Detecting critical changes in mohua *(Mohoua ochrocephala)* populations

Barry Lawrence
Malaghan Road
RD 1
Queenstown

Dawn Palmer
Department of Conservation
PO Box 811
Queenstown

Published by
Department of Conservation
Head Office, PO Box 10-420
Wellington, New Zealand

This report was commissioned by Otago Conservancy.

ISSN 1171-9834

© 2000 Department of Conservation, P.O. Box 10-420, Wellington, New Zealand

Reference to material in this report should be cited thus:

Lawrence, B., Palmer, D., 2000.
Detecting critical changes in mohua *(Mohoua ochrocephala)* populations. *Conservation Advisory Science Notes No. 316,* Department of Conservation, Wellington.

Keywords: mohua, yellowhead, *Mohoua ochrocephala,* population monitoring, distance sampling.

1. Introduction

Mohua (yellowhead) is an endangered bird, which is subject to sudden drops (in the order of 40-50%) in population due to predation following periodic large-scale seeding of beech trees, a beech 'mast'. If a second drop follows without a recovery of population, the mohua population is at very serious risk of local extinction.

There is a need to develop an index of mohua abundance to allow managers a cost-effective method of tracking annual fluctuations and longer-term trends in mohua *(Mohoua ochrocephala)* populations. While O'Donnell (1996) set out a methodology, we wish to determine the effort required to detect a real change of abundance of a predetermined magnitude. In Oct 1998 we undertook a pilot study with the sampling designed to allow the use of the program `Distance' (Buckland et al. 1993) to calculate an index of abundance - an index because we could not conform to all the assumptions to allow `Distance' to calculate an estimate of absolute abundance. In March 2000 we reviewed the method with the help of statistician Jenny Brown. We also looked at standard statistical methods to give us the indices we wanted, with the emphasis placed on resources required to detect a long-term trend and annual fluctuations. This was reviewed again in May with statistician Ian Westbrooke and at the Department's Distance Sampling Workshop with Rosemary Barraclough.

2. Reason for surveys

There are two reasons why surveys are necessary for managing mohua: to check that management action is having an effect, i.e. reducing the predation effects in stoat irruption years (measured by annual survey), or increasing the carrying capacity for mohua (measured by long-term trend); and to determine if more intense management and/or monitoring is necessary to see if the `critical' density (<6 pairs per 100 hectares) is reached with no recovery, or to understand the nature of the annual variation in population compared with the medium- to long-term trend.

Annual fluctuations (as opposed to seasonal changes) in mohua populations occur for a variety of reasons, such as fluctuating predator numbers, winter conditions, variability in food supply, and disease. Annual surveys cannot distinguish these, but they can provide correlations when measures of predators, winter temperatures, and food supply are also being taken. Trends give a measure of longer-term population position, but the critically important relationship is the size of the largest annual variations from the trend line, which give the manager a measure of the population's long-term survival chances. With mohua undergoing periodic 50% annual population falls, there is little room for complacency. While mohua have the capacity to recover rapidly, as in the Dart Valley 1991-93 (B. L. Lawrence unpublished report to DOC 1994), and Eglinton Valley 1991-92 (O'Donnell 1993)), often they do not, as at Hawdon 1987-90 (O'Donnell 1993), Eglinton 1996-1999 (P. Dilks, unpubl. data)). One of the critical measures is to establish a large decline with no recovery.

3. Parameters to be measured

There is a need to determine if we measure territorial groups or individuals or both.

Best practice is taken as outlined in O'Donnell (1996):

"Transects were walked slowly at 0.8-1.0 km/h. All mohua heard within about 200 m were recorded, giving a total number of encounters per transect, and the positions of all mohua were mapped. If mohua were heard, every effort was made to find them and establish how many birds were present. Whenever possible, transects were surveyed in fine weather."

The numbers of territorial groups gives a good measure of the productive status of the population. However, groups encountered in surveys are not necessarily breeding units, or territorial groups. The comparison of the `Distance' analysis with the census mohua populations in the Dart (Appendix 3) highlights the fact that the groups we `encounter' in our surveys can often be just part of a breeding group. Hence the 'Distance' analysis suggests surveys overestimate the number of territorial groups, and underestimate the average territorial group size. As long as this is done consistently it will not be a problem, but low group numbers will still reflect a low population and a truer picture of breeding pairs, while high group numbers will overestimate the number of breeding pairs.

Group size gives an indication of the `spare capacity' in the population, with the qualifications outlined above.

Variation in numbers of individuals can be significantly different from variation in groups. This is illustrated at Lake Sylvan (Table 1), and by considering the coefficient of variation at Hawdon and Millflat (Table 2) for territorial groups and individual mohua. From these examples we might conclude that it is necessary to have measures of both groups and individuals.

Statistically there is no advantage in counting individuals to detect a change, as greater variation outweighs advantages of an increase in numbers, i.e. the CV is larger. Without counting both groups and individuals, distance analysis would not be possible, but by only counting groups, less than half the time is required.

On balance it seems that current best practice should continue (counting both groups and individuals), but that if resources cannot be found to do the repeats necessary the recovery group should seriously look at a change of best practice to counting encountered groups only. Without repeats, annual surveys are of little value.

4. Changes to be assessed

4.1 ANNUAL CHANGE

There is a trade-off in surveying between effort and the size of change we wish to detect. We would certainly wish to detect mohua population collapses in any year. The size of past collapses is recorded in Table 3.

A sampling regime which detects drops of 40% would have picked up all but the reduction in numbers of individual mohua in the Eglinton Valley in 1987, but the greater fall in territorial groups means the event would still be detected.

However, as populations become smaller, more effort is required to detect the same percentage change, i.e. in `Distance' a minimum sample of n = 60-80 is required. Our pilot study had a sample of 78 for an effort of 30 km of transect. If our number of groups halves we will need twice the effort to get a result i.e. 60 km of transect. This is one of the advantages of `Distance', namely that effort can be increased as necessary because you know on the spot if you have sufficient data.

Let us take the Hawdon experience. The average of 12 breeding pairs per 100 ha in 1983 to 1986 (O'Donnell 1996) is typical of four of the five Wakatipu study sites 1999-2000 (Lawrence et al. 2000). After the first stoat irruption in the Hawdon Valley the population fell 50%, with that of pairs falling 50%. The critical aspect of this is that the population had not recovered over the next four years before the next stoat irruption reduced the population to close to zero. After the first collapse the density was reduced to 5 pairs per 100 ha. On our transects that is 1 group per transect. It must also be remembered that detectability declines at lower densities (Mayhew 1999).

We conclude that surveys must be able to detect a 40% decrease in annual populations to be of use and that, given the seriousness of a 40% drop, there should be 0.90 power to detect such a change.

4.2 TRENDS

It could be argued that the long-term trend is over mast cycles and that three cycles are required to establish a trend - nominally 15 years.

In the medium term there is some logic to the concept that we detect a trend over five years, as the average beech mast cycle is five years or so. However, the last and first years can have considerable influence on such a medium trend. Therefore a rule is required that annual data to establish medium trends must not start or finish with the two years following a stoat irruption.

5. Pilot survey

The data were collected along a 6 km transect (broken into 6 x 1 km transects in Table 4) and repeated 5 times (5 different days) during the territorial pre-incubation period. Table 4 provides the data for the survey of mohua groups and Table 5 provides the data for mohua individually.

6. Size and number of samples required

6.1 BETWEEN-YEAR VARIATIONS IN NUMBERS OF MOHUA GROUPS

We have looked at three means of designing the survey, the essential difference being the definition of the sampling unit. The count unit consists of groups of mohua.

(1) The first method is to treat the total effort as one sample, i.e.'Distance' programme in which the count for 30 km of transect is one sample.

(2) The second method is to treat each 1 km transect as a count unit. In this method the pilot is 5 repeats of each of 6 count units, the 6 count units giving one sample. This is analysed as a two-sample 't' test.

(3) The third method is to treat each 1 km transect as a count unit, as above, but analysed as a paired 't' test.

The effort required in the 'Distance' design is discussed in Section 7 below.

To decide between the second and third designs, we separated the sampling error from the location effect by doing a one-factor ANOVA (the factor in this case being location of transect). This established the within-sample error/ variance s^2 as 1.54 compared to a variance due to location effect (between-sample effect) of 6.4. This analysis shows a significant difference between locations, i.e. transects (P = 0.015). For this reason a paired 't' test is appropriate to test if between-year variation is significant. Some scepticism has been expressed at this approach. It therefore needs to be pointed out that if there is a significant location effect and the samples for each year are lumped together for a two-sample 't' test, the between-location variation could well obscure the between-year variation, i.e. a between-year difference is not detected although it is there (a type 11 error has been committed).

Ian Westbrooke's analysis (Appendix 1) gives estimates of effort required as shown in Table 6. These are slightly more conservative than estimates using methods outlined in Zar (1996). There is a basic problem that the spatial distribution of a similar sized population may vary from year to year, but we have no indication of this (we only have one year's data). The "effort required" values in Table 6 are based on the assumption that the spatial distribution of the population is the same year to year. We really have to wait till we have information across time before we can be sure we have correctly estimated the effort to detect a 40% change. In addition the variance must be greater than the value of $s2$, so the power of the test cannot be greater than that found in Table 6. In short, Table 6 is at the optimistic end of the scale.

It has been suggested that analysis by way of log(count +1/2) or Poisson regression would be more exact (see Appendix 2).

6.2 POWER TO DETECT A TREND

We used the program 'Monitor' (Gibbs 1995) to estimate the likelihood of detecting various trends. Key results are given Table 7. Expanded results are given in Table 8. This is a iterative simulation using pilot data. The number of iterations used here to generate these results was 500.

7. Distance sampling

The major advantage of distance sampling is that it corrects for differences in observer detection ability generally, if the observers detect all the birds on the zero line. The latter is the key assumption of distance sampling and it cannot be met. We cannot guarantee to detect all mohua on the zero (transect) line, as they can be hidden in the canopy. However, if observers are consistent in their ability at zero (i.e. the greater variation is at greater differences), distance sampling will retain the ability to correct for observer variation.

Buckland et al. (1993) describe methods to overcome the problem of not detecting all animals on or near the zero line, despite the fact it is the most important assumption.

We may be able to correct for this by calibrating distance surveys against censused populations. This assumes the proportion not detected is constant. Unfortunately we know (Mayhew l999) that as populations of mohua become less dense (e.g. Hurunui), they call less. But Mayhew only analysed 'medium' and 'loud' calls. It may well be that 'soft' chatter and contact calling is more consistent over various levels of mohua density and can be consistently picked up on or near the zero line (< 30 m).

Another real advantage of 'Distance' is that we know the number of encounters required to allow detection curves to be established (60-80). Furthermore, it is possible to hold this number of encounters constant and vary the effort - the length of transect. This enables the field operator to continue along transects till the required number of groups are encountered. After some discussion (R. Barraclough and I. Westbrooke pers. comm.), we believe repeating transects for mohua is not compromising 'Distance' assumptions.

The assumptions of random placement of transects are not met (see Section 9) so no inference can be made of abundance outside the width of the transect. This limitation is accepted, and applies to all designs.

Due to the problematic nature of detecting mohua on the zero line we believe 'Distance' is not an appropriate tool for these transects at present. However, if best practice of locating groups away from the transect and counting them continues, little extra effort is required to record the 'Distance' data (perpendicular distance group to transect). This allows a second quick analysis to check the paired 't' method, and may yet give a reliable index of abundance if further work is done on calibration.

8. Five-minute bird counts

It is important that the best possible use of resources be made while not overloading the surveying team. The opportunity arises to monitor annual and long-term variations in a range of birds at the same time as these transects are being undertaken. It is proposed to undertake a five-minute bird count at the start and finish of each transect, but done to allow 'Distance' analysis of the results. Such a method is described by Barraclough (2000). This is suggested as a pilot only and must be reviewed when the first two years' data are in, with an emphasis on surveyor workload. Prior to the pilot survey commencing a thorough understanding of 'Distance' requirements is necessary, as the pilot must generate the experience necessary to develop a training programme.

It is also important in the Wakatipu valleys, where long-term low-effort stoat control is likely to occur, that kaka are monitored. This is because the experience in the Eglinton Valley (P. Dilks unpublished data) suggests kaka response to this level of management could be significant. Some considerable discussion has occurred. The best option seems to include kaka in the five-minute bird counts and have a running log of all kaka seen or heard during the survey and their approx. location (the nearest km^2). This may be enough to give its a baseline from which to measure any large increase.

9. Consequences of a non-random approach

The topography of valleys in the west of the South Island is typically steep-sided, and routes through the forest must be along the valley if transects are to be practicable. While in bigger valleys a random design may be considered, in the Caples we are limited by the location of forests on steep faces and 10 km of transects in 14 km of forest. The only randomisation possible is along the line of travel. The consequence of this is that the estimates/indices of abundance can only refer to the transect (or area over which the transects were randomised). We have no measure of how these valley floor transects relate to the forest at higher altitude, for instance. For our purposes it is accepted that we are measuring the response of management on two km2 of valley floor forest spread over 14 km only.

10. Recommendations

Recognising that the ability to detect a change (estimate of the power of the survey) is an optimistic estimate, we recommend that:

1. in each valley, 10 transects be repeated 4 times and analysed by paired 't' tests, and a post-hoc power analysis be done each year to check the size of change that may have been missed;

2. to minimise variance, surveys only occur in the three weeks prior to incubation (first three weeks of October in the Wakatipu);

3. mohua groups be found and counted;

4. Mohua Recovery Group review best practice (O'Donnell 1986), i.e. the need to locate and count members of the group;

5. the distance from the mohua group to the transect be recorded to allow a second analysis by `Distance';

6. at the beginning and end of each mohua transect, a five-minute bird count with a `distance snap shot' be done as per Barraclough (2000);

7. all kaka heard at any time during surveying be recorded separately with their approx. location, i.e. to nearest grid square;

8. data be recorded on an xls spread sheet, designed so it can be analysed by `Distance', paired `t' test, and/or techniques such as Poisson regression and simulation methods, without any further data entry;

9. field record sheets reflect the xls sheet for ease of data entry;

10. if total encounters (of mohua groups) are lower than 80, further immediate repeats of the transects be done;

11. after two further years of use, the power of the survey to detect change be reviewed, particularly addressing the power of the five-minute bird counts.

11. Acknowledgements

This report is an attempt to pull together advice and comment from many sources. The statistics advice from Jenny Brown really set us on the right track. Thanks to Ian Westbrooke for his patience with those semi-literate statistically, and Steve Haslett and Greg Arnold for reviewing the statistics. The discussions at the DOC Distance Sampling Workshop, particularly with Rosemary Barraclough, were of enormous value, as were those at the Mohua Recovery Group, particularly comments from Graeme Elliott and Peter Dilks.

12. References

Barraclough, R.K. 2000. Distance Sampling: A discussion document produced for the Department of Conservation. *Science & Research Internal Report* 175, Department of Conservation.

Buckland, S.T.; Anderson, D.R.; Burnham, K. P; Laake, J.L. 1993. *Distance Sampling:Estimating abundance of biological populations.* Chapman & Hall, London.

Gibbs, J.P 1995.Overview of Monitor (version 6.2) Software.James P. Gibbs, Department of Biology, Yale University.

Lawrence, B.L.; van Hal, J,; Low, M.; Mclnnes, K.; McLeod, J. 2000. Mohua densities at observation sites in two valleys. *In:* Lawrence, B.L.; Palmer, D; van Hal,J; Low. M; McInnes, K. McLeod J; Molloy R. Report on field trial of trapping perimeters for stoat control: Dart and Caples Valleys, Upper Wakatipu 1999-2000. Unpublished report to Department of Conservation, Christchurch.

Mayhew, K. 1999. Contrariness to detection in the bush canary. Variation in the detectability of mohua *(Mohoua ochrocephala)* relating to nesting phase different populations and time of day. *University of Otago Wildlife Management Report Number 110.*

O'Donnell, C F J 1986. The Mohua Monitoring Programme. Unpublished report to the NZ Wildlife Service.

O'Donnell, C F J 1993. Mohua (Yellowhead) Recovery Plan. *Threatened Species Recovery Plan No. 6,* Department of Conservation.

O'Donnell, C F J 1996. Monitoring mohua (yellowhead) populations in the South Island, New Zealand. *New Zealand Journal of Zoology* 23:221-228.

Zar, J. H. 1996. *Biostatistical Analysis.* 3rd edition. Prentice-Hall. Inc., Simon & Schuster, Upper Saddle River, New Jersey.

Table 1. Mohua population at Sylvan, Dart Valley, 1995-1999.

Sylvan	1995	1996	1997	1998	1999	CV
Territories	15	14	13	14	14	5%
Individuals	41	40	29	49	46	18%

Table 2. Coefficient of variation (CV) for territorial groups compared to individual mohua.

	Eglinton 1984–1987	Millflat 1995–1999
Territories	CV = 7.5%	CV =12%
Individuals	CV = 17.5	CV =23%

Table 3. Percentage change in historic mohua population collapses.

Place	Date	Territories	Individuals	Source
Eglinton	1987-88	50%	26%	O'Donnell 1996
Eglinton	1990-91	66%	45%	O'Donnell 1996
Eglinton	1995-96	71%	63%	P.Dilks unpublished data
Dart	1990-91	47%	62%	Lawrence 1993 unpub report DOC
Hawdon	1986-97	50%	59%	O'Donnell 1996
Hawdon	1990-91	100%	60%	O'Donnell 1996

Table 4. Pilot survey of groups of mohua.

	CAPLES 1	CAPLES 2	CAPLES 3	CAPLES 4	CAPLES 5	CAPLES 6
Repeats N	5	5	5	5	5	5
LO 95% CI	-0.51	2.52	1.29	0.52	0.18	2.92
MEAN	0.6	4.6	2.4	2.4	1.8	3.6
UP 95% CI	1.71	6.6	3.5	4.2	3.4	4.3
SD	0.89	1.6733	0.89	1.51	1.3038	0.54
SE MEAN	0.4	0.75	0.4	0.68	0.58	0.24

Table 5. Pilot survey of individual mohua.

	CAPLES 1	CAPLES 2	CAPLES 3	CAPLES 4	CAPLES 5	CAPLES 6
Repeats n	5	5	5	5	5	5
LO 95% CI	-0.7560	7.8444	1.4829	1.7166	2.2792	5.9583
MEAN	1.0	14.2	7.2	5.4	8.4	7.8
UP 95%CI	2.7560	20.556	12.917	9.0834	14.521	9.6417
SD	1.4142	5.1186	4.6043	2.9665	4.9295	1.4832

Table 6. Minimum effort required to pick up year-to-year fluctuations in mohua populations.

Sample size n 1 km transects	m repeat counts of each transect	% change to be detected	Power to detect that change
5	4	33%	0.67
5	4	40%	0.85
10	4	33%	0.75
10	4	40%	0.90

Table 7. Effort required to detect trends in numbers of mohua groups.

No. of transects	No. of repeats	No. of years	Power to detect positive annual trends	Overall change	Power to detect negative trends	Overall change
10	4	5	7% Increase = 0.9	40%	9% Decrease = 0.84	48%
8	4	5	9% Increase = 0.91	54%	10% Decrease = 0.76	41%
6	4	5	10% Increase = 0.89	61%	10% Decrease = 0.64	41%
10	4	15	2% Increase = 1.00	35%	2% Decrease = 0.95	14%
8	4	15	2% Increase = 0.97	35%	3% Decrease = 0.97	26%
6	4	15	3% Increase = 1.00	56%	3% Decrease = 0.91	37%

Table 8. Ability to detect trends: results using the `monitor' programme (Gibbs 1995) counting encountered groups of mohua.

No. of transects	No. of repeats	No. of years	Power to detect positive annual trends	Overall trend	Power to detect negative trends	Overall trend
10	4	5	7% Increase = 0.9	40%	9% Decrease = 0.84	48%
8	4	5	9% Increase = 0.91	54%	10% Decrease = 0.76	41%
6	4	5	10% Increase = 0.89	61%	10% Decrease = 0.64	41%
5	4	5	10% Increase = 0.79	61%	10% Decrease = 0.55	41%
10	4	15	2% Increase = 1.00	35%	2% Decrease = 0.95	14%
8	4	15	2% Increase = 0.97	35%	3% Decrease = 0.97	26 %
6	4	15	3% Increase = 1.00	56%	3% Decrease = 0.91	37%
30	1	5	3% Increase = 1.00	16%	3% Decrease = 0.99	14 %
30	1	15	1% Increase = 1.00	16%	1% Decrease = 1.00	14 %
5	4	8	8% Increase = 0.98	85%	10% Decrease = 0.86	57%
6	6	8	6% Increase = 0.99	59%	8% Decrease = 0.95	49%
6	4	8	6% Increase = 0.95	59%	10% Decrease =0.92	57%
6	2	8	8% Increase = 0.95	85%	10% Decrease =0.76	57%

Appendix 1. Mohua transects paired comparison over time

The question is the amount of effort required to pick up year-to-year fluctuations in mohua population, and how this effort can be applied most effectively. We have data from a pilot study on five 1 km transects in the Caples Valley in 1998. Can we use this to estimate the sort of changes in population we might pick up?

Let

$$x_{ij}^{t}$$

be the jth observation j = 1 to m
on the ith transect i = 1 to n
at time t t = 0,1
with mean μ_i^t and constant within transect variance σ^2

Taking $d_i = \bar{x}_i^1 - \bar{x}_i^0$ the difference of the means on each transect over time, with assumed identical expected value of the difference for all the transects

$$\delta = \mu_i^1 - \mu_i^0,$$

Provided m is large enough (how small can we go - 4 or 5?) then by the Central limit theorem:

$$d_i \sim \text{Normal}(\delta, 2\sigma^2/m)$$

Let $\bar{d}$ be the mean of the d_i
Then

$$\bar{d} \sim \text{Normal}(\delta, 2\sigma^2/mn)$$

Under the null hypothesis of no change in the mean of each transect over time

$H_0{:}\delta = 0$

we will reject H_0 if $\bar{d}$ is large. To determine how large, we need an estimate of σ^2, which we will estimate from the pooled within transect sum of squares (e.g. extracted from an ANOVA). However, we need to allow for the loss of n ? degrees of freedom when we estimated the transect means, so we assume that under the null hypothesis:

$$\bar{d} \sim \text{Normal}(0, 2s^2/(m-1)n)$$

where s^2 is variance estimated from the pooled within transect sum of squares (e.g. extracted from an ANOVA)

We use this distribution for $\bar{d}$ to get the critical values for testing H_0

An example:

From 1998 pilot data (1 year only)
We have 5 transects (n = 5), each measured 5 times (m = 5)
And within transects
$s^2 = (1.257)^2$
$=1.58$
estimated variance of $\bar{d}$ = 2* 1.58/20 = 0.158,
giving estimated standard deviation of 0.4,
and a 95% confidence interval for $\bar{d}$ of about ±0.8
(compared with an overall across-transects mean of 2.96)

Power
If we take $H_A:\delta = 1$, that the difference of transect means over time is 1, and use $\alpha = 0.05$
to give a two-sided critical value for $\bar{d}$ of 0.784
then the power P of test if given by
P = prob(rejecting H_0 given that H_A is true)
= prob(N(1,0.4) >0.784)
= 0.67
If we take $H_A:\delta = 1.2$,
P = prob(N(1.2,0.4) >0.784)
= 0.85

From 1998 pilot data, but adjusting to 10 transects with 4 counts
i.e. n=10,m=4
Assuminiz the same within-transects variance
$s^2 = (1.257)^2$
$=1.58$
estimated variance of $\bar{d}$ = 2* 1.58/30 = 0.105,
giving estimated standard deviation of 0.32,
and a 95% confidence interval for $\bar{d}$ of about ±0.64
(compared with an overall across-transects mean of 2.96)

Power
If we take $H_A:\delta=1$, that the difference of transect means over time is 1, and use $\alpha = 0.05$
to give a two-sided critical value for $\bar{d}$ of 0.784
then the power P of test if given by
P = prob(rejecting H_0 given that H_A is true)
= prob(N(1,0.32) >0.784)
= 0.75
If we take $H_A:\delta = 1.2$,
P = prob(N(1.2,0.32) >0.784)
= 0.90

Ian Westbrooke

Appendix 2. Comments on Mohua transects paired comparison over time

Sources of variability

The primary difficulty lies in finding a reasonable estimate of the variance of the change in bird counts for one year to the next.

There are four sources of variability in the changes:

1. Long term spatial variability, for example through changes in nesting patterns, habitat or predation over the year.
2. Short term changes with time, for example through weather or time of day.
3. Observer changes, for example changes in their skill or who they are.
4. Random variability in the birds sighted at the time of counting.

Short term effects and observer differences can be controlled to a reasonable extent by sensible design, because counts can be made by the same people under the same conditions from one year to the next. Random effects add in to any measurement of variability. But the long term spatial component cannot by its very nature be assessed over a short period.

In the draft report's notation what is required is an estimate of $var\, d$ the variability in the change in count between transects. The value used, s^2 from an ANOVA, is an estimate of the variance in the observations within a transect over a short time period. It measures the random component (source 4 above) plus whatever components of short term or observer change are introduced by the sampling scheme [source 2]. Because it ignores the long-term spatial variation it must understate the true value, by how much it is impossible to say.

A model for the observations

To express this more formally, let

x_{ij}^t = j-th observation on i-th transect at time t; $i = 1$ to n; $j = 1$ to m; $t = 0, 1$.

Then x_{ij}^t can be split into components as:

$$x_{ij}^t \quad \mu^t + \alpha_i \pm \beta_i + \gamma_j^t + \varepsilon_{ij}^t$$

μ^t is the bird count at time t, the parameter of interest.

α_i measures differences between the transects (assumed independent) which are constant over time.

β_i measures differences between the transects which change over time t, $+\beta_i$ for t-1 and $-\beta_i$ for $t = 0$. This is source 1 above.

γ_j^t measures short term changes, kept constant over the j-th observation on the five transects. It could represent differences between observers, or time of day or weather differences, for simplicity assumed here to be independent between time 0 and time 1. This is source 2 above.

ε_{ij}^t measures differences between the observations on a specific transect independent between time 0 and time 1 and between different transect. This is source 4 above.

Now $\delta = \mu^1 - \mu^0$ is estimated by

$$\begin{aligned}\bar{\bar{x}}_{..}^{(1)} - \bar{\bar{x}}_{..}^{(0)} &= \frac{1}{n}(\sum_i \bar{x}_i^{(1)} - \frac{1}{n}(\sum_i \bar{x}_i^{(0)}) \\ &= \frac{1}{mn}(\sum_{i,j} x_{ij}^{(1)} - \sum_{i,j} x_{ij}^{(0)}) \\ &= \mu^1 - \mu^0 + 2\beta. + \bar{\gamma}.^1 - \bar{\gamma}.^0 + \bar{\varepsilon}..^1 - \bar{\varepsilon}..^0\end{aligned}$$

So the variance of this estimate depends on the variance of $\beta.$, for which no estimate is available.

Estimating the variance

The draft report measures the variance by the residual from the analysis of variance of $x_{ij}{}^0$. This compares $\alpha_i - \beta_i$ (at a single time the difference between α and β cannot be distinguished) against the variability of $\gamma_j{}^0$ and $\varepsilon_{ij}{}^0$. s^2 omits the variability in β, [source 1 above].

An alternative way of viewing the calculation is to observe that the α_i mean that $\bar{x}_i{}^0$ is not independent of $\bar{x}_i{}^1$, so $var(\bar{x}_i{}^1 - \bar{x}_i{}^0) \bullet 2\, var(\bar{x}_i{}^t)$. More straightforwardly, the difference is between observations on the same transect.

This then illustrates the problem in a formal way. The precision with which δ is estimated depends on effects which cannot be estimated from one year's data. No sophistication of the model can overcome this lack of basic information.

However we can say that the variance must be greater than s^2, so the power of the test cannot be greater than that found here. If power calculations based on s^2 predict that an important change would not be detected the actual data will certainly not show it.

The differences between the five transects, the α_i, reflect the outcome of spatial changes over many years, and therefore will probably overstate variability in changes over one year. An estimate of the variance using the 5 transect means has only 4 degrees of freedom, meaning that a 95% confidence interval is 40% longer than an estimate with 20 degrees of freedom. However using it in a power calculation would provide a lower limit for the power to consider along with the upper limit in the present calculation.

Analysis

As a secondary point the data would be better analysed using log(count + 1/2) as a % change is more likely to be constant across transects and variability is likely to be proportional to counts. This of course does nothing to overcome the basic lack of information.

A refinement giving a rather more exact analysis would be to use Poisson regression as described in Nelder & McCullagh, *Generalized Linear Models,* (1989) Chapter 6

Greg Arnold

Appendix 3. `Distance' analysis compared to census data

Dart 1998

	`Distance', sampled 36 hectares (95% confidence)	Census, covered 200 hectares
Groups per hectare	0.16 (0.05 - 0.38)	0.14
Individuals per hectare	0.38 (0.14 - 1.02)	0.495
Average group size	2.3 (2.0 - 2.6)	3.5

`Distance' underestimation of group size is probably because territorial groups often split while foraging and only by following for some time are they recognised as one territorial group. This is consistent with `Distance' slightly overestimating the number of groups while underestimating the number of individuals. The underestimation of individuals (though still well within the 95% confidence interval) could well reflect missing birds on the zero line.