniques*, 3rd ed. Wiley, New York.
- DICICCIO, T.J. AND ROMANO, J.P. 1988. A review of bootstrap confidence intervals (with discussion). *Journal of the Royal Statistical Society, B* 50: 338–370.
- EFRON, B. 1979a. Bootstrap methods: another look at the jackknife. *The Annals of Statistics* 7: 1–26.
- EFRON, B. 1979b. Computers and the theory of statistics: Thinking the unthinkable. *SIAM Review* 21: 460–480.
- EFRON, B. 1981a. Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. *Biometrika* 68: 589–599.
- EFRON, B. 1981b. Nonparametric standard errors and confidence intervals (with discussion). *Canadian Journal of Statistics* 9: 139–172.
- EFRON, B. 1982. The jackknife, the bootstrap and other resampling plans. National Science Foundation — Conference Board of the Mathematical Sciences, Monograph 38. Society for Industrial and Applied Mathematics. Philadelphia, Pennsylvania.
- EFRON, B. 1985. Bootstrap confidence intervals for a class of parametric problems. *Biometrika* 72: 45–58.
- EFRON, B. 1987. Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82: 171–185.

- EFRON, B. AND GONG, G. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician* 37: 36–48.
- EFRON, B. AND TIBSHIRANI, R.J. 1993. *An introduction to the bootstrap*. Chapman and Hall, London.
- FISHER, R.A. 1930. *The genetical theory of natural selection*. Oxford University Press, Oxford.
- HALL, P. 1988. On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society, B* 50: 35–45.
- JAMES, A.T., WILKINSON, G.N. AND VENABLES, W.N. 1974. Interval estimates for a ratio of means. *Sankhyā: The Indian Journal of Statistics* 36A: 177–183.
- MAYNARD SMITH, J. 1978. *The evolution of sex*. Cambridge University Press, Cambridge.
- MENTIS, M.T. 1972. A review of the life history features of the larger herbivores of Africa. *Lammergeyer* 16:1–89.
- PRATT, J.W. 1968. A normal approximation for binomial,  $F$ , beta, and other common, related tail probabilities, II. *Journal of the American Statistical Association* 63, 1457–1483.
- RICE, J.A. 1988. *Mathematical Statistics and data analysis*. Brookes and Cole, California.
- SCHENKER, N. 1985. Qualms about bootstrap confidence intervals. *Journal of the American Statistical Association* 80: 360–361.
- TRIVERS, R.L. AND WILLARD, D.E. 1973. Natural selection of parental ability to vary sex ratio of offspring. *Science* 179: 90–92.

# A model for the relationship between reproductive mass and vegetative mass in *Solidago altissima*

## ABSTRACT

Plant reproductive mass is an example of data censored from the left; it is never negative and becomes positive only once a certain vegetative mass threshold has been reached. These observations follow the pattern of primary moult indices, except that there is no censoring on the right. In this study we adapt the Underhill-Zucchini moult model (*Ibis* (1988) 130: 358–372 and (1990) 132: 118–123) (effectively a two-sided censored regression model) to handle observations that are only censored from the left. The data of Schmid *et al.* (*The American Naturalist* (1994) 143: 494–507) for *Solidago altissima* L. (tall goldenrod) are used to illustrate the method of maximum likelihood to describe the relationship between reproductive mass and vegetative mass. The approach taken by Underhill and Zucchini adapted to formulate a one-sided censored regression model for the relationship between vegetative mass and reproductive mass problem is shown to be equivalent to the approach taken by Schmid *et al.* (1994). The model parameters from the one approach are transformations of the parameters from the other. Advantages and disadvantages of both algorithms are considered.

## INTRODUCTION

Applications of censored regression models in biology were reviewed by Schmid *et al.* (1994) who discussed mainly botanical examples. In particular, they considered a situation from plant reproductive biology in which the reproductive mass is zero until the vegetative mass reaches a threshold. Their goal was to estimate the mean threshold mass and to find the relationship between reproductive mass and vegetative mass above the threshold. This and their other examples were all of “one-sided” censoring, specifically of left-censored data. The moult problem analysed by Underhill and Zucchini (1988) and Underhill *et al.* (1990)

is an example of a censored regression model with two-sided censoring (i.e. of data censored from the left and the right) because the moult index is constrained to lie between 0 and 1. In the moult problem, the “thresholds” are the mean dates at which moult commences and finishes.

Schmid *et al.* (1994) derived the likelihood function for one-sided censoring, but did not attempt to maximize the likelihood function using differentiation and the Newton-Raphson algorithm. Instead, they made use of the EM (expectation-maximization) algorithm (Dempster *et al.* 1977) to maximize the likelihood function. This paper considers the problem of plant reproduction introduced by Schmid *et al.* (1994) and it develops the method of maximum likelihood algorithm for this problem. The solutions obtained by the EM and maximum likelihood algorithms are compared.

## MATERIAL AND METHODS

### Data description

The data set of Schmid *et al.* (1994) was used to motivate a model for biological data with one-sided censoring in a similar way to that of Underhill and Zucchini (1988) and Underhill *et al.* (1990) for two-sided censoring. The data consist of measurements of reproductive mass (inflorescences) at a given vegetative mass (stem and leaves) of individually grown shoots in the clonal plant *Solidago altissima* L. (tall goldenrod). These plants were grown in an experimental garden of the University of Basel, Switzerland. A sample of 433 plants was randomly selected from 2545 grown plants. Of these, 123 had a reproductive mass of zero because they had not yet reached reproductive maturity, while for the remaining 310 plants both reproductive mass and a positive vegetative mass were observed.

### A model for plant reproductive mass

The formulation of the model for plant reproduction follows the same general procedure as used by Underhill and Zucchini (1988), but adapting the notation and the model to accommodate the present problem.

Let the random variable  $V$  be the vegetative mass at which the threshold is reached for reproduction mass to be observed for a randomly selected plant. Let  $F_V(v)$  be the probability distribution function and  $f_V(v)$  the probability density function of  $V$ . We assume a linear relationship between reproductive mass and vegetative mass thereafter.

We will suppose that  $F_V(v)$  has  $p$  parameters  $\theta_1, \theta_2, \dots, \theta_p$  and for convenience the slope of the line  $\beta$  will also be considered as a parameter of  $F_V(v)$ .

The main quantities of interest are the expectation,  $\mu$ , and the standard deviation,  $\sigma$ , of  $V$ , which give the mean and the spread of vegetative mass at which plants begin to show reproductive mass in the population. We are also interested in the slope of the line,  $\beta$ , which gives the rate of change of reproductive mass with respect to vegetative mass.

The data consist of observations on reproductive mass of random samples of plants at some given vegetative mass. Let  $R|v$  be a random variable giving reproductive mass  $R$  at vegetative mass  $v$  of a plant selected at random from the population. We note that  $R|v$  is a random variable whose probability distribution function  $F_{R|v}(r)$  is dependent on vegetative mass  $v$ .

Our objective is to estimate  $\beta$  and  $F_V(v)$  from observations on the process  $R|v$ . This situation is unusual in that we wish to estimate the distribution of a random variable  $V$  for which no direct observations are available. We will show that we can find a relationship that will make this possible. Suppose that the linear relationship between vegetative mass and reproductive mass is given by  $R|v = \alpha + \beta v$ , and we denote the standard deviation of  $R|v$  by  $\sigma_\epsilon$ . Then  $R|v \leq r$  if and only if  $v \leq r/\beta - \alpha/\beta$ . But  $-\alpha/\beta$  is the  $x$ -intercept which is the random variable  $V$ . Therefore,  $R|v \leq r$  if and only if  $V \geq v - r/\beta$ . Denote  $\lambda = 1/\beta$ , then  $R|v \leq r$  if and only if  $V \geq v - r\lambda$ . Therefore the relationship that provides the basis to estimate  $\beta$  and  $F_V(v)$  from observations on the process  $R|v$  is:

$$R|v \leq r \quad \text{if and only if} \quad V \geq v - r\lambda, \quad r \geq 0.$$

From this it follows that

$$F_{R|v}(r) = 1 - F_V(v - r\lambda), \quad r \geq 0,$$

and

$$f_{R|v}(r) = \lambda f_V(v - r\lambda), \quad r > 0.$$

Maximum likelihood estimates of the parameters of interest, i.e.  $\mu$ ,  $\sigma$ , and  $\lambda$  can be obtained from the appropriate likelihood function. These parameters describe the probability density function of the random variable  $V$ . Schmid *et al.* (1994) looked at the same problem but fitted a model to the random variable  $R|v$ . Because we are assuming a linear relationship between vegetative mass and reproductive mass, there is a one-one relationship

between the parameters of both models. We have already seen that  $\lambda = 1/\beta$  and that  $-\alpha/\beta$  is the  $x$ -intercept which is the mean of  $V$ , that is  $\mu = -\alpha/\beta$ . Also the variance of  $R|v$  is given by  $\beta^2\sigma^2$  because of the linear relationship between  $R|v$  and  $V$ . Because of the invariance property of maximum likelihood estimates (Mood *et al.* 1963), an estimate for the slope ( $\beta$ ) of the line is then simply given by  $\hat{\beta} = 1/\hat{\lambda}$ , the  $y$ -intercept ( $\alpha$ ) is given by  $\hat{\alpha} = -\hat{\mu}/\hat{\lambda}$  and the standard deviation of  $R|v$  ( $\sigma_\epsilon$ ) is given by  $\hat{\sigma}_\epsilon = \hat{\sigma}/\hat{\lambda}$ .

Suppose that the sample consists of  $I$  plants which show no reproductive mass and  $J$  which show reproductive mass. Let the vegetative mass for these observations be  $t = (t_1, t_2, \dots, t_I)$  and  $u = (u_1, u_2, \dots, u_J)$ , respectively, and let  $r = (r_1, r_2, \dots, r_J)$  be the reproductive mass of the  $J$  plants for which strictly positive reproductive mass was observed.

The likelihood function (Cox and Hinkley 1974) of the observations is given by

$$L(\theta; t, u, r) = \prod_{i=1}^I P(t_i) \prod_{j=1}^J q(u_j, r_j), \quad (1)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. The probability functions  $P(t)$  and  $q(u, r)$  are defined as

$$\begin{aligned} P(t) &= \Pr\{R|t = 0\} = 1 - F_T(t), \\ q(u, r) &= \lambda f_U(u - r\lambda). \end{aligned}$$

The maximum likelihood estimates of the parameters are those values of  $\theta$  which maximize (1), or equivalently its logarithm  $\ell_1 = \log L(\theta; t, u, r)$

$$\ell_1 = \sum_{i=1}^I \log(P(t_i)) + \sum_{j=1}^J \log(q(u_j, r_j)).$$

The normal equations are given by setting the partial derivatives equal to zero;

$$\frac{\partial \ell_1}{\partial \theta_a} = \sum_{i=1}^I \frac{1}{P(t_i)} \frac{\partial P(t_i)}{\partial \theta_a} + \sum_{j=1}^J \frac{1}{q(u_j, r_j)} \frac{\partial q(u_j, r_j)}{\partial \theta_a} = 0,$$

for  $a = 1, 2, \dots, p$ .

These equations cannot in general be solved explicitly and therefore the Newton-Raphson iterative algorithm (Ortega and Rheinboldt 1970) (Appendix 1) is used to solve them. For this, the second partial derivatives of  $\ell_1$  are required and these are given by

$$\begin{aligned} \frac{\partial^2 \ell_1}{\partial \theta_a \partial \theta_b} &= \sum_{i=1}^I \frac{1}{P(t_i)^2} \left\{ \frac{\partial^2 P(t_i)}{\partial \theta_a \partial \theta_b} P(t_i) - \frac{\partial P(t_i)}{\partial \theta_a} \frac{\partial P(t_i)}{\partial \theta_b} \right\} \\ &+ \sum_{j=1}^J \frac{1}{q(u_j, r_j)^2} \left\{ \frac{\partial^2 q(u_j, r_j)}{\partial \theta_a \partial \theta_b} q(u_j, r_j) - \frac{\partial q(u_j, r_j)}{\partial \theta_a} \frac{\partial q(u_j, r_j)}{\partial \theta_b} \right\}, \quad a, b = 1, \dots, p. \end{aligned}$$

With these first and second partial derivatives, the Newton-Raphson algorithm can be used to estimate the parameters. The variances of the parameter estimates, and the covariances between them, are given by the elements of the inverse of the  $p \times p$  matrix with entries  $-\partial^2 \ell_1 / \partial \theta_a \partial \theta_b$ . The parameter estimates are denoted  $(\hat{\mu}, \hat{\lambda}, \hat{\sigma})$ . These parameter estimates can then be transformed to obtain the desired parameter estimates  $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon)$ . The following result (Mood *et al.* 1963) can be used to obtain the variances of the converted parameter estimates. Let  $X$  and  $Y$  be two random variables; then

$$\text{Var}(X/Y) \approx \left( \frac{\mu_x}{\mu_y} \right)^2 \left( \frac{\text{Var}(X)}{\mu_x^2} + \frac{\text{Var}(Y)}{\mu_y^2} - 2 \frac{\text{Cov}(X, Y)}{\mu_x \mu_y} \right).$$

### Estimation of parameters

So far no model has been assumed for  $F_V(v)$ , the probability distribution function of the vegetative mass. Once a parametric model has been chosen for  $F_V(v)$ , the above partial derivatives can be computed and maximum likelihood estimates of the parameters, denoted  $\hat{\theta}$ , can then be obtained for the plant reproduction model. The first and second partial derivatives, assuming that vegetative mass is normally distributed are given in Appendix 2. The parameters  $\theta$  become  $(\mu, \lambda, \sigma)$ .

## RESULTS AND DISCUSSION

The estimates of the parameters that describe the relationship between reproductive and vegetative mass are given in Table 1. The parameter estimates obtained here are the same as those obtained using the EM algorithm to estimate the parameters (see Schmid *et al.* 1994, table 2). The estimate for the standard deviation of the threshold parameter (i.e. the vegetative mass beyond which reproductive mass becomes positive) is also given here to give an indication of the spread of the threshold value. Schmid *et al.* (1994) also considered a censored allometric model to account for a non-linear relationship between vegetative and reproductive mass. The model proposed here assumes a linear relationship between vegetative and reproductive mass and does not deal with non-linear relationships, although this extension can readily be made. For the present study of reproductive and vegetative mass in *Solidago altissima*, it has been found that a linear relationship between reproductive and vegetative mass is adequate (Schmid *et al.* 1994).

Although the technique of computing maximum likelihood estimates by the Newton-Raphson algorithm involves the computation of the matrix of second partial derivatives, it

**Table 1.** Estimates (standard errors) of the parameters for the relationship between reproductive and vegetative mass (g) in *Solidago altissima*. The sample size of plants that have a reproductive mass of zero ( $n_0$ ) and of plants that have positive reproductive mass ( $n_1$ ) is also given.

Sample size		Parameter estimates				
$n_0$	$n_1$	$y$ -intercept ( $\hat{\alpha} = -\hat{\mu}/\hat{\lambda}$ )	slope ( $\hat{\beta} = 1/\hat{\lambda}$ )	$x$ -intercept ( $-\hat{\alpha}/\hat{\beta} = \hat{\mu}$ )	std dev of $R v$ ( $\hat{\sigma}_\varepsilon = \hat{\sigma}/\hat{\lambda}$ )	std dev of $V$ ( $\hat{\sigma}$ )
123	310	-7.11 (0.56)	0.60 (0.02)	11.77 (0.61)	5.32 (0.22)	8.80 (0.42)

does however mean that we can obtain standard errors of the estimates as a by-product of the algorithm by computing the inverse of the matrix that consists of the elements  $-\partial^2\ell/\partial\theta_a\partial\theta_b$ , where  $\ell$  is the log-likelihood function. This is a weakness of the EM algorithm approach used by Schmid *et al.* (1994). The standard errors for the parameters are given in Table 1.

## CONCLUSION

The model proposed in this paper for the relationship between vegetative and reproductive mass is an adaptation to the model for primary moult proposed by Underhill and Zucchini (1988). The problem of finding the relationship between vegetative and reproductive mass is complicated by the fact that reproductive mass is only observable (i.e. positive) once some threshold point of vegetative mass is reached. A linear regression model for this problem cannot handle the values of reproductive mass which are zero as these values violate the assumption of a regression model with normally distributed errors and if these values are omitted from the analysis, a negative bias to the estimate of the slope of the regression line is introduced (Schmid *et al.* 1994). The formulation of the problem of analysing plant reproduction is similar to that of primary moult. For the primary moult problem, the moult index is constrained to lie between 0 and 1 while in the plant reproduction problem, reproductive mass is constrained to lie above zero.

This paper serves to demonstrate that the approach taken by Underhill and Zucchini (1988) to formulate a model for data which is constrained on one or two limits is the same as formulating a censored regression model for data of this type proposed by Schmid

*et al.* (1994). Parameter estimates for the model for plant reproduction present here have been estimated by maximizing the likelihood function using differentiation and the Newton-Raphson algorithm. Schmid *et al.* (1994) made use of the EM (expectation-maximization) algorithm (Dempster *et al.* 1977) to maximize the likelihood function. Both algorithms have their advantages and disadvantages (see chapter 6), but one of the advantages of the Newton-Raphson algorithm is that standard errors for the parameter estimates are immediately available once the maximum likelihood estimates have been obtained.

This paper also shows that the two approaches to censored regression are equivalent. Whether the model for plant reproductive mass is formulated by regressing vegetative mass on reproductive mass by the censored regression technique, in which case the reproductive mass is regarded as the random variable and its distribution determined, or whether the distribution of threshold value (i.e. the vegetative mass value beyond which reproductive mass is positive) given the linear relationship between vegetative and reproductive mass, is modelled, the parameter estimates from both approaches are simple transformations of each other. Because of the invariance property of maximum likelihood estimates, once we have obtained parameter estimates by one approach, parameter estimates for the other approach can easily be obtained by making the transformation to the parameter estimates.

The choice of approach to model problems with variables that are discontinuous thus remains a choice of which algorithm to use to obtain maximum likelihood estimates for the parameters. The Newton-Raphson algorithm necessitates the computation of both first partial derivatives and the matrix of second partial derivatives, and, in the case of small data sets, the initial parameter estimates given must be close to the true parameter estimates for convergence to be achieved, but the rate of convergence is fast. The EM-algorithm does not need the first partial derivatives nor the matrix of second partial derivatives but then no standard errors of the estimates are immediately available. The rate of convergence is also slower than that of the Newton-Raphson algorithm.

## REFERENCES

- COX, D.R. AND HINKLEY, D.V. 1974. Theoretical statistics. Chapman and Hall, London.
- DEMPSTER, A.P., LAIRD, N.M. AND RUBIN, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B* 39: 1-38.
- MOOD, A.M., GRAYBILL, F.A. AND BOES, D.C. 1963. Introduction to the theory of statistics. McGraw-Hill, New York.
- ORTEGA, J.M. AND RHEINBOLDT, W.C. 1970. Iterative solution of nonlinear equations in several variables. Academic Press, New York.
- SCHMID, B., POLASEK, W., WEINER, J., KRAUSE, A. AND STOLL, P. 1994. Modelling of discontinuous relationships in biology with censored regression. *The American Naturalist* 143: 494-507.
- UNDERHILL, L.G. AND ZUCCHINI, W. 1988. A model for avian primary moult. *Ibis* 130: 358-372.
- UNDERHILL, L.G., ZUCCHINI, W. AND SUMMERS, R.W. 1990. A model for avian primary moult - data types based on migration strategies and an example using Redshanks *Tringa totanus*. *Ibis* 132: 118-123.

## APPENDIX 1

**Newton-Raphson iterative algorithm to estimate model parameters.**

The Newton-Raphson algorithm is a technique of iteratively solving nonlinear equations of several variables. We define  $\hat{\theta}^{(k)}$  to be the vector of parameter estimates after  $k$  iterations. The algorithm consists of the following steps:

- Step 1:** Calculate initial estimates  $\hat{\theta}^{(0)} = (\hat{\theta}_1^{(0)}, \dots, \hat{\theta}_p^{(0)})$ .
- Step 2:** Compute  $f^{(k)}$  and  $F^{(k)}$ , where  $f^{(k)}$  is the vector of first partial derivatives and  $F^{(k)}$  is the matrix of second partial derivatives, computed at the  $k$ th iteration.
- Step 3:** Compute the vector  $\delta^{(k)}$  which is the solution to the system of  $p$  linear equations

$$F^{(k)}\delta^{(k)} = f^{(k)},$$

where  $p$  represents the number of parameters.

- Step 4:** Set  $\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \delta^{(k)}$ , where  $\hat{\theta}^{(k)}$  contains the parameter estimates at the  $k$ th iteration.
- Step 5:** Test for convergence, for example, if the elements of  $f^{(k)}$  are sufficiently close to zero. If the convergence criterion is met then stop, otherwise increase  $k$  by 1 and return to step 2.

## APPENDIX 2

**First and second partial derivatives for the model to describe the relationship between reproductive mass and vegetative mass**

If we assume that vegetative mass is distributed normally, then the following formulae are required to apply the model for the relationship between reproductive mass and vegetative mass. For the case of the normal distribution, the model parameters are given by  $\theta_1 = \mu$ , the mean vegetative mass,  $\theta_2 = \lambda$ , the inverse of the slope of the line, and  $\theta_3 = \sigma$ , the standard deviation of the mean vegetative mass at which reproductive mass becomes observable. Here  $F_V(v) = \Phi\left(\frac{v - \mu}{\sigma}\right)$ , where  $\Phi(x)$  is the standard normal distribution function, that is

$$\Phi(x) = \int_{-\infty}^x \phi(z) dz,$$

where

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp \frac{-z^2}{2}, \quad -\infty < z < \infty.$$

Define  $g(x) = \frac{(x - \mu)}{\sigma}$ . Then

$$\begin{aligned} P(t) &= 1 - \Phi(g(t)) \\ q(u, r) &= \frac{\lambda}{\sigma} \phi(g(u - r\lambda)). \end{aligned}$$

For the first partial derivatives of  $\ell_1$  we have

$$\begin{aligned} \frac{\partial P(t_i)}{\partial \lambda} &= 0 \\ \frac{\partial P(t_i)}{\partial \mu} &= \frac{1}{\sigma} \phi(g(t_i)) \\ \frac{\partial P(t_i)}{\partial \sigma} &= \frac{g(t_i)}{\sigma} \phi(g(t_i)) \\ \frac{\partial q(u_j, r_j)}{\partial \lambda} &= q(u_j, r_j) \left[ \frac{1}{\lambda} + \frac{r_j g(u_j - r_j \lambda)}{\sigma} \right] \\ \frac{\partial q(u_j, r_j)}{\partial \mu} &= q(u_j, r_j) \frac{g(u_j - r_j \lambda)}{\sigma} \\ \frac{\partial q(u_j, r_j)}{\partial \sigma} &= q(u_j, r_j) \frac{g^2(u_j - r_j \lambda) - 1}{\sigma}. \end{aligned}$$

For the second partial derivatives we have

$$\begin{aligned} \frac{\partial^2 P(t_i)}{\partial \lambda^2} &= 0 \\ \frac{\partial^2 P(t_i)}{\partial \lambda \partial \mu} &= 0 \end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 P(t_i)}{\partial \lambda \partial \sigma} &= 0 \\
\frac{\partial^2 P(t_i)}{\partial \mu^2} &= \frac{g(t_i)}{\sigma^2} \phi(g(t_i)) \\
\frac{\partial^2 P(t_i)}{\partial \mu \partial \sigma} &= \frac{g^2(t_i) - 1}{\sigma^2} \phi(g(t_i)) \\
\frac{\partial^2 P(t_i)}{\partial \sigma^2} &= \frac{g^3(t_i) - 2g(t_i)}{\sigma^2} \phi(g(t_i)) \\
\frac{\partial^2 q(u_j, r_j)}{\partial \lambda^2} &= q(u_j, r_j) \left[ \frac{r_j^2 [g^2(u_j - r_j \lambda) - 1]}{\sigma^2} + \frac{2r_j g(u_j - r_j \lambda)}{\lambda \sigma} \right] \\
\frac{\partial^2 q(u_j, r_j)}{\partial \lambda \partial \mu} &= q(u_j, r_j) \left[ \frac{r_j [g^2(u_j - r_j \lambda) - 1]}{\sigma^2} + \frac{g(u_j - r_j \lambda)}{\lambda \sigma} \right] \\
\frac{\partial^2 q(u_j, r_j)}{\partial \lambda \partial \sigma} &= \frac{q(u_j, r_j)}{\sigma^2} \left[ -3r_j g(u_j - r_j \lambda) + \frac{\sigma [g^2(u_j - r_j \lambda) - 1]}{\lambda} + r_j g^3(u_j - r_j \lambda) \right] \\
\frac{\partial^2 q(u_j, r_j)}{\partial \mu^2} &= q(u_j, r_j) \left[ \frac{g^2(u_j - r_j \lambda) - 1}{\sigma^2} \right] \\
\frac{\partial^2 q(u_j, r_j)}{\partial \mu \partial \sigma} &= q(u_j, r_j) \left[ \frac{-3g(u_j - r_j \lambda) + g^3(u_j - r_j \lambda)}{\sigma^2} \right] \\
\frac{\partial^2 q(u_j, r_j)}{\partial \sigma^2} &= q(u_j, r_j) \left[ \frac{-5g^2(u_j - r_j \lambda) + g^4(u_j - r_j \lambda) + 2}{\sigma^2} \right].
\end{aligned}$$



**The analysis of avian primary moult III:  
inter-annual differences in timing of  
moult of Willow Warblers  
*Phylloscopus trochilus* in  
western Russia**

ABSTRACT

The Underhill-Zucchini moult model (*Ibis* (1988) 130: 358–372 and (1990) 132: 118–123) is extended to enable the parameters of moult in a number of data sets to be estimated simultaneously in such a way that one of the parameters is replicated for each data set. Standard statistical hypothesis testing methods are developed and the Akaike Information Criterion is used to facilitate parsimonious model selection procedures. In the application considered, the mean starting date of moult of Willow Warblers *Phylloscopus trochilus* in the St Petersburg region of western Russia is estimated for each of 11 years, while assuming the remaining parameters to be constant for all data sets. In eight of the 11 years, the mean starting date, for both sexes combined, lay between 9 and 14 July. With exceptions, the mean starting dates for males and females ran closely in parallel, with males starting moult, on average, about 14 days before females.

INTRODUCTION

The model for primary moult developed by Underhill and Zucchini (1988) and Underhill *et al.* (1990) (papers I and II in the series) assumed that all birds sampled have a common set of parameters, regardless of place or year. In the practical application of this model, Underhill and Zucchini (1988) used three parameters, two to define the mean and standard deviation of the starting date of moult and one to define the duration of moult. The moult model of Underhill and Zucchini (1988) and Underhill *et al.* (1990) has been used to estimate

moult parameters in seabirds (Cooper *et al.* 1991), shorebirds (Underhill 1985, Summers *et al.* 1989, Underhill *et al.* 1990, Marks 1993, Marks and Underhill 1994), honeyguides (Underhill *et al.* 1995) and in passerines (Winkler *et al.* 1988, Cooper and Underhill 1991, Underhill *et al.* 1992). In this paper, the third in the series, we consider an extension of the original model, in which one of the parameters is replicated in each data set. In the practical application, we have a separate parameter for the mean starting date of moult for each data set. We develop a strategy consistent with conventional model selection processes to deal with the standard statistical hypothesis testing situations (Linhart and Zucchini 1986). For example, we develop a test for the null hypothesis that the mean starting dates are the same for each data set (e.g. year or locality) against the alternative that moult starts earlier or later in some of the data sets. We also consider guideline strategies for parsimonious selection of model parameters making use of Akaike's (1973) Information Criterion. For example, we show how to select a model with the minimal number of starting date parameters by amalgamating those data sets which have a common starting date parameter.

This extension of the model is demonstrated using moult data on Willow Warblers *Phylloscopus trochilus* in western Russia in which there are data sets collected on an annual basis. This paper is thus also a sequel to the analysis of Willow Warbler primary moult by Underhill *et al.* (1992) in which between-year differences in the timing of primary moult were ignored.

## MATERIAL AND METHODS

### Data description

The data for this paper consist of post-nuptial moult indices for adult Willow Warblers captured in mist-nets at the south-eastern coast of the Lake Ladoga, St Petersburg region, Russia (60°47'N, 32°48'E). Most birds were sexed on wing length, as in Svensson (1992) and Underhill *et al.* (1992). The data contain observations made in 11 years, from 1979 to 1989 (Table 1). Primary moult scores were converted to percentage feather mass grown using the method of Underhill *et al.* (1992). The percentage feather mass grown was used as the moult index.

We could use the Underhill-Zucchini model to estimate the three parameters of moult, combining the observations for all years, as done by Underhill *et al.* (1992). Alternatively,

**Table 1.** Sample sizes for Willow Warbler primary moult data at Lake Ladoga, St Petersburg, Russia.

Year	All <sup>†</sup>			Male			Female		
	Yet to moult	In moult	Ended moult	Yet to moult	In moult	Ended moult	Yet to moult	In moult	Ended moult
1979	39	30	0	12	12	0	27	11	0
1980	30	26	5	7	13	3	23	12	1
1981	44	29	13	14	9	8	30	16	5
1982	56	40	31	23	17	18	33	17	13
1983	57	27	2	15	14	1	42	11	1
1984	25	13	1	6	5	1	19	5	0
1985	34	10	5	9	4	2	25	4	3
1986	29	14	8	11	8	6	18	5	2
1987	29	17	3	5	12	1	24	3	2
1988	35	20	8	12	9	6	23	9	2
1989	37	37	17	10	15	14	27	14	3
Totals	415	263	93	124	118	60	291	107	32

<sup>†</sup>The sex of 39 birds was not recorded.

we could estimate the parameters separately for each year, a total of 33 parameters. The first method would provide information on the average situation, but at the expense of ignoring the possibility and extent of inter-year variation. The second method would provide estimates of all the parameters for each year, but because of smaller sample sizes in each year, the standard errors of the estimates would be larger.

It is a compromise between these two extremes that we develop here. We consider that the main feature of interest in this data set is the possibility that there are inter-year variations in mean starting date. We therefore develop an extension of the Underhill-Zucchini model in which the standard deviation parameter and the duration parameter are both assumed to be constant for all years, and there is a parameter for the mean starting date for each year. For the Willow Warbler example, if the sexes are not separated, there is thus a total of 13 parameters.

Underhill *et al.* (1992) showed that data type 5 of Underhill *et al.* (1990) was the

appropriate data type for post-nuptial Willow Warbler data. Data are of type 5 if birds that have not begun to moult and those in moult are present throughout the sampling period and the birds observed on each day form a random sample from these two groups. With data type 5, birds that have completed moult may either emigrate out of the region immediately or additional birds may immigrate into it (see the motivations in Underhill *et al.* 1990, 1992). The mathematical development of the model for data type 2 is the most straightforward, and we therefore have selected to provide details for this data type and for data type 5, the one applicable to Willow Warblers during post breeding moult. Data are of type 2 if all birds that have not started moult, are in moult and have completed moult are present throughout the sampling period. Data type 2 also makes the assumption that there is no immigration or emigration during the sampling period. The mathematical development for the remaining data types is presented in Appendix 1. Table 2 is a reproduction of table 1 of Underhill *et al.* (1990) that shows a summary of the data types presented in Underhill and Zucchini (1988) and Underhill *et al.* (1990).

**Table 2.** The components of the population from which representative samples are required on each sampling date for the five data types presented in Underhill and Zucchini (1988) and Underhill *et al.* (1990).

Data type	Component of the population for which no			immigration or emigration during the sampling period is assumed
	Not yet moulted	In moult	Completed moult	
1 <sup>†</sup>	yes	yes	yes	all birds
2	yes	yes	yes	all birds
3	no	yes	no	birds in moult
4	no	yes	yes	birds in moult or completed moult
5	yes	yes	no	birds not yet moulted or in moult

<sup>†</sup>For data type 1, it is only necessary to classify birds as 'not yet moulted', 'in moult' or 'completed moult'.

### Extension to the model for avian primary moult

To investigate whether the mean starting date of moult varies in each data set (i.e. from year to year or from place to place), the models proposed by Underhill and Zucchini (1988) and Underhill *et al.* (1990) have to be modified. The same notation as Underhill and Zucchini (1988) and Underhill *et al.* (1990) is adopted here with changes as necessary.

Let the random variable  $T$  be the number of days from some reference date at which a randomly selected bird begins primary moult. Let  $F_T(t)$  be the probability distribution function and  $f_T(t)$  be the probability density function of  $T$ . We assume a linear relationship between moult index and time and that the duration of moult,  $\tau$ , remains constant for all birds.

The main quantities of interest are the expectation,  $\mu$ , and the standard deviation,  $\sigma$ , of  $T$ , which give the mean starting date of moult and the spread of starting dates of moult in the population. We are also interested in the duration of moult,  $\tau$ .

The data consist of observations on the moult indices of random samples of birds captured on some day. Let and  $Y(t)$  be the random variable of the moult index at time  $t$ , whose probability distribution function,  $F_{Y(t)}(y)$ , is dependent on time. Our objective is to estimate  $F_T(t)$  and  $\tau$  from observations on the process  $Y(t)$ . This is possible because of the following relationship (Underhill and Zucchini 1988):

$$Y(t) \leq y \quad \text{if and only if} \quad T \geq t - y\tau, \quad 0 \leq y < 1.$$

From this it follows that

$$F_{Y(t)}(y) = 1 - F_T(t - y\tau), \quad 0 \leq y < 1,$$

and

$$f_{Y(t)}(y) = \tau f_T(t - y\tau), \quad 0 < y < 1.$$

Suppose that the sample consists of  $H$  sets of data. Then for  $h = 1, 2, \dots, H$ , let  $J_h$  denote the number of birds in moult in the  $h$ th data set,  $I_h$  the number of birds which have not commenced moult in the  $h$ th data set, and  $K_h$  the number of birds which have completed moult in the  $h$ th data set. Let the days for each data set on which observations of "birds in moult" were made be

$$u_{11}, u_{12}, \dots, u_{1J_1}, u_{21}, u_{22}, \dots, u_{2J_2}, \dots, u_{H1}, u_{H2}, \dots, u_{HJ_H},$$

the days for each data set on which observations of “birds not in moult” were made be

$$t_{11}, t_{12}, \dots, t_{1J_1}, t_{21}, t_{22}, \dots, t_{2J_2}, \dots, t_{H1}, t_{H2}, \dots, t_{HJ_H},$$

and the days for each data set on which observations of “birds completed moult” were made be

$$v_{11}, v_{12}, \dots, v_{1K_1}, v_{21}, v_{22}, \dots, v_{2K_2}, \dots, v_{H1}, v_{H2}, \dots, v_{HK_H}.$$

Denote these vectors of observations by  $u$ ,  $t$ , and  $v$ , respectively. Let the vector  $y$  denote the moult indices associated with the birds which are actively in moult, so that

$$y = (y_{11}, y_{12}, \dots, y_{1J_1}, y_{21}, y_{22}, \dots, y_{2J_2}, \dots, y_{H1}, y_{H2}, \dots, y_{HJ_H}).$$

The likelihood function (Cox and Hinkley 1974) for these observations for data of type 2 is

$$L(\theta; t, u, y, v) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} P(t_{hi}) \prod_{j=1}^{J_h} q(u_{hj}, y_{hj}) \prod_{k=1}^{K_h} R(v_{hk}) \right), \quad (1)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = H + 2$ . The probability that birds have not commenced moult,  $P(t)$ , the probability that birds have completed moult,  $R(v)$ , and the index of moult for birds in moult,  $q(u, y)$ , are defined as in Underhill and Zucchini (1988):

$$\begin{aligned} P(t) &= \Pr\{Y(t) = 0\} = 1 - F_T(t), \\ R(v) &= \Pr\{Y(v) = 1\} = F_V(v - \tau), \\ q(u, y) &= \tau f_U(u - y\tau). \end{aligned}$$

The maximum likelihood estimates of the parameters are those values of  $\theta$  which maximize (1), or equivalently its logarithm

$$\ell_1 = \log L(\theta; t, u, y, v) = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \log(P(t_{hi})) + \sum_{j=1}^{J_h} \log(q(u_{hj}, y_{hj})) + \sum_{k=1}^{K_h} \log(R(v_{hk})) \right).$$

The normal equations are given by setting the partial derivatives equal to zero;

$$\frac{\partial \ell_1}{\partial \theta_a} = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \frac{1}{P(t_{hi})} \frac{\partial P(t_{hi})}{\partial \theta_a} + \sum_{j=1}^{J_h} \frac{1}{q(u_{hj}, y_{hj})} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} + \sum_{k=1}^{K_h} \frac{1}{R(v_{hk})} \frac{\partial R(v_{hk})}{\partial \theta_a} \right) = 0,$$

for  $a = 1, 2, \dots, p$ .

These equations cannot be solved explicitly and therefore the Newton-Raphson iterative algorithm (Ortega and Rheinboldt 1970) (Appendix 2) is used to solve them. For this, the second partial derivatives of  $\ell_1$  are required and these are given by

$$\begin{aligned} \frac{\partial^2 \ell_1}{\partial \theta_a \partial \theta_b} = & \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \frac{1}{P(t_{hi})^2} \left\{ \frac{\partial^2 P(t_{hi})}{\partial \theta_a \partial \theta_b} P(t_{hi}) - \frac{\partial P(t_{hi})}{\partial \theta_a} \frac{\partial P(t_{hi})}{\partial \theta_b} \right\} \right. \\ & + \sum_{j=1}^{J_h} \frac{1}{q(u_{hj}, y_{hj})^2} \left\{ \frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \theta_a \partial \theta_b} q(u_{hj}, y_{hj}) - \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_b} \right\} \\ & \left. + \sum_{k=1}^{K_h} \frac{1}{R(v_{hk})^2} \left\{ \frac{\partial^2 R(v_{hk})}{\partial \theta_a \partial \theta_b} R(v_{hk}) - \frac{\partial R(v_{hk})}{\partial \theta_a} \frac{\partial R(v_{hk})}{\partial \theta_b} \right\} \right), \quad a, b = 1, \dots, p. \end{aligned}$$

For data of type 5 the likelihood function for these observations is

$$L(\theta; t, u, y) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} \frac{P(t_{hi})}{1 - R(t_{hi})} \prod_{j=1}^{J_h} \frac{q(u_{hj}, y_{hj})}{1 - R(u_{hj})} \right), \quad (2)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters.

Let  $\ell_2 = \log L(\theta; t, u, y)$ . As before, the normal equations are obtained by setting the partial derivatives of  $\ell_2$  to zero;

$$\begin{aligned} \frac{\partial \ell_2}{\partial \theta_a} = & \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \left\{ \frac{1}{P(t_{hi})} \frac{\partial P(t_{hi})}{\partial \theta_a} + \frac{1}{1 - R(t_{hi})} \frac{\partial R(t_{hi})}{\partial \theta_a} \right\} \right. \\ & \left. + \sum_{j=1}^{J_h} \left\{ \frac{1}{q(u_{hj}, y_{hj})} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} + \frac{1}{1 - R(u_{hj})} \frac{\partial R(u_{hj})}{\partial \theta_a} \right\} \right) = 0, \quad a = 1, \dots, p. \end{aligned}$$

Newton-Raphson iteration is required to solve these equations. The second partial derivatives of  $\ell_2$  are

$$\begin{aligned} \frac{\partial^2 \ell_2}{\partial \theta_a \partial \theta_b} = & \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \left\{ \frac{1}{P(t_{hi})^2} \left[ \frac{\partial^2 P(t_{hi})}{\partial \theta_a \partial \theta_b} P(t_{hi}) - \frac{\partial P(t_{hi})}{\partial \theta_a} \frac{\partial P(t_{hi})}{\partial \theta_b} \right] \right. \right. \\ & \left. \left. + \frac{1}{(1 - R(t_{hi}))^2} \left[ \frac{\partial^2 R(t_{hi})}{\partial \theta_a \partial \theta_b} (1 - R(t_{hi})) + \frac{\partial R(t_{hi})}{\partial \theta_a} \frac{\partial R(t_{hi})}{\partial \theta_b} \right] \right\} \right. \\ & + \sum_{j=1}^{J_h} \left\{ \frac{1}{q(u_{hj}, y_{hj})^2} \left[ \frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \theta_a \partial \theta_b} q(u_{hj}, y_{hj}) - \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_b} \right] \right. \\ & \left. \left. + \frac{1}{(1 - R(u_{hj}))^2} \left[ \frac{\partial^2 R(u_{hj})}{\partial \theta_a \partial \theta_b} (1 - R(u_{hj})) + \frac{\partial R(u_{hj})}{\partial \theta_a} \frac{\partial R(u_{hj})}{\partial \theta_b} \right] \right\} \right), \end{aligned}$$

for  $a, b = 1, 2, \dots, p$ .

With these first and second partial derivatives, the Newton-Raphson algorithm can be used to estimate the parameters. The variances of the parameter estimates, and the covariances between them, are given by the elements of the inverse of the  $p \times p$  matrix with entries  $-\partial^2 \ell_2 / \partial \theta_a \partial \theta_b$ , as described in Underhill and Zucchini (1988), where  $\ell$  represents the log-likelihood function.

### Estimation of parameters

So far no model has been assumed for  $F_T(t)$ , the probability distribution function of the starting dates of primary moult. Once a parametric model has been chosen for  $F_T(t)$ , the above partial derivatives can be computed and maximum likelihood estimates of the parameters, denoted  $\hat{\theta}$ , can then be obtained for the primary moult model. We assume that the starting dates are normally distributed, as was done by Underhill and Zucchini (1988), so that, for the  $h$ th data set, the distribution of starting dates has mean  $\mu_h$  and standard deviation  $\sigma$ . The parameters  $\theta$  become  $(\mu_1, \dots, \mu_H, \tau, \sigma)$ . The first and second partial derivatives are given in Appendix 3. The mean completion date of moult for the  $h$ th data set is computed as  $\hat{\mu}_h + \hat{\tau}$ . The standard deviation of the mean completion date is given by  $(\widehat{\text{Var}}(\hat{\mu}_h) + \widehat{\text{Var}}(\hat{\tau}) + 2\widehat{\text{Cov}}(\hat{\mu}_h, \hat{\tau}))^{1/2}$ , with these values obtained from the above matrix.

### Testing the hypothesis of varying mean starting dates of moult

To determine whether the mean starting dates of moult differ between data sets, we test the null hypothesis that the mean starting dates are equal for each data set against the alternative hypothesis that the mean starting dates are not all the same. That is, we want to test

$$H_0 : \mu_h = \mu, \quad h = 1, 2, \dots, H$$

against

$$H_1 : \mu_i \neq \mu_j, \quad \text{for at least one pair } i, j \text{ with } 1 \leq i \neq j \leq H.$$

This hypothesis can be tested using the likelihood ratio test (Wilks 1962). For observations of data type 2, define

$$\lambda = \frac{L(\hat{\mu}, \hat{\tau}, \hat{\sigma}; t, u, y, v)}{L(\hat{\mu}_1, \dots, \hat{\mu}_H, \hat{\tau}, \hat{\sigma}; t, u, y, v)},$$

where the numerator is the likelihood function as derived in equation 5 of Underhill and Zucchini (1988), evaluated at the maximum likelihood estimators  $(\hat{\mu}, \hat{\tau}, \hat{\sigma})$ , i.e. the likelihood function for data type 2 under the null hypothesis of constant mean starting date for all data sets. The denominator is the likelihood function given by equation 1, evaluated at the maximum likelihood estimators  $(\hat{\mu}_1, \dots, \hat{\mu}_H, \hat{\tau}, \hat{\sigma})$ , i.e. assuming the alternative hypothesis of inter-annual differences in the timing of moult. The likelihood ratio,  $\lambda$ , is defined

analogously for data of type 5, in which case the numerator will be given by the likelihood function derived in Underhill *et al.* (1990), and the denominator by equation 2.

The test rejects the null hypothesis for small values of  $\lambda$ , or equivalently for large values of  $-2 \log \lambda$ . Asymptotically, it is well known that  $-2 \log \lambda$  has the chi-squared distribution with  $p - q$  degrees of freedom (Dobson 1983), where  $p$  represents the number of estimated parameters for the model under the alternative hypothesis, and  $q$  represents the number of parameters estimated under the null hypothesis.

### Guidelines for the parsimonious selection of model parameters

We now provide an answer to the question: Is a separate starting date parameter  $\mu_i$  really necessary for each data set? We restrict ourselves to two possible scenarios:

- (1) Does a group of data sets have a common starting date, while the remainder, each with different starting dates, can be considered abnormal?
- (2) Do the data sets fall into three groups, those in which the starting dates are “early”, “normal” or “late”?

Other scenarios can be also handled by the approach to be described, but these two seem to be biologically reasonable. We also restrict the discussion to data type 5, because this is the appropriate data type for Willow Warblers.

Model selection criteria are discussed in Linhart and Zucchini (1986). In this context, the recommended criterion is Akaike’s Information Criterion (AIC) (Akaike 1973), calculated (for data type 5) as

$$\text{AIC} = 2(-\log L(\hat{\theta}; t, u, y) + N),$$

where  $N$  is the number of parameters in the model. The guideline is to choose the model with the smallest value for the AIC. Effectively, the AIC penalizes models with unnecessary, additional parameters. Another application of the AIC in biology, with useful background comments, may be found in Lebreton *et al.* (1992).

In this context, the number of parameters is  $N = H + 2$ , where  $H$  is the number of data sets. The strategy for finding the minimum value for the AIC is to amalgamate data sets and to recalculate the parameter estimates, the likelihood function and the AIC. Because there are  $2^H$  ways of combining  $H$  data sets, it is usually impractical to search for the minimum AIC, which is our motivation for restricting ourselves to scenarios (1) and (2)

described above. Parameter estimates provide guidance about which data sets should be amalgamated, that is we combine data sets with similar parameter estimates.

## RESULTS AND DISCUSSION

For the Willow Warbler data, the overall estimates of the parameters, assuming a constant mean starting date over the 11 years were nearly identical for data types 2 and 5 (Table 3). However, the estimated durations using data type 5 were 3–4 days longer than with data type 2. This is the opposite effect to that observed after post-nuptial moult at seven localities in western Europe in Underhill *et al.* (1992). The estimates of the parameters of moult using the extended model developed in this paper show that for data type 5 (with very similar results for data type 2), the estimated mean starting dates for males and females combined varied over 15 days between 6 July (in 1989) and 21 July (in 1985) (Table 4). However, in six of the 11 years, the estimated mean starting date was within two days of 13 July, the overall mean starting date (Table 3), with 1980, 1985, and 1989 being the furthest from normal. In eight of the 11 years, the mean starting dates lay between 9 July and 14 July. In 1980 and 1985, moult started later than 14 July and in 1989 it started earlier than 9 July. These observations can be tested rigorously using the AIC.

Results for males and females ran closely in parallel (Fig. 1); when males started earlier or later than their average, so did the females. The striking exception was 1987, when males started moulting on 4 July (Table 4), one day earlier than their average, 5 July (Table 3), but females started on 27 July, seven days later than their average, 20 July (Tables 3 and 4). In 1988 males started on time, but females were five days earlier than their average, and in 1989 both sexes were early, but males were relatively earlier than females.

The likelihood ratio test of the null hypothesis of constant mean starting dates was rejected at a high level of significance for the sexes combined and for males ( $P < 0.0001$ ), but for females was rejected at only the 6% significance level ( $P = 0.062$ , data type 5) (Table 5). Thus in all cases there was at least one year with a mean starting date parameter that differed from that of the remaining years. The error bars in Figure 1 for mean starting date of primary moult show that the standard error of the estimates for mean starting date of moult are much larger for females than for males. This explains the apparent contradiction that although the mean starting date for both males and females vary considerably, the hypothesis of constant mean starting dates was rejected at a high level of significance for

**Table 3.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting date of moult was the same for the 11 years of the study.

Parameter estimates	Data type 2			Data type 5		
	All	Male	Female	All	Male	Female
Sample size	771	302	430	678	242	398
Duration ( $\bar{\tau}$ )	40.5 (1.1)	43.3 (1.2)	38.3 (1.7)	44.9 (1.3)	46.6 (1.5)	42.1 (2.0)
Mean starting date ( $\hat{\mu}^\dagger$ )	42.3 (0.7)	34.5 (0.7)	49.9 (1.2)	42.6 (0.8)	34.5 (0.8)	50.1 (1.2)
	12 July	5 July	20 July	13 July	5 July	20 July
Std dev ( $\hat{\sigma}$ )	11.3 (0.4)	7.8 (0.4)	11.7 (0.7)	12.1 (0.5)	8.4 (0.5)	12.1 (0.7)
Mean completion date	82.8 (0.9)	77.8 (1.0)	88.2 (1.4)	87.5 (1.2)	81.1 (1.3)	92.2 (1.7)
	22 Aug	17 Aug	27 Aug	27 Aug	20 Aug	31 Aug

†Mean starting date and mean completion date are in days since 1 June (i.e. day 1 = 1 June) as in Underhill *et al.* (1992).

males but the level of significance was much lower for females.

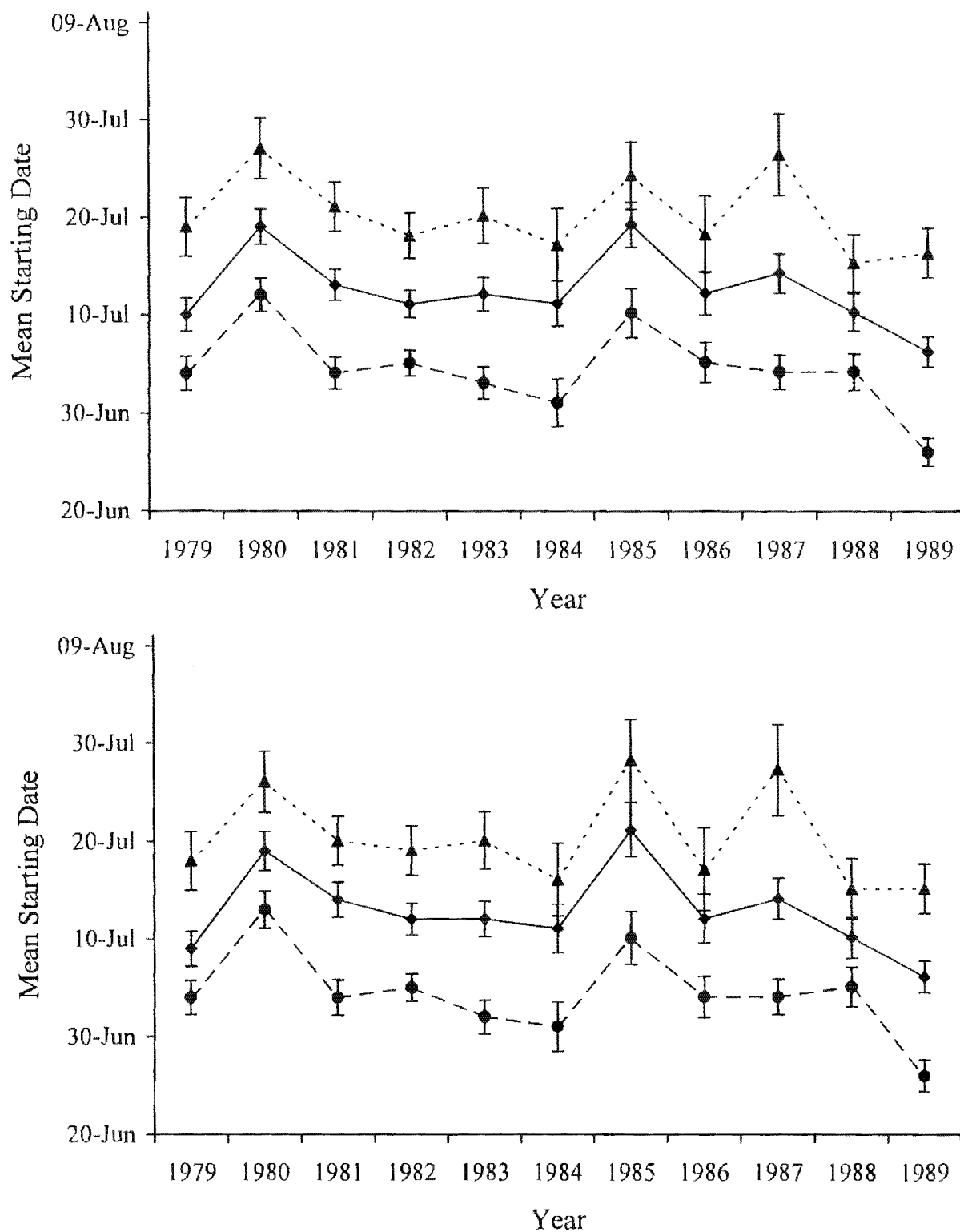
For the data set with males and females combined, and for data type 5, the AIC with common starting dates was 164.0, while if each year had a separate starting date it was 152.8 (Table 6). Using scenario 1, examination of the parameters suggested that the three years 1980, 1985 and 1989 were abnormal (see above). The AIC for this parameterisation was 148.5. For males, the minimum AIC was -1.3, when years 1980, 1985 and 1989 were classified as abnormal, and for females (minimum AIC = 71.3) the abnormal years were 1980, 1985 and 1987 (Table 6). Estimates of the parameters of moult for the models with the minimum AIC values (Table 7) provide values for the mean starting date in "normal" years (4 July for males, 18 July for females and 12 July for both sexes combined). The mean starting dates for "abnormal" years were similar to those estimated when there was a separate parameter for each year (Table 4).

Under scenario 2, and for both sexes combined, the minimum AIC was 147.4 when years 1979, 1988 and 1989 were identified as years when moult started early, 1980 and 1985 as late years, and the remaining years as normal (Table 8). For males, the minimum AIC (-2.0) was achieved when year 1989 was designated as being early, 1980 and 1985 as late, and the

**Table 4.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting dates of moult  $\mu_h$  varied between years,  $h = 1, \dots, 11$ , and a constant duration and standard deviation for all years.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	all	Male	Female
1979	$\hat{\mu}_1^\dagger$	40.3 (1.7)	34.1 (1.7)	49.1 (3.0)	39.4 (1.8)	33.8 (1.7)	47.9 (3.0)
		10 July	4 July	19 July	9 July	4 July	18 July
1980	$\hat{\mu}_2$	49.2 (1.8)	42.3 (1.7)	57.2 (3.1)	49.0 (2.0)	42.6 (1.9)	55.7 (3.1)
		19 July	12 July	27 July	19 July	13 July	26 July
1981	$\hat{\mu}_3$	43.4 (1.6)	33.6 (1.6)	50.8 (2.5)	43.9 (1.8)	33.9 (1.8)	50.2 (2.5)
		13 July	4 July	21 July	14 July	4 July	20 July
1982	$\hat{\mu}_4$	41.3 (1.4)	34.6 (1.3)	47.9 (2.3)	41.9 (1.6)	34.9 (1.4)	48.9 (2.5)
		11 July	5 July	18 July	12 July	5 July	19 July
1983	$\hat{\mu}_5$	41.8 (1.7)	32.7 (1.6)	49.5 (2.8)	41.7 (1.8)	32.3 (1.7)	50.1 (2.9)
		12 July	3 July	20 July	12 July	2 July	20 July
1984	$\hat{\mu}_6$	41.3 (2.3)	30.9 (2.4)	47.4 (3.7)	40.7 (2.5)	30.7 (2.5)	46.2 (3.7)
		11 July	1 July	17 July	11 July	1 July	16 July
1985	$\hat{\mu}_7$	48.6 (2.3)	40.1 (2.5)	53.5 (3.4)	50.8 (2.7)	40.1 (2.7)	58.0 (4.2)
		19 July	10 July	24 July	21 July	10 July	28 July
1986	$\hat{\mu}_8$	42.1 (2.2)	34.6 (2.0)	47.6 (3.9)	42.2 (2.5)	34.2 (2.1)	47.4 (4.2)
		12 July	5 July	18 July	12 July	4 July	17 July
1987	$\hat{\mu}_9$	44.4 (2.0)	34.4 (1.7)	55.9 (4.1)	43.9 (2.1)	33.9 (1.8)	56.8 (4.6)
		14 July	4 July	26 July	14 July	4 July	27 July
1988	$\hat{\mu}_{10}$	39.6 (1.9)	33.9 (1.8)	44.9 (2.9)	40.2 (2.1)	34.5 (2.0)	44.8 (3.1)
		10 July	4 July	15 July	10 July	5 July	15 July
1989	$\hat{\mu}_{11}$	35.8 (1.5)	26.0 (1.4)	45.9 (2.5)	36.4 (1.6)	26.4 (1.6)	45.4 (2.6)
		6 July	26 June	16 July	6 July	26 June	15 July
	Duration ( $\hat{\tau}$ )	40.7 (1.1)	43.9 (1.1)	38.1 (1.8)	44.6 (1.3)	45.8 (1.3)	42.0 (2.0)
	Std dev ( $\hat{\sigma}$ )	10.8 (0.4)	6.6 (0.4)	11.3 (0.6)	11.4 (0.5)	7.0 (0.4)	11.6 (0.7)

†Mean starting date is in days since 1 June (i.e. day 1 = 1 June).



**Figure 1.** Mean annual starting dates of moult for Willow Warbler for data type 2 (top) and data type 5 (bottom). The mean starting dates of moult are given for males (dashed line), females (dotted line) and for all the data (solid line) (see Table 3).

**Table 5.** Likelihood ratio test for variation in mean starting dates in primary moult of the Willow Warbler in different years.

	Data type 2			Data type 5		
	All	Male	Female	All	Male	Female
$-2 \log \lambda$	50.31	60.54	18.33	39.92	47.56	17.60
<i>P</i> -value	< 0.0001	< 0.0001	0.05	< 0.0001	< 0.0001	0.062

remainder as normal. For females, the years 1984, 1986, 1988 and 1989 were considered early, and 1980, 1985 and 1987 as late and the remainder normal (AIC = 68.8) (Table 8). In this example, the minimum AIC values under scenario 2 were smaller than those under scenario 1, but this would not be true in general. Table 9 shows the parameter estimates of moult for the models with minimum AIC values under scenario 2.

**Table 6.** Values of the Akaike's Information Criterion (AIC) for selected models under scenario 1, in which some years are considered abnormal. Years given here are those for which the mean starting dates of moult were allowed to differ from the remaining years, which had a constant mean starting date of moult. Minimum values for the AIC are in bold.

Abnormal years	AIC values		
	All	Male	Female
All	152.8	4.1	76.4
1979, 1980, 1985, 1988, 1989	148.8	0.6	73.6
1979, 1980, 1985, 1989	<b>148.3</b>	-0.3	73.6
1980, 1985, 1987, 1988	153.3	7.9	71.8
1980, 1984, 1985, 1989	149.4	-1.1	73.4
1980, 1985, 1989	148.5	<b>-1.3</b>	72.7
1980, 1985, 1987	152.3	7.3	<b>71.3</b>
1980, 1985	152.4	6.5	72.6
1980, 1987	158.9	10.0	73.5
1980, 1989	153.5	0.5	74.2
1985, 1987	159.5	17.1	73.6
1985, 1989	153.7	7.4	74.1
None	164.0	17.4	75.8

**Table 7.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lagoda Lake, St Petersburg, Russia, under scenario 1 and for data type 5. The estimated mean starting date in “normal” years is denoted by  $\hat{\mu}$ ; subscripts are used for years in which the estimated mean starting date was significantly different from  $\hat{\mu}$ .

Parameter estimates	All	Male	Female
$\hat{\mu}^\dagger$	42.2 (0.9)	33.8 (0.7)	48.1 (1.3)
	12 July	4 July	18 July
$\hat{\mu}_{1979}$	39.4 (1.8)	–	–
	9 July	–	–
$\hat{\mu}_{1980}$	49.0 (2.0)	42.6 (1.9)	56.0 (3.2)
	19 July	13 July	26 July
$\hat{\mu}_{1985}$	50.8 (2.7)	40.1 (2.7)	58.2 (4.2)
	21 July	10 July	28 July
$\hat{\mu}_{1987}$	–	–	57.1 (4.6)
	–	–	27 July
$\hat{\mu}_{1989}$	36.4 (1.6)	26.5 (1.6)	–
	6 July	26 June	–
Duration ( $\hat{\tau}$ )	44.7 (1.3)	45.8 (1.3)	41.8 (2.0)
Standard deviation ( $\hat{\sigma}$ )	11.4 (0.5)	7.0 (0.4)	11.7 (0.7)

$\dagger$ Mean starting date is in days since 1 June (i.e. day 1 = 1 June).

**Table 8.** Values of the AIC for selected models under scenario 2, in which years are considered as being “early”, “normal” or “late”. The years in which the estimated mean starting date of moult was “early” or “late” are listed. Minimum values for the AIC are in bold.

Designation	Years <sup>†</sup>	AIC values		
		All	Male	Female
Early	1989	147.4	<b>-2.0</b>	71.8
Late	1980, 1985			
Early	1984, 1986, 1988, 1989	151.8	9.0	<b>68.8</b>
Late	1980, 1985, 1987			
Early	1979, 1989	<b>147.3</b>	4.6	71.0
Late	1980, 1985			

**Table 9.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lagoda Lake, St Petersburg, Russia, under scenario 2 and for data type 5. The estimated mean starting date in “early” years is denoted by  $\hat{\mu}_e$ , in “normal” years by  $\hat{\mu}_n$  and in “late” years by  $\hat{\mu}_l$ .

Parameter estimates	All	Male	Female
$\hat{\mu}_e^{\dagger}$	38.4 (1.1)	26.5 (1.6)	45.8 (1.8)
	8 July	27 June	16 July
$\hat{\mu}_n$	42.5 (0.9)	33.8 (0.8)	49.4 (1.5)
	13 July	4 July	19 July
$\hat{\mu}_l$	49.7 (1.6)	41.9 (1.6)	56.7 (2.3)
	20 July	12 July	27 July
Duration ( $\hat{\tau}$ )	44.5 (1.3)	45.8 (1.3)	41.8 (1.9)
Standard deviation ( $\hat{\sigma}$ )	11.4 (0.5)	7.1 (0.4)	11.6 (0.7)

<sup>†</sup>Mean starting dates are in days since 1 June (i.e. day 1 = 1 June).

## CONCLUSION

The model for primary moult proposed in this paper is an extension of the model of Underhill and Zucchini (1988) and Underhill *et al.* (1990) which allows a flexible and statistically coherent approach to the analysis of variation in the timing of moult in different data sets. The extension developed here relates to the possibility that the mean starting date of moult varies between years or between localities.

Clearly, other extensions of the original moult model are possible. Another potentially useful extension would be to allow both duration and mean starting date to vary between data sets. However, the fitting of such models by the method of maximum likelihood will require large amounts of “well-behaved” data. This is because there is a tendency, as shown by Underhill and Zucchini (1988), for the maximum likelihood estimates of mean starting date and duration to be negatively correlated with each other. This means that almost the same maximum value for the likelihood function can be achieved by delaying the mean starting date by, say one day, and decreasing the duration by one day. This problem will be aggravated as the number of pairs of correlated parameters increases. A more valuable extension would be models that enable explanatory variables to be incorporated. For example, the mean starting date of moult in year  $h$  could be modelled as

$$\mu_h = \mu + \beta_1 x_{1h} + \dots + \beta_p x_{ph} + \epsilon,$$

where  $x_{1h}, \dots, x_{ph}$  are the values of  $p$  explanatory variables in year  $h$  (temperature, rainfall, breeding productivity, parasite levels, etc) and  $\beta_1, \dots, \beta_p$  are regression coefficients to be estimated. Methods to test whether candidate explanatory variables would make a significant difference to predicting the mean starting date could be developed. This extension would present no particular technical difficulties.

Underhill *et al.* (1992) estimated the parameters of primary moult for Willow Warblers at Lammi, southern Finland, at almost the same latitude as Lake Ladoga, 61°N. The mean starting dates and the durations of moult at the two localities were similar: the mean starting dates were 11 July and 12 July at Lammi and Lake Ladoga, respectively, and the durations were 43 and 45 days, respectively (Underhill *et al.* 1992, table 3; Table 7).

The race of Willow Warbler in the study area is the same as that at Lammi, the subspecies *acredula*. For this northern race, Underhill *et al.* (1992) postulated that starting date of primary moult does not show delay with increasing latitude northwards, as was demonstrated for the nominate race *trochilus*. The data from Lake Ladoga, at latitude 61°N, and

close to the southern limits of race *acredula*, fit this hypothesis (see Underhill *et al.* 1992, fig. 2).

The difference between the estimates of the parameters using data types 2 and 5 went in the opposite direction to those obtained for post-nuptial moult in western Europe by Underhill *et al.* (1992) where, for seven localities, the durations of moult estimated using data type 2 were longer than for data type 5. This was interpreted as evidence of emigration of birds which had completed moult. At Lake Ladoga, the duration of moult from data type 2 was shorter than for data type 5, as observed for pre-nuptial moult in Uganda (Underhill *et al.* 1992) and the same explanation is proposed: an influx of birds which had completed moult occurred at a time when the local population was still moulting. These birds are likely to be birds which had completed moult at localities further north, had commenced migration, and were passing through the study area at Lake Ladoga.

The statistical model for the analysis of primary moult, as formulated by Underhill and Zucchini (1988), is an example of a censored regression model. Applications of censored regression models in biology were reviewed by Schmid *et al.* (1994) who discussed mainly botanical examples. In particular, they considered a situation from plant reproductive biology in which the reproductive mass is zero until the vegetative mass reaches a threshold. Their goal was to estimate this mean threshold mass and to relate the reproductive mass to vegetative mass. This and their other examples were all of "one-sided" censoring. The moult problem is an example of a censored regression model with two-sided censoring because the moult index is constrained to lie between 0 and 1. In the moult problem, the "thresholds" are the mean dates at which moult commences and finishes.

Schmid *et al.* (1994) derived the analogous likelihood function for one-sided censoring to the function that Underhill and Zucchini (1988) used for two-sided censoring, but did not attempt to maximize the likelihood function using differentiation and the Newton-Raphson algorithm. Instead, they made use of the EM (expectation-maximization) algorithm (Dempster *et al.* 1977) to maximize the likelihood function. Further studies to compare the numerical efficiencies of the two algorithms would be valuable.

## REFERENCES

- AKAIKE, H. 1973. Information theory and an extension of the maximum likelihood principle. *In* Petrov, B.C. and Csáki, F. (Eds). International symposium on information theory, 2nd ed: 267–281. Akadémiai Kiadó, Budapest.
- COOPER, J. AND UNDERHILL, L.G. 1991. Breeding, mass and primary moult of European Starlings *Sturnus vulgaris* at Dassen Island, South Africa. *Ostrich* 62: 1–7.
- COOPER, J., UNDERHILL, L.G. AND AVERY, G. 1991. Primary moult and transequatorial migration of the Sooty Shearwater. *Condor* 93: 724–730.
- COX, D.R. AND HINKLEY, D.V. 1974. Theoretical statistics. Chapman and Hall, London.
- DEMPSTER, A.P., LAIRD, N.M. AND RUBIN, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B* 39: 1–38.
- DOBSON, A.J. 1983. An introduction to statistical modelling. Chapman and Hall, London.
- LEBRETON, J.-D., BURNHAM, K.P., CLOBERT, J. AND ANDERSON, D.R. 1992. Modeling survival and test biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* 62: 67–118.
- LINHART, H. AND ZUCCHINI, W. 1986. Model selection. Wiley, New York.
- MARKS, J.S. 1993. Molt of Bristle-thighed Curlews in the Northwestern Hawaiian Islands. *Auk* 110: 573–587.
- MARKS, J.S. AND UNDERHILL, L.G. 1994. Molt, migration and mass of a handicapped Bristle-thighed Curlew *Numenius tahitiensis*. *Ardea* 82: 153–156.
- ORTEGA, J.M. AND RHEINBOLDT, W.C. 1970. Iterative solution of nonlinear equations in several variables. Academic Press, New York.
- SCHMID, B., POLASEK, W., WEINER, J., KRAUSE, A. AND STOLL, P. 1994. Modelling of discontinuous relationships in biology with censored regression. *American Naturalist* 143: 494–507.
- SUMMERS, R.W., UNDERHILL, L.G., CLINNING, C.F. AND NICOLL, M. 1989. Populations, migrations, biometrics and moult of the Turnstone *Arenaria i. interpres* on the

- east Atlantic coastline, with special reference to the Siberian population. *Ardea* 77: 145–168.
- SVENSSON, L. 1992. Identification guide to European Passerines, 4th ed. Lars Svensson, Stockholm.
- UNDERHILL, L.G. 1985. Estimating the parameters for primary moult — a new statistical model. *Wader Study Group Bulletin* 44: 27–29.
- UNDERHILL, L.G. AND ZUCCHINI, W. 1988. A model for avian primary moult. *Ibis* 130: 358–372.
- UNDERHILL, L.G., ZUCCHINI, W. AND SUMMERS, R.W. 1990. A model for avian primary moult — data types based on migration strategies and an example using the Redshank *Tringa totanus*. *Ibis* 132: 118–123.
- UNDERHILL, L.G., PRÛS-JONES, R.P., DOWSETT, R.J., HERROELEN, P., JOHNSON, D.N., LAWN, M.R., NORMAN, S.C., PEARSON, D.J. AND TREE, A.J. 1992. The biannual primary moult of Willow Warblers *Phylloscopus trochilus* in Europe and Africa. *Ibis* 134: 286–297.
- UNDERHILL, L.G., UNDERHILL, G.D., MARTIN, C.G.C. AND FRASER, M.W. 1995. Primary moult, wing-length and mass of the Lesser Honeyguide *Indicator minor*. *Bulletin of the British Ornithologists' Club* 115: 229–234.
- WILKS, S.S. 1962. *Mathematical statistics*. Wiley, New York.
- WINKLER, R., DAUNICHT, W.D. AND UNDERHILL, L.G. 1988. Die Grossgefiedermauser von Alpendohle *Pyrrhocorax graculus* und Alpenkrahe *P. pyrrhocorax*. [The flight feather moult of the Alpine Chough and the Chough.] *Ornithologische Beobachter* 85: 245–249.

## APPENDIX 1

**Primary moult models for data type 1, type 3 and type 4.**

Although the data set examined in this study falls under the specification of type 2 and type 5 data, all models proposed by Underhill and Zucchini (1988) and Underhill *et al.* (1990) have been modified to incorporate the possibility that the model parameters may vary from year to year. In this appendix the theory for observations for data of type 1, type 3 and type 4 is given.

Given the notation as stated before, the likelihood function for those observations for data of type 1, i.e. when each bird is classified as “not yet started moult”, “in moult” or “completed moult”, is given by

$$L(\theta; t, u, v) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} P(t_{hi}) \prod_{j=1}^{J_h} Q(u_{hj}) \prod_{k=1}^{K_h} R(v_{hk}) \right), \quad (3)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = H + 2$ . The probability functions  $P(t)$ ,  $R(v)$  are defined as before and  $Q(u)$  is defined as

$$Q(u) = \Pr\{0 \leq Y(u) \leq 1\} = F_U(u) - F_U(u - \tau).$$

The maximum likelihood estimates of the parameters are those values of  $\theta$  which maximize (3), or equivalently its logarithm

$$\ell_3 = \log L(\theta; t, u, v) = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \log(P(t_{hi})) + \sum_{j=1}^{J_h} \log(Q(u_{hj})) + \sum_{k=1}^{K_h} \log(R(v_{hk})) \right).$$

The normal equations are given by setting the partial derivatives equal to zero;

$$\frac{\partial \ell_3}{\partial \theta_a} = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \frac{1}{P(t_{hi})} \frac{\partial P(t_{hi})}{\partial \theta_a} + \sum_{j=1}^{J_h} \frac{1}{Q(u_{hj})} \frac{\partial Q(u_{hj})}{\partial \theta_a} + \sum_{k=1}^{K_h} \frac{1}{R(v_{hk})} \frac{\partial R(v_{hk})}{\partial \theta_a} \right) = 0,$$

for  $a = 1, 2, \dots, p$ .

These equations cannot be solved explicitly and therefore the Newton-Raphson iteration method is used to solve them. For this, the second partial derivatives of  $\ell_3$  are required and these are given by

$$\frac{\partial^2 \ell_3}{\partial \theta_a \partial \theta_b} = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \frac{1}{P(t_{hi})^2} \left\{ \frac{\partial^2 P(t_{hi})}{\partial \theta_a \partial \theta_b} P(t_{hi}) - \frac{\partial P(t_{hi})}{\partial \theta_a} \frac{\partial P(t_{hi})}{\partial \theta_b} \right\} \right)$$

$$\begin{aligned}
 & + \sum_{j=1}^{J_h} \frac{1}{Q(u_{hj})^2} \left\{ \frac{\partial^2 Q(u_{hj})}{\partial \theta_a \partial \theta_b} Q(u_{hj}) - \frac{\partial Q(u_{hj})}{\partial \theta_a} \frac{\partial Q(u_{hj})}{\partial \theta_b} \right\} \\
 & + \sum_{k=1}^{K_h} \frac{1}{R(v_{hk})^2} \left\{ \frac{\partial^2 R(v_{hk})}{\partial \theta_a \partial \theta_b} R(v_{hk}) - \frac{\partial R(v_{hk})}{\partial \theta_a} \frac{\partial R(v_{hk})}{\partial \theta_b} \right\} \Bigg), \quad a, b = 1, \dots, p.
 \end{aligned}$$

For data of type 3, i.e. when only the moult indices of birds actively in moult are available, the likelihood function for these observations is

$$L(\theta; u, y) = \prod_{h=1}^H \prod_{j=1}^{J_h} \left( \frac{q(u_{hj}, y_{hj})}{Q(u_{hj})} \right), \quad (4)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters.

Let  $\ell_4 = \log L(\theta; u, y)$ . As before, the normal equations are obtained by setting the partial derivatives of  $\ell_4$  equal to zero;

$$\frac{\partial \ell_4}{\partial \theta_a} = \sum_{h=1}^H \sum_{j=1}^{J_h} \left\{ \frac{1}{q(u_{hj}, y_{hj})} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} - \frac{1}{Q(u_{hj})} \frac{\partial Q(u_{hj})}{\partial \theta_a} \right\} = 0, \quad a = 1, \dots, p.$$

Newton-Raphson iteration is required to solve these equations. The second partial derivatives of  $\ell_4$  are

$$\begin{aligned}
 \frac{\partial^2 \ell_4}{\partial \theta_a \partial \theta_b} & = \sum_{h=1}^H \sum_{j=1}^{J_h} \left\{ \frac{1}{q(u_{hj}, y_{hj})^2} \left[ \frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \theta_a \partial \theta_b} q(u_{hj}, y_{hj}) - \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_b} \right] \right. \\
 & \quad \left. - \frac{1}{Q(u_{hj})^2} \left[ \frac{\partial^2 Q(u_{hj})}{\partial \theta_a \partial \theta_b} Q(u_{hj}) - \frac{\partial Q(u_{hj})}{\partial \theta_a} \frac{\partial Q(u_{hj})}{\partial \theta_b} \right] \right\}, \quad a, b = 1, \dots, p.
 \end{aligned}$$

For data of type 4, i.e. when only birds that are in moult or that have completed moult are available, the likelihood function for these observations is

$$L(\theta; u, v, y) = \prod_{h=1}^H \left( \prod_{j=1}^{J_h} \frac{q(u_{hj}, y_{hj})}{1 - P(u_{hj})} \prod_{k=1}^{K_h} \frac{R(v_{hk})}{1 - P(v_{hk})} \right) \quad (5)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters.

Let  $\ell_5 = \log L(\theta; u, v, y)$ . As before, the normal equations are obtained by setting the partial derivatives of  $\ell_5$  to zero;

$$\begin{aligned}
 \frac{\partial \ell_5}{\partial \theta_a} & = \sum_{h=1}^H \left( \sum_{j=1}^{J_h} \left\{ \frac{1}{q(u_{hj}, y_{hj})} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} + \frac{1}{1 - P(u_{hj})} \frac{\partial P(u_{hj})}{\partial \theta_a} \right\} \right. \\
 & \quad \left. + \sum_{k=1}^{K_h} \left\{ \frac{1}{R(v_{hk})} \frac{\partial R(v_{hk})}{\partial \theta_a} + \frac{1}{1 - P(v_{hk})} \frac{\partial P(v_{hk})}{\partial \theta_a} \right\} \right) = 0, \quad a = 1, \dots, p.
 \end{aligned}$$

Newton-Raphson iteration is required to solve these equations. The second partial derivatives of  $\ell_5$  are

$$\frac{\partial^2 \ell_5}{\partial \theta_a \partial \theta_b} = \sum_{h=1}^H \left( \sum_{j=1}^{J_h} \left\{ \frac{1}{q(u_{hj}, y_{hj})^2} \left[ \frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \theta_a \partial \theta_b} q(u_{hj}, y_{hj}) - \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_a} \frac{\partial q(u_{hj}, y_{hj})}{\partial \theta_b} \right] \right. \right.$$

$$\begin{aligned}
& + \frac{1}{(1 - P(u_{hj}))^2} \left[ \frac{\partial^2 P(u_{hj})}{\partial \theta_a \partial \theta_b} (1 - P(u_{hj})) + \frac{\partial P(u_{hj})}{\partial \theta_a} \frac{\partial P(u_{hj})}{\partial \theta_b} \right] \Bigg\} \\
& + \sum_{k=1}^{K_h} \left\{ \frac{1}{R(v_{hk})^2} \left[ \frac{\partial^2 R(v_{hk})}{\partial \theta_a \partial \theta_b} R(v_{hk}) - \frac{\partial R(v_{hk})}{\partial \theta_a} \frac{\partial R(v_{hk})}{\partial \theta_b} \right] \right. \\
& \left. + \frac{1}{(1 - P(v_{hk}))^2} \left[ \frac{\partial^2 P(v_{hk})}{\partial \theta_a \partial \theta_b} (1 - P(v_{hk})) + \frac{\partial P(v_{hk})}{\partial \theta_a} \frac{\partial P(v_{hk})}{\partial \theta_b} \right] \right\}, \quad a, b = 1, \dots, p.
\end{aligned}$$

With these first and second partial derivatives, the Newton-Raphson algorithm can be used to estimate the parameters (Appendix 2). The variances of the parameter estimates, and the covariances between them, are given by the elements of the inverse of the  $p \times p$  matrix with entries  $-\partial^2 \ell / \partial \theta_a \partial \theta_b$ , where  $\ell$  represents the log-likelihood function.

The necessary first and second partial derivatives to estimate the model parameters for data of type 4 are given in Appendix 3. For data of type 1 and type 3, the following results are also necessary.

Assuming that the starting dates of primary moult are normally distributed and defining  $g(x_h) = \frac{(x_h - \mu_h)}{\sigma}$ , then

$$Q(u_h) = \Phi(g(u_h)) - \Phi(g(u_h - \tau)).$$

For the first partial derivatives of  $\ell_3$  and  $\ell_4$  we have

$$\begin{aligned}
\frac{\partial Q(u_{hj})}{\partial \tau} &= \frac{1}{\sigma} \phi(g(u_{hj} - \tau)) \\
\frac{\partial Q(u_{hj})}{\partial \mu_s} &= \begin{cases} \frac{1}{\sigma} \phi(g(u_{sj} - \tau)) - \frac{1}{\sigma} \phi(g(u_{sj})) & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\
\frac{\partial Q(u_{hj})}{\partial \sigma} &= \frac{g(u_{hj} - \tau)}{\sigma} \phi(g(u_{hj} - \tau)) - \frac{g(u_{hj})}{\sigma} \phi(g(u_{hj})).
\end{aligned}$$

For the second partial derivatives we have

$$\begin{aligned}
\frac{\partial^2 Q(u_{hj})}{\partial \tau^2} &= \frac{g(u_{hj} - \tau)}{\sigma^2} \phi(g(u_{hj} - \tau)) \\
\frac{\partial^2 Q(u_{hj})}{\partial \tau \partial \mu_s} &= \begin{cases} \frac{g(u_{sj} - \tau)}{\sigma^2} \phi(g(u_{sj} - \tau)) & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\
\frac{\partial^2 Q(u_{hj})}{\partial \tau \partial \sigma} &= \frac{g^2(u_{hj} - \tau) - 1}{\sigma^2} \phi(g(u_{hj} - \tau)) \\
\frac{\partial^2 Q(u_{hj})}{\partial \mu_s \partial \mu_r} &= \begin{cases} \frac{g(u_{sj} - \tau)}{\sigma^2} \phi(g(u_{sj} - \tau)) - \frac{g(u_{sj})}{\sigma^2} \phi(g(u_{sj})) & r = s \\ 0 & r \neq s \end{cases}, \quad s, r = 1, \dots, H \\
\frac{\partial^2 Q(u_{hj})}{\partial \mu_s \partial \sigma} &= \begin{cases} \frac{g^2(u_{sj} - \tau) - 1}{\sigma^2} \phi(g(u_{sj} - \tau)) - \frac{g^2(u_{sj}) - 1}{\sigma^2} \phi(g(u_{sj})) & s = h \\ 0 & s \neq h \end{cases},
\end{aligned}$$

$$\frac{\partial^2 Q(u_{hj})}{\partial \sigma^2} = \begin{array}{c} \text{for } s = 1, \dots, H \\ \frac{g^3(u_{hj} - \tau) - 2g(u_{hj} - \tau)}{\sigma^2} \phi(g(u_{hj} - \tau)) - \frac{g^3(u_{hj}) - 2g(u_{hj})}{\sigma^2} \phi(g(u_{hj})) \end{array}.$$

## APPENDIX 2

**Newton-Raphson iterative algorithm to estimate model parameters.**

The Newton-Raphson algorithm is a technique of iteratively solving nonlinear equations of several variables. We define  $\hat{\theta}^{(k)}$  to be the vector of parameter estimates after  $k$  iterations. The algorithm consists of the following steps:

- Step 1:** Calculate initial estimates  $\hat{\theta}^{(0)} = (\hat{\theta}_1^{(0)}, \dots, \hat{\theta}_p^{(0)})$ .
- Step 2:** Compute  $f^{(k)}$  and  $F^{(k)}$ , where  $f^{(k)}$  is the vector of first partial derivatives and  $F^{(k)}$  is the matrix of second partial derivatives, computed at the  $k$ th iteration.
- Step 3:** Compute the vector  $\delta^{(k)}$  which is the solution to the system of  $p$  linear equations

$$F^{(k)}\delta^{(k)} = f^{(k)},$$

where  $p$  represents the number of parameters.

- Step 4:** Set  $\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \delta^{(k)}$ , where  $\hat{\theta}^{(k)}$  contains the parameter estimates at the  $k$ th iteration.
- Step 5:** Test for convergence, for example, if the elements of  $f^{(k)}$  are sufficiently close to zero. If the convergence criterion is met then stop, otherwise increase  $k$  by 1 and return to step 2.

## APPENDIX 3

**First and second partial derivatives for primary moult model.**

If we assume that the starting dates are distributed normally, then the following formulae are required to apply the moult model. For the case of the normal distribution, the model parameters are given by  $\theta_h = \mu_h$ ,  $h = 1, 2, \dots, H$ , the mean starting dates,  $\theta_{H+1} = \tau$ , the duration of moult, and  $\theta_{H+2} = \sigma$ , the standard deviation of the starting date. Here  $F_T(t) = \Phi\left(\frac{t - \mu_h}{\sigma}\right)$ , where  $\Phi(x)$  is the standard normal distribution function, that is

$$\Phi(x) = \int_{-\infty}^x \phi(z) dz,$$

where

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp \frac{-z^2}{2}, \quad -\infty < z < \infty.$$

Define  $g(x_h) = \frac{(x_h - \mu_h)}{\sigma}$ . Then

$$P(t_h) = 1 - \Phi(g(t_h))$$

$$R(v_h) = \Phi(g(v_h - \tau))$$

$$q(u_h, y_h) = \frac{\tau}{\sigma} \phi(g(u_h - y_h \tau)).$$

For the first partial derivatives of  $\ell_1$  and  $\ell_2$  we have

$$\begin{aligned} \frac{\partial P(t_{hi})}{\partial \tau} &= 0 \\ \frac{\partial P(t_{hi})}{\partial \mu_s} &= \begin{cases} \frac{1}{\sigma} \phi(g(t_{si})) & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\ \frac{\partial P(t_{hi})}{\partial \sigma} &= \frac{g(t_{hi})}{\sigma} \phi(g(t_{hi})) \\ \frac{\partial R(v_{hk})}{\partial \tau} &= -\frac{1}{\sigma} \phi(g(v_{hk} - \tau)) \\ \frac{\partial R(v_{hk})}{\partial \mu_s} &= \begin{cases} -\frac{1}{\sigma} \phi(g(v_{sk} - \tau)) & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\ \frac{\partial R(v_{hk})}{\partial \sigma} &= -\frac{g(v_{hk} - \tau)}{\sigma} \phi(g(v_{hk} - \tau)) \\ \frac{\partial q(u_{hj}, y_{hj})}{\partial \tau} &= q(u_{hj}, y_{hj}) \left[ \frac{1}{\tau} + \frac{y_{hj} g(u_{hj} - y_{hj} \tau)}{\sigma} \right] \\ \frac{\partial q(u_{hj}, y_{hj})}{\partial \mu_s} &= \begin{cases} q(u_{sj}, y_{sj}) \frac{g(u_{sj} - y_{sj} \tau)}{\sigma} & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\ \frac{\partial q(u_{hj}, y_{hj})}{\partial \sigma} &= q(u_{hj}, y_{hj}) \frac{g^2(u_{hj} - y_{hj} \tau) - 1}{\sigma}. \end{aligned}$$

For the second partial derivatives we have

$$\begin{aligned}
\frac{\partial^2 P(t_{hi})}{\partial \tau^2} &= 0 \\
\frac{\partial^2 P(t_{hi})}{\partial \tau \partial \mu_s} &= 0, \quad s = 1, \dots, H \\
\frac{\partial^2 P(t_{hi})}{\partial \tau \partial \sigma} &= 0 \\
\frac{\partial^2 P(t_{hi})}{\partial \mu_s \partial \mu_r} &= \begin{cases} \frac{g(t_{si})}{\sigma^2} \phi(g(t_{si})) & r = s \\ 0 & r \neq s \end{cases}, \quad s, r = 1, \dots, H \\
\frac{\partial^2 P(t_{hi})}{\partial \mu_s \partial \sigma} &= \begin{cases} \frac{g^2(t_{si}) - 1}{\sigma^2} \phi(g(t_{si})) & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\
\frac{\partial^2 P(t_{hi})}{\partial \sigma^2} &= \frac{g^3(t_{hi}) - 2g(t_{hi})}{\sigma^2} \phi(g(t_{hi})) \\
\frac{\partial^2 R(v_{hk})}{\partial \tau^2} &= -\frac{g(v_{hk} - \tau)}{\sigma^2} \phi(g(v_{hk} - \tau)) \\
\frac{\partial^2 R(v_{hk})}{\partial \tau \partial \mu_s} &= \begin{cases} -\frac{g(v_{sk} - \tau)}{\sigma^2} \phi(g(v_{sk} - \tau)) & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\
\frac{\partial^2 R(v_{hk})}{\partial \tau \partial \sigma} &= -\frac{g^2(v_{hk} - \tau) - 1}{\sigma^2} \phi(g(v_{hk} - \tau)) \\
\frac{\partial^2 R(v_{hk})}{\partial \mu_s \partial \mu_r} &= \begin{cases} -\frac{g(v_{sk} - \tau)}{\sigma^2} \phi(g(v_{sk} - \tau)) & r = s \\ 0 & r \neq s \end{cases}, \quad s, r = 1, \dots, H \\
\frac{\partial^2 R(v_{hk})}{\partial \mu_s \partial \sigma} &= \begin{cases} -\frac{g^2(v_{sk} - \tau) - 1}{\sigma^2} \phi(g(v_{sk} - \tau)) & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H \\
\frac{\partial^2 R(v_{hk})}{\partial \sigma^2} &= -\frac{g^3(v_{hk} - \tau) - 2g(v_{hk} - \tau)}{\sigma^2} \phi(g(v_{hk} - \tau)) \\
\frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \tau^2} &= q(u_{hj}, y_{hj}) \left[ \frac{y_{hj}^2 [g^2(u_{hj} - y_{hj}\tau) - 1]}{\sigma^2} + \frac{2y_{hj}g(u_{hj} - y_{hj}\tau)}{\tau\sigma} \right] \\
\frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \tau \partial \mu_s} &= \begin{cases} q(u_{sj}, y_{sj}) \left[ \frac{y_{sj} [g^2(u_{sj} - y_{sj}\tau) - 1]}{\sigma^2} + \frac{g(u_{sj} - y_{sj}\tau)}{\tau\sigma} \right] & s = h \\ 0 & s \neq h \end{cases}, \\
&\text{for } s = 1, \dots, H \\
\frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \tau \partial \sigma} &= \frac{q(u_{hj}, y_{hj})}{\sigma^2} \left[ -3y_{hj}g(u_{hj} - y_{hj}\tau) + \frac{\sigma [g^2(u_{hj} - y_{hj}\tau) - 1]}{\tau} + y_{hj}g^3(u_{hj} - y_{hj}\tau) \right] \\
\frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \mu_s \partial \mu_r} &= \begin{cases} q(u_{sj}, y_{sj}) \left[ \frac{g^2(u_{sj} - y_{sj}\tau) - 1}{\sigma^2} \right] & s = r \\ 0 & s \neq r \end{cases}, \quad s, r = 1, \dots, H
\end{aligned}$$

$$\frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \mu_s \partial \sigma} = \begin{cases} q(u_{sj}, y_{sj}) \left[ \frac{-3g(u_{sj} - y_{sj}\tau) + g^3(u_{sj} - y_{sj}\tau)}{\sigma^2} \right] & s = h \\ 0 & s \neq h \end{cases}, \quad s = 1, \dots, H$$

$$\frac{\partial^2 q(u_{hj}, y_{hj})}{\partial \sigma^2} = q(u_{hj}, y_{hj}) \left[ \frac{-5g^2(u_{hj} - y_{hj}\tau) + g^4(u_{hj} - y_{hj}\tau) + 2}{\sigma^2} \right].$$



## The analysis of avian primary moult IV: flexible model fitting using the EM algorithm

### ABSTRACT

Various extensions of the Underhill-Zucchini moult model (*Ibis* (1988) 130: 358–372 and (1990) 132: 118–123) are developed to allow a flexible and statistically coherent approach to the analysis of variation in the timing and duration of moult in different data sets. The moult indices are constrained to lie between 0 and 1 and as such they are an example of censored observations from both the left and the right. The techniques of censored regression and the expectation-maximization (EM) algorithm (Schmid *et al.*, *The American Naturalist* (1994) 143: 494–507) are used to facilitate the extensions to the moult model. Standard statistical hypothesis testing methods are developed to test the goodness of fit of each model. In the application considered, four models are fitted to the primary moult of Willow Warblers *Phylloscopus trochilus* in the St Petersburg region of western Russia. Each model makes different assumptions about the parameters of moult by allowing none, some or all the parameters to vary between years or localities. Under the assumption of constant standard deviation, the appropriate model for moult when the sex of the birds is ignored, assumes a constant duration of moult but variable annual mean starting dates of moult. For males and females, the appropriate moult model assumes variable annual mean starting dates and duration of primary moult. With exceptions, the mean annual starting dates for males and females run closely in parallel, with males starting moult, on average, about 14 days before females. With a few exceptions, annual duration of moult is shorter in males than in females (about 12 days shorter).

### INTRODUCTION

As an initial extension to the models for primary moult first suggested by Underhill and Zucchini (1988) and Underhill *et al.* (1990), only the mean starting dates of moult were considered to vary from year to year in chapter 5. As a further extension, it is conceivable

that both the mean starting date of moult and the duration of moult might vary from year to year. This also permits us to investigate the plausibility of the assumption of constant duration of moult throughout the years. In fact, it would also be of interest to test the assumption of a constant standard deviation throughout the years. This chapter describes extensions to the model for primary moult to enable testing of hypotheses about the parameters of the model for primary moult.

The statistical model for the analysis of primary moult, as formulated by Underhill and Zucchini (1988), is an example of a censored regression model. Applications of censored regression models in biology were reviewed by Schmid *et al.* (1994) who discussed mainly botanical examples. In particular, they considered a situation from plant reproductive biology in which the reproductive mass is zero until the vegetative mass reaches a threshold. Their goal was to estimate this mean threshold mass and to relate the reproductive mass to vegetative mass. This and their other examples were all of "one-sided" censoring. The moult problem is an example of a censored regression model with two-sided censoring because the moult index is constrained to lie between 0 and 1. In the moult problem, the "thresholds" are the mean dates at which moult commences and finishes.

Schmid *et al.* (1994) derived the analogous likelihood function for one-sided censoring to the function that Underhill and Zucchini (1988) used for two-sided censoring, but did not attempt to maximize the likelihood function using differentiation and the Newton-Raphson algorithm (Ortega and Rheinboldt 1970). Instead, they made use of the EM (expectation-maximization) algorithm (Dempster *et al.* 1977, Little and Rubin 1987) to maximise the likelihood function. This approach is applied to the present problem of extending the model for primary moult.

## MATERIAL AND METHODS

The data for this study are the same as that used in the extension of the Underhill-Zucchini moult model (Underhill and Zucchini 1988, Underhill *et al.* 1990). That is, the data set consist of post-nuptial moult indices for adult Willow Warblers *Phylloscopus trochilus* captured in mist-nets at the south-eastern coast of the Lake Ladoga, St Petersburg region, Russia (60°47'N, 32°48'E). Most birds were sexed on wing length, as in Svensson (1992) and Underhill *et al.* (1992). The data contain observations made in 11 years, from 1979 to 1989 (Table 1). Primary moult scores were converted to percentage feather mass grown us-

ing the method of Underhill *et al.* (1992). The percentage feather mass grown (as described by Underhill *et al.* 1992) was used as the moult index.

**Table 1.** Sample sizes for Willow Warbler primary moult data at Lake Ladoga, St Petersburg, Russia.

Year	All†			Male			Female		
	Yet to moult	In moult	Ended moult	Yet to moult	In moult	Ended moult	Yet to moult	In moult	Ended moult
1979	39	30	0	12	12	0	27	11	0
1980	30	26	5	7	13	3	23	12	1
1981	44	29	13	14	9	8	30	16	5
1982	56	40	31	23	17	18	33	17	13
1983	57	27	2	15	14	1	42	11	1
1984	25	13	1	6	5	1	19	5	0
1985	34	10	5	9	4	2	25	4	3
1986	29	14	8	11	8	6	18	5	2
1987	29	17	3	5	12	1	24	3	2
1988	35	20	8	12	9	6	23	9	2
1989	37	37	17	10	15	14	27	14	3
Totals	415	263	93	124	118	60	291	107	32

†The sex of 39 birds was not recorded.

The model for primary moult consists of three parameters, that is, the mean starting date, the duration and the standard deviation of the starting date of moult. For data of primary moult accumulated over a number of data sets (i.e. over several years or from various places), one would like to be able to answer a number of questions, namely:

1. Do all the parameters in the model for primary moult remain constant across the data sets?
2. Does the standard deviation of the mean starting date remain constant over the data sets?
3. Does the duration of moult remain constant?

4. Does the mean starting date of moult vary from data set to data set and if they vary, then how do they vary?

Hypotheses about the parameters of primary moult can be tested by specifying each hypothesis in terms of a model and then finding some measure for comparing each model. To answer the above questions, four models will be fitted to the primary moult data. The "minimal" model (model 1) is the model in which all the parameters of primary moult remain constant throughout the data sets. Model 2 is the model in which both the standard deviation of mean starting date and the duration of moult remain constant throughout the data sets, but the mean starting date differs between data sets. Model 3 is the model in which the standard deviation of mean starting date is a constant but mean starting date and duration of moult vary from data set to data set. The "fullest" model (referred to as model 4) of interest will be the one in which all parameters change in each data set.

In the application to the Willow Warbler data, to allow the duration of moult to vary from year to year implies an increment in the number of parameters to be estimated from 13, in the case where only the mean starting dates vary yearly, to 23, if both vary. The Newton-Raphson algorithm to obtain maximum likelihood estimates of the model parameters requires the calculation and programming of the first partial derivatives and the matrix of second partial derivatives. The adaptation to the model for primary moult mentioned above, with an increment of 10 parameters, requires the computation of an extra 11 first partial derivatives and the number of second partial derivatives increases from 91 to 276 (since the matrix of second partial derivatives is symmetrical). With a view to facilitate the procedure of obtaining maximum likelihood estimates of the parameters for the extended moult model an alternative approach of using censored regression techniques and the expectation-maximization (EM) algorithm to solve the likelihood function is investigated (Schmid *et al.* 1994). A brief description of the Newton-Raphson and EM algorithms, together with some advantages and disadvantages of both techniques is given in Appendix 1.

Observations are said to be censored if the value of the random variable under investigation is unobserved for some items in the sample (Johnson *et al.* 1985). The moult index in this study has been scaled so that it is confined between 0 (not yet started moult) and 1 (completed moult). Thus those observations for which "no moult index" can be observed because birds have either finished moulting or have not started to moult are considered to be censored data. If moult has not yet started, then these observations are defined to be censored from the left, while if moult has completed, then these observations are defined to

be censored from the right. Marrying the definition of censored data to the different data types distinguished by Underhill and Zucchini (1988) and Underhill *et al.* (1990), it follows that data type 2 (Table 2) is censored from the right and the left. Both data type 4 and data type 5 are censored from the left and the right, as the moult index is still confined to lie between 0 and 1, but for data type 4, the censored observations on the left are not observed because birds that have not started moult are still on migration and have not arrived in the sampling area, while for data type 5, the censored observations on the right are not observed because birds migrate out of the sampling area soon after they have completed moult.

**Table 2.** The components of the population from which representative samples are required on each sampling date for the five data types presented in Underhill and Zucchini (1988) and Underhill *et al.* (1990).

Data type	Component of the population for which no			immigration or emigration during the sampling period is assumed
	Not yet moulted	In moult	Completed moult	
1†	yes	yes	yes	all birds
2	yes	yes	yes	all birds
3	no	yes	no	birds in moult
4	no	yes	yes	birds in moult or completed moult
5	yes	yes	no	birds not yet moulted or in moult

†For data type 1, it is only necessary to classify birds as 'not yet moulted', 'in moult' or 'completed moult'.

The censored regression models developed for primary moult will only deal with data type 5 of Underhill *et al.* (1990) and data type 2 of Underhill *et al.* (1988) for the same reasons as discussed in the previous chapter.

### Censored regression models for avian primary moult

In the previous chapter, the model for primary moult is constructed by estimating the probability distribution function of the random variable  $T$ , where  $T$  represents the date at which a randomly selected bird starts primary moult, and the duration of moult, assuming

that the moult index increases linearly from the start to the end of moult. Thus the parameters of interest for primary moult are the mean starting date of moult ( $\mu$ ), the standard deviation of the starting date of moult ( $\sigma$ ), and the duration of moult ( $\tau$ ). Although there are no actual observations of the random variable  $T$ , a relationship between the random variable  $T$  and the random variable  $Y(t)$ , the moult index, makes it possible to estimate the probability distribution function,  $F_T(t)$  from observations on the random variable  $Y(t)$  (see Underhill and Zucchini 1988). In the censored regression procedure, the relationship between the random variable  $Y(t)$ , the moult index, and  $T$ , time, is modelled as a simple linear regression but the information available in the censored points is incorporated into the model (Johnson and Kotz 1970, Wolynetz 1979a, 1979b, Aitkin 1981, Schneider and Weissfeld 1986, Polasek and Krause 1994, Schmid *et al.* 1994). The parameters of a linear regression model are the slope of the regression line ( $\beta$ ), the  $y$ -intercept ( $\alpha$ ) and the standard deviation of the regression ( $\sigma_\epsilon$ ). These parameters can easily be related to the parameters of the previous formulation of the model for primary moult in that the  $x$ -intercept ( $-\alpha/\beta$ ) of the regression line is the same as the mean starting date ( $\mu$ ) and the duration of moult is simply given by  $\tau = 1/\beta$ . The standard deviation of the mean starting date ( $\sigma$ ) can be related to the standard deviation of the regression ( $\sigma_\epsilon$ ) by the fact that  $\sigma = \sigma_\epsilon/\beta$  for a one-to-one relationship. The invariance property of maximum likelihood estimation states that if  $\omega(\theta)$  is any function of the parameters  $\theta$ , then the maximum likelihood estimator of  $\omega$  is

$$\hat{\omega} = \omega(\hat{\theta}).$$

As a consequence of this property we can obtain maximum likelihood estimates for the parameterization of the primary moult model that is most convenient and then use the invariance property to obtain maximum likelihood estimates for another parameterization (Cox and Hinkley 1974).

## Censored regression models for data type 2

### Model 1

The development of the censored regression model for primary moult will be described in detail for data type 2 and model 1, that is for data censored from the left and the right and for a constant mean starting date of moult, a constant standard deviation of the

mean starting date of moult, and a constant duration of moult (equivalently, a common  $x$ -intercept, a constant standard deviation of the regression line, and a common slope).

Let  $d = (d_1, d_2, \dots, d_n)$ , where  $n$  is the total number of observations, denote the vector of days in which observations on primary moult are made. Adopting the same notation as in Underhill and Zucchini (1988), this vector can be partitioned into a vector that consist of days in which birds have not commenced moult denoted by  $t = (t_1, t_2, \dots, t_I)$ , one in which birds are in moult denoted by  $u = (u_1, u_2, \dots, u_J)$  and one in which birds have completed moult denoted by  $v = (v_1, v_2, \dots, v_K)$ . For those birds which are actively in moult, the vector of moult indices is denoted as  $y = (y_1, y_2, \dots, y_J)$ .

Define a normally distributed latent random variable  $y^*$ , where

$$y_l^* = \alpha + \beta d_l + \varepsilon_l, \quad \varepsilon_l \sim N(0, \sigma_\varepsilon^2), \quad l = 1, 2, \dots, n,$$

such that negative values of the dependent latent variable have a moult index  $y_l$  equal to zero, values greater than one have a moult index equal to one, otherwise it is equal to the observed moult index  $y_l$ . That is, the relation between the observed dependent variable  $y$  and the latent variable  $y^*$  is given by

$$y_l = \begin{cases} 1 & \text{if } y_l^* \geq 1 & \text{(censored on the right)} \\ y_l^* & \text{if } 0 < y_l^* < 1 & \text{(uncensored)} \\ 0 & \text{if } y_l^* \leq 0 & \text{(censored on the left),} \end{cases}$$

for  $l = 1, 2, \dots, n$ .

The likelihood function of the censored regression for model 1 for data type 2 is given by

$$L(\theta; t, u, y, v) = \prod_{i=1}^I \left\{ 1 - \Phi \left( \frac{\alpha + \beta t_i}{\sigma_\varepsilon} \right) \right\} \prod_{j=1}^J \left\{ \frac{1}{\sigma_\varepsilon} \phi \left( \frac{y_j - (\alpha + \beta u_j)}{\sigma_\varepsilon} \right) \right\} \prod_{k=1}^K \left\{ 1 - \Phi \left( \frac{1 - (\alpha + \beta v_k)}{\sigma_\varepsilon} \right) \right\}, \quad (1)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = 3$ . The functions  $\phi(x)$  and  $\Phi(x)$  are the density and distribution function of a standard normal distribution respectively.

The maximum likelihood estimates of the parameters are those values of  $\theta$  which maximize (1), or equivalently its logarithm

$$\ell_1 = \sum_{i=1}^I \log \left( 1 - \Phi \left( \frac{\alpha + \beta t_i}{\sigma_\varepsilon} \right) \right) + \sum_{j=1}^J \left\{ \log \left( \frac{1}{\sigma_\varepsilon} \right) + \log \left( \phi \left( \frac{y_j - (\alpha + \beta u_j)}{\sigma_\varepsilon} \right) \right) \right\}$$

$$+ \sum_{k=1}^K \log \left( 1 - \Phi \left( \frac{1 - (\alpha + \beta v_k)}{\sigma_\varepsilon} \right) \right), \quad (2)$$

where  $\ell_1 = \log L(\theta; t, u, y, v)$ .

This function however, cannot be maximized explicitly and therefore has to be solved numerically. The expectation-maximization (EM) iterative algorithm (Dempster *et al.* 1977) is used to maximize (2). The E-step consists of computing the conditional expectation of the censored values given the observed data and the current estimated parameters, and then replacing the censored values of the dependent variable by the conditional expectations which are given by

$$\begin{aligned} E \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] &= \hat{y}_i - \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right), \quad i = 1, 2, \dots, I \\ E \left[ y_k^* \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] &= \hat{y}_k + \hat{\sigma}_\varepsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\varepsilon} \right), \quad k = 1, 2, \dots, K, \end{aligned}$$

where  $h(x) = \frac{\phi(x)}{1 - \Phi(x)}$  and  $\hat{y}_i = \hat{\alpha} + \hat{\beta}t_i$  for birds that have not started moult and  $\hat{y}_k = \hat{\alpha} + \hat{\beta}v_k$  for birds that have completed moult.

The M-step consists of maximizing the log-likelihood function for the complete data, that is for the data obtained in the E-step in which the censored values have been replaced by their conditional expectation. This is the ordinary least-squares estimates of the linear regression model. The estimate for the variance is (Schmid *et al.* 1994)

$$n\hat{\sigma}_\varepsilon^2 = \left[ \sum_{i=1}^I \text{Var} \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] + \sum_{l=1}^n (y_l - \hat{y}_l)^2 + \sum_{k=1}^K \text{Var} \left[ y_k^* \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] \right],$$

where

$$\begin{aligned} \text{Var} \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] &= E \left[ (y_i - \hat{y}_i)^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] \\ &= \hat{\sigma}_\varepsilon^2 + \hat{y}_i \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right) - \hat{\sigma}_\varepsilon^2 h^2 \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right) \\ \text{Var} \left[ y_k^* \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] &= E \left[ (y_k - \hat{y}_k)^2 \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] \\ &= \hat{\sigma}_\varepsilon^2 + \hat{\sigma}_\varepsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\varepsilon} \right) - \hat{y}_k \hat{\sigma}_\varepsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\varepsilon} \right) - \hat{\sigma}_\varepsilon^2 h^2 \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\varepsilon} \right). \end{aligned}$$

The derivation of the conditional means and the conditional variances of the censored normal distribution for data type 2 and data type 5 are given in Appendix 2.

The two steps of the EM algorithm (E and M steps) are repeated until some convergence criterion is reached. As the computation of the matrix of second partial derivatives is not necessary when using the EM algorithm for parameter estimation, approximate standard

errors of the estimates are not automatically available as with the Newton-Raphson algorithm. However, approximate standard errors can be obtained by computing the second partial derivatives numerically (Dennis and Schnabel 1983, Press *et al.* 1992). The variances of the parameter estimates, and the covariances between them, are then given by the elements of the inverse of the  $p \times p$  matrix with entries  $-\partial^2 \ell / \partial \theta_a \partial \theta_b$ , where  $\ell$  represents the log-likelihood function. The parameter estimates dealt with in the present formulation are  $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon)$ . These parameter estimates can then be transformed to obtain the desired parameter estimates  $(\hat{\mu}, \hat{\tau}, \hat{\sigma})$ , and the following result (Mood *et al.* 1963) can be used to obtain the variances of the converted parameter estimates. Let  $X$  and  $Y$  be two random variables; then

$$\text{Var}(X/Y) \approx \left( \frac{\mu_x}{\mu_y} \right)^2 \left( \frac{\text{Var}(X)}{\mu_x^2} + \frac{\text{Var}(Y)}{\mu_y^2} - 2 \frac{\text{Cov}(X, Y)}{\mu_x \mu_y} \right). \quad (3)$$

## Model 2

As an extension to model 1 we consider the case when the mean starting date of primary moult varies annually, or equivalently we consider the regression line with variable  $y$ -intercepts for each year. Define the vector  $d_h = (d_{h1}, d_{h2}, \dots, d_{hn_h})$ , where  $n_h$  represents the total number of observations in each of the  $H$  data sets, and  $h = 1, 2, \dots, H$ . Adopting the same notation as in the previous chapter, this vector can be partitioned into a vector that consist of days in which birds have not commenced moult which is denoted by

$$t = t_{11}, t_{12}, \dots, t_{1J_1}, t_{21}, t_{22}, \dots, t_{2J_2}, \dots, t_{H1}, t_{H2}, \dots, t_{HJ_H},$$

one vector in which birds are in moult, denoted by

$$u = u_{11}, u_{12}, \dots, u_{1J_1}, u_{21}, u_{22}, \dots, u_{2J_2}, \dots, u_{H1}, u_{H2}, \dots, u_{HJ_H},$$

and one vector in which birds have completed moult, denoted by

$$v = v_{11}, v_{12}, \dots, v_{1K_1}, v_{21}, v_{22}, \dots, v_{2K_2}, \dots, v_{H1}, v_{H2}, \dots, v_{HK_H}.$$

For those birds which are actively in moult, the vector of moult indices is denoted as

$$y = (y_{11}, y_{12}, \dots, y_{1J_1}, y_{21}, y_{22}, \dots, y_{2J_2}, \dots, y_{H1}, y_{H2}, \dots, y_{HJ_H}).$$

Define a normally distributed latent random variable  $y^*$ , where

$$y_{hl}^* = \alpha_h + \beta d_{hl} + \varepsilon_{hl}, \quad \varepsilon_{hl} \sim N(0, \sigma_\varepsilon^2), \quad l = 1, 2, \dots, n_h, \quad h = 1, 2, \dots, H.$$

The relation between the observed dependent variable  $y$  and the latent variable  $y^*$  is given by

$$y_{hl} = \begin{cases} 1 & \text{if } y_{hl}^* \geq 1 & \text{(censored on the right)} \\ y_{hl}^* & \text{if } 0 < y_{hl}^* < 1 & \text{(uncensored)} \\ 0 & \text{if } y_{hl}^* \leq 0 & \text{(censored on the left),} \end{cases}$$

for  $l = 1, 2, \dots, n_h$  and  $h = 1, 2, \dots, H$ .

The likelihood function of the censored regression for model 2 is

$$L(\theta; t, u, y, v) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} \left\{ 1 - \Phi \left( \frac{\alpha_h + \beta t_{hi}}{\sigma_\varepsilon} \right) \right\} \prod_{j=1}^{J_h} \left\{ \frac{1}{\sigma_\varepsilon} \phi \left( \frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon} \right) \right\} \right. \\ \left. \prod_{k=1}^{K_h} \left\{ 1 - \Phi \left( \frac{1 - (\alpha_h + \beta v_{hk})}{\sigma_\varepsilon} \right) \right\} \right), \quad (4)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = H + 2$ .

Maximum likelihood estimates of the parameters  $\theta$  are again obtained by maximizing the log-likelihood function (i.e. the logarithm of equation 4) using the EM algorithm. For this we need the conditional means and the conditional variances of the censored normal distribution which are given by

$$\begin{aligned} E[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] &= \widehat{y}_{hi} - \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) \\ E[y_{hk}^* | y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] &= \widehat{y}_{hk} + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) \\ \text{Var}[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) \\ \text{Var}[y_{hk}^* | y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] &= \widehat{\sigma}_\varepsilon^2 + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{y}_{hk} \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) \\ &\quad - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right), \end{aligned}$$

where  $\widehat{y}_{hi} = \widehat{\alpha}_h + \widehat{\beta} t_{hi}$  for birds that have not commenced moult and  $\widehat{y}_{hk} = \widehat{\alpha}_h + \widehat{\beta} v_{hk}$  for birds that have completed moult, for  $i = 1, 2, \dots, I_h$ ,  $k = 1, 2, \dots, K_h$  and  $h = 1, 2, \dots, H$ .

The E and M-steps are the same as in the preceding section using the according conditional expectations and variances, where the estimate for the variance is now given by

$$n \widehat{\sigma}_\varepsilon^2 = \sum_{h=1}^H \left[ \sum_{i=1}^{I_h} \text{Var}[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] + \sum_{l=1}^{n_h} (y_{hl} - \widehat{y}_{hl})^2 \right. \\ \left. + \sum_{k=1}^{K_h} \text{Var}[y_{hk}^* | y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] \right].$$

**Model 3**

In model 3 both the mean starting date and the duration of primary moult differ from year to year, while the standard deviation of mean starting date is constant. That is in the regression formulation, both the  $y$ -intercept and the slope vary in each year. In this instance the normally distributed latent random variable  $y^*$  is defined as

$$y_{hl}^* = \alpha_h + \beta_h d_{hl} + \varepsilon_{hl}, \quad \varepsilon_{hl} \sim N(0, \sigma_\varepsilon^2), \quad l = 1, 2, \dots, n_h, \quad h = 1, 2, \dots, H.$$

The relation between the observed dependent variable  $y$  and the latent variable  $y^*$  is the same as for model 2.

The likelihood function of the censored regression for model 3 is

$$L(\theta; t, u, y, v) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} \left\{ 1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\varepsilon} \right) \right\} \prod_{j=1}^{J_h} \left\{ \frac{1}{\sigma_\varepsilon} \phi \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon} \right) \right\} \right. \\ \left. \prod_{k=1}^{K_h} \left\{ 1 - \Phi \left( \frac{1 - (\alpha_h + \beta_h v_{hk})}{\sigma_\varepsilon} \right) \right\} \right), \quad (5)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters and  $p$ , the total number of parameters in the model, is given by  $p = 2H + 1$ .

Maximum likelihood estimates of the parameters  $\theta$  are again obtained by maximizing the log-likelihood function (i.e. the logarithm of equation 5) using the EM algorithm. For this we need the conditional means and the conditional variances of the censored normal distribution which are given by

$$\begin{aligned} E[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] &= \widehat{y}_{hi} - \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) \\ E[y_{hk}^* | y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] &= \widehat{y}_{hk} + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) \\ \text{Var}[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) \\ \text{Var}[y_{hk}^* | y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] &= \widehat{\sigma}_\varepsilon^2 + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{y}_{hk} \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) \\ &\quad - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right), \end{aligned}$$

where  $\widehat{y}_{hi} = \widehat{\alpha}_h + \widehat{\beta}_h t_{hi}$  for birds that have not commenced moult and  $\widehat{y}_{hk} = \widehat{\alpha}_h + \widehat{\beta}_h v_{hk}$  for birds that have completed moult for  $i = 1, 2, \dots, I_h$ ,  $k = 1, 2, \dots, K_h$ , and  $h = 1, 2, \dots, H$ .

The E and M-steps are the same as in the preceding section using the according conditional expectations and variances, where the estimate for the variance is now given by

$$n \widehat{\sigma}_\varepsilon^2 = \sum_{h=1}^H \left[ \sum_{i=1}^{I_h} \text{Var}[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] + \sum_{l=1}^{n_h} (y_{hl} - \widehat{y}_{hl})^2 \right]$$

$$+ \sum_{k=1}^{K_h} \text{Var} \left[ y_{hk}^* \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right].$$

#### Model 4

In model 4 all the parameters of primary moult vary annually. In essence, this model is model 1 applied to each individual data set. For completeness, the formulation of the likelihood function for model 4 is given here but the procedure to obtain maximum likelihood estimates of the parameters is that of model 1 applied to each of the  $H$  data sets.

Define the latent random variable  $y^*$  as

$$y_{hl}^* = \alpha_h + \beta_h d_{hl} + \varepsilon_{hl}, \quad \varepsilon_{hl} \sim N(0, \sigma_{\varepsilon_h}^2), \quad l = 1, 2, \dots, n_h, \quad h = 1, 2, \dots, H,$$

where the relation between the observed dependent variable  $y$  and the latent variable  $y^*$  is the same as that for model 2.

The likelihood function of the censored regression for model 4 for data type 2 is

$$L(\theta; t, u, y, v) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} \left\{ 1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_{\varepsilon_h}} \right) \right\} \prod_{j=1}^{J_h} \left\{ \frac{1}{\sigma_{\varepsilon_h}} \phi \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_{\varepsilon_h}} \right) \right\} \right. \\ \left. \prod_{k=1}^{K_h} \left\{ 1 - \Phi \left( \frac{1 - (\alpha_h + \beta_h v_{hk})}{\sigma_{\varepsilon_h}} \right) \right\} \right), \quad (6)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = 3H$ .

### Censored regression models for data type 5

#### Model 1

With data type 5 we have the following scenario: birds are present in the sampling area prior to starting, and during moult, but migrate soon after they have completed moult. Thus, although the primary moult indices still fall under the category of observations censored from the left and the right, we no longer have a random sample of birds that have completed moult.

Define the vectors  $d$ ,  $t$ ,  $u$ , and  $y$  as in model 1 for data type 2 and define a normally distributed latent random variable  $y^*$  such that

$$y_l^* = \alpha + \beta d_l + \varepsilon_l, \quad \varepsilon_l \sim N(0, \sigma_\varepsilon^2), \quad l = 1, 2, \dots, n,$$

and is related to the observed dependent variable  $y$  by:

$$y_i = \begin{cases} y_i^* & \text{if } 0 < y_i^* < 1 & \text{(uncensored)} \\ 0 & \text{if } y_i^* \leq 0 & \text{(censored on the left).} \end{cases}$$

The likelihood function of the censored regression for model 1 is

$$L(\theta; t, u, y) = \prod_{i=1}^I \left\{ \frac{1 - \Phi\left(\frac{\alpha + \beta t_i}{\sigma_\epsilon}\right)}{\Phi\left(\frac{1 - (\alpha + \beta t_i)}{\sigma_\epsilon}\right)} \right\} \prod_{j=1}^J \left\{ \frac{\phi\left(\frac{y_j - (\alpha + \beta u_j)}{\sigma_\epsilon}\right)}{\sigma_\epsilon \Phi\left(\frac{1 - (\alpha + \beta u_j)}{\sigma_\epsilon}\right)} \right\}, \quad (7)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = 3$ .

The maximum likelihood estimates of the parameters are those values of  $\theta$  which maximize (7), or equivalently its logarithm

$$\begin{aligned} \ell_4 = & \sum_{i=1}^I \left( \log \left( 1 - \Phi \left( \frac{\alpha + \beta t_i}{\sigma_\epsilon} \right) \right) - \log \Phi \left( \frac{1 - (\alpha + \beta t_i)}{\sigma_\epsilon} \right) \right) \\ & + \sum_{j=1}^J \left( \log \frac{1}{\sigma_\epsilon} + \log \left( \phi \left( \frac{y_j - (\alpha + \beta u_j)}{\sigma_\epsilon} \right) \right) - \log \Phi \left( \frac{1 - (\alpha + \beta u_j)}{\sigma_\epsilon} \right) \right), \quad (8) \end{aligned}$$

where  $\ell_4 = \log L(\theta; t, u, y)$ .

This function cannot be solved explicitly and therefore has to be solved numerically. The expectation-maximization (EM) iterative algorithm is used to maximize (8). The conditional expectation and the conditional variance of the censored values given the observed data and the current estimated parameters are

$$\begin{aligned} E \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] &= \hat{y}_i - \hat{\sigma}_\epsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\epsilon} \right) \\ \text{Var} \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] &= E \left[ (y_i - \hat{y}_i)^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] \\ &= \hat{\sigma}_\epsilon^2 + \hat{y}_i \hat{\sigma}_\epsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\epsilon} \right) - \hat{\sigma}_\epsilon^2 h^2 \left( \frac{\hat{y}_i}{\hat{\sigma}_\epsilon} \right), \end{aligned}$$

where  $h(x) = \frac{\phi(x)}{1 - \Phi(x)}$ ,  $\hat{y}_i = \hat{\alpha} + \hat{\beta} t_i$ , and  $i = 1, 2, \dots, I$ .

In the E-step censored values of the dependent variable are replaced by their conditional expectation given the observed values and new parameter estimates are obtained in the M-step which consists of maximizing the log-likelihood function for the complete data by ordinary least-squares. The estimate for the variance is

$$n \hat{\sigma}_\epsilon^2 = \left[ \sum_{i=1}^I \text{Var} \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] + \sum_{l=1}^n (y_l - \hat{y}_l)^2 \right].$$

### Model 2

With model 2 we consider the case when the mean starting date of primary moult varies annually, or equivalently we consider the regression line with variable  $y$ -intercepts for each year. Define the vectors  $d$ ,  $t$ ,  $u$ , and  $y$  as in model 2 for data type 2 and define the normally distributed latent random variable  $y^*$  as

$$y_{hl}^* = \alpha_h + \beta d_{hl} + \varepsilon_{hl}, \quad \varepsilon_{hl} \sim N(0, \sigma_\varepsilon^2), \quad l = 1, 2, \dots, n_h, \quad h = 1, 2, \dots, H.$$

The latent variable  $y^*$  is related to the dependent variable  $y$  by

$$y_{hl} = \begin{cases} y_{hl}^* & \text{if } 0 < y_{hl}^* < 1 & \text{(uncensored)} \\ 0 & \text{if } y_{hl}^* \leq 0 & \text{(censored on the left)}. \end{cases}$$

The likelihood function of the censored regression for model 2 is

$$L(\theta; t, u, y) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} \left\{ \frac{1 - \Phi\left(\frac{\alpha_h + \beta t_{hi}}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{1 - (\alpha_h + \beta t_{hi})}{\sigma_\varepsilon}\right)} \right\} \prod_{j=1}^{J_h} \left\{ \frac{\phi\left(\frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon}\right)}{\sigma_\varepsilon \Phi\left(\frac{1 - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon}\right)} \right\} \right), \quad (9)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters and  $p$ , the total number of parameters in the model, is given by  $p = H + 2$ .

Maximum likelihood estimates of the parameters  $\theta$  are again obtained by maximizing the log-likelihood function (i.e. the logarithm of equation 9) using the EM algorithm. For this we need the conditional mean and the conditional variance of the censored normal distribution which are given by

$$\begin{aligned} E[y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] &= \widehat{y}_{hi} - \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) \\ \text{Var}[y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right), \end{aligned}$$

where  $\widehat{y}_{hi} = \widehat{\alpha}_h + \widehat{\beta} t_{hi}$ ,  $i = 1, 2, \dots, I_h$ , and  $h = 1, 2, \dots, H$ .

The E and M-steps are the same as in the preceding section using the according conditional expectations and variances, where the estimate for the variance is now given by

$$n \widehat{\sigma}_\varepsilon^2 = \sum_{h=1}^H \left[ \sum_{i=1}^{I_h} \text{Var}[y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon] + \sum_{l=1}^{n_h} (y_{hl} - \widehat{y}_{hl})^2 \right].$$

### Model 3

In model 3 both the mean starting date and the duration of primary moult differ from year to year. That is in the regression formulation, both the  $y$ -intercept and the slope of

the line vary from year to year. In this instance the normally distributed latent random variable  $y^*$  is defined as

$$y_{hl}^* = \alpha_h + \beta_h d_{hl} + \varepsilon_{hl}, \quad \varepsilon_{hl} \sim N(0, \sigma_\varepsilon^2), \quad l = 1, 2, \dots, n_h, \quad h = 1, 2, \dots, H,$$

and it is related to the observed variable  $y$  as in model 2 for data type 5.

The likelihood function of the censored regression for model 3 is

$$L(\theta; t, u, y) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} \left\{ \frac{1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\varepsilon} \right)}{\Phi \left( \frac{1 - (\alpha_h + \beta_h t_{hi})}{\sigma_\varepsilon} \right)} \right\} \prod_{j=1}^{J_h} \left\{ \frac{\phi \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon} \right)}{\sigma_\varepsilon \Phi \left( \frac{1 - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon} \right)} \right\} \right), \quad (10)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = 2H + 1$ .

Maximum likelihood estimates of the parameters  $\theta$  are again obtained by maximizing the log-likelihood function (i.e. the logarithm of equation 10) using the EM algorithm. For this we need the conditional mean and the conditional variance of the censored normal distribution which are given by

$$\begin{aligned} E[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] &= \widehat{y}_{hi} - \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) \\ \text{Var}[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right), \end{aligned}$$

where  $\widehat{y}_{hi} = \widehat{\alpha}_h + \widehat{\beta}_h t_{hi}$ ,  $i = 1, 2, \dots, I_h$ , and  $h = 1, 2, \dots, H$ .

The E and M-steps are the same as in the preceding section using the according conditional expectations and variances where the estimate for the variance is now given by

$$n\widehat{\sigma}_\varepsilon^2 = \sum_{h=1}^H \left[ \sum_{i=1}^{I_h} \text{Var}[y_{hi}^* | y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon] + \sum_{l=1}^{n_h} (y_{hl} - \widehat{y}_{hl})^2 \right].$$

#### Model 4

In model 4 all the parameters of primary moult vary from year to year. In essence, model 4 is the same as model 1 for data type 5 applied to each of the  $H$  data sets individually. For completeness, the formulation of the likelihood function is given here but the procedure to obtain maximum likelihood estimates of the model parameters is the same as for model 1 and will not be repeated here.

The latent random variable  $y^*$  is defined as

$$y_{hl}^* = \alpha_h + \beta_h d_{hl} + \varepsilon_{hl}, \quad \varepsilon_{hl} \sim N(0, \sigma_{\varepsilon_h}^2), \quad l = 1, 2, \dots, n_h, \quad h = 1, 2, \dots, H,$$

where the relationship between  $y^*$  and the observed variable  $y$  is the same as for model 2.

The likelihood function of the censored regression for model 4 for data type 5 is

$$L(\theta; t, u, y) = \prod_{h=1}^H \left( \prod_{i=1}^{I_h} \left\{ \frac{1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_{\epsilon h}} \right)}{\Phi \left( \frac{1 - (\alpha_h + \beta_h t_{hi})}{\sigma_{\epsilon h}} \right)} \right\} \prod_{j=1}^{J_h} \left\{ \frac{\phi \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_{\epsilon h}} \right)}{\sigma_{\epsilon h} \Phi \left( \frac{1 - (\alpha_h + \beta_h u_{hj})}{\sigma_{\epsilon h}} \right)} \right\} \right), \quad (11)$$

where  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is the vector of parameters. Here  $p$  denotes the total number of parameters in the model, which in this application, is given by  $p = 3H$ .

### Estimation of parameters

The above models can be easily implemented with the statistical software GENSTAT (Payne *et al.* 1994). Algorithms to implement the various models for data types 2 and 5 are given in Appendix 3. The GENSTAT programs to implement the various models for data types 2 and 5 are given in Appendix 4. These algorithms are used to obtain maximum likelihood estimates of the parameters, denoted  $\hat{\theta}$ , for the primary moult models. For model 1 the parameters  $\theta$  are  $(\alpha, \beta, \sigma_\epsilon)$  which, because of the invariance property of maximum likelihood estimation, can be converted to the parameters  $(\mu, \tau, \sigma)$  of Underhill and Zucchini (1988), where  $\mu = -\alpha/\beta$ ,  $\tau = 1/\beta$ , and  $\sigma = \sigma_\epsilon/\beta$ . For model 2 the parameters  $\theta$  become  $(\alpha_1, \dots, \alpha_H, \beta, \sigma_\epsilon)$  which can be converted to the parameters  $(\mu_1, \dots, \mu_H, \tau, \sigma)$ , where  $\mu_i = -\alpha_i/\beta$ ,  $i = 1, \dots, H$ ,  $\tau = 1/\beta$  and  $\sigma = \sigma_\epsilon/\beta$ . For model 3 the parameters  $\theta$  become  $(\alpha_1, \dots, \alpha_H, \beta_1, \dots, \beta_H, \sigma_\epsilon)$ . The parameters  $(\alpha_1, \dots, \alpha_H, \beta_1, \dots, \beta_H)$  can be converted to the parameters  $(\mu_1, \dots, \mu_H, \tau_1, \dots, \tau_H)$  where  $\mu_i = -\alpha_i/\beta_i$  and  $\tau_i = 1/\beta_i$  for  $i = 1, \dots, H$ . In this case a direct unique conversion from  $\sigma_\epsilon$  to  $\sigma$  cannot be performed because there is no longer a one-one relationship between  $\sigma_\epsilon$  and  $\sigma$  because we now have multiple slopes. For model 4 the parameters  $\theta$  are given by  $(\alpha_1, \dots, \alpha_H, \beta_1, \dots, \beta_H, \sigma_{\epsilon_1}, \dots, \sigma_{\epsilon_H})$ . These parameters can be converted to the parameters  $(\mu_1, \dots, \mu_H, \tau_1, \dots, \tau_H, \sigma_1, \dots, \sigma_H)$ , where  $\mu_i = -\alpha_i/\beta_i$ ,  $\tau_i = 1/\beta_i$  and  $\sigma_i = \sigma_{\epsilon_i}/\beta_i$  for  $i = 1, \dots, H$ .

Denote the mean completion date of moult for model 1 by  $\delta$  and by  $\delta_h$ ,  $h = 1, 2, \dots, H$ , for models 2, 3 and 4. Then an estimate for the mean completion date of moult for model 1 is computed as  $\hat{\delta} = \hat{\mu} + \hat{\tau}$ , for the  $h$ th data set of model 2 it is computed as  $\hat{\delta}_h = \hat{\mu}_h + \hat{\tau}$  and for model 3 and model 4 as  $\hat{\delta}_h = \hat{\mu}_h + \hat{\tau}_h$ . The standard error of the mean completion date of primary moult for model 1 is given by  $(\widehat{\text{Var}}(\hat{\mu}) + \widehat{\text{Var}}(\hat{\tau}) + 2\widehat{\text{Cov}}(\hat{\mu}, \hat{\tau}))^{1/2}$ , for model 2 by  $(\widehat{\text{Var}}(\hat{\mu}_h) + \widehat{\text{Var}}(\hat{\tau}) + 2\widehat{\text{Cov}}(\hat{\mu}_h, \hat{\tau}))^{1/2}$  and for model 3 and model 4 is given by  $(\widehat{\text{Var}}(\hat{\mu}_h) + \widehat{\text{Var}}(\hat{\tau}_h) + 2\widehat{\text{Cov}}(\hat{\mu}_h, \hat{\tau}_h))^{1/2}$ , with these values obtained from the inverse of the

$p \times p$  matrix with entries  $-\partial^2 \ell / \partial \theta_a \partial \theta_b$ , where  $\ell$  represents the respective log-likelihood function for the particular model. To compute the standard error of mean completion date of moult as described above, the log-likelihood function needs to be in terms of the parameters  $\mu$ ,  $\tau$ , and  $\sigma$  (considering model 1 as an example). However, in the present formulation, the log-likelihood function is in terms of the parameters  $\alpha$ ,  $\beta$ , and  $\sigma_\epsilon$ . Using equation 3, the variances of  $\mu$  and  $\tau$  can easily be computed but the covariance between  $\mu$  and  $\tau$  is not easily obtained from what is available. However, the mean completion date of moult can be computed as  $\hat{\delta} = \hat{\mu} + \hat{\tau} = \frac{-\hat{\alpha}}{\hat{\beta}} + \frac{1}{\hat{\beta}}$  and therefore an estimate for the standard error of the estimate for mean completion date of moult is given by

$$\left(\widehat{\text{Var}}(\hat{\delta})\right)^{-1} = \left(\widehat{\text{Var}}\left(\frac{1 - \hat{\alpha}}{\hat{\beta}}\right)\right)^{-1}$$

and equation 3 can be used to compute this.

### Comparison of models

To determine properties of the parameters of primary moult, we construct a null hypothesis about the characteristic of the parameters of interest and test it against a more general alternative hypothesis. One approach to achieve this aim consists of specifying each hypothesis in terms of a model and measuring the goodness of fit of each model. That is, we want to test

$$H_0 : \theta_0 = [\theta_1, \dots, \theta_q]^T$$

against

$$H_1 : \theta_1 = [\theta_1, \dots, \theta_p]^T$$

where  $q < p < n$ .

This hypothesis can be tested using the likelihood ratio test (Wilks 1962),

$$\lambda = \frac{L(\hat{\theta}_0; y)}{L(\hat{\theta}_1; y)},$$

where the numerator is the likelihood function of the model that specifies the null hypothesis, evaluated at the maximum likelihood estimators  $\hat{\theta}_0$ . The denominator is the likelihood function of the model that specifies the alternative hypothesis, evaluated at the maximum likelihood estimators  $\hat{\theta}_1$ .

The test rejects the null hypothesis for small values of  $\lambda$ , or equivalently for large values of  $-2 \log \lambda$ . Asymptotically, it is well known that  $-2 \log \lambda$  has the chi-squared distribution with  $p - q$  degrees of freedom (Dobson 1983), where  $p$  represents the number of estimated parameters for the model under the alternative hypothesis, and  $q$  represents the number of parameters estimated under the null hypothesis.

Let  $\Theta_j$  denote the vector of parameters of the  $j$ th model for primary moult which have been described in the previous sections for data types 2 and 5. That is, define

$$\begin{aligned}\Theta_1 &= [\sigma, \mu, \tau]^T \\ \Theta_2 &= [\sigma, \mu_h, \tau]^T \\ \Theta_3 &= [\sigma, \mu_h, \tau_h]^T \\ \Theta_4 &= [\sigma_h, \mu_h, \tau_h]^T,\end{aligned}$$

where  $h = 1, 2, \dots, H$  and  $H$  is the number of data sets. From the specified models various assumptions about the parameters of primary moult can then be tested.

To determine whether the parameters of primary moult remain constant throughout the years, we test

$$H_0 : \Theta_1 = [\sigma, \mu, \tau]^T$$

against

$$H_1 : \Theta_4 = [\sigma_h, \mu_h, \tau_h]^T,$$

where,  $q = 3$  and  $p = 33$ . Rejection of the null hypothesis leads to the conclusion that for at least one pair  $i, j$   $\sigma_i \neq \sigma_j$ , where  $1 \leq i \neq j \leq H$ , and/or for at least one pair  $i, j$   $\mu_i \neq \mu_j$ , where  $1 \leq i \neq j \leq H$ , and/or for at least one pair  $i, j$   $\tau_i \neq \tau_j$ , where  $1 \leq i \neq j \leq H$ . The next step is to determine which, if any, of the model parameters does not differ significantly from year to year. The first logical assumption is to hypothesize that the variance remains constant throughout the data sets. For this we want to test

$$H_0 : \Theta_3 = [\sigma, \mu_h, \tau_h]^T$$

against

$$H_1 : \Theta_4 = [\sigma_h, \mu_h, \tau_h]^T,$$

where, for this application,  $q = 23$  and  $p = 33$ . From this point on, we assume that the variance is equal for all data sets because we consider that it is unlikely and with

no meaningful biological interpretation to have the scenario where the mean starting date and/or duration of moult remain constant over the years but the standard deviation of mean starting date varies annually. To make sure that the rejection of the first hypothesis was not due to the variance differing significantly in at least one of the data sets only, we can test the hypothesis that there is a significant difference in either the mean starting data and/or the duration of moult, assuming that the variance remains constant, that is we test

$$H_0 : \Theta_1 = [\sigma, \mu, \tau]^T$$

against

$$H_1 : \Theta_3 = [\sigma, \mu_h, \tau_h]^T,$$

where, for this application,  $q = 3$  and  $p = 23$ . If the hypothesis of constant variance is accepted, then of course this hypothesis would be redundant in that we would come to the same conclusion as in the first hypothesis test.

To determine whether the duration of moult differs significantly between data sets, we test the null hypothesis that the duration of moult remains constant for the data sets against the alternative hypothesis that the duration of moult changes between the data sets. No restriction is made on the mean starting dates of moult. That is, we want to test

$$H_0 : \Theta_2 = [\sigma, \mu_h, \tau]^T$$

against

$$H_1 : \Theta_3 = [\sigma, \mu_h, \tau_h]^T,$$

where, for this application,  $q = 13$  and  $p = 23$ .

Finally, if we can assume that the duration of moult does not vary between the data sets, the next hypothesis of interest is whether the mean starting dates of moult differ between the data sets, that is, we test the null hypothesis that the mean starting dates are equal for each data set against the alternative hypothesis that the mean starting dates are not all the same. In other words we want to test

$$H_0 : \Theta_1 = [\sigma, \mu, \tau]^T$$

against

$$H_1 : \Theta_2 = [\sigma, \mu_h, \tau]^T,$$

where, for this application,  $q = 3$  and  $p = 13$ .

The above hypotheses can be tested by computing the likelihood ratio test as described above. The likelihood functions for the models that specify the null hypothesis and the alternative hypothesis are evaluated at the maximum likelihood estimators for the respective models and for the respective data type.

## RESULTS AND DISCUSSION

Examining the parameter estimates assuming a constant mean starting date over the 11 years obtained using the censored regression technique (Table 3) against the parameter estimates obtained using the model of Underhill and Zucchini (1988) (table 3 of the previous chapter), it can be seen that both model formulations give the same model parameter estimates. The same can be seen when examining the parameter estimates assuming that the mean starting date varies annually (Table 4 versus table 4 of previous chapter). This agrees with the results obtained in chapter 4, in which it was shown that the parameter estimates of the model to describe the relationship between reproductive mass and vegetative mass (an example of data censored from the left) are equivalent whether the EM algorithm or the Newton-Raphson algorithm is used to maximize the likelihood function. Although the results for these two models are the same as the results given in the previous chapter, they are repeated here for easy reference and to have the results of all the different models for primary moult discussed together.

For the Willow Warbler data, the overall estimates of the parameters, assuming a constant mean starting date, a constant duration of moult and a constant standard deviation of starting moult over the 11 years (model 1) were nearly identical for data type 2 and data type 5 (Table 3). However, the estimated durations using data type 5 were 3–4 days longer than with data type 2. This is the opposite effect to that observed after post-nuptial moult at seven localities in western Europe in Underhill *et al.* (1992). The estimates of the parameters of moult using the extended model that allows the mean starting date to vary annually (model 2) show that for data type 5 (with very similar results for data type 2), the estimated mean starting dates for males and females combined varied over 15 days between 6 July (in 1989) and 21 July (in 1985) (Table 4). However, in six of the 11 years, the estimated mean starting date was within two days of 13 July, the overall mean starting date (Table 3), with 1980, 1985, and 1989 being the furthest from normal. In eight of the

**Table 3.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting date  $\mu$ , duration  $\tau$  and standard deviation  $\sigma$  of moult remain the same for the 11 years of the study, i.e. model 1. Parameter estimates (standard errors) of the censored regression formulation are also given, that is, the  $y$ -intercept  $\alpha$ , the slope  $\beta$ , and the standard deviation of the regression line  $\sigma_\epsilon$ . Estimates (standard errors) for the mean completion date of moult  $\delta$  ( $\delta = \mu + \tau$ ) are also given.

	Data type 2			Data type 5		
	All	Male	Female	All	Male	Female
$n$	771	302	430	678	242	398
$\hat{\tau}$	40.5 (1.0)	43.3 (1.2)	38.4 (3.5)	45.0 (1.5)	46.6 (1.5)	42.1 (2.6)
$\hat{\mu}^\dagger$	42.3 (0.7)	34.5 (0.7)	49.9 (2.2)	42.6 (0.8)	34.5 (0.8)	50.1 (1.5)
	12 July	5 July	20 July	13 July	5 July	20 July
$\hat{\sigma}$	11.4 (0.7)	7.8 (0.3)	11.7 (0.7)	12.1 (0.5)	8.4 (0.5)	12.1 (0.7)
$\hat{\alpha}$	-1.04 (0.001)	-0.80 (0.001)	-1.30 (0.002)	-0.95 (0.001)	-0.74 (0.001)	-1.19 (0.001)
$\hat{\beta}$	0.025 (0.04)	0.023 (0.04)	0.026 (0.17)	0.022 (0.04)	0.021 (0.04)	0.024 (0.10)
$\hat{\sigma}_\epsilon$	0.28 (0.01)	0.18 (0.01)	0.31 (0.04)	0.27 (0.01)	0.18 (0.01)	0.29 (0.02)
$\hat{\delta}^\dagger$	82.8 (0.9)	77.8 (1.0)	88.3 (1.9)	87.6 (1.2)	81.1 (1.3)	92.2 (1.9)
	22 Aug	17 Aug	27 Aug	27 Aug	20 Aug	31 Aug

†Mean starting date and mean completion date are in days since 1 June (i.e. day 1 = 1 June) as in Underhill *et al.* (1992).

11 years, the mean starting dates lay between 9 July and 14 July. In 1980 and 1985, moult started later than 14 July and in 1989 it started earlier than 9 July.

Results for males and females for model 2 ran closely in parallel (Fig. 1); when males started earlier or later than their average, so did the females. The striking exception was 1987, when males started moulting on 4 July (Table 4), one day earlier than their average, 5 July (Table 3), but females started on 27 July, seven days later than their average, 20 July (Tables 3 and 4). In 1988 males started on time, but females were five days earlier than their average, and in 1989 both sexes were early, but males were relatively earlier than females. The same pattern for mean annual completion dates of moult as for mean annual starting dates of moult was found in males and females (Fig. 1).

**Table 4.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, under the assumption that the mean starting date of moult  $\mu_h$  vary between years,  $h = 1, \dots, 11$ , and a constant duration of moult  $\tau$  and a constant standard deviation  $\sigma$  for all years, i.e. model 2. Estimates (standard errors) of mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) and the standard deviation of the regression line  $\sigma_\epsilon$  are also given.

Year	Esti- mates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\mu}_1^\dagger$	40.3 (1.7) 10 July	34.1 (1.6) 4 July	49.1 (2.9) 19 July	39.4 (1.7) 9 July	33.8 (1.7) 4 July	47.9 (2.9) 18 July
1980	$\hat{\mu}_2$	49.2 (1.8) 19 July	42.3 (1.7) 12 July	57.2 (2.7) 27 July	49.0 (2.0) 19 July	42.6 (1.9) 13 July	55.7 (3.1) 26 July
1981	$\hat{\mu}_3$	43.4 (1.5) 13 July	33.6 (1.6) 4 July	50.8 (2.6) 21 July	43.9 (1.7) 14 July	33.9 (1.8) 4 July	50.2 (2.4) 20 July
1982	$\hat{\mu}_4$	41.3 (1.3) 11 July	34.6 (1.3) 5 July	47.9 (2.3) 18 July	41.9 (1.6) 12 July	34.9 (1.5) 5 July	48.9 (2.7) 19 July
1983	$\hat{\mu}_5$	41.8 (1.7) 12 July	32.7 (1.5) 3 July	49.5 (2.6) 20 July	41.7 (1.8) 12 July	32.3 (1.7) 2 July	50.1 (2.9) 20 July
1984	$\hat{\mu}_6$	41.3 (2.4) 11 July	30.9 (2.3) 1 July	47.4 (4.1) 17 July	40.7 (2.5) 11 July	30.7 (2.5) 1 July	46.2 (3.7) 16 July
1985	$\hat{\mu}_7$	48.6 (2.2) 19 July	40.1 (2.4) 10 July	53.5 (3.4) 24 July	50.8 (2.8) 21 July	40.1 (2.7) 10 July	58.0 (4.0) 28 July
1986	$\hat{\mu}_8$	42.1 (2.1) 12 July	34.6 (2.0) 5 July	47.6 (4.5) 18 July	42.2 (2.6) 12 July	34.2 (2.1) 4 July	47.4 (4.2) 17 July
1987	$\hat{\mu}_9$	44.4 (2.1) 14 July	34.4 (1.7) 4 July	55.9 (4.3) 26 July	43.9 (2.2) 14 July	33.9 (1.8) 4 July	56.8 (5.2) 27 July
1988	$\hat{\mu}_{10}$	39.6 (1.8) 10 July	33.9 (1.8) 4 July	44.9 (3.0) 15 July	40.2 (2.2) 10 July	34.5 (2.0) 5 July	44.8 (3.0) 15 July
1989	$\hat{\mu}_{11}$	35.8 (1.4) 6 July	26.0 (1.4) 26 June	45.9 (2.4) 16 July	36.4 (1.5) 6 July	26.5 (1.6) 27 June	45.4 (2.5) 15 July
	$\hat{\tau}$	40.7 (1.0)	43.9 (1.1)	38.1 (1.7)	44.6 (1.2)	45.8 (1.3)	42.0 (1.9)
	$\hat{\sigma}$	10.8 (0.4)	6.6 (0.4)	11.3 (0.6)	11.4 (0.5)	7.0 (0.4)	11.6 (0.7)
	$\hat{\sigma}_\epsilon$	0.26 (0.01)	0.15 (0.01)	0.30 (0.02)	0.26 (0.01)	0.15 (0.01)	0.28 (0.02)

**Table 4 (continued).** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, under the assumption that the mean starting date of moult  $\mu_h$  vary between years,  $h = 1, \dots, 11$ , and a constant duration of moult  $\tau$  and a constant standard deviation  $\sigma$  for all years, i.e. model 2. Estimates (standard errors) of mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) and the standard deviation of the regression line  $\sigma_z$  are also given.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\delta}_1^\dagger$	81.0 (1.8) 21 Aug	78.0 (1.8) 17 Aug	87.2 (3.1) 26 Aug	84.0 (2.0) 23 Aug	79.6 (2.0) 19 Aug	89.9 (3.2) 29 Aug
1980	$\hat{\delta}_2$	89.9 (1.8) 29 Aug	86.2 (1.7) 25 Aug	95.3 (2.8) 3 Sep	93.6 (2.1) 2 Sep	88.4 (2.0) 27 Aug	97.7 (3.0) 6 Sep
1981	$\hat{\delta}_3$	84.1 (1.6) 23 Aug	77.5 (1.7) 17 Aug	88.9 (2.6) 28 Aug	88.5 (2.0) 28 Aug	79.7 (2.0) 19 Aug	92.2 (2.9) 31 Aug
1982	$\hat{\delta}_4$	82.0 (1.4) 21 Aug	78.5 (1.4) 18 Aug	86.0 (2.2) 25 Aug	86.5 (1.7) 26 Aug	80.7 (1.7) 20 Aug	90.9 (2.6) 30 Aug
1983	$\hat{\delta}_5$	82.5 (1.8) 22 Aug	76.6 (1.7) 16 Aug	87.6 (3.1) 27 Aug	86.3 (2.1) 25 Aug	78.1 (1.9) 17 Aug	92.1 (3.3) 31 Aug
1984	$\hat{\delta}_6$	82.0 (2.4) 21 Aug	74.8 (2.4) 14 Aug	85.5 (4.0) 25 Aug	85.3 (2.7) 24 Aug	76.5 (2.7) 16 Aug	88.2 (3.9) 27 Aug
1985	$\hat{\delta}_7$	89.3 (2.3) 28 Aug	84.0 (2.4) 23 Aug	91.6 (3.5) 31 Aug	95.4 (2.9) 3 Sep	85.9 (2.8) 25 Aug	100.0 (4.5) 8 Sep
1986	$\hat{\delta}_8$	82.8 (2.2) 22 Aug	78.5 (2.0) 18 Aug	85.7 (4.2) 25 Aug	86.8 (2.6) 26 Aug	80.0 (2.2) 19 Aug	89.4 (4.3) 28 Aug
1987	$\hat{\delta}_9$	85.1 (2.1) 24 Aug	78.3 (1.9) 17 Aug	94.0 (4.2) 2 Sep	88.5 (2.3) 28 Aug	79.7 (2.0) 19 Aug	98.8 (5.0) 7 Sep
1988	$\hat{\delta}_{10}$	80.3 (1.9) 19 Aug	77.8 (2.0) 17 Aug	83.0 (3.0) 22 Aug	84.8 (2.3) 24 Aug	80.3 (2.3) 19 Aug	86.8 (3.3) 26 Aug
1989	$\hat{\delta}_{11}$	76.5 (1.5) 16 Aug	69.9 (1.5) 9 Aug	84.0 (2.6) 23 Aug	81.0 (1.8) 20 Aug	72.3 (1.8) 11 Aug	87.4 (2.8) 26 Aug

†Mean starting date and mean completion date are in days since 1 June (i.e. day 1 = 1 June)

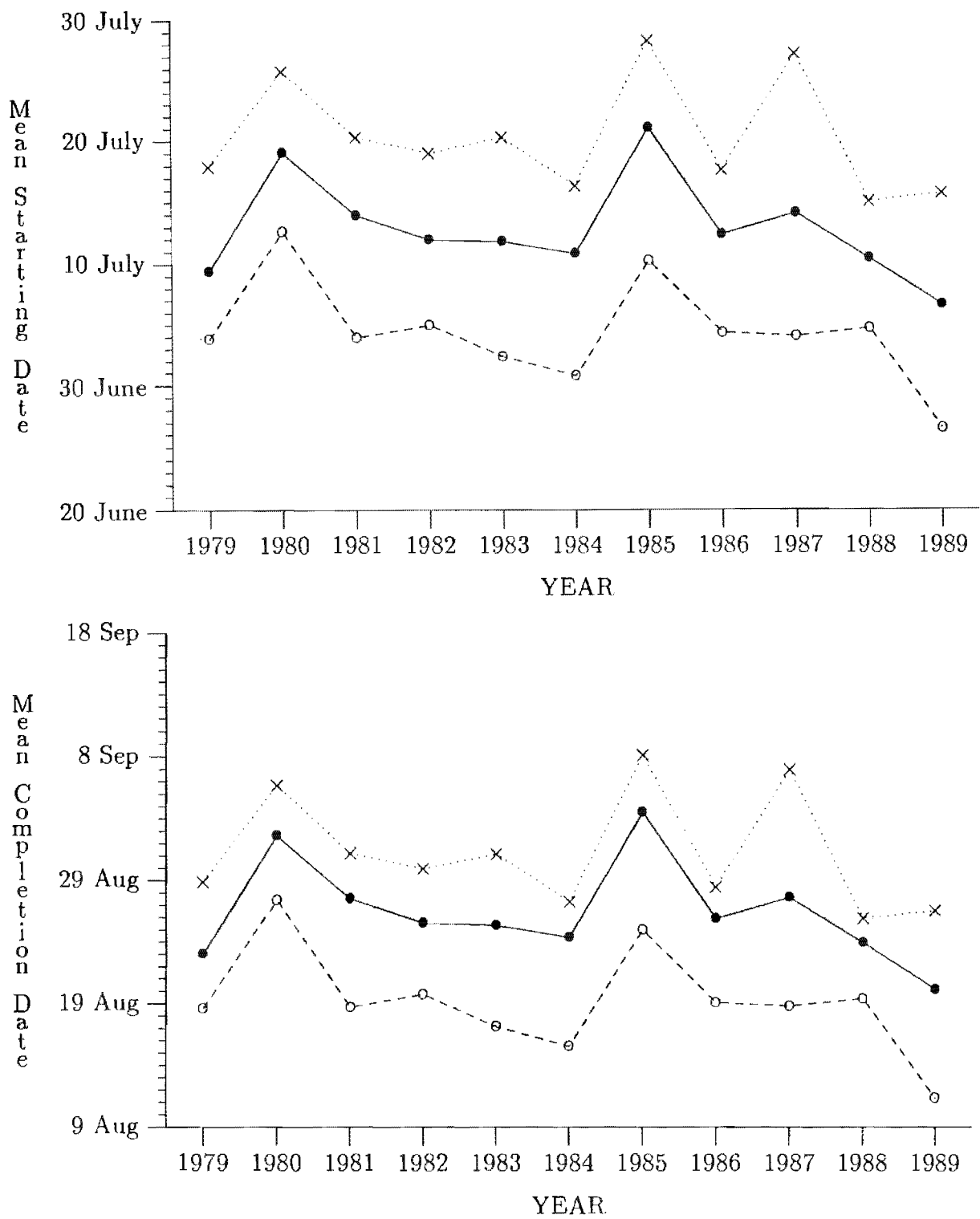


Figure 1. Mean annual starting dates (top) and mean annual completion dates (bottom) of moult for Willow Warbler for males (dashed line), females (dotted line), and for all the data (solid line) for model 2 and data type 5 (see Table 4).

Estimates of primary moult parameters for the model assuming that the standard deviations are constant over the years, while the mean starting date and duration of moult vary annually (model 3) are given in Table 5. Again the mean starting dates over the 11 years were nearly identical for data type 2 and type 5 but the estimated durations for data type 5 were for most years a few days longer. An exception occurred with males in that the duration of moult for data type 5 was within 0–2 days of that for data type 2 and only for the years 1980 and 1988 was the duration of moult for data type 5 much longer than that for data type 2 (5 and 10 days respectively). The estimated mean starting dates for the complete data varied over 15 days between 7 July (1989) and 22 July (1985). This range is the same as that obtained with model 2 (where the duration of moult is also assumed to be constant over the years) with the dates shifted by one day. In fact the estimates for mean starting date obtained for model 2 and for model 3 were very similar (see Fig. 2). The duration of moult for the complete data ranged from 32 (1985) to 50 (1980) days. The shortest duration of moult corresponds to the year when moult started the latest (22 July). In six of the 11 years, the duration of moult was within 3 days of 45, the overall duration estimate (Table 3), and 10 of the 11 years were within 5 days of the overall mean. The furthest from normal was in 1985 when duration of moult was 13 days shorter than the mean.

The mean starting dates of males were always earlier than that of females for data type 5 (the results for data type 2 were very similar) (Table 5 and Fig. 3). The results for males and females generally ran closely in parallel in that if males started later or earlier than their overall mean, so did the females. Exceptions occurred in 1983, when males started moult on 29 June (Table 5), six days earlier than their average, 5 July (Table 3), but females started on 20 July, the same day as their average (Tables 3 and 5) and in 1987 males started moult on 5 July, the same day as their average, but females started moult on 1 August (Table 5), 12 days later than their average. In 1989, both males and females started moult earlier than their average, but males were eight days earlier, while females were only two days earlier than their average.

Figure 3 shows that except for the year 1981, the duration of moult of females was usually much shorter (12–19 days) than that of males. Exceptions to this occurred in 1979, 1982 and 1986 when the duration of moult of females was within a day of that of males. In 1981 the opposite relationship was observed in that the duration of moult was 20 days longer in females than in males (Table 5). In five of the 11 years, the difference in the duration of

**Table 5.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting dates  $\mu_h$  and durations  $\tau_h$  vary between years,  $h = 1, \dots, 11$ , and a constant standard deviation  $\sigma$ , i.e. model 3. Estimates (standard errors) of the mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) and standard deviation of regression line  $\sigma_\epsilon$  are also given.

Year	Esti- mates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\tau}_1$	45.6 (4.0)	49.6 (6.7)	48.7 (6.1)	46.1 (3.3)	49.8 (6.3)	49.5 (6.1)
1980	$\hat{\tau}_2$	46.8 (2.8)	52.9 (3.4)	42.8 (4.6)	50.2 (3.2)	58.1 (4.4)	44.3 (4.7)
1981	$\hat{\tau}_3$	43.3 (2.7)	33.9 (2.3)	50.0 (4.5)	48.5 (3.4)	34.1 (3.1)	53.9 (5.5)
1982	$\hat{\tau}_4$	37.7 (1.8)	40.6 (2.0)	36.2 (3.3)	41.5 (2.0)	41.4 (2.2)	40.8 (4.3)
1983	$\hat{\tau}_5$	44.4 (3.7)	55.5 (3.8)	34.2 (6.0)	47.4 (5.6)	56.7 (4.0)	39.8 (7.7)
1984	$\hat{\tau}_6$	38.9 (4.2)	43.4 (4.4)	31.2 (5.3)	40.3 (4.1)	43.8 (4.6)	31.7 (5.2)
1985	$\hat{\tau}_7$	28.6 (4.1)	35.2 (3.3)	18.6 (3.7)	31.6 (4.1)	35.4 (3.6)	22.7 (6.2)
1986	$\hat{\tau}_8$	43.8 (3.1)	48.0 (3.1)	43.9 (5.5)	48.2 (4.0)	49.0 (3.3)	48.1 (6.8)
1987	$\hat{\tau}_9$	39.3 (2.6)	39.0 (3.4)	25.1 (14.6)	40.4 (6.5)	39.0 (3.4)	27.3 (10.9)
1988	$\hat{\tau}_{10}$	40.3 (3.4)	48.6 (3.3)	37.0 (4.6)	45.3 (4.6)	58.6 (7.4)	40.4 (5.5)
1989	$\hat{\tau}_{11}$	38.2 (6.9)	41.8 (2.5)	29.8 (3.7)	43.1 (2.6)	43.7 (2.9)	31.4 (4.0)
1979	$\hat{\mu}_1^\dagger$	39.2 (1.9)	32.8 (2.0)	47.0 (3.2)	39.0 (1.7)	32.7 (1.9)	46.5 (3.2)
		9 July	3 July	17 July	9 July	3 July	17 July
1980	$\hat{\mu}_2$	46.7 (2.0)	38.8 (2.0)	54.3 (3.7)	47.1 (2.1)	38.9 (2.2)	54.2 (3.6)
		17 July	9 July	24 July	17 July	9 July	24 July
1981	$\hat{\mu}_3$	42.5 (1.8)	36.4 (4.0)	47.8 (2.8)	43.0 (1.9)	36.3 (1.7)	48.0 (2.9)
		13 July	6 July	18 July	13 July	6 July	18 July
1982	$\hat{\mu}_4$	42.6 (1.4)	35.5 (1.3)	48.6 (2.5)	42.6 (1.4)	35.4 (1.3)	48.8 (2.6)
		13 July	6 July	19 July	13 July	5 July	19 July
1983	$\hat{\mu}_5$	41.2 (1.8)	29.4 (1.9)	49.3 (3.8)	41.4 (2.1)	29.3 (1.9)	49.7 (2.7)
		11 July	29 June	19 July	11 July	29 June	20 July
1984	$\hat{\mu}_6$	41.5 (2.4)	30.6 (2.3)	48.2 (3.4)	41.3 (2.3)	30.5 (2.2)	47.8 (3.3)
		12 July	1 July	18 July	11 July	1 July	18 July
1985	$\hat{\mu}_7$	51.8 (2.6)	43.7 (2.5)	58.3 (2.7)	52.1 (2.3)	43.6 (2.7)	58.3 (3.0)
		22 July	14 July	28 July	22 July	14 July	28 July
1986	$\hat{\mu}_8$	40.8 (2.4)	32.6 (2.2)	45.3 (4.1)	41.3 (2.3)	32.5 (2.1)	45.8 (4.3)
		11 July	3 July	15 July	11 July	3 July	16 July
1987	$\hat{\mu}_9$	44.4 (1.7)	35.4 (1.6)	62.9 (15.0)	44.2 (2.6)	35.4 (1.6)	62.3 (10.3)
		14 July	5 July	2 Aug	14 July	5 July	1 Aug
1988	$\hat{\mu}_{10}$	39.6 (2.1)	32.7 (1.8)	44.6 (3.0)	40.0 (2.2)	33.2 (2.2)	44.6 (3.1)
		10 July	3 July	15 July	10 July	3 July	15 July
1989	$\hat{\mu}_{11}$	36.8 (4.3)	26.6 (1.5)	48.5 (2.4)	36.7 (1.5)	26.6 (1.4)	48.2 (2.4)
		7 July	27 June	19 July	7 July	27 June	18 July
	$\hat{\sigma}_\epsilon$	0.26 (0.01)	0.13 (0.01)	0.27 (0.02)	0.25 (0.01)	0.13 (0.02)	0.26 (0.02)

**Table 5 (continued).** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting dates  $\mu_h$  and durations  $\tau_h$  vary between years,  $h = 1, \dots, 11$ , and a constant standard deviation  $\sigma$ , i.e. model 3. Estimates (standard errors) of the mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) and standard deviation of regression line  $\sigma_\varepsilon$  are also given.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\delta}_1^\dagger$	84.8 (3.7)	82.4 (5.7)	95.7 (6.2)	85.1 (3.4)	82.5 (5.5)	96.0 (6.1)
		24 Aug	21 Aug	4 Sep	24 Aug	22 Aug	4 Sep
1980	$\hat{\delta}_2$	93.5 (2.7)	91.7 (2.7)	97.1 (3.6)	97.3 (3.2)	97.0 (3.6)	98.5 (3.7)
		2 Sep	31 Aug	5 Sep	5 Sep	5 Sep	7 Sep
1981	$\hat{\delta}_3$	85.8 (2.4)	70.3 (2.1)	97.8 (4.4)	91.5 (3.4)	70.4 (2.3)	101.9 (5.5)
		25 Aug	9 Aug	6 Sep	31 Aug	9 Aug	10 Sep
1982	$\hat{\delta}_4$	80.3 (1.6)	76.1 (1.7)	84.8 (2.5)	84.1 (2.1)	76.8 (1.9)	89.6 (3.6)
		19 Aug	15 Aug	24 Aug	23 Aug	16 Aug	29 Aug
1983	$\hat{\delta}_5$	85.6 (3.8)	84.9 (3.2)	83.5 (5.7)	88.8 (5.2)	86.0 (3.6)	89.5 (7.8)
		25 Aug	24 Aug	24 Aug	28 Aug	25 Aug	29 Aug
1984	$\hat{\delta}_6$	80.4 (3.9)	74.0 (4.0)	79.4 (4.8)	81.6 (4.1)	74.3 (4.2)	79.5 (4.6)
		19 Aug	13 Aug	19 Aug	21 Aug	13 Aug	19 Aug
1985	$\hat{\delta}_7$	80.4 (2.9)	78.9 (2.6)	76.9 (3.0)	83.7 (3.8)	79.0 (2.7)	81.0 (5.6)
		19 Aug	18 Aug	16 Aug	23 Aug	18 Aug	20 Aug
1986	$\hat{\delta}_8$	84.6 (3.0)	80.6 (2.5)	89.2 (5.5)	89.5 (4.2)	81.5 (2.8)	93.9 (7.1)
		24 Aug	20 Aug	28 Aug	29 Aug	20 Aug	2 Sep
1987	$\hat{\delta}_9$	83.7 (3.2)	74.4 (2.8)	88.0 (3.4)	84.6 (5.2)	74.4 (2.8)	89.6 (4.2)
		24 Aug	13 Aug	27 Aug	24 Aug	13 Aug	29 Aug
1988	$\hat{\delta}_{10}$	79.9 (2.9)	81.3 (3.0)	81.6 (4.1)	85.3 (4.4)	91.8 (7.2)	85.0 (5.2)
		19 Aug	20 Aug	21 Aug	24 Aug	31 Aug	24 Aug
1989	$\hat{\delta}_{11}$	75.0 (3.2)	68.4 (2.0)	78.3 (2.8)	79.8 (2.6)	70.3 (2.5)	79.6 (3.2)
		14 Aug	7 Aug	17 Aug	19 Aug	9 Aug	19 Aug

†Mean starting date and mean completion date are in days since 1 June (i.e. day 1 = 1 June)

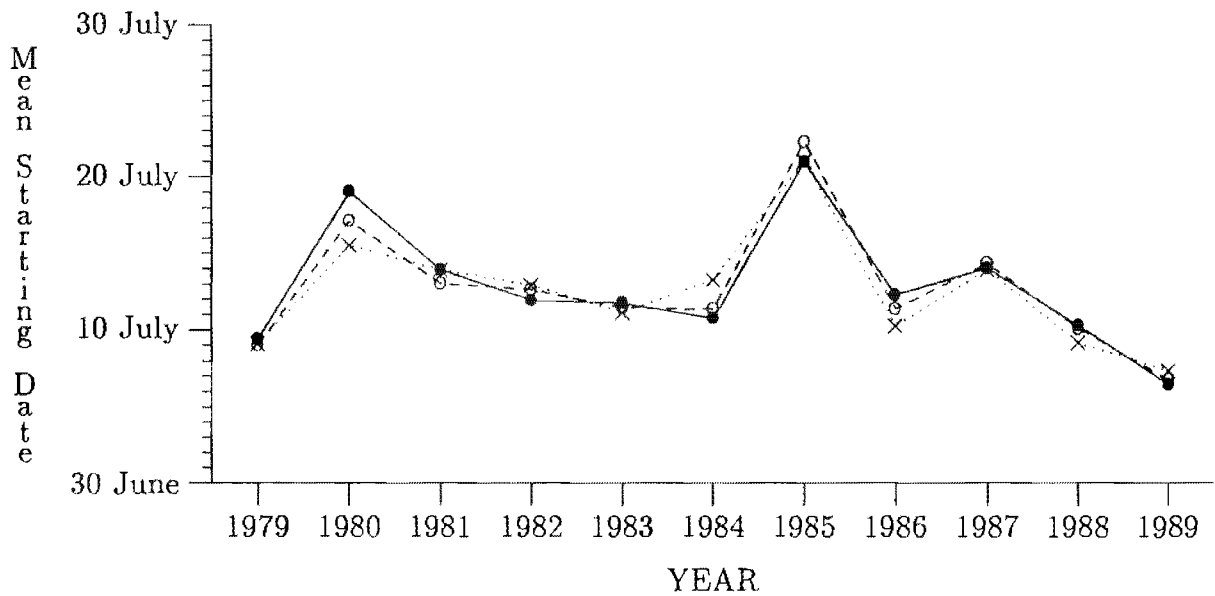
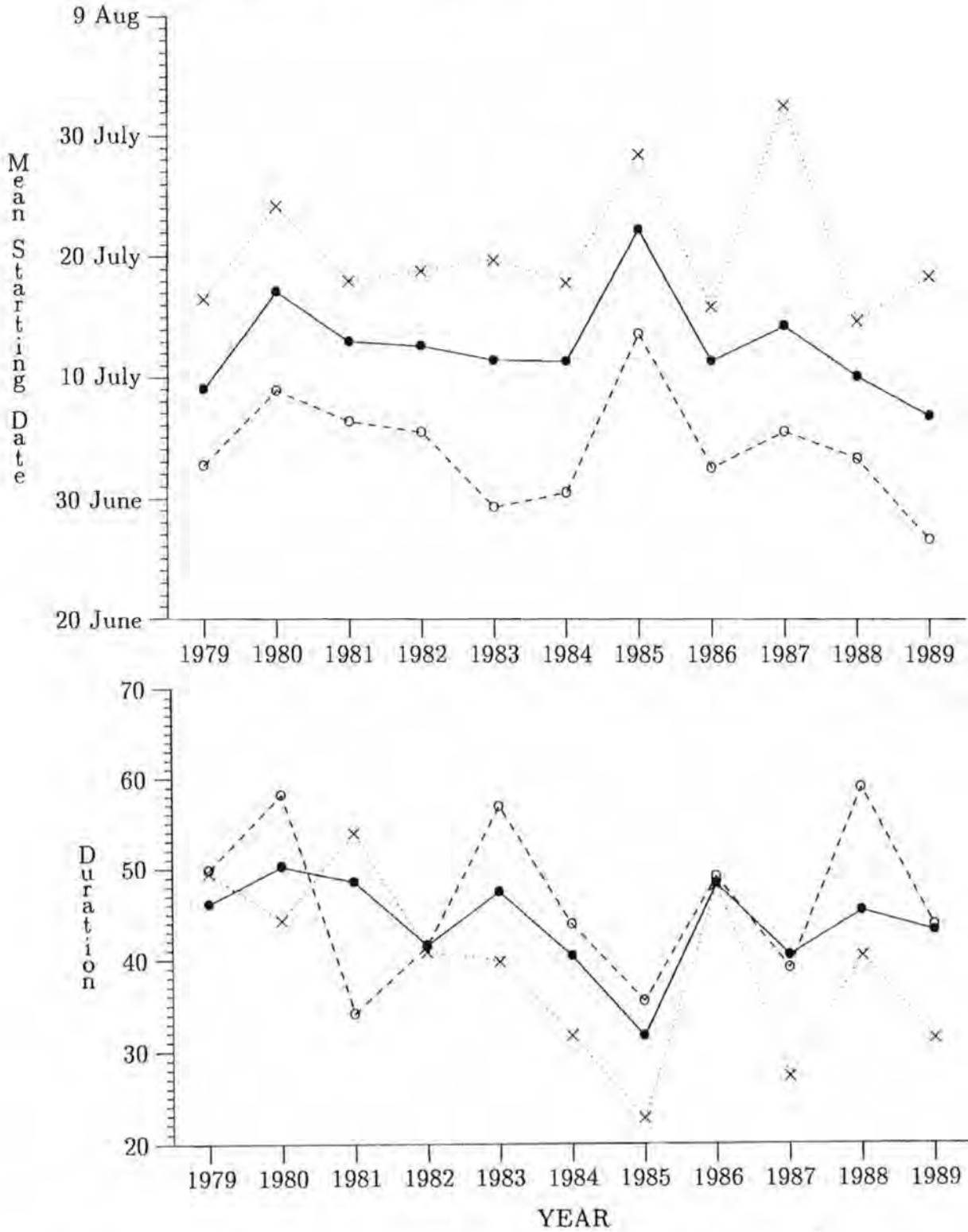


Figure 2. Comparison of mean annual starting dates of moult for Willow Warbler between model 2 (solid line), model 3 (dashed line) and model 4 (dotted line) for all the data, for data type 5 (see Tables 4, 5 and 6).

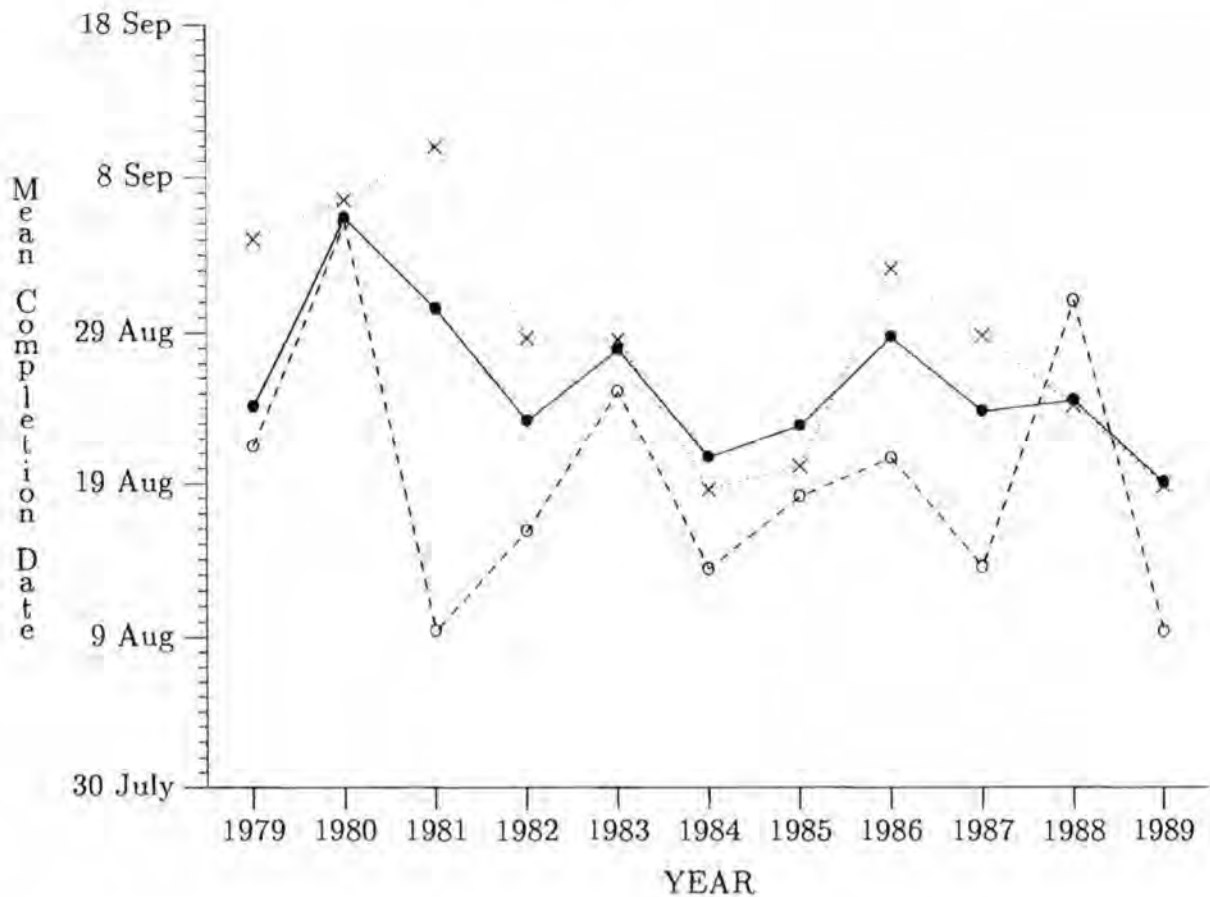
moult between males and females was more than 9 days. In 1981, the duration of moult in males was 34 days (Table 5), 13 days shorter than their average, 47 days (Table 3), but for the females the duration of moult was 54 days, 12 days longer than their average, 42 days (Tables 3 and 5). In 1983, duration of moult in males was 57 days, ten days longer than their average, but for females it was 40 days, 2 days shorter than their average. In 1988, duration of moult in males was 59 days, 12 days longer than their average, but for females it was 40 days, 2 days shorter than their average.

For most years the mean completion date of moult for data type 5 was a few days later than that for data type 2 (Table 5). This, together with the fact that duration of moult was longer for data type 5, suggested the possibility of emigration of birds that had completed moult into the sampling area. Mean completion date for the complete data ranged from 19 August (1989) to 5 September (1980) with eight of the 11 years being within four days of the overall mean, 27 August (Table 3). Females completed moult later than males with the exception that in 1988 females completed moult 7 days before males (Table 5 and Fig. 4). In 6 of the 11 years the difference between the mean completion date of males and females was more than 9 days. In 1981, females completed moult 32 days later than males. In general, if males completed moult earlier or later than their average, so did the females.



**Figure 3.** Mean annual starting dates (top) and duration (bottom) of moult for Willow Warbler for males (dashed line), females (dotted line), and for all the data (solid line) for model 3 and data type 5 (see Table 5).

Large exceptions occurred in 1981 when males completed moult on 9 August (Table 5), 11 days earlier than their average, 20 August (Table 3) and females completed moult on 10 September, 10 days later than their average, 31 August (Tables 3 and 5). In 1988 males completed moult on 31 August, 11 days later than their average and females completed moult on 24 August, 7 days earlier than their average.



**Figure 4.** Comparison of mean annual completion dates of moult for Willow Warbler between males (dashed line), females (dotted line) and for all the data (solid line) for model 3, for data type 5 (see Table 5).

Table 6 shows the estimates of the primary moult parameters under the assumption that the mean starting date, duration and standard deviation of moult vary annually (model 4). A table with these parameter estimates obtained using the Newton-Raphson algorithm to maximize the likelihood function is shown in Appendix 5 to demonstrate that the parameter estimates obtained using either algorithm are identical. The parameter estimates for the mean starting date were very similar to those obtained from the model for primary moult

with the assumption of constant standard deviation (model 3) and for the model with the assumption of constant duration of moult and constant standard deviation (model 2) when the data was combined (Fig. 2). The same was true for the mean starting dates of moult for males and females (Fig. 5), except that in 1987, females started moult on 17 July, 15 days earlier than their estimate from model 3, 1 August (Table 5), 10 days earlier than their estimate from model 2, 27 July (Table 4), but it was within 3 days of their overall mean starting date of moult, 20 July (Table 3).

The mean starting dates of moult of males were always earlier than that of females for data type 5 (with very similar results for data type 2) (Table 6 and Fig. 6). The results for males and females generally ran closely in parallel in that if males started later or earlier than their overall mean, so did the females. Exceptions occurred in 1981, when males started moult on 6 July (Table 6), one day later than their average, 5 July (Table 3), but females started moult on 17 July, three days earlier than their average, 20 July (Tables 3 and 6), and in 1984, males started moult on 30 June, five days earlier than their average, but females started moult on 21 July, one day later than their overall mean.

The estimates for duration of moult for model 3 and model 4 were very similar when the data was combined and for males (Fig. 7). For females (Fig. 8) there was an exception in 1987, when the duration of moult for model 4 was 43 days (Table 6), 16 days longer than that for model 3, 27 days (Table 5), but was within 1 day of the overall mean duration of moult, 42 days (Table 3). The duration of moult in females was usually shorter than that of males (Table 6 and Fig. 9). Exceptions occurred in 1979, 1981 and 1987, when the duration of moult in females was 8, 21 and 3 days longer than in males respectively. In 1982 and 1986 the duration of moult in females was only two to three days shorter than that of males, while on average, the duration of moult of females was 13 days shorter than that of males.

Figures 10-12 show the comparison of the parameter estimates for mean completion dates between model 2, model 3 and model 4 for all the data, for males and for females respectively. It is evident that the parameter estimates of mean completion dates of moult were very similar for model 3 and model 4. Therefore the discussion of the results for model 4 will resemble those already discussed for model 3. An interesting result obtained here is that although the parameter estimates for mean starting date of moult and duration of moult differed substantially between model 3 and model 4 for 1987, the mean completion date of moult for both models was 29 August. In fact, for all cases of Willow Warbler, making the extra assumption that the standard deviation of mean starting date varies annually,

**Table 6.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting dates  $\mu_h$ , durations  $\tau_h$  and standard deviations  $\sigma_h$  vary between years,  $h = 1, \dots, 11$ , i.e. model 4. Estimates (standard errors) of the mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) and standard deviations of the regression line  $\sigma_{\epsilon h}$  are also given.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\tau}_1$	45.8 (3.5)	44.4 (7.1)	51.9 (11.3)	45.8 (3.5)	44.4 (7.1)	51.9 (11.3)
1980	$\hat{\tau}_2$	48.2 (3.5)	51.1 (4.1)	42.9 (4.5)	52.1 (3.8)	57.3 (5.3)	44.4 (7.2)
1981	$\hat{\tau}_3$	41.8 (3.4)	34.6 (2.3)	52.0 (5.1)	47.5 (3.2)	34.6 (2.6)	55.8 (4.7)
1982	$\hat{\tau}_4$	37.6 (2.2)	41.7 (2.0)	33.9 (4.1)	41.3 (4.2)	42.0 (2.1)	38.9 (3.4)
1983	$\hat{\tau}_5$	44.7 (4.3)	55.1 (4.2)	34.7 (7.7)	48.0 (4.7)	56.3 (4.2)	43.0 (6.1)
1984	$\hat{\tau}_6$	36.8 (9.7)	44.0 (3.9)	27.8 (3.0)	37.9 (9.1)	44.1 (4.2)	27.8 (3.0)
1985	$\hat{\tau}_7$	30.0 (2.5)	37.9 (2.2)	21.7 (1.3)	33.3 (3.0)	37.9 (2.5)	24.2 (3.0)
1986	$\hat{\tau}_8$	45.7 (4.6)	50.0 (2.9)	44.3 (5.8)	49.5 (5.2)	50.3 (3.0)	48.3 (9.9)
1987	$\hat{\tau}_9$	40.2 (3.8)	39.7 (2.4)	42.3 (3.0)	40.9 (4.2)	39.7 (2.4)	42.7 (2.2)
1988	$\hat{\tau}_{10}$	41.7 (3.0)	53.0 (2.7)	34.2 (3.4)	46.6 (6.2)	63.8 (3.3)	37.3 (4.6)
1989	$\hat{\tau}_{11}$	37.2 (3.2)	41.7 (2.7)	29.7 (4.1)	42.3 (7.2)	43.5 (3.1)	30.9 (5.6)
1979	$\hat{\mu}_1^\dagger$	39.1 (1.9)	35.2 (2.6)	44.8 (5.9)	39.1 (1.9)	35.2 (2.6)	44.8 (5.9)
		9 July	5 July	15 July	9 July	5 July	15 July
1980	$\hat{\mu}_2$	45.6 (2.5)	39.8 (2.5)	54.2 (3.7)	45.5 (2.5)	39.5 (2.6)	54.1 (5.9)
		16 July	10 July	24 July	16 July	10 July	24 July
1981	$\hat{\mu}_3$	43.5 (2.3)	35.8 (1.4)	46.4 (3.1)	43.9 (2.0)	35.8 (1.6)	46.6 (2.8)
		14 July	6 July	16 July	14 July	6 July	17 July
1982	$\hat{\mu}_4$	42.6 (1.6)	34.8 (1.3)	50.0 (3.0)	42.9 (2.7)	34.9 (1.3)	50.2 (2.6)
		13 July	5 July	20 July	13 July	5 July	20 July
1983	$\hat{\mu}_5$	41.1 (2.1)	29.7 (2.2)	49.1 (2.8)	41.1 (2.2)	29.6 (2.1)	48.7 (2.6)
		11 July	30 June	19 July	11 July	30 June	19 July
1984	$\hat{\mu}_6$	43.1 (5.4)	30.2 (2.1)	51.1 (3.3)	43.2 (4.5)	30.2 (2.2)	51.1 (3.3)
		13 July	30 June	21 July	13 July	30 June	21 July
1985	$\hat{\mu}_7$	51.0 (1.9)	41.3 (1.7)	57.3 (1.0)	50.9 (2.0)	41.3 (1.9)	57.1 (1.3)
		21 July	11 July	27 July	21 July	11 July	27 July
1986	$\hat{\mu}_8$	39.5 (3.3)	31.1 (2.0)	45.0 (4.4)	40.2 (3.6)	31.3 (2.1)	45.7 (6.8)
		10 July	1 July	15 July	10 July	1 July	16 July
1987	$\hat{\mu}_9$	43.7 (2.2)	34.9 (1.2)	47.6 (2.9)	43.8 (2.4)	34.9 (1.2)	47.4 (2.1)
		14 July	5 July	18 July	14 July	5 July	17 July
1988	$\hat{\mu}_{10}$	38.8 (1.9)	31.1 (1.3)	46.5 (2.8)	39.1 (3.2)	30.3 (1.1)	46.8 (3.4)
		9 July	1 July	17 July	9 July	30 June	17 July
1989	$\hat{\mu}_{11}$	37.3 (2.1)	26.7 (1.6)	48.6 (2.7)	37.3 (3.5)	26.7 (1.6)	48.6 (3.3)
		7 July	27 June	19 July	7 July	27 June	19 July

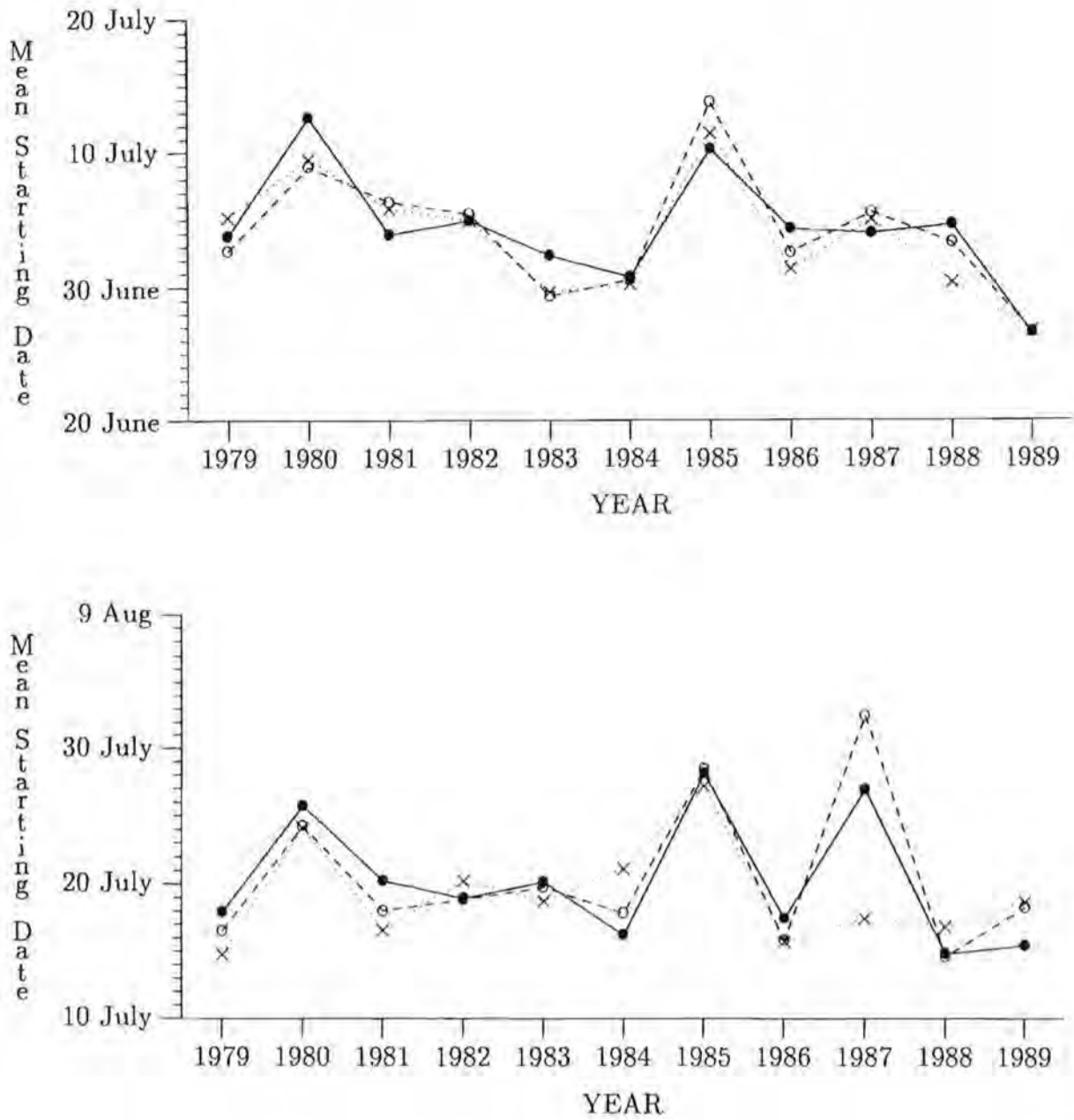
**Table 6 (continued).** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting dates  $\mu_h$ , durations  $\tau_h$  and standard deviations  $\sigma_h$  vary between years,  $h = 1, \dots, 11$ , i.e. model 4. Estimates (standard errors) of the mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) and standard deviations of the regression line  $\sigma_{\epsilon h}$  are also given.

Year	Estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\sigma}_1$	11.7 (1.5)	9.2 (2.0)	11.6 (2.4)	11.7 (1.5)	9.2 (2.0)	11.6 (2.4)
1980	$\hat{\sigma}_2$	11.0 (1.4)	8.7 (1.6)	11.6 (2.0)	11.0 (1.4)	8.7 (1.7)	11.4 (2.1)
1981	$\hat{\sigma}_3$	12.1 (1.4)	3.9 (0.8)	12.3 (1.9)	12.9 (1.6)	3.9 (0.9)	12.6 (2.0)
1982	$\hat{\sigma}_4$	9.8 (0.9)	4.6 (0.7)	10.8 (1.5)	10.6 (1.1)	4.8 (0.7)	11.5 (1.7)
1983	$\hat{\sigma}_5$	11.3 (1.4)	7.6 (1.4)	9.1 (1.9)	11.6 (1.4)	7.7 (1.4)	9.6 (1.8)
1984	$\hat{\sigma}_6$	11.4 (2.1)	5.1 (1.5)	10.4 (2.8)	11.7 (2.5)	5.2 (1.7)	10.4 (2.8)
1985	$\hat{\sigma}_7$	6.6 (1.2)	2.2 (0.7)	2.2 (0.8)	6.9 (1.3)	2.2 (0.7)	2.2 (0.8)
1986	$\hat{\sigma}_8$	10.4 (1.6)	4.9 (1.1)	11.7 (2.7)	11.3 (1.9)	5.1 (1.2)	12.4 (3.2)
1987	$\hat{\sigma}_9$	9.5 (1.5)	3.7 (0.8)	0.71 (0.3)	9.7 (1.6)	3.7 (0.8)	0.69 (2.8)
1988	$\hat{\sigma}_{10}$	9.7 (1.2)	4.0 (0.8)	11.2 (2.1)	10.5 (1.6)	3.1 (0.7)	11.9 (2.4)
1989	$\hat{\sigma}_{11}$	10.4 (1.1)	5.6 (0.9)	8.2 (1.5)	11.5 (1.7)	5.9 (1.1)	8.4 (1.7)
1979	$\hat{\sigma}_{\epsilon 1}$	0.25 (0.03)	0.21 (0.04)	0.22 (0.08)	0.25 (0.03)	0.21 (0.04)	0.22 (0.08)
1980	$\hat{\sigma}_{\epsilon 2}$	0.23 (0.03)	0.17 (0.03)	0.27 (0.05)	0.21 (0.03)	0.15 (0.03)	0.26 (0.07)
1981	$\hat{\sigma}_{\epsilon 3}$	0.29 (0.04)	0.11 (0.03)	0.24 (0.04)	0.27 (0.03)	0.11 (0.03)	0.23 (0.04)
1982	$\hat{\sigma}_{\epsilon 4}$	0.26 (0.03)	0.11 (0.02)	0.32 (0.06)	0.26 (0.04)	0.11 (0.02)	0.30 (0.05)
1983	$\hat{\sigma}_{\epsilon 5}$	0.25 (0.03)	0.14 (0.03)	0.26 (0.06)	0.24 (0.03)	0.14 (0.03)	0.22 (0.04)
1984	$\hat{\sigma}_{\epsilon 6}$	0.31 (0.10)	0.12 (0.04)	0.37 (0.08)	0.31 (0.07)	0.12 (0.04)	0.37 (0.08)
1985	$\hat{\sigma}_{\epsilon 7}$	0.22 (0.05)	0.06 (0.02)	0.10 (0.04)	0.21 (0.04)	0.06 (0.02)	0.09 (0.03)
1986	$\hat{\sigma}_{\epsilon 8}$	0.23 (0.05)	0.10 (0.02)	0.27 (0.07)	0.23 (0.05)	0.10 (0.03)	0.26 (0.08)
1987	$\hat{\sigma}_{\epsilon 9}$	0.24 (0.04)	0.09 (0.02)	0.02 (0.01)	0.24 (0.04)	0.09 (0.02)	0.02 (0.01)
1988	$\hat{\sigma}_{\epsilon 10}$	0.23 (0.03)	0.08 (0.02)	0.33 (0.06)	0.22 (0.05)	0.05 (0.01)	0.32 (0.06)
1989	$\hat{\sigma}_{\epsilon 11}$	0.28 (0.03)	0.13 (0.02)	0.28 (0.05)	0.27 (0.04)	0.14 (0.03)	0.27 (0.05)

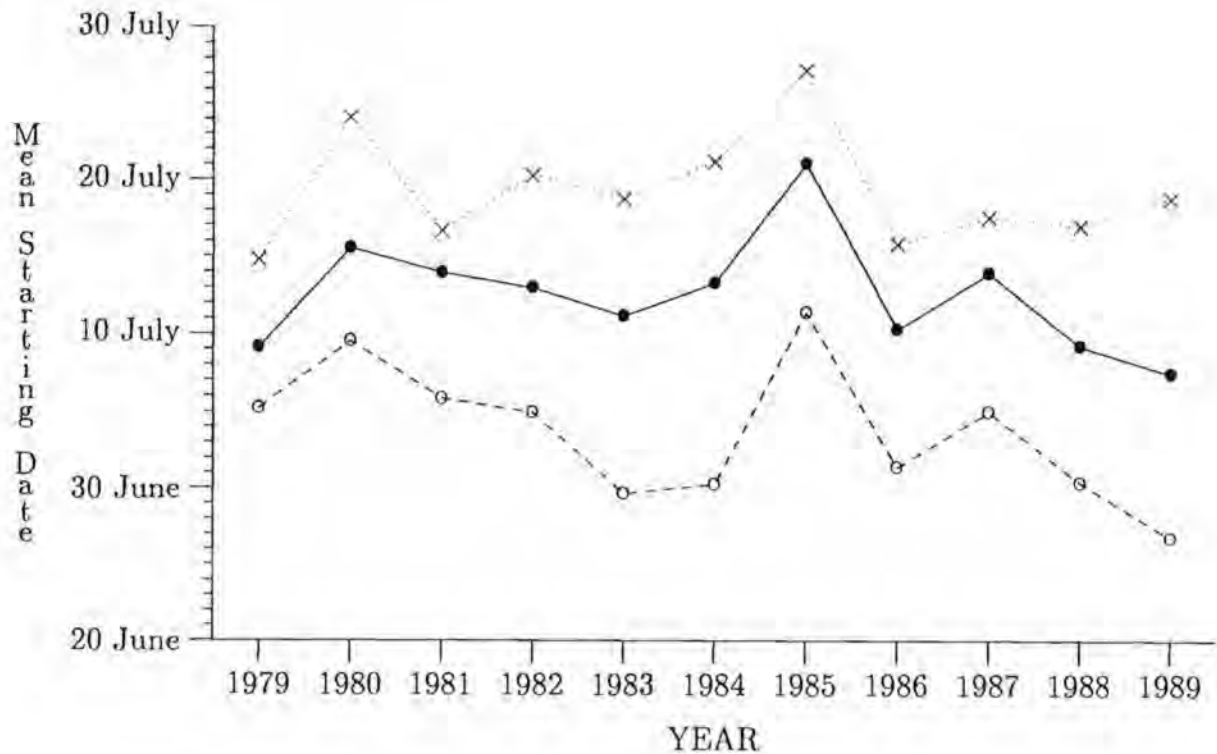
**Table 6 (continued).** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, assuming that the mean starting dates  $\mu_h$ , durations  $\tau_h$  and standard deviations  $\sigma_h$  vary between years,  $h = 1, \dots, 11$ , i.e. model 4. Estimates (standard errors) of the mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) and standard deviations of the regression line  $\sigma_{\epsilon h}$  are also given.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\delta}_1^\dagger$	84.9 (3.5)	79.6 (6.5)	96.7 (7.2)	84.9 (3.5)	79.6 (6.5)	96.7 (7.2)
		24 Aug	19 Aug	5 Sep	24 Aug	19 Aug	5 Sep
1980	$\hat{\delta}_2$	93.8 (2.6)	90.9 (3.4)	97.1 (3.5)	97.6 (2.9)	96.8 (4.4)	98.5 (3.8)
		2 Sep	30 Aug	5 Sep	6 Sep	5 Sep	7 Sep
1981	$\hat{\delta}_3$	85.3 (2.7)	70.4 (1.9)	98.4 (4.2)	91.4 (3.4)	70.4 (1.9)	102.4 (4.8)
		24 Aug	9 Aug	6 Sep	30 Aug	9 Aug	10 Sep
1982	$\hat{\delta}_4$	80.2 (1.7)	76.5 (1.5)	83.9 (2.8)	84.2 (2.6)	76.9 (1.7)	89.1 (3.6)
		19 Aug	16 Aug	23 Aug	23 Aug	16 Aug	28 Aug
1983	$\hat{\delta}_5$	85.8 (3.9)	84.8 (3.4)	83.8 (6.7)	89.1 (4.4)	85.9 (3.7)	91.7 (6.3)
		25 Aug	24 Aug	23 Aug	28 Aug	25 Aug	31 Aug
1984	$\hat{\delta}_6$	79.9 (5.9)	74.2 (3.5)	78.9 (5.2)	81.1 (6.5)	74.3 (3.8)	78.9 (5.2)
		19 Aug	13 Aug	18 Aug	20 Aug	13 Aug	18 Aug
1985	$\hat{\delta}_7$	81.0 (2.3)	79.2 (1.4)	79.0 (1.3)	84.2 (3.0)	79.2 (1.4)	81.3 (2.4)
		20 Aug	18 Aug	18 Aug	23 Aug	18 Aug	20 Aug
1986	$\hat{\delta}_8$	85.2 (2.6)	81.1 (2.1)	89.3 (5.4)	89.7 (4.1)	81.6 (2.3)	94.0 (7.5)
		24 Aug	20 Aug	28 Aug	29 Aug	21 Aug	2 Sep
1987	$\hat{\delta}_9$	83.9 (3.4)	74.6 (2.0)	89.9 (0.4)	84.7 (3.8)	74.6 (2.0)	90.1 (0.4)
		23 Aug	14 Aug	29 Aug	24 Aug	14 Aug	29 Aug
1988	$\hat{\delta}_{10}$	80.5 (2.7)	84.1 (2.2)	80.7 (4.2)	85.7 (4.5)	94.1 (3.1)	84.1 (5.3)
		20 Aug	23 Aug	20 Aug	25 Aug	2 Sep	23 Aug
1989	$\hat{\delta}_{11}$	74.5 (2.2)	68.4 (2.1)	78.3 (2.9)	79.6 (4.5)	70.2 (2.7)	79.5 (3.7)
		14 Aug	7 Aug	17 Aug	19 Aug	9 Aug	19 Aug

†Mean starting date and mean completion date are in days since 1 June (i.e. day 1 = 1 June)



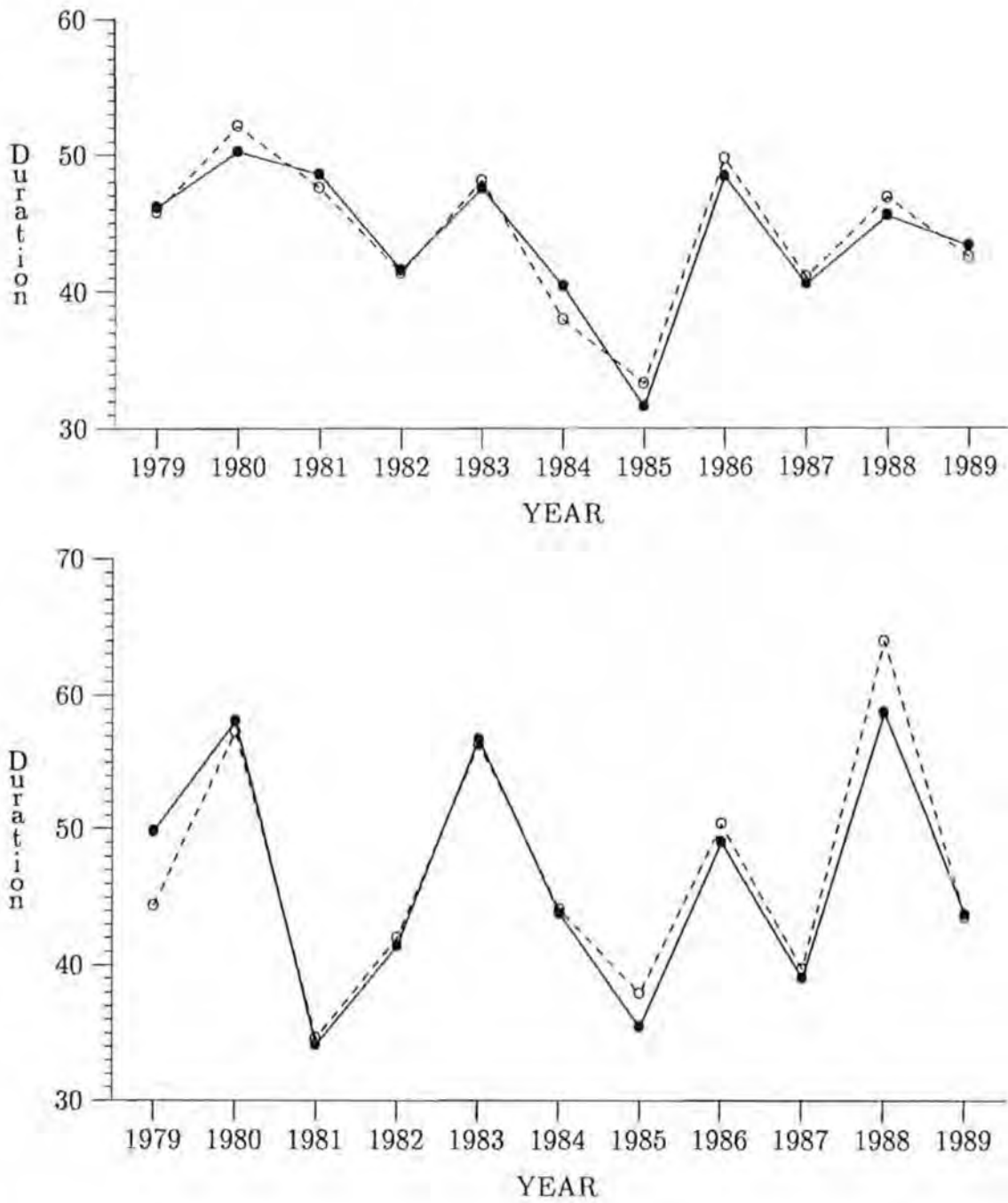
**Figure 5.** Comparison of mean annual starting dates of moult for Willow Warbler between model 2 (solid line), model 3 (dashed line) and model 4 (dotted line) for data type 5. The mean starting dates of moult are given for males (top) and females (bottom) (see Tables 4, 5 and 6).



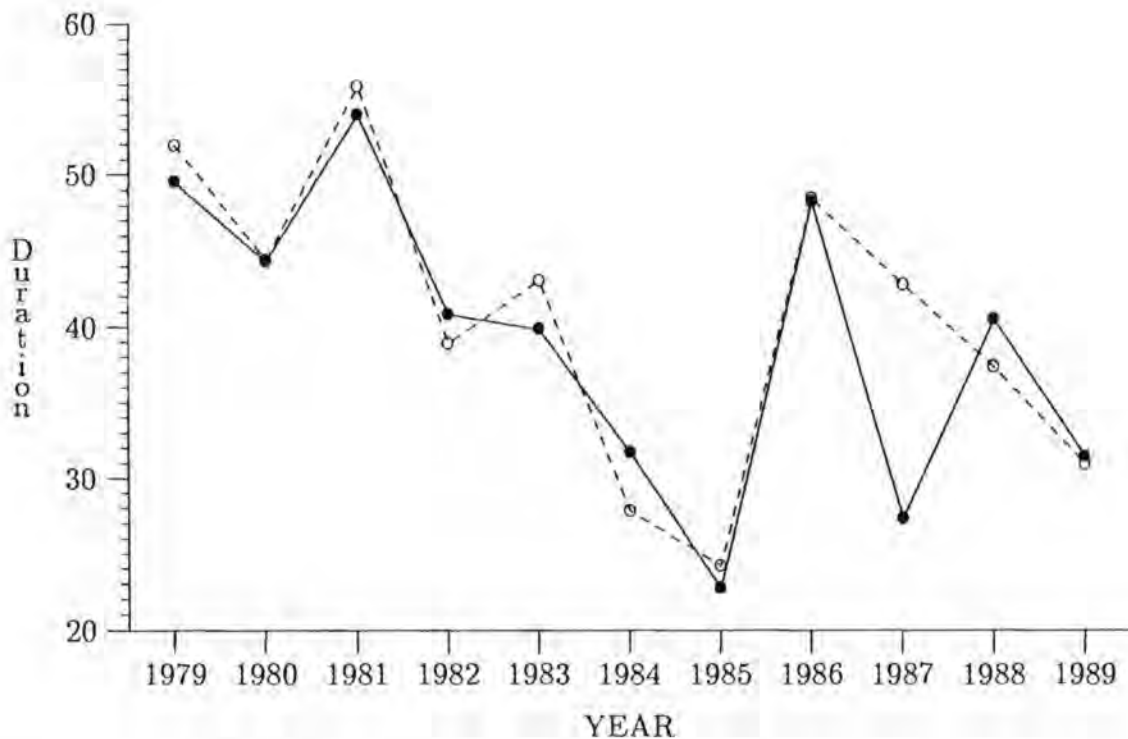
**Figure 6.** Mean annual starting dates of moult for Willow Warbler for model 4 for data type 5. The mean starting dates of moult are given for all the data (solid line), males (dashed line) and females (dotted line) (see Table 6).

mostly altered the parameter estimates of mean starting date and duration of moult from those obtained when the assumption of a constant standard deviation of mean starting date of moult was made, but the estimates for mean completion dates for both models were very similar. This shows that, at least for the present application, the estimates for mean completion dates of moult remained approximately the same whether model 3 or model 4 was fitted to the data, but the estimates for mean starting dates and durations of moult differed according to which model was fitted, but in such a way that the difference was balanced out so that the mean completion dates remained the same.

The mean completion dates of moult of females for model 4 were generally later than that of males (Table 6 and Fig. 13), with the exception that in 1988 males completed moult ten days later than females. In 1980 and 1985, the difference in completion of moult between males and females was only two days, while for most years, females completed moult on average 12 days later than males. In 1981 females completed moult 32 days later than males.



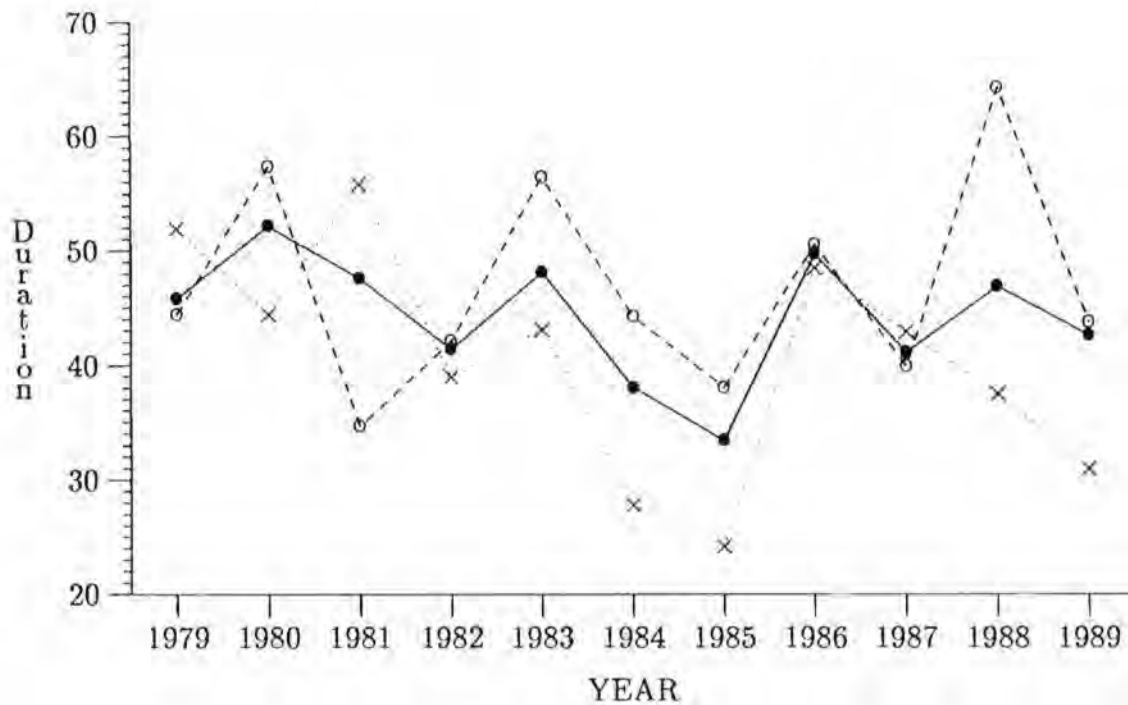
**Figure 7.** Comparison of annual durations of moult between model 3 (solid line) and model 4 (dashed line) for Willow Warbler for data type 5. The durations of moult are given for all the data (top) and males (bottom) (see Tables 5 and 6).



**Figure 8.** Comparison of annual durations of moult between model 3 (solid line) and model 4 (dashed line) for Willow Warbler for females, for data type 5 (see Tables 5 and 6).

From the different models fitted to primary moult, several hypotheses about the parameters were tested. In an hierarchical way, the first hypothesis tested about the parameters of primary moult was whether the mean starting date, the duration and the variance of moult remained constant throughout the years. The null hypothesis of constant model parameters for primary moult was rejected at a high level of significance for the sexes combined, males and females ( $P = 0.008$ ,  $<0.00001$ , and  $0.001$  respectively, for data type 5) (Table 7).

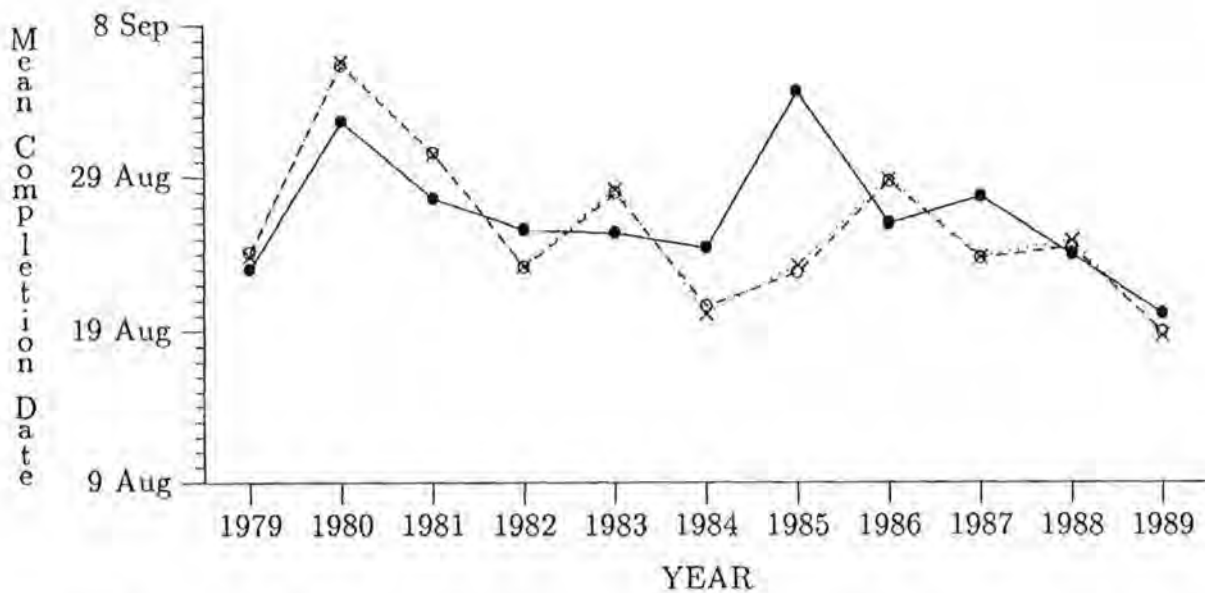
The next hypothesis made about the parameters of primary moult was to assume that the variance of moult remained the same across the years but that the mean starting date and duration of moult varied annually. This hypothesis was accepted when the sexes were combined ( $P = 0.93$ ), but for both males and females the null hypothesis was rejected ( $P = 0.003$  and  $0.02$  respectively) for data type 5 and with similar results for data type 2. Although, given the results of the last hypothesis test on the parameters of primary moult for males and females, any further hypothesis about the parameters of moult should include the fact that the variances change with time, an examination of the parameter estimates



**Figure 9.** Comparison of annual durations of moult between all the data (solid line), males (dashed line) and females (dotted line) for Willow Warbler for model 4, for data type 5 (see Table 6).

of the variance when allowed to vary annually, led to the conclusion that the years 1985 and 1988 might be the cause for a significant difference found for males, when the variance for these years was less than that of other years (Table 6). The variances in 1979, 1980 and 1983 were also higher than that of other years. For females, the years 1985 and 1987 had very small variances, 2.2 and 0.69 respectively, as opposed to the other years which lay in the region of 8.4 and 12.6. When the data was split into the different sexes, the sample size was reduced considerably. Splitting the data by years further reduced the sample size available to estimate the primary moult parameters. For males, there were only four observations of birds in moult for the year 1985 (Table 1), a year that has a relatively small variance. For females, there were only four observations of birds in moult in 1985 and three observations in 1987, both years that had very small standard deviations.

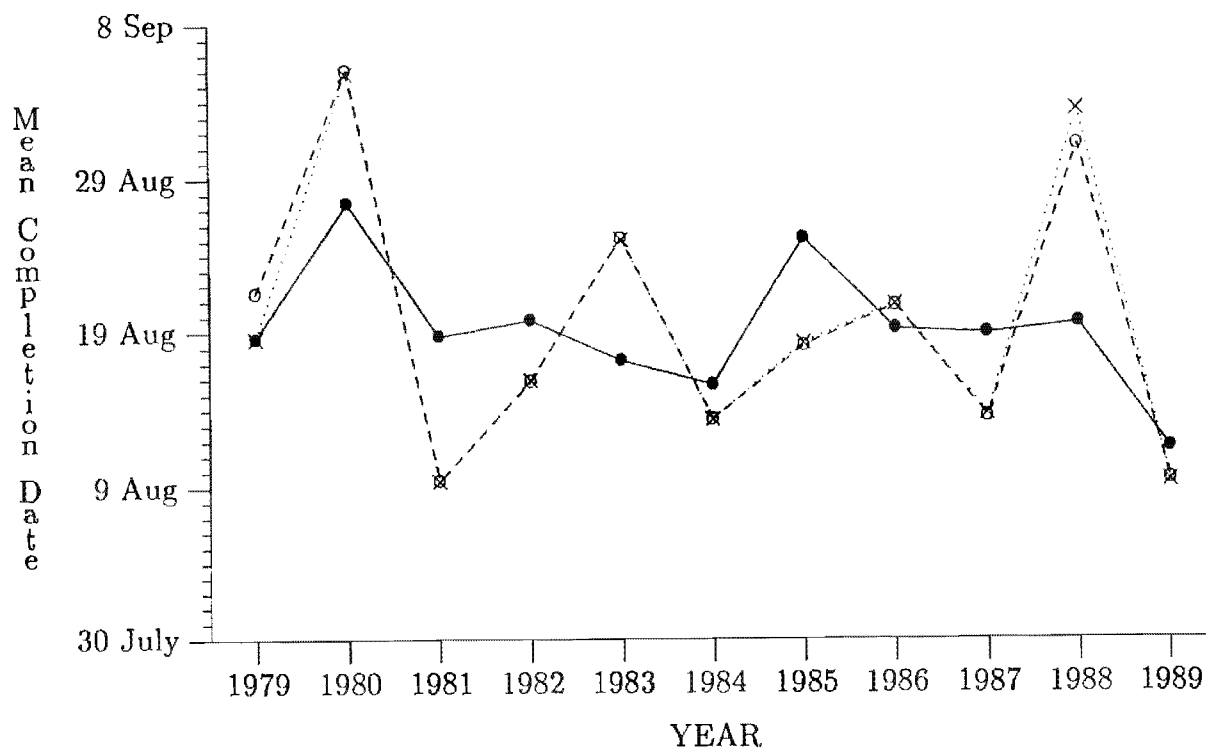
Thus, considering that it is unlikely that the mean starting date and/or duration of moult should remain constant over the years when the standard deviation varies and that, because when the sexes were combined with more observations being available per year



**Figure 10.** Comparison of mean annual completion dates of moult for Willow Warbler between model 2 (solid line), model 3 (dashed line), and model 4 (dotted line) for all the data, for data type 5 (see Tables 4, 5 and 6).

for the different stages of moult, there was no significant difference between the standard deviations over the years, it was considered that either all parameters vary from year to year, in which case model 4 is applicable for Willow Warbler males and females, or it can be assumed that the standard deviations remains constant over the years and further hypotheses are tested to determine whether the mean starting date and/or duration of primary moult vary from year to year. As a precaution to see that the only significant difference was in the standard deviations, the null hypothesis that the mean starting date and the duration of moult are constant over the years versus the alternative hypothesis that at least the mean starting date in at least one year varies significantly from the other years or that the duration of moult in at least one year differs significantly from the other years, was tested. As expected, for all categories of the data set, the null hypothesis was rejected ( $P = 0.0005$  for the complete data,  $P < 0.00001$  for males and  $P = 0.008$  for females).

The next step in testing assumptions about the moult parameters was to test the hypothesis that the duration of moult remained constant throughout the years but not necessarily the mean starting date. For data type 5, this hypothesis was accepted for the complete data ( $P = 0.63$ ), but for males this hypothesis was rejected at a high level of significance ( $P < 0.0001$ ). The hypothesis was also rejected for females, but at a lower level

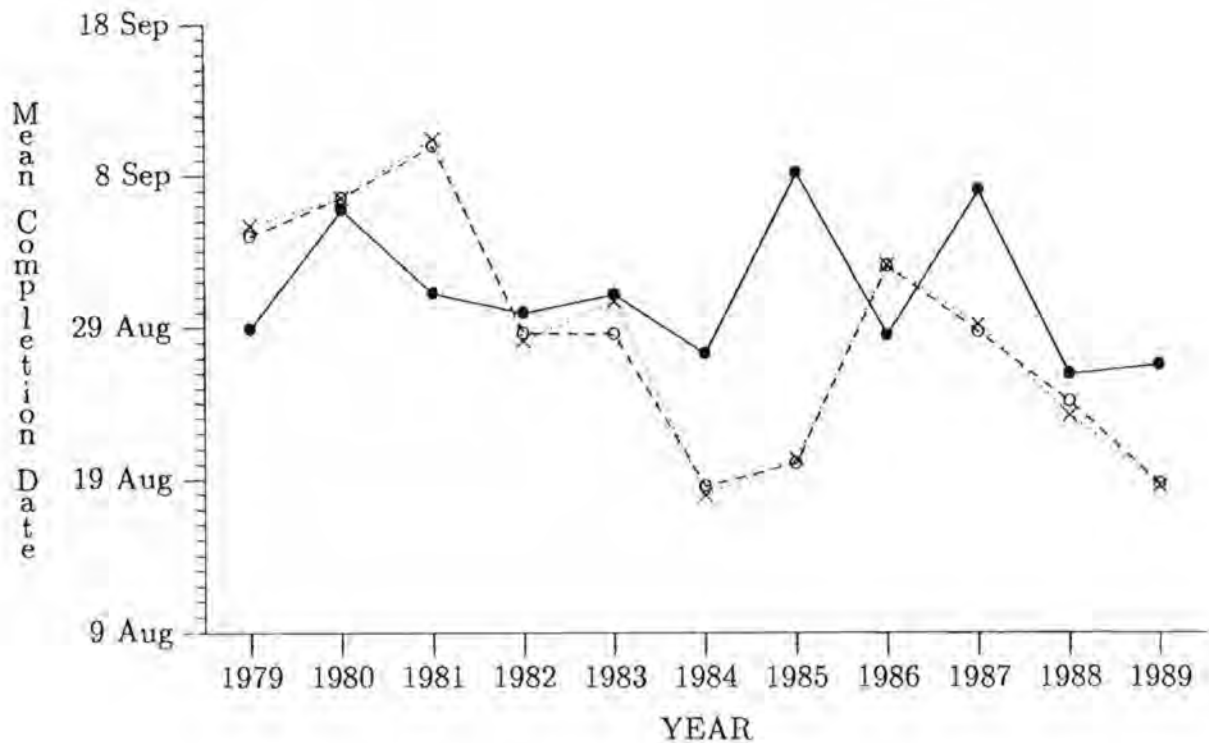


**Figure 11.** Comparison of mean annual completion dates of moult for Willow Warbler between model 2 (solid line), model 3 (dashed line), and model 4 (dotted line) for males, for data type 5 (see Tables 4, 5 and 6).

of significance ( $P = 0.02$ ). For data type 2 the results were similar for males and females but for all the data, the hypothesis was also rejected ( $P = 0.01$ ).

The likelihood ratio test of the null hypothesis of constant mean starting dates was rejected at a high level of significance for the sexes combined and for males ( $P < 0.0001$ ), but for females was rejected only at the 6% significance level ( $P = 0.06$ , data type 5) (Table 7). Thus in all cases there was at least one year with a mean starting date parameter that differed from that of the remaining years.

Under the assumption of constant standard deviation of mean starting date of moult, for all the data, the appropriate model for the complete data was model 2 for data type 5, that is, constant duration of moult and variance but in at least one year the mean starting date of moult differs significantly from the other years. For data type 2, model 3 was the appropriate model, that is, a constant variance but in at least one year the mean starting date of moult differs significantly from the other years and in at least one year the duration of moult differs significantly from the other years. For both males and females the appropriate



**Figure 12.** Comparison of mean annual completion dates of moult for Willow Warbler between model 2 (solid line), model 3 (dashed line), and model 4 (dotted line) for females, for data type 5 (see Tables 4, 5 and 6).

model for primary moult was model 3, that is, both the mean starting date and duration of primary moult differ significantly from the other years in at least one year. This was true for both data type 2 and data type 5. If the assumption of constant variance was not made for males and females, then model 4 was the appropriate model. That is, in at least one year, the standard deviation of mean starting date of moult differs significantly from the other years, in at least one year the mean starting date of moult differs significantly from the other years, and in at least one year the duration of moult differs significantly from the other years.

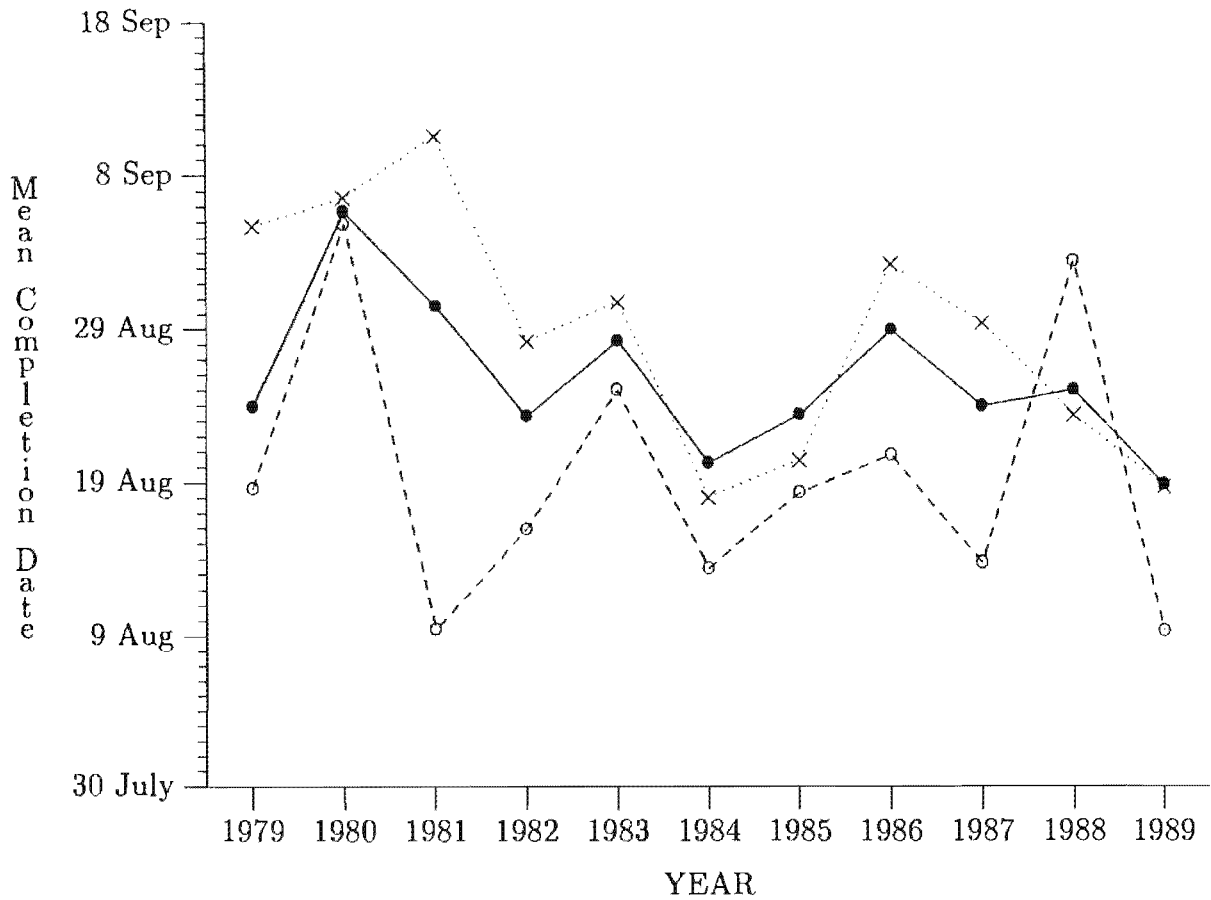


Figure 13. Comparison of mean annual completion dates of moult between all the data (solid line), males (dashed line) and females (dotted line) for Willow Warbler for model 4, for data type 5 (see Table 6).

**Table 7.** Likelihood ratio test for hypotheses on the parameters of primary moult of Willow Warblers at Lake Lagoda, St Petersburg, Russia. The likelihood ratio test statistic  $-2 \log \lambda$  is given by the difference  $D_i - D_j$ , where  $i$  and  $j$  represent the  $i$ th and  $j$ th model respectively, and  $D_i = -2\ell_i$ , where  $\ell_i$  represents the log-likelihood function of the  $i$ th model. The order of the likelihood ratio test statistic given in the table is:  $D_1 - D_2$ ,  $D_2 - D_3$ ,  $D_3 - D_4$ ,  $D_1 - D_3$ , and  $D_1 - D_4$ .

Source of variation	df	Data type 2			Data type 5		
		$D_i = -2\ell_i^\dagger$	$-2 \log \lambda$	$P$ -value	$D_i = -2\ell_i^\dagger$	$-2 \log \lambda$	$P$ -value
<b>All</b>							
$\Theta_1 = [\sigma, \mu, \tau]$	3	419.8	50.4	< 0.00001	177.7	39.6	< 0.0001
$\Theta_2 = [\sigma, \mu_h, \tau]$	13	369.4	22.8	0.01	138.1	8.0	0.63
$\Theta_3 = [\sigma, \mu_h, \tau_h]$	23	346.6	4.4	0.93	130.1	4.4	0.93
$\Theta_4 = [\sigma_h, \mu_h, \tau_h]$	33	342.2	73.2	< 0.00001	125.7	47.6	0.0005
			77.6	< 0.0001		52.0	0.008
<b>Male</b>							
$\Theta_1 = [\sigma, \mu, \tau]$	3	61.3	60.5	< 0.00001	-26.5	47.5	< 0.0001
$\Theta_2 = [\sigma, \mu_h, \tau]$	13	0.7	44.0	< 0.0001	-74.0	36.4	< 0.0001
$\Theta_3 = [\sigma, \mu_h, \tau_h]$	23	-43.3	21.2	0.02	-110.4	26.6	0.003
$\Theta_4 = [\sigma_h, \mu_h, \tau_h]$	33	-64.5	104.6	< 0.00001	-137.0	83.9	< 0.00001
			125.8	< 0.00001		110.5	< 0.00001
<b>Female</b>							
$\Theta_1 = [\sigma, \mu, \tau]$	3	188.3	18.3	0.05	67.8	17.5	0.06
$\Theta_2 = [\sigma, \mu_h, \tau]$	13	170.0	36.9	< 0.0001	50.2	21.0	0.02
$\Theta_3 = [\sigma, \mu_h, \tau_h]$	23	133.1	21.3	0.02	29.2	20.8	0.02
$\Theta_4 = [\sigma_h, \mu_h, \tau_h]$	33	111.8	55.2	< 0.0001	8.4	38.5	0.008
			76.5	< 0.0001		59.3	0.001

†The value of  $i$  corresponds to the  $i$ th model

## CONCLUSION

The models for primary moult proposed in this paper are extensions of the model of Underhill and Zucchini (1988) and Underhill *et al.* (1990) which allows a flexible and statistically coherent approach to the analysis of variation in the timing and duration of moult in different data sets.

The extensions developed here relate to the possibility that the mean starting date of moult, duration and standard deviation vary between years or between localities. Various extensions are developed to allow some or all the parameters of moult to vary between years or localities. The fitting of such models by using the Newton-Raphson algorithm to obtain maximum likelihood estimates require large amounts of "well-behaved" data. This is because there is a tendency, as shown by Underhill and Zucchini (1988), for the maximum likelihood estimates of mean starting date and duration to be negatively correlated with each other. This means that almost the same maximum value for the likelihood function can be achieved by delaying the mean starting date by, say one day, and decreasing the duration by one day. This problem will be aggravated as the number of pairs of correlated parameters increases. The moult problem is an example of a censored regression model with two-sided censoring because the moult index is constrained to lie between 0 and 1 and formulating the moult model as a censored regression, the EM (expectation-maximization) algorithm to maximize the likelihood function can be used which greatly facilitates extensions to the primary moult model as the computation of first and second partial derivatives are no longer needed by the algorithm.

Underhill *et al.* (1992) estimated the parameters of primary moult for Willow Warblers at Lammi, southern Finland, at almost the same latitude as Lake Ladoga, 61°N. The mean starting dates and the durations of moult at the two localities were similar: the mean starting dates were 11 July and 13 July at Lammi and Lake Ladoga, respectively and the durations were 43 and 45 days, respectively (Underhill *et al.* 1992, table 3; Table 3).

The race of Willow Warbler in the study area is the same as that at Lammi, the subspecies *acredula*. For this northern race, Underhill *et al.* (1992) postulated that starting date of primary moult does not show delay with increasing latitude northwards, as was demonstrated for the nominate race *trochilus*. The data from Lake Ladoga, at latitude 61°N, and close to the southern limits of race *acredula*, fit this hypothesis (see Underhill *et al.* 1992, fig. 2).

The difference between the estimates of the parameters using data type 2 and type 5 went in the opposite direction to those obtained for post-nuptial moult in western Europe by Underhill *et al.* (1992) where, for seven localities, the durations of moult estimated using data type 2 were longer than for data type 5. This was interpreted as evidence of emigration of birds which had completed moult. At Lake Ladoga, the duration of moult from data type 2 was shorter than for data type 5, as observed for pre-nuptial moult in Uganda (Underhill *et al.* 1992) and the same explanation is proposed: an influx of birds which had completed moult occurred at a time when the local population was still moulting. These birds are likely to be birds which had completed moult at localities further north, had commenced migration, and were passing through the study area at Lake Ladoga.

The standard deviation of primary moult did not differ significantly between years when the sex of the birds was ignored and all the data combined. Although the likelihood test showed a significant annual variation in the standard deviation of primary moult for both males and females, it was considered that the implications of having the standard deviation of primary moult varying from year to year but a constant duration or mean starting date would have no practicable meaning, in which case the appropriate model for primary moult for males and for females would be model 4, that is when mean starting date, duration and standard deviation vary annually. If, on the other hand, we considered that for some years the sample sizes of birds in moult were rather small and that when the data was combined no significant difference in the standard deviation between years was observed, then the observed significant annual variation in the standard deviation of males and females might be attributable to the small sample sizes, we can further test for significant difference in the other parameters of primary moult of males and females, assuming that there was no significant difference in the standard deviation.

Under the assumption of constant standard deviation, for data type 5, the appropriate model for primary moult ignoring the sex of the birds was model 2, that is, constant duration but variable annual mean starting dates of moult. For data type 2, model 3 was the model suggested for primary moult for the combined data, that is, variable mean starting dates and durations of moult. For both males and females, regardless of data type, model 3 was the model applicable for primary moult, that is, variable mean starting dates and duration of primary moult over the years.

## REFERENCES

- AITKIN, M. 1981. A note on the regression analysis of censored data. *Technometrics* 23: 161–163.
- COX, D.R. AND HINKLEY, D.V. 1974. *Theoretical statistics*. Chapman and Hall, London.
- DEMPSTER, A.P., LAIRD, N.M. AND RUBIN, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B* 39: 1–38.
- DENNIS, J.E. AND SCHNABEL, R.B. 1983. *Numerical methods of unconstrained optimization and nonlinear equations*. Prentice-Hall, New Jersey.
- DOBSON, A.J. 1983. *An introduction to statistical modelling*. Chapman and Hall, London.
- JOHNSON, N.L. AND KOTZ, S. 1970. *Distributions in statistics: Continuous univariate distributions — 1*. Wiley, New York.
- JOHNSON, S., KOTZ, N.L. AND READ, C.B. 1985. *Encyclopedia of statistical sciences*. Wiley, New York.
- LITTLE, R.J.A. AND RUBIN, D.B. 1987. *Statistical analysis with missing data*. Wiley, New York.
- MOOD, A.M., GRAYBILL, F.A. AND BOES, D.C. 1963. *Introduction to the theory of statistics*. McGraw-Hill, New York.
- ORTEGA, J.M. AND RHEINBOLDT, W.C. 1970. *Iterative solution of nonlinear equations in several variables*. Academic Press, New York.
- PAYNE, R.W., LANE, P.W., AINSLEY, A.E., BICKNELL, K.E., DIGBY, P.G.N., HARDING, S.A., LEECH, P.K., SIMPSON, H.R., TODD, A.D., VERRIER, P.J. AND WHITE, R.P. 1987. *GENSTAT 5 reference manual*. Clarendon, Oxford.
- POLASEK, W. AND KRAUSE, A. 1994. The hierarchical Tobit model: a case study in Bayesian computing. *OR Spektrum* 16: 145–154.
- PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T. AND FLANNERY, B.P. 1992. *Numerical recipes in Fortran. The art of scientific computing*, 2nd ed. Cambridge University Press, Cambridge.

- SCHMID, B., POLASEK, W., WEINER, J., KRAUSE, A. AND STOLL, P. 1994. Modelling of discontinuous relationships in biology with censored regression. *American Naturalist* 143: 494–507.
- SCHNEIDER, H. AND WEISSFELD, L. 1986. Estimation in linear models with censored data. *Biometrika* 73: 741–745.
- SVENSSON, L. 1992. Identification guide to European Passerines, 4th ed. Lars Svensson, Stockholm.
- UNDERHILL, L.G. AND ZUCCHINI, W. 1988. A model for avian primary moult. *Ibis* 130: 358–372.
- UNDERHILL, L.G., ZUCCHINI, W. AND SUMMERS, R.W. 1990. A model for avian primary moult — data types based on migration strategies and an example using Redshanks *Tringa totanus*. *Ibis* 132: 118–123.
- UNDERHILL, L.G., PRÛS-JONES, R.P., DOWSETT, R.J., HERROELEN, P., JOHNSON, D.N., LAWN, M.R., NORMAN, S.C., PEARSON, D.J. AND TREE, A.J. 1992. The biannual primary moult of Willow Warblers *Phylloscopus trochilus* in Europe and Africa. *Ibis* 134: 286–297.
- WILKS, S.S. 1962. *Mathematical statistics*. Wiley, New York.
- WOLYNETZ, M.S. 1979a. AS 138: maximum likelihood estimation from confined and censored normal data. *Applied Statistics* 28: 185–195.
- WOLYNETZ, M.S. 1979b. AS 139: maximum likelihood estimation in a linear model from confined and censored normal data. *Applied Statistics* 28: 195–206.

## APPENDIX 1

## The Newton-Raphson and EM algorithms

## The Newton-Raphson Algorithm

The Newton-Raphson algorithm (Ortega and Rheinboldt 1970) is a technique of iteratively solving nonlinear equations of several variables. We define  $\hat{\theta}^{(k)}$  to be the vector of parameter estimates after  $k$  iterations. The algorithm consists of the following steps:

- Step 1:** Calculate (or input) initial estimates of the parameters,  $\hat{\theta}^{(0)} = (\hat{\theta}_1^{(0)}, \dots, \hat{\theta}_p^{(0)})$ .
- Step 2:** Compute  $f^{(k)}$  and  $F^{(k)}$ , where  $f^{(k)}$  is the vector of first partial derivatives and  $F^{(k)}$  is the matrix of second partial derivatives, computed at the  $k$ th iteration.
- Step 3:** Compute the vector  $\delta^{(k)}$  which is the solution to the system of  $p$  linear equations

$$F^{(k)}\delta^{(k)} = f^{(k)},$$

where  $p$  is the number of parameters in the model.

- Step 4:** Set  $\theta^{(k+1)} = \theta^{(k)} - \delta^{(k)}$ , where  $\theta^{(k)}$  contains the parameter estimates computed at the  $k$ th iteration.
- Step 5:** Test for convergence, for example, if the elements of  $f^{(k)}$  are sufficiently close to zero. If the convergence criterion is met then stop, otherwise increase  $k$  by 1 and return to step 2.

This procedure has some disadvantages, namely, the necessity to calculate the matrix of second partial derivatives of the log-likelihood function. As the log-likelihood function becomes more complex and as the number of parameters increases, the more complicated the second partial derivatives become and of higher dimension, thus requiring thorough algebraic calculations and efficient programming. Secondly, initial estimates have to be sufficiently close to the solutions to be able to achieve convergence. This means that sometimes one has to experiment with different initial values before convergence is obtained. As the function that is to be maximized becomes more complex in terms of the number of variables, the more initial values have to be “guessed” accurately. Also the smaller the sample size, the more accurate the initial values have to be to the “true” values for the algorithm to converge.

In the application to the moult models, allowing for the parameters of primary moult to vary from year to year exacerbates this problem as we are in effect reducing the sample size to estimate the various parameters.

An advantage of this method is that once “good” initial values are obtained as starting points, the rate of convergence is quite rapid and very few iterations are necessary before the algorithm achieves convergence. Another advantage is that once the matrix of the second partial derivatives is available, standard errors of the parameters are easily obtainable from the inverse of the  $p \times p$  matrix with entries  $-\partial^2 \ell / \partial \theta_a \partial \theta_b$ , where  $\ell$  represents the log-likelihood function.

### The EM algorithm

The EM (Expectation-Maximization) algorithm (Dempster *et al.* 1977, Little and Rubin 1987) is a method which iteratively computes maximum likelihood estimates in the presence of missing observations and which does not necessitate the computation of the matrix of second derivatives. The basic idea of the EM algorithm is:

1. Replace missing values by estimated values.
2. Estimate the parameters.
3. Re-estimate the missing values assuming the new parameter estimates are correct.
4. Re-estimate parameters, and so on until convergence.

Mathematically, let  $Z$  be a complete data set matrix of  $n$  observations. We assume that the data is generated by a model described by a density function  $f(Z|\phi)$  indexed by unknown parameter  $\phi$ . Given the model and parameter vector  $\phi$ ,  $f(Z|\phi)$  is a function of  $Z$ , that is, of the observations.

Let  $Z = (Z_{\text{obs}}, Z_{\text{mis}})$  where  $Z_{\text{obs}}$  denotes the observed values of  $Z$  and  $Z_{\text{mis}}$  denotes the missing values of  $Z$ . Let  $f(Z|\phi) = f(Z_{\text{obs}}, Z_{\text{mis}}|\phi)$  denote the density function of the joint distribution of  $Z_{\text{obs}}$  and  $Z_{\text{mis}}$ . To obtain the marginal probability density of  $Z_{\text{obs}}$ , the missing data  $Z_{\text{mis}}$  is integrated out. That is,

$$f(Z_{\text{obs}}|\phi) = \int f(Z_{\text{obs}}, Z_{\text{mis}}|\phi) dZ_{\text{mis}}.$$

The likelihood function of  $\phi$  based on  $Z_{\text{obs}}$  is defined to be any function of  $\phi$  proportional to  $f(Z_{\text{obs}}|\phi)$ :

$$L(\phi, Z_{\text{obs}}) \propto f(Z_{\text{obs}}|\phi).$$

In situations where values are missing at random,  $L(\phi, Z_{\text{obs}})$  is called the true likelihood of  $\phi$  based on the observed data  $Z_{\text{obs}}$ . By making use of the complete data specification  $f(Z|\phi)$ , the EM algorithm is used to estimate the parameter  $\phi$  which maximizes  $f(Z_{\text{obs}}|\phi)$ . In other words, we try to maximize the likelihood function

$$L(\phi, Z_{\text{obs}}) = \int f(Z_{\text{obs}}, Z_{\text{mis}}|\phi) dZ_{\text{mis}},$$

with respect to  $\phi$ .

The EM algorithm has a useful and simple interpretation when the complete data  $Z$  has a distribution from the regular exponential family defined by

$$f(Z|\phi) = \frac{b(Z) \exp(\phi t(Z)^T)}{a(\phi)},$$

where

$\phi$  denotes a  $(1 \times p)$  vector of parameters,

$t(Z)$  denotes a  $(1 \times p)$  vector of complete data sufficient statistics, and  $a$  and  $b$  are functions of  $\phi$  and  $Z$  respectively.

For the normal distribution case, the sufficient statistics are given by

$$t(Z) = \begin{pmatrix} \sum_{i=1}^n Z_i \\ \sum_{i=1}^n Z_i^2 \end{pmatrix}.$$

Because the statistic  $t(Z)$  is sufficient for the parameter  $\phi$ , it therefore has all the relevant information contained in  $Z$  for inference about the parameters.

### The E step and the M step of EM.

Each iteration of the EM algorithm involves two steps which are called the expectation step (E step) and the maximization step (M step). The steps given below may be applied if the distribution belongs to the exponential family of distribution.

Suppose that  $\phi^{(k)}$  denotes the current value of  $\phi$  after  $k$  cycles of the algorithm. The next cycle involves the following two steps:

**E step:** At the  $(k + 1)$  cycle, the E step is the computation of the conditional expectation of the complete data sufficient statistics given:

- i) the observed data  $Z_{\text{obs}} = (Z_{\text{obs},1}, \dots, Z_{\text{obs},n})$  and

ii) the estimated value of the parameter from the  $k$ th cycle.

That is, we compute

$$t^{(k)} = E \left[ t(Z) | Z_{\text{obs}}, \phi^{(k)} \right].$$

**M step:** At the  $(k+1)$  cycle, the M step is the maximization of the complete data likelihood function in which the complete data sufficient statistics  $t(Z)$  has been replaced by its conditional expectation obtained in the E step. We set the derivatives of the complete data likelihood function to zero and determine  $\phi^{(k+1)}$ , i.e. as the solution of the equation

$$E(t(Z)|\phi) = t^{(k)},$$

which defines the maximum likelihood estimator of  $\phi$  under the assumption that the distribution comes from a class of the exponential family.

An advantage of the EM algorithm is that it converges reliably, but the disadvantage is that its rate of convergence can be very slow especially if there is a lot of missing data. Another advantage of the EM algorithm is that the computation of the first partial derivatives and the matrix of second partial derivatives is not necessary. The disadvantage of this is that we do not automatically obtain approximate estimates for the standard errors of the parameter estimates.

## APPENDIX 2

## Mean and variance of censored normal distribution for primary moult models

The conditional means and conditional variances of the censored normal distribution for the various models for primary moult are given here. The primary moult models for data type 2 are considered first, followed by the models for primary moult for data type 5. Model 4 is not given here as this model is in essence model 1 fitted to each individual  $h$  data set, where  $h = 1, 2, \dots, H$ . In what follows the standard deviation of the mean starting date of moult is constant over the years unless otherwise specified.

## Data type 2

(a) Model 1: constant starting date and duration of primary moult (common  $y$ -intercept and slope)

The log-likelihood function for model 1 of data type 2 is given by

$$\begin{aligned} \ell_1 = & \sum_{i=1}^I \log \left( 1 - \Phi \left( \frac{\alpha + \beta t_i}{\sigma_\epsilon} \right) \right) + \sum_{j=1}^J \left\{ \log \left( \frac{1}{\sigma_\epsilon} \right) + \log \left( \phi \left( \frac{y_j - (\alpha + \beta u_j)}{\sigma_\epsilon} \right) \right) \right\} \\ & + \sum_{k=1}^K \log \left( 1 - \Phi \left( \frac{1 - (\alpha + \beta v_k)}{\sigma_\epsilon} \right) \right). \end{aligned}$$

Let  $\mu_l$  be the generic name for  $\alpha + \beta d_l$ , where the vector  $d$  is partitioned into vectors  $t$ ,  $u$  and  $v$ , which represent the days for which birds have not started moult, are in moult and have finished moult respectively. Define  $h(x) = \frac{\phi(x)}{1 - \Phi(x)}$  and  $\hat{y}_i = \hat{\alpha} + \hat{\beta}t_i$  for birds that have not started moult and  $\hat{y}_k = \hat{\alpha} + \hat{\beta}v_k$  for birds that have completed moult. Then the first partial derivative of  $\ell_1$  with respect to  $\mu$  is given by

$$\begin{aligned} \frac{\partial \ell_1}{\partial \mu} = & \sum_{i=1}^I \frac{-\phi \left( \frac{\alpha + \beta t_i}{\sigma_\epsilon} \right)}{\sigma_\epsilon \left( 1 - \Phi \left( \frac{\alpha + \beta t_i}{\sigma_\epsilon} \right) \right)} + \sum_{j=1}^J \frac{y_j - (\alpha + \beta u_j)}{\sigma_\epsilon^2} \\ & + \sum_{k=1}^K \frac{\phi \left( \frac{1 - (\alpha + \beta v_k)}{\sigma_\epsilon} \right)}{\sigma_\epsilon \left( 1 - \Phi \left( \frac{1 - (\alpha + \beta v_k)}{\sigma_\epsilon} \right) \right)} \\ = & \sum_{i=1}^I \frac{-1}{\sigma_\epsilon} h \left( \frac{\mu_i}{\sigma_\epsilon} \right) + \sum_{j=1}^J \frac{y_j - (\alpha + \beta u_j)}{\sigma_\epsilon^2} + \sum_{k=1}^K \frac{1}{\sigma_\epsilon} h \left( \frac{1 - \mu_k}{\sigma_\epsilon} \right). \end{aligned}$$

Therefore

$$E \left[ (y_i^* - \hat{y}_i) \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] = -\hat{\sigma}_\epsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\epsilon} \right),$$

which implies that

$$E \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] = \hat{y}_i - \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right), \quad i = 1, 2, \dots, I.$$

Also we have

$$E \left[ (y_k^* - \hat{y}_k) \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] = \hat{\sigma}_\varepsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\varepsilon} \right),$$

therefore

$$E \left[ y_k^* \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] = \hat{y}_k + \hat{\sigma}_\varepsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\varepsilon} \right), \quad k = 1, 2, \dots, K.$$

The partial derivative of  $\ell_1$  with respect to  $\sigma_\varepsilon$  is given by

$$\begin{aligned} \frac{\partial \ell_1}{\partial \sigma_\varepsilon} &= \sum_{i=1}^I \frac{\alpha + \beta t_i}{\sigma_\varepsilon^2} h \left( \frac{\mu_i}{\sigma_\varepsilon} \right) + \sum_{j=1}^J \left\{ \frac{(y_j - (\alpha + \beta u_j))^2}{\sigma_\varepsilon^3} - \frac{1}{\sigma_\varepsilon} \right\} \\ &\quad + \sum_{k=1}^K \frac{1 - (\alpha + \beta v_k)}{\sigma_\varepsilon^2} h \left( \frac{1 - \mu_k}{\sigma_\varepsilon} \right). \end{aligned}$$

Therefore, we have that

$$E \left[ (y_i^* - \hat{y}_i)^2 - \hat{\sigma}_\varepsilon^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] = \hat{y}_i \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right),$$

and so

$$\begin{aligned} E \left[ (y_i^*)^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] &= \hat{\sigma}_\varepsilon^2 + \hat{y}_i \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right) + 2\hat{y}_i E \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] - \hat{y}_i^2 \\ &= \hat{\sigma}_\varepsilon^2 - \hat{y}_i \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right) + \hat{y}_i^2. \end{aligned}$$

Therefore the conditional variance for birds not yet in moult given the observed moult indices is given by

$$\begin{aligned} \text{Var} \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] &= E \left[ (y_i^*)^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] - \left( E \left[ y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] \right)^2 \\ &= \hat{\sigma}_\varepsilon^2 - \hat{y}_i \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right) + \hat{y}_i^2 - \left( \hat{y}_i - \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right) \right)^2 \\ &= \hat{\sigma}_\varepsilon^2 + \hat{y}_i \hat{\sigma}_\varepsilon h \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right) - \hat{\sigma}_\varepsilon^2 h^2 \left( \frac{\hat{y}_i}{\hat{\sigma}_\varepsilon} \right). \end{aligned}$$

Similarly, we have

$$E \left[ (y_k^* - \hat{y}_k)^2 - \hat{\sigma}_\varepsilon^2 \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\varepsilon \right] = (1 - \hat{y}_k) \hat{\sigma}_\varepsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\varepsilon} \right),$$

and so

$$\begin{aligned} E \left[ (y_k^*)^2 \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] &= \hat{\sigma}_\epsilon^2 + (1 - \hat{y}_k) \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right) - \hat{y}_k^2 \\ &\quad + 2\hat{y}_k E \left[ y_k^* \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] \\ &= \hat{\sigma}_\epsilon^2 + \hat{y}_k \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right) + \hat{y}_k^2 + \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right). \end{aligned}$$

Thus the conditional variance for birds that have completed moult given the observed moult indecies is

$$\begin{aligned} \text{Var} \left[ y_k^* \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] &= E \left[ (y_k^*)^2 \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] - \left( E \left[ y_k^* \mid y_k = 1; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon \right] \right)^2 \\ &= \hat{\sigma}_\epsilon^2 + \hat{y}_k \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right) + \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right) + \hat{y}_k^2 \\ &\quad - \left( \hat{y}_k + \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right) \right)^2 \\ &= \hat{\sigma}_\epsilon^2 + \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right) - \hat{y}_k \hat{\sigma}_\epsilon h \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right) - \hat{\sigma}_\epsilon^2 h^2 \left( \frac{1 - \hat{y}_k}{\hat{\sigma}_\epsilon} \right). \end{aligned}$$

**(b) Model 2: different starting dates and constant duration of primary moult (different  $y$ -intercepts and common slope)**

Denote the log-likelihood function for model 2 by  $\ell_2 = \log L(\theta; t, u, y, v)$ , which is given by

$$\begin{aligned} \ell_2 &= \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \log \left( 1 - \Phi \left( \frac{\alpha_h + \beta t_{hi}}{\sigma_\epsilon} \right) \right) + \sum_{j=1}^{J_h} \left\{ \log \left( \frac{1}{\sigma_\epsilon} \right) \right. \right. \\ &\quad \left. \left. + \log \left( \phi \left( \frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\epsilon} \right) \right) \right\} + \sum_{k=1}^{K_h} \log \left( 1 - \Phi \left( \frac{1 - (\alpha_h + \beta v_{hk})}{\sigma_\epsilon} \right) \right) \right). \end{aligned}$$

Define  $\mu_{hi}$  to be the generic name for  $\alpha_h + \beta d_{hi}$ , where the vector  $d$  is partitioned into vectors  $t, u$  and  $v$  as before and  $h = 1, 2, \dots, H$ . Define  $\hat{y}_{hi} = \hat{\alpha}_h + \hat{\beta} t_{hi}$  for birds that have not started moult and  $\hat{y}_{hk} = \hat{\alpha}_h + \hat{\beta} v_{hk}$  for birds that have completed moult.

Then the first partial derivative of  $\ell_2$  with respect to  $\mu_h$  is given by

$$\begin{aligned} \frac{\partial \ell_2}{\partial \mu_h} &= \sum_{i=1}^{I_h} \frac{-\phi \left( \frac{\alpha_h + \beta t_{hi}}{\sigma_\epsilon} \right)}{\sigma_\epsilon \left( 1 - \Phi \left( \frac{\alpha_h + \beta t_{hi}}{\sigma_\epsilon} \right) \right)} + \sum_{j=1}^{J_h} \frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\epsilon^2} \\ &\quad + \sum_{k=1}^{K_h} \frac{\phi \left( \frac{1 - (\alpha_h + \beta v_{hk})}{\sigma_\epsilon} \right)}{\sigma_\epsilon \left( 1 - \Phi \left( \frac{1 - (\alpha_h + \beta v_{hk})}{\sigma_\epsilon} \right) \right)} \\ &= \sum_{i=1}^{I_h} \frac{-1}{\sigma_\epsilon} h \left( \frac{\mu_{hi}}{\sigma_\epsilon} \right) + \sum_{j=1}^{J_h} \frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\epsilon^2} + \sum_{k=1}^{K_h} \frac{1}{\sigma_\epsilon} h \left( \frac{1 - \mu_{hk}}{\sigma_\epsilon} \right), \end{aligned}$$

for  $h = 1, 2, \dots, H$ . Then the conditional expectations of the censored regression are

$$\begin{aligned} E \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] &= \widehat{y}_{hi} - \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) \\ E \left[ y_{hk}^* \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] &= \widehat{y}_{hk} + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right), \end{aligned}$$

where  $i = 1, 2, \dots, I_h$ ,  $k = 1, 2, \dots, K_h$  and  $h = 1, 2, \dots, H$ .

The partial derivative of  $\ell_2$  with respect to  $\sigma_\varepsilon$  is

$$\begin{aligned} \frac{\partial \ell_2}{\partial \sigma_\varepsilon} &= \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \frac{\alpha_h + \beta t_{hi}}{\sigma_\varepsilon^2} h \left( \frac{\mu_{hi}}{\sigma_\varepsilon} \right) + \sum_{j=1}^{J_h} \left\{ \frac{(y_{hi} - (\alpha_h + \beta u_{hj}))^2}{\sigma_\varepsilon^3} - \frac{1}{\sigma_\varepsilon} \right\} \right. \\ &\quad \left. + \sum_{k=1}^{K_h} \frac{1 - (\alpha_h + \beta v_{hk})}{\sigma_\varepsilon^2} h \left( \frac{1 - \mu_{hk}}{\sigma_\varepsilon} \right) \right). \end{aligned}$$

Then

$$E \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] = \widehat{\sigma}_\varepsilon^2 - \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) + \widehat{y}_{hi}^2,$$

and

$$E \left[ (y_{hk}^*)^2 \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] = \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hk} \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) + \widehat{y}_{hk}^2 + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right).$$

Therefore the conditional variances of the censored regression are given by

$$\begin{aligned} \text{Var} \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] &= E \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] \\ &\quad - \left( E \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] \right)^2 \\ &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right), \end{aligned}$$

and

$$\begin{aligned} \text{Var} \left[ y_{hk}^* \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] &= E \left[ (y_{hk}^*)^2 \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] \\ &\quad - \left( E \left[ y_{hk}^* \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] \right)^2 \\ &= \widehat{\sigma}_\varepsilon^2 + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{y}_{hk} \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) \\ &\quad - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right). \end{aligned}$$

(c) **Model 3: different starting dates and durations of primary moult (different  $y$ -intercepts and slopes)**

Let the log-likelihood function for model 3 be denoted by  $\ell_3 = \log L(\theta; t, u, y, v)$ . Then

$$\ell_3 = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \log \left( 1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\epsilon} \right) \right) + \sum_{j=1}^{J_h} \left\{ \log \left( \frac{1}{\sigma_\epsilon} \right) + \log \left( \phi \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\epsilon} \right) \right) \right\} + \sum_{k=1}^{K_h} \log \left( 1 - \Phi \left( \frac{1 - (\alpha_h + \beta_h v_{hk})}{\sigma_\epsilon} \right) \right) \right).$$

Define  $\mu_{hi}$  to be the generic name for  $\alpha_h + \beta_h d_{hi}$ , where the vector  $d$  is partitioned into vectors  $t$ ,  $u$  and  $v$ , which represent the days for which birds have not started moult, are in moult and have finished moult respectively and  $h = 1, 2, \dots, H$ . Define  $\widehat{y}_{hi} = \widehat{\alpha}_h + \widehat{\beta}_h t_{hi}$  for birds that have not started moult and  $\widehat{y}_{hk} = \widehat{\alpha}_h + \widehat{\beta}_h v_{hk}$  for birds that have completed moult. Then the first partial derivative of  $\ell_3$  with respect to  $\mu_h$  is given by

$$\begin{aligned} \frac{\partial \ell_3}{\partial \mu_h} &= \sum_{i=1}^{I_h} \frac{-\phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\epsilon} \right)}{\sigma_\epsilon \left( 1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\epsilon} \right) \right)} + \sum_{j=1}^{J_h} \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\epsilon^2} \\ &\quad + \sum_{k=1}^{K_h} \frac{\phi \left( \frac{1 - (\alpha_h + \beta_h v_{hk})}{\sigma_\epsilon} \right)}{\sigma_\epsilon \left( 1 - \Phi \left( \frac{1 - (\alpha_h + \beta_h v_{hk})}{\sigma_\epsilon} \right) \right)} \\ &= \sum_{i=1}^{I_h} \frac{-1}{\sigma_\epsilon} h \left( \frac{\mu_{hi}}{\sigma_\epsilon} \right) + \sum_{j=1}^{J_h} \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\epsilon^2} + \sum_{k=1}^{K_h} \frac{1}{\sigma_\epsilon} h \left( \frac{1 - \mu_{hk}}{\sigma_\epsilon} \right), \end{aligned}$$

where  $h = 1, 2, \dots, H$ . Then the conditional expectations of the censored regression are given by

$$\begin{aligned} E \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\epsilon \right] &= \widehat{y}_{hi} - \widehat{\sigma}_\epsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\epsilon} \right) \\ E \left[ y_{hk}^* \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\epsilon \right] &= \widehat{y}_{hk} + \widehat{\sigma}_\epsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\epsilon} \right), \end{aligned}$$

where  $i = 1, 2, \dots, I_h$ ,  $k = 1, 2, \dots, K_h$  and  $h = 1, 2, \dots, H$ .

The partial derivative of  $\ell_3$  with respect to  $\sigma_\epsilon$  is

$$\begin{aligned} \frac{\partial \ell_3}{\partial \sigma_\epsilon} &= \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\epsilon^2} h \left( \frac{\mu_{hi}}{\sigma_\epsilon} \right) + \sum_{j=1}^{J_h} \left\{ \frac{(y_{hj} - (\alpha_h + \beta_h u_{hj}))^2}{\sigma_\epsilon^3} - \frac{1}{\sigma_\epsilon} \right\} \right. \\ &\quad \left. + \sum_{k=1}^{K_h} \frac{1 - (\alpha_h + \beta_h v_{hk})}{\sigma_\epsilon^2} h \left( \frac{1 - \mu_{hk}}{\sigma_\epsilon} \right) \right). \end{aligned}$$

Thus

$$E \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] = \widehat{\sigma}_\varepsilon^2 - \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) + \widehat{y}_{hi}^2,$$

and

$$E \left[ (y_{hk}^*)^2 \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] = \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hk} \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) + \widehat{y}_{hk}^2 + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right).$$

Therefore the conditional variances of the censored regression are given by

$$\begin{aligned} \text{Var} \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] &= E \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \\ &\quad - \left( E \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \right)^2 \\ &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right), \end{aligned}$$

and

$$\begin{aligned} \text{Var} \left[ y_{hk}^* \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] &= E \left[ (y_{hk}^*)^2 \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \\ &\quad - \left( E \left[ y_{hk}^* \mid y_{hk} = 1; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \right)^2 \\ &= \widehat{\sigma}_\varepsilon^2 + \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{y}_{hk} \widehat{\sigma}_\varepsilon h \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right) \\ &\quad - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{1 - \widehat{y}_{hk}}{\widehat{\sigma}_\varepsilon} \right). \end{aligned}$$

## Data type 5

- (a) **Model 1: constant starting date and duration of primary moult (common  $y$ -intercept and slope)**

The log-likelihood function for model 1 of data type 5 is given by

$$\begin{aligned} \ell_4 &= \sum_{i=1}^I \left( \log \left( 1 - \Phi \left( \frac{\alpha + \beta t_i}{\sigma_\varepsilon} \right) \right) - \log \left( \Phi \left( \frac{1 - (\alpha + \beta t_i)}{\sigma_\varepsilon} \right) \right) \right) \\ &\quad + \sum_{j=1}^J \left( \log \left( \frac{1}{\sigma_\varepsilon} \right) + \log \left( \phi \left( \frac{y_j - (\alpha + \beta u_j)}{\sigma_\varepsilon} \right) \right) - \log \left( \Phi \left( \frac{1 - (\alpha + \beta u_j)}{\sigma_\varepsilon} \right) \right) \right). \end{aligned}$$

Let  $\mu_l$  be the generic name for  $\alpha + \beta d_l$ , where the vector  $d$  is partitioned into vectors  $t$  and  $u$ , which represent the days for which birds have not started moult and those that are in moult respectively. Define the functions  $h(x) = \frac{\phi(x)}{1 - \Phi(x)}$  and  $f(x) = \frac{\phi(x)}{\Phi(x)}$ .

Let  $\hat{y}_i = \hat{\alpha} + \hat{\beta}t_i$  for birds that have not started moult. Then the first partial derivative of  $\ell_4$  with respect to  $\mu$  is given by

$$\begin{aligned} \frac{\partial \ell_4}{\partial \mu} &= \sum_{i=1}^I \left( \frac{-\phi\left(\frac{\alpha + \beta t_i}{\sigma_\epsilon}\right)}{\sigma_\epsilon \left(1 - \Phi\left(\frac{\alpha + \beta t_i}{\sigma_\epsilon}\right)\right)} + \frac{\phi\left(\frac{1 - (\alpha + \beta t_i)}{\sigma_\epsilon}\right)}{\sigma_\epsilon \Phi\left(\frac{1 - (\alpha + \beta t_i)}{\sigma_\epsilon}\right)} \right) \\ &\quad + \sum_{j=1}^J \left( \frac{y_j - (\alpha + \beta u_j)}{\sigma_\epsilon^2} + \frac{\phi\left(\frac{1 - (\alpha + \beta u_j)}{\sigma_\epsilon}\right)}{\sigma_\epsilon \Phi\left(\frac{1 - (\alpha + \beta u_j)}{\sigma_\epsilon}\right)} \right) \\ &= \sum_{i=1}^I \left( \frac{-1}{\sigma_\epsilon} h\left(\frac{\mu_i}{\sigma_\epsilon}\right) + \frac{1}{\sigma_\epsilon} f\left(\frac{1 - \mu_i}{\sigma_\epsilon}\right) \right) + \sum_{j=1}^J \left( \frac{y_j - (\alpha + \beta u_j)}{\sigma_\epsilon^2} + \frac{1}{\sigma_\epsilon} f\left(\frac{1 - \mu_j}{\sigma_\epsilon}\right) \right). \end{aligned}$$

Therefore

$$E\left[(y_i^* - \hat{y}_i) \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right] = -\hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right),$$

which implies that

$$E\left[y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right] = \hat{y}_i - \hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right), \quad i = 1, 2, \dots, I.$$

The partial derivative of  $\ell_4$  with respect to  $\sigma_\epsilon$  is

$$\begin{aligned} \frac{\partial \ell_4}{\partial \sigma_\epsilon} &= \sum_{i=1}^I \left( \frac{\alpha + \beta t_i}{\sigma_\epsilon^2} h\left(\frac{\mu_i}{\sigma_\epsilon}\right) + \frac{1 - (\alpha + \beta t_i)}{\sigma_\epsilon^2} f\left(\frac{1 - \mu_i}{\sigma_\epsilon}\right) \right) \\ &\quad + \sum_{j=1}^J \left( \frac{(y_j - (\alpha + \beta u_j))^2}{\sigma_\epsilon^3} + \frac{1 - (\alpha + \beta u_j)}{\sigma_\epsilon^2} f\left(\frac{1 - \mu_j}{\sigma_\epsilon}\right) - \frac{1}{\sigma_\epsilon} \right). \end{aligned}$$

Therefore, we have that

$$E\left[(y_i^* - \hat{y}_i)^2 - \hat{\sigma}_\epsilon^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right] = \hat{y}_i \hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right),$$

and so

$$\begin{aligned} E\left[(y_i^*)^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right] &= \hat{\sigma}_\epsilon^2 + \hat{y}_i \hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right) + 2\hat{y}_i E\left[y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right] - \hat{y}_i^2 \\ &= \hat{\sigma}_\epsilon^2 - \hat{y}_i \hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right) + \hat{y}_i^2. \end{aligned}$$

Thus the conditional variance for birds not yet in moult given the observed moult indices is given by

$$\begin{aligned} \text{Var}\left[y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right] &= E\left[(y_i^*)^2 \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right] - \left(E\left[y_i^* \mid y_i = 0; \hat{\alpha}, \hat{\beta}, \hat{\sigma}_\epsilon\right]\right)^2 \\ &= \hat{\sigma}_\epsilon^2 - \hat{y}_i \hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right) + \hat{y}_i^2 - \left(\hat{y}_i - \hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right)\right)^2 \\ &= \hat{\sigma}_\epsilon^2 + \hat{y}_i \hat{\sigma}_\epsilon h\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right) - \hat{\sigma}_\epsilon^2 h^2\left(\frac{\hat{y}_i}{\hat{\sigma}_\epsilon}\right). \end{aligned}$$

(b) **Model 2: different starting dates and constant duration of primary moult (different  $y$ -intercepts and common slope)**

Denote the log-likelihood function for model 2 by  $\ell_5 = \log L(\theta; t, u, y)$ . Then

$$\ell_5 = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \left\{ \log \left( 1 - \Phi \left( \frac{\alpha_h + \beta t_{hi}}{\sigma_\varepsilon} \right) \right) - \log \left( \Phi \left( \frac{1 - (\alpha_h + \beta t_{hi})}{\sigma_\varepsilon} \right) \right) \right\} \right. \\ \left. + \sum_{j=1}^{J_h} \left\{ \log \left( \frac{1}{\sigma_\varepsilon} \right) + \log \left( \phi \left( \frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon} \right) \right) - \log \left( \Phi \left( \frac{1 - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon} \right) \right) \right\} \right).$$

Define  $\mu_{hi}$  to be the generic name for  $\alpha_h + \beta d_{hi}$ , where the vector  $d$  is partitioned into vectors  $t$  and  $u$  as before, where  $h = 1, 2, \dots, H$ . Define  $\widehat{y}_{hi} = \widehat{\alpha}_h + \widehat{\beta} t_{hi}$  for birds that have not yet started moult. Then the first partial derivative of  $\ell_5$  with respect to  $\mu_h$  is given by

$$\frac{\partial \ell_5}{\partial \mu_h} = \sum_{i=1}^{I_h} \left( \frac{-\phi \left( \frac{\alpha_h + \beta t_{hi}}{\sigma_\varepsilon} \right)}{\sigma_\varepsilon \left( 1 - \Phi \left( \frac{\alpha_h + \beta t_{hi}}{\sigma_\varepsilon} \right) \right)} + \frac{\phi \left( \frac{1 - (\alpha_h + \beta t_{hi})}{\sigma_\varepsilon} \right)}{\sigma_\varepsilon \Phi \left( \frac{1 - (\alpha_h + \beta t_{hi})}{\sigma_\varepsilon} \right)} \right) \\ + \sum_{j=1}^{J_h} \left( \frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon^2} + \frac{\phi \left( \frac{1 - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon} \right)}{\sigma_\varepsilon \Phi \left( \frac{1 - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon} \right)} \right) \\ = \sum_{i=1}^{I_h} \left( \frac{-1}{\sigma_\varepsilon} h \left( \frac{\mu_{hi}}{\sigma_\varepsilon} \right) + \frac{1}{\sigma_\varepsilon} f \left( \frac{1 - \mu_{hj}}{\sigma_\varepsilon} \right) \right) + \sum_{j=1}^{J_h} \left( \frac{y_{hj} - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon^2} \right. \\ \left. + \frac{1}{\sigma_\varepsilon} f \left( \frac{1 - \mu_{hj}}{\sigma_\varepsilon} \right) \right).$$

Then the conditional expectation of the censored regression is

$$E \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] = \widehat{y}_{hi} - \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right),$$

for  $i = 1, 2, \dots, I_h$  and  $h = 1, 2, \dots, H$ .

The partial derivative of  $\ell_5$  with respect to  $\sigma_\varepsilon$  is

$$\frac{\partial \ell_5}{\partial \sigma_\varepsilon} = \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \left\{ \frac{\alpha_h + \beta t_{hi}}{\sigma_\varepsilon^2} h \left( \frac{\mu_{hi}}{\sigma_\varepsilon} \right) + \frac{1 - (\alpha_h + \beta t_{hi})}{\sigma_\varepsilon^2} f \left( \frac{1 - \mu_{hi}}{\sigma_\varepsilon} \right) \right\} \right. \\ \left. + \sum_{j=1}^{J_h} \left\{ \frac{(y_{hi} - (\alpha_h + \beta u_{hj}))^2}{\sigma_\varepsilon^3} + \frac{1 - (\alpha_h + \beta u_{hj})}{\sigma_\varepsilon^2} f \left( \frac{1 - \mu_{hj}}{\sigma_\varepsilon} \right) - \frac{1}{\sigma_\varepsilon} \right\} \right).$$

Therefore

$$E \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}, \widehat{\sigma}_\varepsilon \right] = \widehat{\sigma}_\varepsilon^2 - \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) + \widehat{y}_{hi}^2,$$

and the conditional variance of the censored regression is given by

$$\begin{aligned} \text{Var} \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] &= \text{E} \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \\ &\quad - \left( \text{E} \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \right)^2 \\ &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right). \end{aligned}$$

**(c) Model 3: different starting dates and durations of primary moult (different  $y$ -intercepts and slopes)**

Let the log-likelihood function for model 3 be denoted by  $\ell_6 = \log L(\theta; t, u, y)$ , then

$$\begin{aligned} \ell_6 &= \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \left\{ \log \left( 1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\varepsilon} \right) \right) - \log \left( \Phi \left( \frac{1 - (\alpha_h + \beta_h t_{hi})}{\sigma_\varepsilon} \right) \right) \right\} \right. \\ &\quad \left. + \sum_{j=1}^{J_h} \left\{ \log \left( \frac{1}{\sigma_\varepsilon} \right) + \log \left( \phi \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon} \right) \right) \right. \right. \\ &\quad \left. \left. - \log \left( \Phi \left( \frac{1 - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon} \right) \right) \right\} \right). \end{aligned}$$

Define  $\mu_{hi}$  to be the generic name for  $\alpha_h + \beta_h d_{hi}$ , where the vector  $d$  is partitioned into vectors  $t$  and  $u$  which represent the days for which birds have not started moult and those actively in moult respectively, and  $h = 1, 2, \dots, H$ . Define  $\widehat{y}_{hi} = \widehat{\alpha}_h + \widehat{\beta}_h t_{hi}$  for birds that have not started moult. Then the first partial derivative of  $\ell_6$  with respect to  $\mu_h$  is given by

$$\begin{aligned} \frac{\partial \ell_6}{\partial \mu_h} &= \sum_{i=1}^{I_h} \left( \frac{-\phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\varepsilon} \right)}{\sigma_\varepsilon \left( 1 - \Phi \left( \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\varepsilon} \right) \right)} + \frac{\phi \left( \frac{1 - (\alpha_h + \beta_h t_{hi})}{\sigma_\varepsilon} \right)}{\sigma_\varepsilon \Phi \left( \frac{1 - (\alpha_h + \beta_h t_{hi})}{\sigma_\varepsilon} \right)} \right) \\ &\quad + \sum_{j=1}^{J_h} \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon^2} + \frac{\phi \left( \frac{1 - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon} \right)}{\sigma_\varepsilon \Phi \left( \frac{1 - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon} \right)} \right) \\ &= \sum_{i=1}^{I_h} \left( \frac{-1}{\sigma_\varepsilon} h \left( \frac{\mu_{hi}}{\sigma_\varepsilon} \right) + \frac{1}{\sigma_\varepsilon} f \left( \frac{1 - \mu_{hj}}{\sigma_\varepsilon} \right) \right) + \sum_{j=1}^{J_h} \left( \frac{y_{hj} - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon^2} \right. \\ &\quad \left. + \frac{1}{\sigma_\varepsilon} f \left( \frac{1 - \mu_{hj}}{\sigma_\varepsilon} \right) \right). \end{aligned}$$

Then the conditional expectation of the censored regression is given by

$$\text{E} \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] = \widehat{y}_{hi} - \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right),$$

for  $i = 1, 2, \dots, I_h$  and  $h = 1, 2, \dots, H$ .

The partial derivative of  $\ell_6$  with respect to  $\sigma_\varepsilon$  is

$$\begin{aligned} \frac{\partial \ell_6}{\partial \sigma_\varepsilon} = & \sum_{h=1}^H \left( \sum_{i=1}^{I_h} \left\{ \frac{\alpha_h + \beta_h t_{hi}}{\sigma_\varepsilon^2} h \left( \frac{\mu_{hi}}{\sigma_\varepsilon} \right) + \frac{1 - (\alpha_h + \beta_h t_{hi})}{\sigma_\varepsilon^2} f \left( \frac{1 - \mu_{hi}}{\sigma_\varepsilon} \right) \right\} \right. \\ & \left. + \sum_{j=1}^{J_h} \left\{ \frac{(y_{hj} - (\alpha_h + \beta_h u_{hj}))^2}{\sigma_\varepsilon^3} + \frac{1 - (\alpha_h + \beta_h u_{hj})}{\sigma_\varepsilon^2} f \left( \frac{1 - \mu_{hj}}{\sigma_\varepsilon} \right) - \frac{1}{\sigma_\varepsilon} \right\} \right). \end{aligned}$$

Then

$$E \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] = \widehat{\sigma}_\varepsilon^2 - \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) + \widehat{y}_{hi}^2,$$

and the conditional variance of the censored regression is given by

$$\begin{aligned} \text{Var} \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] &= E \left[ (y_{hi}^*)^2 \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \\ &\quad - \left( E \left[ y_{hi}^* \mid y_{hi} = 0; \widehat{\alpha}_h, \widehat{\beta}_h, \widehat{\sigma}_\varepsilon \right] \right)^2 \\ &= \widehat{\sigma}_\varepsilon^2 + \widehat{y}_{hi} \widehat{\sigma}_\varepsilon h \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right) - \widehat{\sigma}_\varepsilon^2 h^2 \left( \frac{\widehat{y}_{hi}}{\widehat{\sigma}_\varepsilon} \right). \end{aligned}$$

## APPENDIX 3

## Algorithms for censored regression models

The algorithm of Schmid *et al.* (1994) has been adopted here, with the appropriate changes to our own requirements. The statistical package GENSTAT (Payne *et al.* 1987) was used to program the following algorithms. Unless otherwise specified, the standard deviation of the mean starting date of moult is constant.

## Data type 2

(a) Model 1: constant starting date and duration of primary moult (common  $y$ -intercept and slope)

1. Define vectors  $t$  and  $y$  of length  $n$  for the independent and the dependent variable respectively.
2. Define the vectors  $w_1$  and  $w_2$  such that

$$w_1 = \begin{cases} 0 & \text{if } y = 0 \\ 1 & \text{otherwise,} \end{cases}$$

and

$$w_2 = \begin{cases} 1 & \text{if } y = 1 \\ 0 & \text{otherwise.} \end{cases}$$

3. Fit the linear regression  $y = a + bt$  (by ordinary least squares) for values for which  $0 \leq y \leq 1$ .
4. Assign  $A = a$ ,  $B = b$  and  $s_1 = \sqrt{rms}$ , where  $a$  and  $b$  are the initial parameter estimates obtained in Step 3, and  $rms$  is the residual mean square.
5. Assign  $f_v =$  fitted values, for all  $n$  units.
6. Define  $z = \frac{f_v}{s_1}$ , the standerdized fitted values.
7. Define the vector  $f_z = \Phi(z)$ , i.e.  $f_z$  contains the values of the standard normal cumulative distribution function for each value of  $z$ .
8. Compute  $p_z = 0.3986 \exp(-z^2/2)$ , i.e. the normal density values of  $z$ .
9. Compute  $h_z = \frac{s_1 p_z}{(1 - f_z)}$ ;  $h_z = 0$  if division by zero.

10. Define  $x = \frac{1 - f_v}{s_1}$ , the standardized (1 - fitted values).
11. Define the vector  $f_x = \Phi(x)$ , i.e. the standard normal cumulative distribution function for each value of  $x$ .
12. Compute  $p_x = 0.3986 \exp(-x^2/2)$ , i.e. the normal density values of  $x$ .
13. Compute  $h_x = \frac{s_1 p_x}{(1 - f_x)}$ ;  $h_x = 0$  if division by zero.
14. Assign  $n_1 = w_1 y + (1 - w_1)(f_v - h_x)$ .
15. Assign  $n_2 = (1 - w_2)n_1 + w_2(f_v + h_x)$ .
16. Let  $v_1 = (1 - w_1)(s_1^2 + f_v h_x - h_x^2)$ .
17. Let  $v_3 = w_2(s_1^2 + h_x - f_v h_x - h_x^2)$ .
18. Fit the linear regression  $n_2 = a + bt$  (by ordinary least squares) on all  $n$  units and assign new fitted values to  $f_v$ .
19. Compare the new regression parameters  $a$  and  $b$  with the old parameters  $A$  and  $B$ . If the absolute difference between old and new parameters is greater than some criterion (e.g. 0.0001), proceed to Step 20; otherwise convergence is achieved and proceed to Step 25.
20. Assign new parameter estimates to old ones, i.e.  $A = a$  and  $B = b$ .
21. Let  $v_2 = (n_2 - f_v)(n_2 - f_v)$ .
22. Define  $s_2 = \frac{1}{n} (\sum(v_1) + \sum(v_2) + \sum(v_3))$ , i.e.  $s_2$  is the sum of the sum of all values in  $v_1$ ,  $v_2$  and  $v_3$ .
23. If  $s_2 > 0$ , let  $s_1 = \sqrt{s_2}$ .
24. Return to Step 6.
25. Compute  $v_2$  with recent fitted values, i.e.  $v_2 = (n_2 - f_v)(n_2 - f_v)$  and update  $s_2$  such that  $s_2 = \frac{1}{n} (\sum(v_1) + \sum(v_2) + \sum(v_3))$ . Compute  $s_1 = \sqrt{s_2}$  if  $s_2 > 0$ .
26. Convert regression coefficient estimates to mean starting data setup, i.e.  $\mu = -a/b$ ,  $\tau = 1/b$  and  $\sigma = s_1/b$ .
27. Compute the log-likelihood function.

(b) **Model 2: different starting dates and constant duration of primary moult (different  $y$ -intercepts and common slope)**

This algorithm to implement model 2 remains the same as the previous algorithm except for certain adaptations to some of the steps to include the fact that we are now fitting a regression line with different  $y$ -intercepts for each year. Only the steps that have changed are given here.

1. Step 1: Define vectors  $t$  and  $y$  of length  $n$ , where  $n = n_1 + \dots + n_H$ , for the independent and the dependent variable respectively and factor  $F$  representing the  $H$  different data sets.
2. Step 3: Fit the linear regression  $y = a_h + bt$ ,  $h = 1, 2, \dots, H$  (by ordinary least squares) for values for which  $0 \leq y \leq 1$ .
3. Step 4: Assign  $A_1 = a_1, \dots, A_H = a_H, B = b$  and  $s_1 = \sqrt{rms}$ , where  $a_h$  and  $b$  are the initial parameter estimates obtained in Step 3, and  $rms$  is the residual mean square.
4. Step 18: Fit the linear regression  $n_2 = a_h + bt$ ,  $h = 1, 2, \dots, H$  (by ordinary least squares) on all  $n$  units and assign new fitted values to  $f_v$ .
5. Step 20: Assign new parameter estimates to old ones, i.e.  $A_h = a_h$ ,  $h = 1, 2, \dots, H$  and  $B = b$ .
6. Step 26: Convert regression coefficient estimates to mean starting data setup, i.e.  $\mu_h = -a_h/b$ ,  $h = 1, 2, \dots, H$ ,  $\tau = 1/b$  and  $\sigma = s_1/b$ .

(c) **Model 3: different starting dates and durations of primary moult (different  $y$ -intercepts and slopes)**

Again only the steps which differ from the algorithm for model 1 are given here.

1. Step 1: Define vectors  $t$  and  $y$  of length  $n$ , where  $n = n_1 + \dots + n_H$ , for the independent and the dependent variable respectively and factor  $F$  representing the  $H$  different data sets.
2. Step 3: Fit the linear regression  $y = a_h + b_h t$ ,  $h = 1, 2, \dots, H$  (by ordinary least squares) for values for which  $0 \leq y \leq 1$ .

3. Step 4: Assign  $A_1 = a_1, \dots, A_H = a_H, B_1 = b_1, \dots, B_H = b_H$  and  $s_1 = \sqrt{rms}$ , where  $a_h$  and  $b_h$  are the initial parameter estimates obtained in Step 3, and  $rms$  is the residual mean square.
4. Step 18: Fit the linear regression  $n_2 = a_h + b_h t, \quad h = 1, 2, \dots, H$  (by ordinary least squares) on all  $n$  units and assign new fitted values to  $f_v$ .
5. Step 20: Assign new parameter estimates to old ones, i.e.  $A_h = a_h, B_h = b_h, \quad h = 1, 2, \dots, H$ .
6. Step 26: Convert regression coefficient estimates to mean starting data setup, i.e.  $\mu_h = -a_h/b_h, \tau_h = 1/b_h, \quad h = 1, 2, \dots, H$  (cannot get  $\sigma$  from  $\sigma_\epsilon$  because of multiple slopes).

### Data type 5

#### (a) Model 1: constant starting date and duration of primary moult (common $y$ -intercept and slope)

1. Define vectors  $t$  and  $y$  of length  $n$  for the independent and the dependent variable respectively.
2. Define the vectors  $w_1$  such that

$$w_1 = \begin{cases} 0 & \text{if } y = 0 \\ 1 & \text{otherwise.} \end{cases}$$

3. Fit the linear regression  $y = a + bt$  (by ordinary least squares) for values for which  $0 \leq y \leq 1$ .
4. Assign  $A = a, B = b$  and  $s_1 = \sqrt{rms}$ , where  $a$  and  $b$  are the initial parameter estimates obtained in Step 3, and  $rms$  is the residual mean square.
5. Assign  $f_v =$  fitted values, for all  $n$  units.
6. Define  $z = \frac{f_v}{s_1}$ , the standerized fitted values.
7. Define the vector  $f_z = \Phi(z)$ , i.e.  $f_z$  contains the values of the standard normal cumulative distribution function for each value of  $z$ .
8. Compute  $p_z = 0.3986 \exp(-z^2/2)$ , i.e. the normal density values of  $z$ .
9. Compute  $h_z = \frac{s_1 p_z}{(1 - f_z)}$ ;  $h_z = 0$  if division by zero.

10. Assign  $n_1 = w_1 y + (1 - w_1)(f_v - h_z)$ .
11. Let  $v_1 = (1 - w_1)(s_1^2 + f_v h_z - h_z^2)$ .
12. Fit the linear regression  $n_1 = a + bt$  (by ordinary least squares) on all  $n$  units and assign new fitted values to  $f_v$ .
13. Compare the new regression parameters  $a$  and  $b$  with the old parameters  $A$  and  $B$ . If the absolute difference between old and new parameters is greater than some criterion (e.g. 0.0001), proceed to Step 14; otherwise convergence is achieved and proceed to Step 19.
14. Assign new parameter estimates to old ones, i.e.  $A = a$  and  $B = b$ .
15. Let  $v_2 = (n_1 - f_v)(n_1 - f_v)$ .
16. Define  $s_2 = \frac{1}{n} (\sum(v_1) + \sum(v_2))$ , i.e.  $s_2$  is the sum of the sum of all values in  $v_1$  and  $v_2$ .
17. If  $s_2 > 0$ , let  $s_1 = \sqrt{s_2}$ .
18. Return to Step 6.
19. Compute  $v_2$  with recent fitted values, i.e.  $v_2 = (n_1 - f_v)(n_1 - f_v)$  and update  $s_2$  such that  $s_2 = \frac{1}{n} (\sum(v_1) + \sum(v_2))$ . Compute  $s_1 = \sqrt{s_2}$  if  $s_2 > 0$ .
20. Convert regression coefficient estimates to mean starting data setup, i.e.  $\mu = -a/b$ ,  $\tau = 1/b$  and  $\sigma = s_1/b$ .
21. Compute the log-likelihood function.

**(b) Model 2: different starting dates and constant duration of primary moult (different  $y$ -intercepts and common slope)**

The algorithm to implement model 2 for data type 5 remains the same as that one for model 1 for data type 5, except for certain adaptations to some of the steps to include the fact that we are now fitting a regression line with different  $y$ -intercepts for each year. Only the steps that have changed are given here.

1. Step 1: Define vectors  $t$  and  $y$  of length  $n$ , where  $n = n_1 + \dots + n_H$ , for the independent and the dependent variable respectively and factor  $F$  representing the  $h$  different data sets.

2. Step 3: Fit the linear regression  $y = a_h + bt$ ,  $h = 1, 2, \dots, H$  (by ordinary least squares) for values for which  $0 \leq y \leq 1$ .
3. Step 4: Assign  $A_1 = a_1, \dots, A_H = a_H$ ,  $B = b$  and  $s_1 = \sqrt{rms}$ , where  $a_h$  and  $b$  are the initial parameter estimates obtained in Step 3, and  $rms$  is the residual mean square.
4. Step 12: Fit the linear regression  $n_1 = a_h + bt$ ,  $h = 1, 2, \dots, H$  (by ordinary least squares) on all  $n$  units and assign new fitted values to  $f_v$ .
5. Step 14: Assign new parameter estimates to old ones, i.e.  $A_h = a_h$ ,  $h = 1, 2, \dots, H$  and  $B = b$ .
6. Step 20: Convert regression coefficient estimates to mean starting data setup, i.e.  $\mu_h = -a_h/b$ ,  $h = 1, 2, \dots, H$ ,  $\tau = 1/b$  and  $\sigma = s_1/b$ .

(c) **Model 3: different starting dates and durations of primary moult (different  $y$ -intercepts and slopes)**

Again only the steps which differ from the algorithm for model 1 are given here.

1. Step 1: Define vectors  $t$  and  $y$  of length  $n$ , where  $n = n_1 + \dots + n_H$ , for the independent and the dependent variable respectively and factor  $F$  representing the  $H$  different data sets.
2. Step 3: Fit the linear regression  $y = a_h + b_h t$ ,  $h = 1, 2, \dots, H$  (by ordinary least squares) for values for which  $0 \leq y \leq 1$ .
3. Step 4: Assign  $A_1 = a_1, \dots, A_H = a_H$ ,  $B_1 = b_1, \dots, B_H = b_H$  and  $s_1 = \sqrt{rms}$ , where  $a_h$  and  $b_h$  are the initial parameter estimates obtained in Step 3, and  $rms$  is the residual mean square.
4. Step 12: Fit the linear regression  $n_1 = a_h + b_h t$ ,  $h = 1, 2, \dots, H$  (by ordinary least squares) on all  $n$  units and assign new fitted values to  $f_v$ .
5. Step 14: Assign new parameter estimates to old ones, i.e.  $A_h = a_h$ ,  $B_h = b_h$ ,  $h = 1, 2, \dots, H$ .
6. Step 20: Convert regression coefficient estimates to mean starting data setup, i.e.  $\mu_h = -a_h/b_h$ ,  $\tau_h = 1/b_h$ ,  $h = 1, 2, \dots, H$  ( $\sigma$  cannot be estimated from the estimate of  $\sigma_\epsilon$  because of multiple slopes).

## APPENDIX 4

**GENSTAT programs for censored regression models**

The algorithm of Schmid *et al.* (1994) has been adopted here, with the appropriate changes to our own requirements. The statistical package GENSTAT (Payne *et al.* 1987) was used to program the following algorithms. Unless otherwise specified, the standard deviation of the mean starting date of moult is constant.

**Data type 2****(a) Model 1: constant starting date and duration of primary moult (common  $y$ -intercept and slope)**

```

"Replace n by the sample size"
Units[nvalues=n]
Open name='datafile.dat';channel=2
Variate day,moult
Variate [values=n(0)] y
Read[Print=*;channel=2;end=*] day,moult
Close channel=2
Calculate w1 = moult>0
& w2 = moult>=1
Restrict day,moult; condition=(moult.gt.0).and.(moult.lt.1)
Model moult
Fit [Print=model,summary,estimates;Fprobability=yes;Tprobability=yes] day
Rkeep estimates=est;deviance=dev
Restrict day,moult
Scalar alpha,beta
Calculate alpha = est$[1]
& beta = est$[2]
& fv = alpha+beta*day
& s1 = Sqrt(dev/(n-2))
"Replace nmax by the maximum number of iterations"
For [ntimes=nmax]

```

```

Calculate z = fv/s1
& fz = normal(z)
& pz = 0.3989*exp(-(z**2)/2)
& hz = (s1*pz)/(1-fz)
& hz = Mvreplace (hz;y)
& x = (1-fv)/s1
& fx = normal(x)
& px = 0.3989*exp(-(x**2)/2)
& hx = (s1*px)/(1-fx)
& hx = Mvreplace (hx;y)
& n1 = w1*moult+(1-w1)*(fv-hz)
& n2 = (1-w2)*n1+(w2)*(fv+hx)
& v1 = (1-w1)*(s1**2+fv*hz-hz**2)
& v3 = (w2)*(s1**2+hx-fv*hx-hx**2)
Model n2
Fit [Print=model,summary,estimates;Fprobability=yes;Tprobability=yes] day
Rkeep fittedvalues=fv;estimates=est;deviance=dev
Scalar a,b
Calculate a = est$[1]
& b = est$[2]
Exit [exp='Convergence'] abs((alpha-a)<0.0001).and.abs((beta-b)<0.0001)
If (abs(alpha-a)>0.0001).or.(abs(beta-b)>0.0001)
  Calculate alpha = a
  & beta = b
Endif
Calculate v2 = (n2-fv)**2
& s2 = (sum(v1)+sum(v2)+sum(v3))/n
If s2>0
  Calculate s1 = sqrt(s2)
Endif
Endfor
Calculate v2 = (n2-fv)**2
& s2 = (sum(v1)+sum(v2)+sum(v3))/n

```

```
If s2>0
  Calculate s1 = sqrt(s2)
Endif
Calculate mean = -a/b
& tau = 1/b
& sig = s1/b
Print mean,tau,s1,sig
"Compute the full log likelihood"
Calculate z = fv/s1
& fz = normal(z)
& lfz = log(1-fz)
& lfz = (1-w1)*lfz
& x1 = (moult/s1-z)
& px1 = 0.3989*exp(-(x1**2)/2)
& lpx1 = log(px1)
& lpx1 = w1*lpx1
& lpx1 = (1-w2)*lpx1
& x2 = (1-fv)/s1
& fx2 = normal(x2)
& lfx2 = log(1-fx2)
& lfx2 = w2*lfx2
& va = log(1/s1)
& va = w1*va
& va = (1-w2)*va
& lik = sum(lfz)+sum(lpx1)+sum(lfx2)+sum(va)
Print lik
Stop
```

(b) **Model 2: different starting dates and constant duration of primary moult (different  $y$ -intercepts and common slope)**

“Replace  $n$  by sample size”

“In this program it is assumed that there are 4 years of data starting in year 79”

Units[nvalues= $n$ ]

Open name='datafile.dat';channel=2

Variate day,moult

Factor year

Variate [values= $n(0)$ ]  $y$

Read[Print=\*;channel=2;end=\*] day,moult,year;Frepresentation=2(\*),levels

Close channel=2

Calculate  $w1 = \text{moult} > 0$

&  $w2 = \text{moult} \geq 1$

Restrict day,moult; condition=(moult.gt.0).and.(moult.lt.1)

Model moult

Terms [full=yes] year\*day

Fit [Print=model,summary,estimates;constant=omit;Fprobability=yes; \

  Tprobability=yes] year+day

Rkeep estimates=est;deviance=dev

Restrict day,moult

Scalar al1,al2,al3,al4,beta

Calculate al1 = est\$[1]

& al2 = est\$[2]

& al3 = est\$[3]

& al4 = est\$[4]

& beta = est\$[5]

Restrict day,fv;condition=year.eq.79

Calculate fv = al1+beta\*day

Restrict day,fv

Restrict day,fv;condition=year.eq.80

Calculate fv = al2+beta\*day

Restrict day,fv

```

Restrict day,fv;condition=year.eq.81
Calculate fv = a13+beta*day
Restrict day,fv
Restrict day,fv;condition=year.eq.82
Calculate fv = a14+beta*day
Restrict day,fv
Calculate s1 = Sqrt(dev/(n-2))
"Replace nmax by the maximum number of iterations"
For [ntimes=nmax]
  Calculate z = fv/s1
  & fz = normal(z)
  & pz = 0.3989*exp(-(z**2)/2)
  & hz = (s1*pz)/(1-fz)
  & hz = Mvreplace (hz;y)
  & x = (1-fv)/s1
  & fx = normal(x)
  & px = 0.3989*exp(-(x**2)/2)
  & hx = (s1*px)/(1-fx)
  & hx = Mvreplace (hx;y)
  & n1 = w1*moult+(1-w1)*(fv-hz)
  & n2 = (1-w2)*n1+(w2)*(fv+hx)
  & v1 = (1-w1)*(s1**2+fv*hz-hz**2)
  & v3 = (w2)*(s1**2+hx-fv*hx-hx**2)
Model n2
Terms [full=yes] year*day
Fit [Print=model,summary,estimates;constant=omit;Fprobability=yes; \
  Tprobability=yes] year+day
Rkeep fittedvalues=fv;estimates=est;deviance=dev
Scalar a1,a2,a3,a4,b
Calculate a1 = est$[1]
& a2 = est$[2]
& a3 = est$[3]
& a4 = est$[4]

```

```

& b = est$[5]
Exit [exp='Convergence'] abs((a1-a1)<0.0001).and.abs((a2-a2)<0.0001) \
    .and.abs((beta-b)<0.0001).and.abs((a3-a3)<0.0001).and.abs((a4-a4)<0.0001)
If (abs(a1-a1)>0.0001).or.(abs(a2-a2)>0.0001).or.(abs(a3-a3)>0.0001) \
    .or.(abs(a4-a4)>0.0001).or.(abs(beta-b)>0.0001)
    Calculate a1 = a1
    & a2 = a2
    & a3 = a3
    & a4 = a4
    & beta = b
  Endif
  Calculate v2 = (n2-fv)**2
  & s2 = (sum(v1)+sum(v2)+sum(v3))/n
  If s2>0
    Calculate s1 = sqrt(s2)
  Endif
Endfor
Calculate v2 = (n2-fv)**2
& s2 = (sum(v1)+sum(v2)+sum(v3))/n
If s2>0
  Calculate s1 = sqrt(s2)
Endif
Calculate mean1 = -a1/b
& mean2 = -a2/b
& mean3 = -a3/b
& mean4 = -a4/b
& tau = 1/b
& sig = s1/b
Print mean1,mean2,mean3,mean4,tau,s1,sig
"Compute the full log likelihood"
Calculate z = fv/s1
& fz = normal(z)
& lfz = log(1-fz)

```

```

& lfz = (1-w1)*lfz
& x1 = (moult/s1-z)
& px1 = 0.3989*exp(-(x1**2)/2)
& lpx1 = log(px1)
& lpx1 = w1*lpx1
& lpx1 = (1-w2)*lpx1
& x2 = (1-fv)/s1
& fx2 = normal(x2)
& lfx2 = log(1-fx2)
& lfx2 = w2*lfx2
& va = log(1/s1)
& va = w1*va
& va = (1-w2)*va
& lik = sum(lfz)+sum(lpx1)+sum(lfx2)+sum(va)
Print lik
Stop

```

(c) **Model 3: different starting dates and durations of primary moult (different  $y$ -intercepts and slopes)**

```

“Replace n by sample size”
“In this program it is assumed that there is 4 years of data starting in year 79”
Units[nvalues=n]
Open name='datafile.dat';channel=2
Variate day,moult
Factor year
Variate [values=n(0)] y
Read[Print=*;channel=2;end=*] day,moult,year;Frepresentation=2(*),levels
Close channel=2
Calculate w1 = moult>0
& w2 = moult>=1
Restrict day,moult; condition=(moult.gt.0).and.(moult.lt.1)

```

```

Model moult
Terms [full=yes] year*day
Fit [Print=model,summary,estimates;constant=omit;Fprobability=yes; \
    Tprobability=yes] year+year.day
Rkeep estimates=est;deviance=dev
Restrict day,moult
Scalar al1,al2,al3,al4,be1,be2,b3,b4
Calculate al1 = est$[1]
& al2 = est$[2]
& al3 = est$[3]
& al4 = est$[4]
& be1 = est$[5]
& be2 = est$[6]
& be3 = est$[7]
& be4 = est$[8]
Restrict day,fv;condition=year.eq.79
Calculate fv = al1+be1*day
Restrict day,fv
Restrict day,fv;condition=year.eq.80
Calculate fv = al2+be2*day
Restrict day,fv
Restrict day,fv;condition=year.eq.81
Calculate fv = al3+be3*day
Restrict day,fv
Restrict day,fv;condition=year.eq.82
Calculate fv = al4+be4*day
Restrict day,fv
Calculate s1 = Sqrt(dev/(n-2))
"Replace nmax by the maximum number of iterations"
For [ntimes=nmax]
    Calculate z = fv/s1
    & fz = normal(z)
    & pz = 0.3989*exp(-(z**2)/2)

```

```

& hz = (s1*pz)/(1-fz)
& hz = Mvreplace (hz;y)
& x = (1-fv)/s1
& fx = normal(x)
& px = 0.3989*exp(-(x**2)/2)
& hx = (s1*px)/(1-fx)
& hx = Mvreplace (hx;y)
& n1 = w1*moult+(1-w1)*(fv-hz)
& n2 = (1-w2)*n1+(w2)*(fv+hx)
& v1 = (1-w1)*(s1**2+fv*hz-hz**2)
& v3 = (w2)*(s1**2+hx-fv*hx-hx**2)
Model n2
Terms [full=yes] year*day
Fit [Print=model,summary,estimates;constant=omit;Fprobability=yes; \
  Tprobability=yes] year+year.day
Rkeep fittedvalues=fv;estimates=est;deviance=dev
Scalar a1,a2,a3,a4,b1,b2,b3,b4
Calculate a1 = est$[1]
& a2 = est$[2]
& a3 = est$[3]
& a4 = est$[4]
& b1 = est$[5]
& b2 = est$[6]
& b3 = est$[7]
& b4 = est$[8]
Scalar d1,d2,d3,d4,d5,d6,d7,d8
Calculate d1 = abs(a1-a1)
& d2 = abs(a2-a2)
& d3 = abs(a3-a3)
& d4 = abs(a4-a4)
& d5 = abs(b1-b1)
& d6 = abs(b2-b2)
& d7 = abs(b3-b3)

```

```

& d8 = abs(be4-b4)
Exit [exp='Convergence'] (d1<0.0001).and.(d2<0.0001).and.(d3<0.0001).and. \
    (d4<0.0001).and.(d5<0.0001).and.(d6<0.0001).and.(d7<0.0001).and.(d8<0.0001)
If (d1>0.0001).or.(d2>0.0001).or.(d3>0.0001).or.(d4>0.0001).or. \
    (d5>0.0001).or.(d6>0.0001).or.(d7>0.0001).or.(d8>0.0001).or.
    Calculate a1 = a1
    & a2 = a2
    & a3 = a3
    & a4 = a4
    & be1 = b1
    & be2 = b2
    & be3 = b3
    & be4 = b4
Endif
Calculate v2 = (n2-fv)**2
& s2 = (sum(v1)+sum(v2)+sum(v3))/n
If s2>0
    Calculate s1 = sqrt(s2)
Endif
Endfor
Calculate v2 = (n2-fv)**2
& s2 = (sum(v1)+sum(v2)+sum(v3))/n
If s2>0
    Calculate s1 = sqrt(s2)
Endif
Calculate mean1 = -a1/b1
& mean2 = -a2/b2
& mean3 = -a3/b3
& mean4 = -a4/b4
& tau1 = 1/b1
& tau2 = 1/b2
& tau3 = 1/b3
& tau4 = 1/b4

```

```

Print mean1,mean2,mean3,mean4,tau1,tau2,tau3,tau4,s
"Compute the full log likelihood"
Calculate z = fv/s1
& fz = normal(z)
& lfz = log(1-fz)
& lfz = (1-w1)*lfz
& x1 = (moult/s1-z)
& px1 = 0.3989*exp(-(x1**2)/2)
& lpx1 = log(px1)
& lpx1 = w1*lpx1
& lpx1 = (1-w2)*lpx1
& x2 = (1-fv)/s1
& fx2 = normal(x2)
& lfx2 = log(1-fx2)
& lfx2 = w2*lfx2
& va = log(1/s1)
& va = w1*va
& va = (1-w2)*va
& lik = sum(lfz)+sum(lpx1)+sum(lfx2)+sum(va)
Print lik
Stop

```

### Data type 5

- (a) **Model 1: constant starting date and duration of primary moult (common  $y$ -intercept and slope)**

```

"Replace n by the sample size"
Units[nvalues=n]
Open name='datafile.dat';channel=2
Variate day,moult
Variate [values=n(0)] y

```

```

Read[Print=*;channel=2;end=*] day,moult
Close channel=2
Calculate w1 = moult>0
Restrict day,moult; condition=moult.ne.0)
Model moult
Fit [Print=model,summmary,estimates;Fprobability=yes;Tprobability=yes] day
Rkeep estimates=est;deviance=dev
Restrict day,moult
Scalar alpha,beta
Calculate alpha = est$[1]
& beta = est$[2]
& fv = alpha+beta*day
& s1 = Sqrt(dev/(n-2))
“Replace nmax by the maximum number of iterations”
For [ntimes=nmax]
  Calculate z = fv/s1
  & fz = normal(z)
  & pz = 0.3989*exp(-(z**2)/2)
  & hz = (s1*pz)/(1-fz)
  & hz = Mvreplace (hz;y)
  & n1 = w1*moult+(1-w1)*(fv-hz)
  & v1 = (1-w1)*(s1**2+fv*hz-hz**2)
  Model n1
  Fit [Print=model,summmary,estimates;Fprobability=yes;Tprobability=yes] day
  Rkeep fittedvalues=fv;estimates=est;deviance=dev
  Scalar a,b
  Calculate a = est$[1]
  & b = est$[2]
  Exit [exp='Convergence'] abs((alpha-a)<0.0001).and.abs((beta-b)<0.0001)
  If (abs(alpha-a)>0.0001).or.(abs(beta-b)>0.0001)
    Calculate alpha = a
    & beta = b
  Endif

```

```

    Calculate v2 = (n1-fv)**2
    & s2 = (sum(v1)+sum(v2))/n
    If s2>0
        Calculate s1 = sqrt(s2)
    Endif
Endfor
Calculate v2 = (n1-fv)**2
& s2 = (sum(v1)+sum(v2))/n
If s2>0
    Calculate s1 = sqrt(s2)
Endif
Calculate mean = -a/b
& tau = 1/b
& sig = s1/b
Print mean,tau,s1,sig
"Compute the full log likelihood"
Calculate z = fv/s1
& fz = normal(z)
& lfz = log(1-fz)
& lfz = (1-w1)*lfz
& x1 = (moult/s1-z)
& px1 = 0.3989*exp(-(x1**2)/2)
& lpx1 = log(px1)
& lpx1 = w1*lpx1
& x2 = (1-fv)/s1
& fx2 = normal(x2)
& lfx2 = log(fx2)
& va = log(1/s1)
& va = w1*va
& lik = sum(lfz)+sum(lpx1)-sum(lfx2)+sum(va)
Print lik
Stop

```

(b) **Model 2: different starting dates and constant duration of primary moult (different  $y$ -intercepts and common slope)**

“Replace  $n$  by sample size”

“In this program it is assumed that there are 4 years of data starting in year 79”

Units[nvalues= $n$ ]

Open name='datafile.dat';channel=2

Variate day,moult

Factor year

Variate [values= $n(0)$ ]  $y$

Read[Print=\*;channel=2;end=\*] day,moult,year;Frepresentation=2(\*),levels

Close channel=2

Calculate  $w1 = \text{moult} > 0$

Restrict day,moult; condition=(moult.ne.0)

Model moult

Terms [full=yes] year\*day

Fit [Print=model,summmary,estimates;constant=omit;Fprobability=yes; \

  Tprobability=yes] year+day

Rkeep estimates=est;deviance=dev

Restrict day,moult

Scalar al1,al2,al3,al4,beta

Calculate  $al1 = \text{est}[1]$

&  $al2 = \text{est}[2]$

&  $al3 = \text{est}[3]$

&  $al4 = \text{est}[4]$

&  $beta = \text{est}[5]$

Restrict day,fv;condition=year.eq.79

Calculate  $fv = al1 + beta * day$

Restrict day,fv

Restrict day,fv;condition=year.eq.80

Calculate  $fv = al2 + beta * day$

Restrict day,fv

Restrict day,fv;condition=year.eq.81

```

Calculate fv = al3+beta*day
Restrict day,fv
Restrict day,fv;condition=year.eq.82
Calculate fv = al4+beta*day
Restrict day,fv
Calculate s1 = Sqrt(dev/(n-2))
"Replace nmax by the maximum number of iterations"
For [ntimes=nmax]
  Calculate z = fv/s1
  & fz = normal(z)
  & pz = 0.3989*exp(-(z**2)/2)
  & hz = (s1*pz)/(1-fz)
  & hz = Mvreplace (hz;y)
  & n1 = w1*moult+(1-w1)*(fv-hz)
  & v1 = (1-w1)*(s1**2+fv*hz-hz**2)
  Model n1
  Terms [full=yes] year*day
  Fit [Print=model,summmary,estimates;constant=omit;Fprobability=yes; \
    Tprobability=yes] year+day
  Rkeep fittedvalues=fv;estimates=est;deviance=dev
  Scalar a1,a2,a3,a4,b
  Calculate a1 = est$[1]
  & a2 = est$[2]
  & a3 = est$[3]
  & a4 = est$[4]
  & b = est$[5]
  Exit [exp='Convergence'] abs((al1-a1)<0.0001).and.abs((al2-a2)<0.0001)\
    .and.abs((beta-b)<0.0001).and.abs((al3-a3)<0.0001).and.abs((al4-a4)<0.0001)
  If (abs(al1-a1)>0.0001).or.(abs(al2-a2)>0.0001).or.(abs(al3-a3)>0.0001)\
    .or.(abs(al4-a4)>0.0001).or.(abs(beta-b)>0.0001)
  Calculate al1 = a1
  & al2 = a2
  & al3 = a3

```

```

    & a14 = a4
    & beta = b
  Endif
  Calculate v2 = (n1-fv)**2
  & s2 = (sum(v1)+sum(v2))/n
  If s2>0
    Calculate s1 = sqrt(s2)
  Endif
Endfor
Calculate v2 = (n1-fv)**2
& s2 = (sum(v1)+sum(v2))/n
If s2>0
  Calculate s1 = sqrt(s2)
Endif
Calculate mean1 = -a1/b
& mean2 = -a2/b
& mean3 = -a3/b
& mean4 = -a4/b
& tau = 1/b
& sig = s1/b
Print mean1,mean2,mean3,mean4,tau,s1,sig
"Compute the full log likelihood"
Calculate z = fv/s1
& fz = normal(z)
& lfz = log(1-fz)
& lfz = (1-w1)*lfz
& x1 = (moult/s1-z)
& px1 = 0.3989*exp(-(x1**2)/2)
& lpx1 = log(px1)
& lpx1 = w1*lpx1
& x2 = (1-fv)/s1
& fx2 = normal(x2)
& lfx2 = log(fx2)

```

```

& va = log(1/s1)
& va = w1*va
& lik = sum(lfz)+sum(lpx1)-sum(lfx2)+sum(va)
Print lik
Stop

```

(c) **Model 3: different starting dates and durations of primary moult (different y-intercepts and slopes)**

```

"Replace n by sample size"
"In this program it is assumed that there is 4 years of data starting in year 79"
Units[nvalues=n]
Open name='datafile.dat';channel=2
Variate day,moult
Factor year
Variate [values=n(0)] y
Read[Print=*;channel=2;end=*] day,moult,year;Frepresentation=2(*),levels
Close channel=2
Calculate w1 = moult>0
Restrict day,moult; condition=(moult.ne.0)
Model moult
Terms [full=yes] year*day
Fit [Print=model,summary,estimates;constant=omit;Fprobability=yes; \
  Tprobability=yes] year+year.day
Rkeep estimates=est;deviance=dev
Restrict day,moult
Scalar al1,al2,al3,al4,be1,be2,b3,b4
Calculate al1 = est$[1]
& al2 = est$[2]
& al3 = est$[3]
& al4 = est$[4]
& be1 = est$[5]

```

```

& be2 = est$[6]
& be3 = est$[7]
& be4 = est$[8]
Restrict day,fv;condition=year.eq.79
Calculate fv = a1+be1*day
Restrict day,fv
Restrict day,fv;condition=year.eq.80
Calculate fv = a2+be2*day
Restrict day,fv
Restrict day,fv;condition=year.eq.81
Calculate fv = a3+be3*day
Restrict day,fv
Restrict day,fv;condition=year.eq.82
Calculate fv = a4+be4*day
Restrict day,fv
Calculate s1 = Sqrt(dev/(n-2))
"Replace nmax by the maximum number of iterations"
For [ntimes=nmax]
  Calculate z = fv/s1
  & fz = normal(z)
  & pz = 0.3989*exp(-(z**2)/2)
  & hz = (s1*pz)/(1-fz)
  & hz = Mvreplace (hz;y)
  & n1 = w1*moult+(1-w1)*(fv-hz)
  & v1 = (1-w1)*(s1**2+fv*hz-hz**2)
Model n1
Terms [full=yes] year*day
Fit [Print=model,summmary,estimates;constant=omit;Fprobability=yes; \
  Tprobability=yes] year+year.day
Rkeep fittedvalues=fv;estimates=est;deviance=dev
Scalar a1,a2,a3,a4,b1,b2,b3,b4
Calculate a1 = est$[1]
& a2 = est$[2]

```

```

& a3 = est$[3]
& a4 = est$[4]
& b1 = est$[5]
& b2 = est$[6]
& b3 = est$[7]
& b4 = est$[8]
Scalar d1,d2,d3,d4,d5,d6,d7,d8
Calculate d1 = abs(a1-a1)
& d2 = abs(a2-a2)
& d3 = abs(a3-a3)
& d4 = abs(a4-a4)
& d5 = abs(b1-b1)
& d6 = abs(b2-b2)
& d7 = abs(b3-b3)
& d8 = abs(b4-b4)
Exit [exp='Convergence'] (d1<0.0001).and.(d2<0.0001).and.(d3<0.0001).and. \
  (d4<0.0001).and.(d5<0.0001).and.(d6<0.0001).and.(d7<0.0001).and.(d8<0.0001)
If (d1>0.0001).or.(d2>0.0001).or.(d3>0.0001).or.(d4>0.0001).or. \
  (d5>0.0001).or.(d6>0.0001).or.(d7>0.0001).or.(d8>0.0001).or,
  Calculate a1 = a1
  & a2 = a2
  & a3 = a3
  & a4 = a4
  & b1 = b1
  & b2 = b2
  & b3 = b3
  & b4 = b4
Endif
Calculate v2 = (n1-fv)**2
& s2 = (sum(v1)+sum(v2))/n
If s2>0
  Calculate s1 = sqrt(s2)
Endif

```

```

Endfor
Calculate v2 = (n1-fv)**2
& s2 = (sum(v1)+sum(v2))/n
If s2>0
    Calculate s1 = sqrt(s2)
Endif
Calculate mean1 = -a1/b1
& mean2 = -a2/b2
& mean3 = -a3/b3
& mean4 = -a4/b4
& tau1 = 1/b1
& tau2 = 1/b2
& tau3 = 1/b3
& tau4 = 1/b4
Print mean1,mean2,mean3,mean4,tau1,tau2,tau3,tau4,s
"Compute the full log likelihood"
Calculate z = fv/s1
& fz = normal(z)
& lfz = log(1-fz)
& lfz = (1-w1)*lfz
& x1 = (moult/s1-z)
& px1 = 0.3989*exp(-(x1**2)/2)
& lpx1 = log(px1)
& lpx1 = w1*lpx1
& x2 = (1-fv)/s1
& fx2 = normal(x2)
& lfx2 = log(fx2)
& va = log(1/s1)
& va = w1*va
& lik = sum(lfz)+sum(lpx1)-sum(lfx2)+sum(va)
Print lik
Stop

```

## APPENDIX 5

## Parameter estimates for model 4, obtained using the Newton-Raphson algorithm

**Table A1.** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, under the assumption that the mean starting dates  $\mu_h$ , durations  $\tau_h$  and standard deviations  $\sigma_h$  vary between years,  $h = 1, \dots, 11$ , using the Newton-Raphson algorithm to obtain maximum likelihood estimates for the moult parameters. Estimates (standard errors) of mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) of primary moult are also given.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\mu}_1^\dagger$	39.1 (2.2) 9 July	35.2 (2.7) 5 July	44.8 (3.7) 15 July	39.1 (2.2) 9 July	35.2 (2.7) 5 July	44.8 (3.7) 15 July
1980	$\hat{\mu}_2$	45.6 (2.5) 16 July	39.8 (2.7) 10 July	54.2 (4.6) 24 July	45.5 (2.5) 16 July	39.5 (2.6) 10 July	54.1 (4.6) 24 July
1981	$\hat{\mu}_3$	43.5 (2.3) 14 July	35.8 (1.5) 6 July	46.5 (3.1) 17 July	43.9 (2.4) 14 July	35.8 (1.5) 6 July	46.6 (3.2) 17 July
1982	$\hat{\mu}_4$	42.6 (1.8) 13 July	34.8 (1.3) 5 July	50.0 (3.0) 20 July	42.9 (1.9) 13 July	34.9 (1.3) 5 July	50.2 (3.2) 20 July
1983	$\hat{\mu}_5$	41.1 (2.1) 11 July	29.7 (2.1) 30 June	49.1 (2.7) 19 July	41.1 (2.2) 11 July	29.6 (2.2) 30 June	48.7 (2.9) 19 July
1984	$\hat{\mu}_6$	43.1 (3.2) 13 July	30.2 (2.3) 30 June	51.1 (5.0) 21 July	43.2 (3.3) 13 July	30.2 (2.3) 30 June	51.1 (5.0) 21 July
1985	$\hat{\mu}_7$	51.0 (2.3) 21 July	41.3 (1.7) 11 July	57.3 (1.3) 27 July	51.0 (2.4) 21 July	41.3 (1.7) 11 July	57.1 (1.3) 27 July
1986	$\hat{\mu}_8$	39.6 (3.0) 10 July	31.1 (2.0) 1 July	45.1 (5.7) 15 July	40.2 (3.3) 10 July	31.3 (2.1) 1 July	45.7 (6.2) 16 July
1987	$\hat{\mu}_9$	43.7 (2.3) 14 July	34.9 (1.2) 5 July	47.6 (2.6) 18 July	43.8 (2.4) 14 July	34.9 (1.2) 5 July	47.4 (2.0) 17 July
1988	$\hat{\mu}_{10}$	38.8 (2.2) 9 July	31.2 (1.3) 1 July	46.5 (4.0) 17 July	39.1 (2.4) 9 July	30.3 (1.1) 30 June	46.8 (4.2) 17 July
1989	$\hat{\mu}_{11}$	37.3 (1.9) 7 July	26.7 (1.5) 27 June	48.6 (2.7) 19 July	37.3 (2.0) 7 July	26.7 (1.6) 27 June	48.6 (2.7) 17 July

**Table A1 (continued).** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, under the assumption that the mean starting dates  $\mu_h$ , durations  $\tau_h$  and standard deviations  $\sigma_h$  vary between years,  $h = 1, \dots, 11$ , using the Newton-Raphson algorithm to obtain maximum likelihood estimates for the moult parameters. Estimates (standard errors) of mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) of primary moult are also given.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\tau}_1$	45.8 (4.3)	44.4 (7.7)	51.9 (6.6)	45.8 (4.3)	44.4 (7.7)	51.9 (6.6)
1980	$\hat{\tau}_2$	48.2 (3.5)	51.1 (4.4)	42.9 (5.6)	52.0 (3.7)	57.3 (5.2)	44.4 (5.7)
1981	$\hat{\tau}_3$	41.8 (3.4)	34.6 (2.5)	52.0 (5.0)	47.4 (4.2)	34.6 (2.5)	55.8 (5.5)
1982	$\hat{\tau}_4$	37.6 (2.4)	41.7 (2.0)	33.9 (4.1)	41.2 (2.9)	42.0 (2.1)	38.9 (4.9)
1983	$\hat{\tau}_5$	44.7 (4.3)	55.1 (4.1)	34.7 (6.9)	48.0 (4.9)	56.2 (4.4)	43.0 (9.1)
1984	$\hat{\tau}_6$	36.8 (5.1)	44.0 (4.2)	27.8 (6.8)	37.8 (5.4)	44.1 (4.3)	27.8 (6.8)
1985	$\hat{\tau}_7$	30.0 (3.5)	37.8 (2.3)	21.7 (2.1)	33.2 (4.1)	37.8 (2.3)	24.2 (2.7)
1986	$\hat{\tau}_8$	45.7 (4.2)	50.0 (2.9)	44.3 (7.7)	49.4 (4.8)	50.3 (3.0)	48.2 (8.5)
1987	$\hat{\tau}_9$	40.2 (3.9)	39.7 (2.5)	42.3 (2.7)	40.9 (4.2)	39.7 (2.5)	42.7 (2.1)
1988	$\hat{\tau}_{10}$	41.6 (3.5)	53.0 (2.7)	34.2 (5.9)	46.6 (4.5)	63.8 (3.4)	37.2 (6.8)
1989	$\hat{\tau}_{11}$	37.2 (2.8)	41.7 (2.6)	29.6 (4.1)	42.2 (3.7)	43.5 (3.1)	30.9 (4.4)
1979	$\hat{\sigma}_1$	11.7 (1.5)	9.2 (2.0)	11.6 (2.2)	11.7 (1.5)	9.2 (2.0)	11.6 (2.2)
1980	$\hat{\sigma}_2$	11.0 (1.4)	8.7 (1.7)	11.6 (2.0)	11.0 (1.4)	8.7 (1.7)	11.4 (2.0)
1981	$\hat{\sigma}_3$	12.1 (1.4)	3.9 (0.9)	12.3 (1.9)	12.9 (1.6)	3.9 (0.9)	12.6 (2.1)
1982	$\hat{\sigma}_4$	9.8 (0.9)	4.6 (0.7)	10.8 (1.5)	10.6 (1.1)	4.8 (0.7)	11.5 (1.8)
1983	$\hat{\sigma}_5$	11.3 (1.4)	7.6 (1.4)	9.1 (1.9)	11.6 (1.5)	7.7 (1.4)	9.6 (2.0)
1984	$\hat{\sigma}_6$	11.4 (2.1)	5.1 (1.5)	10.4 (2.9)	11.7 (2.2)	5.2 (1.6)	10.4 (3.0)
1985	$\hat{\sigma}_7$	6.6 (1.2)	2.3 (0.8)	2.2 (0.8)	6.9 (1.4)	2.3 (0.8)	2.2 (0.8)
1986	$\hat{\sigma}_8$	10.4 (1.6)	4.9 (1.1)	11.8 (2.9)	11.3 (1.9)	5.1 (1.2)	12.4 (3.3)
1987	$\hat{\sigma}_9$	9.5 (1.5)	3.7 (0.8)	0.71 (0.3)	9.7 (1.6)	3.7 (0.8)	0.69 (0.3)
1988	$\hat{\sigma}_{10}$	9.7 (1.3)	4.0 (0.9)	11.2 (2.2)	10.5 (1.5)	3.2 (0.7)	11.9 (2.5)
1989	$\hat{\sigma}_{11}$	10.4 (1.1)	5.6 (0.9)	8.2 (1.5)	11.5 (1.4)	5.9 (1.1)	8.5 (1.6)

**Table A1 (continued).** Estimates (standard errors) of the parameters of primary moult (days) of Willow Warblers at Lake Ladoga, St Petersburg, Russia, under the assumption that the mean starting dates  $\mu_h$ , durations  $\tau_h$  and standard deviations  $\sigma_h$  vary between years,  $h = 1, \dots, 11$ , using the Newton-Raphson algorithm to obtain maximum likelihood estimates for the moult parameters. Estimates (standard errors) of mean completion dates  $\delta_h$  ( $\delta_h = \mu_h + \tau_h$ ) of primary moult are also given.

Year	Parameter estimates	Data type 2			Data type 5		
		All	Male	Female	All	Male	Female
1979	$\hat{\delta}_1^\dagger$	84.9 (5.8)	79.6 (6.9)	96.7 (5.7)	84.9 (3.7)	79.6 (6.9)	96.7 (5.7)
		24 Aug	19 Aug	5 Sep	24 Aug	19 Aug	5 Sep
1980	$\hat{\delta}_2$	93.8 (2.6)	90.9 (3.5)	97.1 (3.6)	97.5 (2.9)	96.8 (4.4)	98.5 (3.7)
		2 Sep	30 Aug	5 Sep	6 Sep	5 Sep	7 Sep
1981	$\hat{\delta}_3$	85.3 (2.7)	70.4 (1.9)	98.5 (4.2)	91.3 (3.7)	70.4 (1.9)	102.4 (5.0)
		24 Aug	9 Aug	7 Sep	30 Aug	9 Aug	10 Sep
1982	$\hat{\delta}_4$	80.2 (1.7)	76.5 (1.5)	83.9 (2.9)	84.1 (2.3)	76.9 (1.7)	89.1 (3.9)
		19 Aug	16 Aug	23 Aug	23 Aug	16 Aug	28 Aug
1983	$\hat{\delta}_5$	85.8 (3.9)	84.8 (3.4)	83.8 (6.1)	89.1 (4.5)	85.8 (3.7)	91.7 (8.4)
		25 Aug	24 Aug	23 Aug	28 Aug	25 Aug	31 Aug
1984	$\hat{\delta}_6$	79.9 (4.4)	74.2 (3.6)	78.9 (5.8)	81.0 (4.9)	74.3 (3.8)	78.9 (5.8)
		19 Aug	13 Aug	18 Aug	20 Aug	13 Aug	18 Aug
1985	$\hat{\delta}_7$	81.0 (2.6)	79.1 (1.4)	79.0 (1.5)	84.2 (4.7)	79.1 (1.4)	81.3 (2.2)
		20 Aug	18 Aug	18 Aug	23 Aug	18 Aug	20 Aug
1986	$\hat{\delta}_8$	85.3 (2.9)	81.1 (2.1)	89.4 (5.6)	89.6 (4.0)	81.6 (3.1)	93.9 (10.5)
		24 Aug	20 Aug	28 Aug	29 Aug	21 Aug	2 Sep
1987	$\hat{\delta}_9$	83.9 (3.4)	74.6 (2.0)	89.9 (0.4)	84.7 (3.8)	74.6 (2.0)	90.1 (0.4)
		23 Aug	14 Aug	29 Aug	24 Aug	14 Aug	29 Aug
1988	$\hat{\delta}_{10}$	80.4 (2.8)	84.2 (2.2)	80.7 (4.7)	85.7 (4.0)	94.1 (3.1)	84.0 (5.8)
		20 Aug	23 Aug	20 Aug	25 Aug	2 Sep	23 Aug
1989	$\hat{\delta}_{11}$	74.5 (2.1)	68.4 (2.0)	78.2 (2.9)	79.5 (3.1)	70.2 (2.7)	79.5 (3.3)
		14 Aug	7 Aug	17 Aug	19 Aug	9 Aug	19 Aug

†Mean starting date and mean completion date are in days since 1 June (i.e. day 1 = 1 June)

