ESTIMATING THE PARAMETERS FOR PRIMARY MOULT – A NEW STATISTICAL MODEL

by L.G. Underhill

INTRODUCTION

It has long been recognized that the regression methods that have been used to estimate the parameters of primary moult are unsatisfactory (see for example Pimm (1976)). Summers *et al.* (1983) compared seven regression methods, most of which have recently been used in the analysis of moult, for estimating the starting date, completion date and duration of moult. Their estimates of starting date varied from 29 June to 31 July (32 days), of completion date from 2 to 24 October (22 days), and of duration from 72 to 109 (37) days for the Redshank *Tringa totanus*. There is obviously a need for a standard and statistically sound method.

ORDINARY LINEAR REGRESSION

The simplest technique for estimating the parameters of moult (starting and completion dates and duration) is to fit a linear regression of moult score on date, using all birds actively moulting. The reason for the poor performance of linear regression in this case is straightfoward. One of the underlying assumptions of least squares regression is that the variance (variability) of the dependent variable (moult score) is the same for all values of the independent variable (date). This assumption is grossly violated, since active primary moult scores lie between 1 and (usually) 49, so that near the commencement and conclusion of moult there is less variability in moult score than during the middle of the moult period (Figure 1). The technical term for lack of constant variance is heteroscedasticity. The regression line runs diagonally across the long axis of the parallelogram that encloses the scatter of points, effectively giving the starting and completion date of the first and last birds respectively in the population, rather than the average bird (Figure 1), (see also Summers *et al.* 1983). Most of the other methods considered by Summers *et al.* (1983) are *ad hoc* attempts that have been devised to overcome heteroscedasticity.

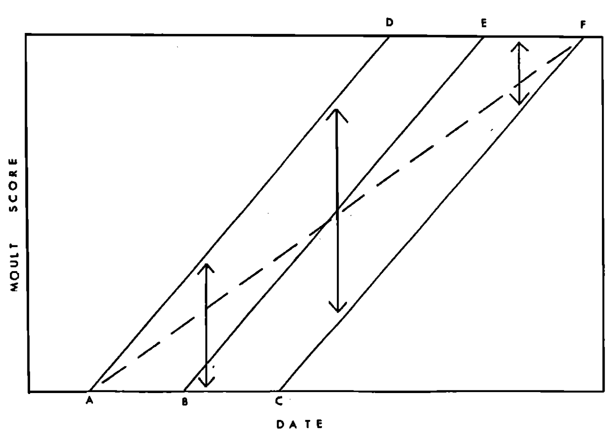

Figure 1. In a plot of moult score against date, the points generally lie within a parellogram ACFD. The "average bird" is depicted by the line BE. The arrows demonstrate that the variability of moult score is not constant for all dates (see text). Ordinary linear regression gives the dashed line, indicating the starting and completion dates for the first and last birds, respectively, in the population, rather than the average bird.

One approach which does apparently eliminate the problem of heteroscedasticity is to reverse the roles of date and moult score, treating moult score as the independent variable and date as the dependent variable (Pimm 1976). This is logically absurd, as there is no sense in which date depends on moult score. The bird-ringer chooses the dates on which to catch birds, and observes the scatter of moult scores in the sample caught on these dates. (Unless birds are caught at regular intervals throughout the moult period, the heteroscedasticity problem itself is still not solved.)

Several other methods of Summers *et al.* (1983) regress either mean or median moult scores on date, or date on mean or median moult score. All these methods are attempts to force the date to fit the assumptions of the standard regression model. The correct approach is the opposite one – first to devise a statistical model which is as realistic as possible, and then to develop the appropriate mathematics to estimate the parameters of the model. This is the approach that has been adopted here.

A NEW MODEL FOR MOULT ASSESSMENT

Underhill and Zucchini (1984) have developed a model, specifically designed for moult data. The model makes the following three assumptions:

1. The starting date for moult is normally distributed about some mean starting date T with standard deviation s. (Other distributions for the starting date may be assumed, but the mathematics then become much more difficult.)

2. Moult score (or some function or transformation of moult score) increases linearly with time since the starting date, so that the duration D for each bird is the same. This assumption is implicit in all applications of linear regression to moult.

3. The population is present from the time the first birds commence moult until the completion of moult. (It is possible to weaken this assumption.)

The model is thus characterized by three parameters: the average starting date for the population (T), the variability of the starting date, measured by the standard deviation (s), and the duration of moult, assumed to be the same for all birds (D). The statistical and mathematical formulation of the model is relatively complex. The method of maximum likelihood is used to estimate the three parameters. The method also finds confidence intervals for the parameter estimates. The value of a confidence interval is that it enables the precision with which the parameter has been estimated to be judged. A long confidence interval indicates untrustworthy estimation of the parameter; in general, the shorter the confidence interval, the more reliably the parameter has been estimated. The numerical algorithm (the computing procedure) uses an iterative technique to obtain successively better estimates of the parameters, and requires substantial computing power. Occasionally the algorithm fails to converge to a solution. This occurs when the data are sparse and inconsistent. A FORTRAN programme implementing the model has been written, and is available from the author. A full statistical description of the model will be submitted to a statistical journal (Underhill and Zucchini in prep.).

All the regression models considered by Summers *et al.* (1983) make use only of those birds actively moulting, i.e. having a primary moult score between 1 and 49. No use is made of the information contained in those birds which on a particular date had either not yet started, or had already completed, moult. Provided assumption 3 holds, at least approximately, the new model is able to extract this valuable information. Assumption 3 can be adapted for populations in which birds arrive, commence moult, remain until moult is complete, and then depart. Other variations in moult strategy could also be accommodated by modifications to the model.

The types of transformation envisaged under assumption 2 are those that make the relationship of moult score with date more nearly linear. Summers *et al.* (1983) propose converting primary moult score to percentage feather mass grown, a sensible transformation when the primaries are not all of approximately equal size. This is done in the example below. Another useful transformation is the power-series family of transformations (Box and Cox 1964) that consist of raising the dependent variable (moult score) to a power. Since moult tends to slow down towards its end, a power between one and two may help to make the relationship more linear.

AN EXAMPLE

1758 adult Sanderlings *Calidris alba* have been caught and examined for primary moult during ringing operations of the Western Cape Wader Study Group between 1970 and 1984 in southern Africa, mainly at Langebaan Lagoon, Kommetjie and Olifants River mouth (in the south western Cape) and at Walvis Bay Lagoon (Walvis Bay Enclave). In southern Africa there is relatively little variation in Sanderling numbers from the time of arrival in October until their departure in early May, and there is little evidence of movement between localities (Pringle and Cooper 1977, McLachlan *et al.* 1980, Summers *et al.* in prep.). Thus, for this population, assumption 3 is satisfied.

Moult scores were converted to percentage of feather mass grown, using the technique described by Summers *et al.* (1983) and Waltner

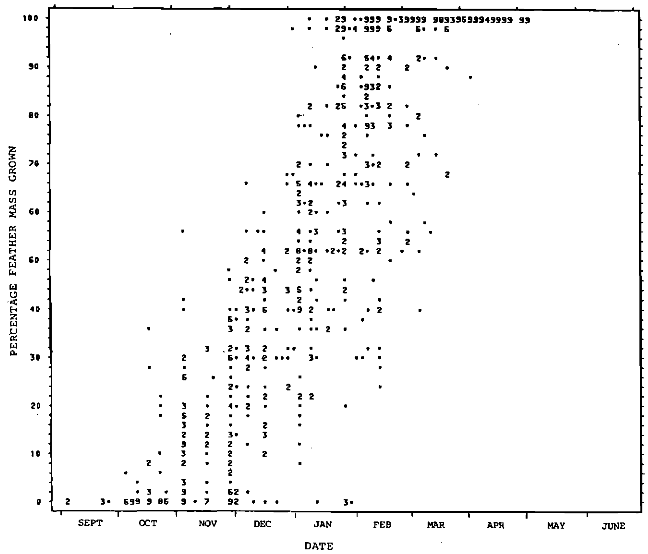

Figure 2. Plot of the percantage of primary feather mass grown against date for 1758 Sanderlings. Dots represent a single record at a point, and numbers indicate coincident points, with 9 representing "9 or more" coincident points. The overall trend is linear.

et al. (in prep.). Thus (assumption 2) we are assuming that the rate of growth of feather matter is constant. The plot of percentage feather mass against date shows a linear relationship, so the assumtion is justified (Figure 2). 672 of the Sanderlings were actively moulting, 269 had not started moult, and 817 had completed moult. The numerical algorithm converged and the parameter estimates (and their 95% confidence intervals) obtained were: average starting date T = 10 November (9 to 11 November), standard deviation s = 20.7 days (19.0 to 22.4), and duration D = 98 days (96.5 to 99.5). The estimated completion date for the average bird (T+D) is thus 16 February. The data in Table 1 show that these are reasonable estimates. Note that for all 5-day periods prior to 10 November, more than 50% of the birds had not started moult, and less than 50% after this date. Likewise, all 5-day periods prior to 16 February have less than 50% completed moult, but over 50% after this date. The confidence intervals for the starting and completion dates for the population as a whole (estimated by T $\pm$ 1.96s) suggest that 95% have completed between 7 January and 28 March. These confidence intervals also relate to the data (Table 1).

There are sufficient Sanderling moult data for some years and localities to be able to make year-to-year and locality-to-locality comparisons. These will form part of a further paper (Waltner *et al.* in prep.).

CONCLUSION

The purpose of this paper has been to bring to the attention of those analysing moult data the existence of a statistical model and computer programme designed specifically to estimate the parameters of moult. Like all models it makes simplyfying assumptions. The underlying assumptions of whatever models are applied need to be considered seriously. As Summers *et al.* (1983) point out, it is invalid to make

Table 1. Percentages of Sanderlings at different moult stages in 5-day intervals.

DATE	NOT YET STARTED	IN MOULT	COMPLETED MOULT	N
Sep 4-8	100	0	0	2
9-13				0
14-18				0
19-23	100	0	0	3
24-28	100	0	0	1
29-2	100	0	0	1
Oct 3-7	93	7	0	15
8-12	95	5	0	104
13-17	73	27	0	37
18-22	62	38	0	13
23-27	83	17	0	6
28-1				0
Nov 2-6	59	40	0	142
7-11	100	0	0	1
12-16	22	78	0	23
17-21	0	100	0	1
22-26	25	75	0	8
27-1	15	86	0	62
Dec 2-6	0	100	0	29
7-11	20	80	0	5
12-16	3	97	0	40
17-21	25	75	0	4
22-26	0	100	0	2
27-31	0	100	0	20
Jan 1-5	0	100	0	57
6-10	0	98	2	50
11-15	0	100	0	10
16-20	0	80	20	5
21-25	0	77	23	82
26-30	17	83	0	24
31-4	0	89	11	19
Feb 5-9	0	69	31	137
10-14	0	66	34	67
15-19	0	29	71	66
20-24	0	20	80	5
25-1	0	23	77	47
Mar 2-6	0	19	81	63
7-11	0	17	83	30
12-16	0	25	75	12
17-21	0	3	97	318
22-26	0	0	100	20
27-31	0	0	100	7
Apr 1-5	0	2	98	55
6-10	0	0	100	18
11-15	0	0	100	34
16-20	0	0	100	16
21-25	0	0	100	50
26-30	0	0	100	11
May 1-5	0	0	100	35

comparisons between estimates of starting and completion dates and duration of moult if the assumptions of the models by which they have been estimated differ. Before we are in a position to compare the parameters of moult between different species, or within a species but at different localities, it will be necessary to process the data through the same model. This raises the problem of accessing the raw data of moult studies which have been published. It seems sensible that journals should insist, when publishing a paper about moult, that a copy of the raw data should be deposited with them to allow for future analyses.

Our model (Underhill and Zucchini 1984, in prep.) is unlikely to prove the definite model for the analysis of primary moult. However it is a first step in the direction of custom-built models.

ACKNOWLEDGEMENTS

The Sanderling data have been collected by members of the Western Cape Wader Study Group. Charles Clinning and Paul Martin also contributed data. Ron Summers commented on an earlier draft and the Editor made helpful suggestions. The paper was written while on study leave, supported by the University of Cape Town and the S.A. Council for Scientific and Industrial Research, in the Statistics Department at Rothamsted Experimental Station, Harpenden, U.K. I express my thanks to all three institutions.

REFERENCES

Box,G.E.P. and Cox,D.R. 1964. An analysis of transformations (with discussion). *Journal of the Royal Statistical Society* A 143: 383-430.

McLachlan,A., Wooldridge,T., Schramm,M. and Kuhn,M. 1980. Seasonal abundance, biomass, and feeding of shore birds on sandy beaches in the eastern Cape, South Africa. *Ostrich* 51: 44-52.

Pringle,J.S. and Cooper,J. 1978. Wader populations (Charadrii) of the marine littoral of the Cape Peninsula, South Africa. *Ostrich* 48: 98-105.

Pimm,S. 1976. Estimation of the duration of molt. *Condor* 78: 550.

Summers,R.W., Swann,R.L. and Nicholl,M. 1983. The effects of methods on estimates of primary moult duration in the Redshank *Tringa totanus*. *Bird Study* 30: 149-156.

Summers,R.W., Underhill,L.G., Waltner,M. and Whitelaw,D.A. in prep. Population, biometrics and movements of the Sanderling *Calidris alba* in southern Africa. To be submitted to *Ostrich*.

Underhill,L.G. and Zucchini,W. 1984. Regression with a bounded dependent variable, motivated by avian moult. (Summary of paper presented at the annual conference of the South African Statistical Society, Stellenbosch, 1984.) *South African Statistical Journal* 18: 203-204.

Underhill,L.G. and Zucchini,W. in prep. A statistical model for primary moult.

Waltner,M., Underhill,L.G. and Summers,R.W. in prep. The moult of the Sanderling *Calidris alba* in southern Africa.

L.G.Underhill, Department of Mathematical Statistics, University of Cape Town, Rondebosch, 7700 South Africa.