BTO Research Report No. 463

Towards Developing Thresholds For Waterbirds That Take Into Account Turnover

Authors

Phil Atkinson[1], Remi Choquet[2], Morten Frederiksen[3], Simon Gillings[1], Roger Pradel[2] & Mark M. Rehfisch[1]

[1] BTO, The Nunnery, Thetford, Norfolk IP24 2PU, UK.
[2] Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, 1919 route de Mende, F-34293 Montpellier Cedex 5, France.
[3] CEH Banchory, Hill of Brathens, Banchory, Aberdeenshire, AB31 4BW, Scotland.

A report to Natural England

July 2007

British Trust for Ornithology, The Nunnery, Thetford, Norfolk, IP24 2PU
Registered Charity No. 216652

British Trust for Ornithology

Towards Developing Thresholds For Waterbirds That Take Into Account Turnover

BTO Research Report No. 463

Phil Atkinson, Remi Choquet, Morten Frederiksen,
Simon Gillings, Roger Pradel & Mark M. Rehfisch

Published in July 2007 by the British Trust for Ornithology
The Nunnery, Thetford, Norfolk, IP24 2PU, UK

ISBN 978-1-906204-22-8

CONTENTS

Page No.

LIST OF TABLES

Page No.

Executive Summary

1. To attain international importance and thus protection as a Ramsar site or as a Special Protection Area (SPA) a wetland site must either "regularly" support at least 20,000 waterbirds or seabirds, or 1% of the individuals of a population of a species or subspecies of waterbird.

2. In most cases, sites have been designated by using the maxima of individual counts. These counts will underestimate volume (i.e. total number) of birds passing through the site if turnover of birds occurs.

3. Using count data, observations of individually marked birds and survival and recruitment mark-recapture models, we present three different methods (V1, V2 & V3) implemented in the StopOver Duration Analysis or SODA program (Choquet & Pradel 2007) for estimating the total volume of birds passing through a site. We use simulated data to determine their performance using both biased and unbiased data. Specifically, we tested whether the estimates of volume were biased where the following parameters varied: proportion of birds marked, daily resighting rate, timing of arrival, proportion of transients in the population, heterogeneity in the resighting rates (i.e. some individuals with a high or low resighting rate), variation in arrival and stopover time and count error.

4. With a relatively simple dataset (single arrival, no biases), the proportion of individuals marked had little effect on the reliability of the resulting volume estimates for both V1 and V3. Estimates of volume from V2 were always overestimated. The major factor that caused a small positive bias in V1 and V3 was the resighting probability. Lower resighting probabilities caused a small positive bias in the volume estimates.

5. Resighting heterogeneity (i.e. some birds more likely to be seen than others) caused a substantial positive bias for all estimators. Transience (i.e. some birds stopping over for shorter time than others) caused no bias in V1 and V3, but a strong negative bias in V2.Transience seemed to reduce the positive bias due to heterogeneity in V1 and V3 when both were present. The use of trap-dependent models (i.e. those that allow individuals to have differential recapture rates) showed some promise for V3 as little bias in the volume estimate was observed when there was a moderate amount of variation in individuals' resighting rates.

6. V1 & V3 performed well under scenarios of varying arrival and stopover duration as well as where error in the counts was introduced. V2 was consistently biased (see Table 4.1)

7. The V3 method performed well and consistently had the highest precision; it is the method we recommend to use to estimate volume. It is important that goodness of fit tests are used to determine biases in the data and appropriate models are used in Program SODA. Although some biases in the data have little effect on the resulting volume estimates, care must be taken when setting up a study to reduce bias. We present eight different ways of ensuring that bias is reduced during the collection of data.

8. Practical ways to deal with biases are discussed. Recommendations (see section 4.2 for further details) are to: (i) Count at the same time as reading colour rings; (ii) Count at approximately one-third of the length of stay interval, e.g. if the species is thought to stay ten days on a site during passage then count every 5 days; (iii) aim to resight > 30 individuals during every count period, although preferably more; obtain as far as is possible representative samples of the population being studied; (iv) the timing of marking of the study species, the number of sites included, and the timing of counts is discussed.

1. INTRODUCTION

1.1 Background and Context

To attain international importance and thus protection as a Ramsar site, under the Convention on Wetlands of International Importance especially as Waterfowl Habitat (the Ramsar Convention: Ramsar, 1988), or as a Special Protection Area (SPA), under the EC Directive 79/409 on the Conservation of Wild Birds (the "Birds Directive"), a wetland site must meet at least one of two conditions. It must "regularly" support at least 20,000 waterbirds or seabirds (Criterion 5), or 1% of the individuals of a population of a species or subspecies of waterbird (Criterion 6).

Over 138 countries have adopted these criteria and the use of the 1% criterion has been, and will continue to be, an effective tool for the identification of important wetlands for wintering waterbirds (Fuller & Langslow 1986) but it has shortcomings. Single or multiple counts of wintering or breeding birds that do not move between sites may provide an adequate estimate of the total number of birds using sites during that period. However, in many cases and especially at staging areas, there will be a flux of birds entering and leaving the site. An understanding of turnover will lead to the improvement of site designation. Wetlands that regularly hold 1% or more of the national or international population of a species of waterbird at any one time will be designated as being important for the species and will be afforded legislative protection, whereas wetlands that may be equally important for the species, but that are only visited by 1% or more of its national or international population over the duration of spring passage, autumn passage or winter, will not.

One of the major barriers to measuring turnover in waterbird populations has been a poor understanding amongst shorebird biologists of both field and statistical methodology, constraints and potential biases that can occur when estimating length of stay and total numbers. Various attempts have been made to estimate length of stay, including mark-recapture techniques (Schaub *et al.* 2001) and radio-tracking (Farmer & Durbian 2006). These have been further developed to estimate the total population, or volume, of birds passing through a site by linking length of stay and the total number of bird days (Frederiksen *et al.* 2001) such that total numbers are calculated by dividing the total number of bird days by the mean length of stay of individuals.

Despite having important implications for site designation, turnover issues have received relatively little attention from the conservation community. There are various reasons for this, including the need for regular counts and sightings of individually marked birds, which are labour intensive and, until recently, the lack of dedicated software. Methods have been available to estimate length of stay and the number of birds passing through a site (Frederiksen *et al.* 2001, Schaub *et al*. 2001), but these are computationally complex and often beyond the abilities of those without detailed statistical knowledge. Improving access and ease of use of such methods would be of great value principally to two wider communities. It would help conservationists ensure that the number of individual birds that sites support can be estimated and thus allow suitable sites to be designated, and it would help biologists answer important biological questions about migration and wintering ecology.

1.2 Previous Methods Used to Estimate Length of Stay and Numbers of Birds Using a Site

Three terms are central to the discussion of turnover. "Bird days" (BD) is the sum of counts over all days, "volume" (V) is the total number of birds using a site, and average "length of stay" (LOS) or stopover is how many days each bird uses the site. The basic relationship between these terms is:

$$BD = V \times LOS$$

Of particular importance to waterbird monitoring and site designation in particular is that if turnover is low (LOS is high and/or arrivals and departures are synchronised), then the peak counts give a realistic estimate of the total number of birds using a site or volume. However, if turnover is high

(LOS is low and/or arrivals are staggered) then the peak counts can seriously underestimate the total number of birds using a site.

In capture-mark-recapture studies, there are well-established models to determine survival and recruitment into populations (Pradel 1996). In the case of turnover, survival and recruitment do not refer to mortality and births, rather to the probability that a bird present on one occasion was not present on the previous occasion ('recruitment' or 'immigration') or the probability that that individual leaves the population on the subsequent occasion ('mortality' or 'emigration'). Real mortality is assumed to be negligible over the period of interest. From these two parameters it is possible to determine the average LOS of individuals present at each occasion. The SODA (StopOver Duration Analysis) program (Schaub *et al.* 2001) provides a means of estimating LOS of individuals.

To estimate the total numbers of birds using a site, it is necessary to add in count information. Frederiksen *et al.* (2001) developed a method to estimate the total 'volume' of birds passing through the site. They collected (i) count information and (ii) observations of individually colour-marked geese on a regular basis. In summary, they knew the number of birds at the start of the study, estimated the rate of emigration of individuals ('mortality') and by knowing the number of individuals present the previous day the number of recruits could then be determined by subtraction. These, summed over the entire period of observation, gave an estimate of the total numbers passing through.

The Frederiksen method has been in the scientific literature since 2001 but it has been used only rarely, partly because of a lack of easy-to-use software, and there is a need for further development of the methodology described below and also the production of software to make estimating turnover more widely available.

2. METHODS

This project extended existing software (Program SODA, Stop Over Duration Analysis, Version 2.1; Choquet & Pradel 2007) to incorporate both resighting and count data.

2.1 Developments of the SODA Program to Include Volume Estimation

Three different approaches were included in SODA v 2.1 to estimate total volume.

2.1.1 Volume estimator 1 (V1)

This estimator, developed by Frederiksen *et al*. (2001), is essentially a chaining method that calculates volume as a function of the initial count and then uses the survival models and counts to estimate the number of immigrants and emigrants between each interval. The number of recruits is calculated by multiplying the survival rate by the previous count to estimate the number left from the previous interval and then subtracting this from the new count to estimate the number of new recruits.

$$V_1 = C_1 + \sum_{i=2}^{k} (C_i - C_{i-1}\Phi_{i-1})$$

Where C_i = count at time i, Φ_i = survival between time i and time $i+1$, k = the number of time periods.

However, there will be error associated with the counts and it is important to determine this in some way through actions such as repeat counts per session or small inter-count periods. In practice, the total number of birds passing through a site and the associated count error is best calculated using data from repeated counts that allow a maximum likelihood method to be used to obtain the count error estimate. It is likely that, as with other chaining methods, this method may suffer from a 'random walk' problem where count error is significant, so that in successive time periods the error may be compounded.

2.1.2 Volume estimator 2 (V2)

This is a development of the Frederiksen method and relies on the fact that estimating length of stopover is well established using the SODA program (Schaub *et al*. 2001). If birds are counted at the same time then it is intuitive that the total number of birds passing through the site is the total number of bird days divided by the average length of stay. Calculating a robust mean length of stay for all birds will be difficult, and a better estimator might be:

$$V_2 = \sum_{i=1}^{k} \frac{C_i}{LOS_i}$$

where V= Volume, C_i = the numbers of birds at time i and LOS_i = the average length of stay of individuals at time i. These are summed over the period of observation.

2.1.3 Volume estimator 3 (V3)

$$V_3 = \left(\sum_{i=1}^{k} C_i \right) \frac{\sum_{i=1}^{k} \beta_i \prod_{j=1}^{i-1} \lambda_j}{\sum_{i=1}^{k} \prod_{j=1}^{i-1} \lambda_j} \; with \; \beta_1 = 1$$

This new approach uses the colour-mark data to estimate arrival as well departure rates and accounts for uncertainty in the count data and then merges the two to get optimal estimates. As a basis, this approach uses the survival and recruitment models developed by Pradel (1996), which SODA estimates of LOS are based on, to calculate λ, the rate of change in the size of the population.

The rate of change λ is estimated by determining the rate of immigration (β) and emigration ($1-\Phi$) into the population between time periods (Pradel 1996). Census information is then be used to scale these changes so that total numbers can be estimated.

2.1.4 Assumptions

Each method relies on several assumptions which may or may not be met under field conditions. These include that censuses should be unbiased (Jolly 1965, Seber 1965, Pradel 1996) and that the marked individuals are representative of the population as a whole. Counts of waterbirds tend to be accurate at low numbers (e.g. tens of birds) but at larger numbers there is an often increasing variance associated with the counts which can come about as a result of rounding error or the fact that large flocks of birds are difficult to count accurately and numbers tend to be underestimated (Kersten *et al*. 1981, Rappoldt *et al*. 1985). The second is more difficult to quantify.

Birds do not tend to all behave in a similar manner and therefore there is a great deal of unexplained variation associated with individual behaviour. Ideally, ways should be found to stratify this if possible. For example, in spring male Icelandic Black-tailed Godwit *Limosa limosa* tend to stop-over for two days less than females probably due to competition for breeding territories (Gunnarsson *et al*. 2006). This type of stratification may be easily identified and explained but individuals may also adopt different wintering or migration strategies. Pradel *et al*. (1997) found evidence for both trap-dependence (colour-marked birds seen more often than would be expected) and transience (birds marked but not seen again) occurring in a wintering population of Teal *Anas crecca* signifying that some birds were effectively residents whereas some passed through the site and only spent a short time there. This heterogeneity in resighting data is often viewed as a 'nuisance' parameter but in fact is often biologically insightful. These issues can be identified in the data using goodness-of-fit tests in programs such as U-CARE and appropriate models built in Program SODA to deal with some of these issues.

2.2 Simulation Study

Simulated datasets were used to test the degree of bias and precision in the estimates of volume produced by the three models. We specifically tested the effects of changing the proportion of individuals marked, the daily resighting rate, differential arrival, length of stay, count error as well as the effects of transients in the population (an excess of birds being seen once and not again) and resighting heterogeneity (individuals having different resighting rates, e.g. certain individuals being more likely to be seen than others).

Simulations were structured around a 50-day period and a population of 20,000 individuals. Birds were divided into an early arrival group and a late arrival group and the proportion of birds marked and daily probability of resighting were varied. An individual's arrival date (at 0001 hrs) and staging duration were determined from random Normal distributions and used to dictate a departure date (at 2359 hrs). Resighting histories were generated by randomly resighting individuals (at 1200 hrs) on each day using a random Uniform function. Counts were daily and accurate.

The mean arrival date was day 14 (variance = 2 days) for the early group and day 25 (variance = 2 days) for the late group and both had a mean of staging duration of 14 days (variance = 2 days). For each different scenario, we produced five different sets of simulated data. Thus under the differing scenarios, the simulation provided the five estimates of the number of individuals present at the site and the number resighted per day, the staging duration, time since arrival and to departure of each individual. The resighting histories provided by each simulation were analysed using program SODA to obtain mark-recapture estimates of staging duration, V1, V2 and V3 which were compared with the true value, i.e. 20000. Simulation resighting rates were constant, fully time-dependent survival and recruitment models were specified to mimic the analyses undertaken with field data. One hundred boot-strapped estimates of total staging duration were produced from which medians and interquartile ranges were derived.

2.3 Estimating Volume Using Real Data

2.3.1 Red Knot *Calidris canutus* passing through Delaware Bay on spring passage

As part of an ongoing study in Delaware Bay, shorebirds have been systematically caught on bay beaches of New Jersey and Delaware states every c.5 days from early May to early June since 1997. All individuals have biometrics taken and are fitted with standard United States Geological Survey metal bands (rings). Also, since spring 2003, each Red Knot has been fitted with a plastic flag bearing an alphanumeric code allowing individual recognition (see Clark *et al*. 2005). Throughout the spring 2003 field season, 1385 Red Knot were caught and individually-marked, meaning that the return of birds in spring 2004 afforded the first opportunity to estimate staging duration, turnover and passage population size.

In May-June 2004, teams of observers surveyed all areas where Red Knot regularly occur within Delaware Bay. Near complete coverage of the Bay was achieved approximately every two days. All individually-marked birds were noted, plus information such as body condition, activity and flock size. Bay-wide resighting histories were produced for each recording period and analysed in Program MARK (version 4.1; White & Burnham 1999) and SODA (versions 2.1; Choquet & Pradel 2007). The analysis requires that the data meet certain assumptions. We specifically tested (using U-Care; Choquet *et al*. 2005) and found no evidence for trap-dependence or transience in the data. SODA allows the fitting of models with time-dependence and constant survival and reporting rates but no means of assessing model fit. Therefore, the 'live recapture' and 'recruitment and survival' models in MARK were used to determine the form of models to be used in SODA. The need for inclusion of constant (noted as subscript c) or time-dependent (noted as subscript t) survival (S), reporting (p) and recruitment rates (f) was determined by using the Akaike Information Criterion (AIC) as a basis for model selection, as recommended by Burnham and Anderson (1998). The AIC was used as it selects the most parsimonious model which best explains the data, but uses the fewest parameters. Differences in AIC of >2 between models are generally considered 'significant', although AIC is not a formal testing procedure, but is rather a measure of the strength of support of any given model. For the survival only model, the fully time-dependent model (S_t p_t) had an AIC 10.07 lower than the next best fit model and so this was used in the survival modelling. We included these in the survival and recruitment models and tested whether recruitment was time-dependent or constant. S_t p_t f_c was a slightly better fit than S_t p_t f_t with an AIC difference of 1.96. A likelihood-ratio test between these two models was not significant at the $P < 0.05$ level. ($\chi^2_8 = 15.008$, $P = 0.059$). Although the constant recruitment model was a slightly better fit, the time-dependent model was a more biologically realistic scenario as counts of birds within the bay showed different size pulses of arrivals between days. We therefore used the fully time-dependent model for the SODA analyses, and computed 100 boot-strapped estimates of staging parameters.

2.3.2 Spring stopover of Pink-footed Geese in N. Norway

Pink-footed geese *Anser brachyrhynchus* breeding in Svalbard, winter in Denmark, The Netherlands and Belgium. During spring and autumn migration, they stage at various sites along the Norwegian coast. The final staging site in spring is at Vesterålen in northern Norway (approx. 69° N, 15°30' E). Here, more or less daily counts of the number of geese present have been carried out during the staging period in May in most years since 1991. At the same time, geese individually marked with coded neck collars have been resighted. Marking takes place in early spring in Denmark. There is substantial management interest in knowing how large a proportion of the total population uses this staging site.

For this test, data sets from 2001 and 2002, when both counts and resightings were available on all days, were provided by Dr Jesper Madsen, National Environmental Research Institute, Denmark. Resighting histories of marked geese were extracted, and goodness of fit to a time-dependent model was tested in U-CARE. A set of basic models was then run in the program MARK to assess whether time-dependence in daily survival, recruitment and resighting parameters was needed. The data were then analysed in SODA 2.1 using the model indicated by MARK, with and without an assumed count error (coefficient of variation of 10%). Five hundred bootstrap replicates were used for the capture-recapture data, with 10 simulated count series for each in count error scenarios.

Scenario:	I	II	III	IV	V	VI	VII
Simulation testing:	Number of individuals marked and daily resighting rate	Resighting heterogeneity	Presence of transients	Resighting heterogeneity & presence of transients	Variation in arrival time	Count error	Variable stopover times
Arrival & stopover							
Proportion arriving early:	1	1	1	1	0.5	1	0.5
Mean early cohort arrival date (variance):	14 (2)	14 (2)	14 (2)	14 (2)	14 (2)	14 (2)	14 (2)
Mean late cohort arrival date (variance):					25 (2)	n/a	25 (2)
Mean stopover (variance)	14 (0)	14 (0)	14 (0)	14 (0)	14 (0)	14 (0)	14 (3)
Resighting histories							
Proportion ringed	0.03, 0.06, 0.12, 0.25	0.12	0.12	0.12	0.12	0.12	0.12
Daily resighting probability	0.25, 0.5, 0.75, 1	1	1	1	0.75	0.75	0.75
Biases and count error							
Proportion of transients	0	0	0.2, 0.4	0.2,0.4	0	0	0
Resighting heterogeneity (proportion of individuals with low daily resighting probability (0.15)). The high resighting probability was 1	0	0.2, 0.5	0	0.2, 0.5	0	0	0
Poisson error associated with counts	no	no	no	no	no	yes	no
Total number of simulations	80	10	10	20	5	5	5

Table 2.1 Summary of the simulations performed to test the reliability of the V1, V2, V3 volume estimators against biased and non-biased datasets where the total volume was known.

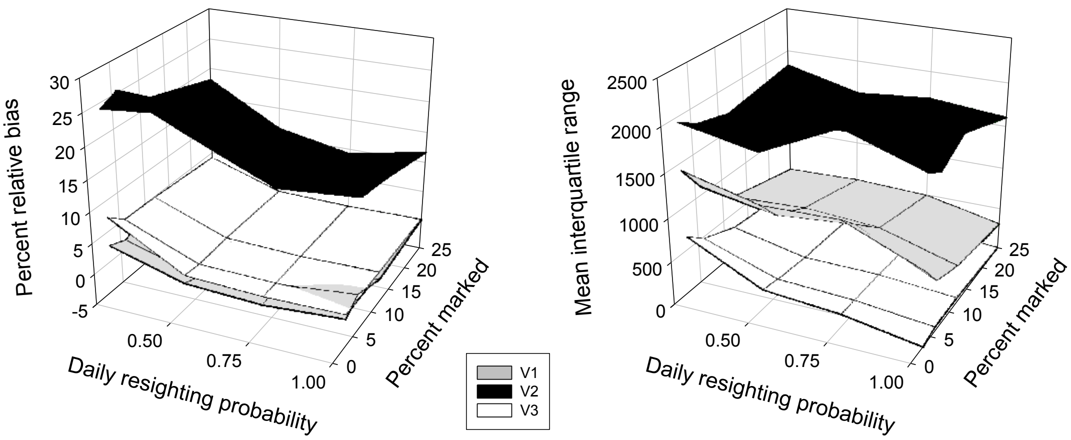
3. RESULTS

3.1 Simulation Study

In general, V3 provided the most precise estimates of volume, i.e. with the lowest interquartile range. However, SODA occasionally produces outlying V3 estimates one degree of magnitude or more greater than the median. This problem can largely be circumvented by using non-parametric summary statistics (median and interquartile range) rather than parametric ones (mean and standard error). Given that V3 estimates also were unbiased under most scenarios (see below), this estimator is probably the most robust of the three tested here.

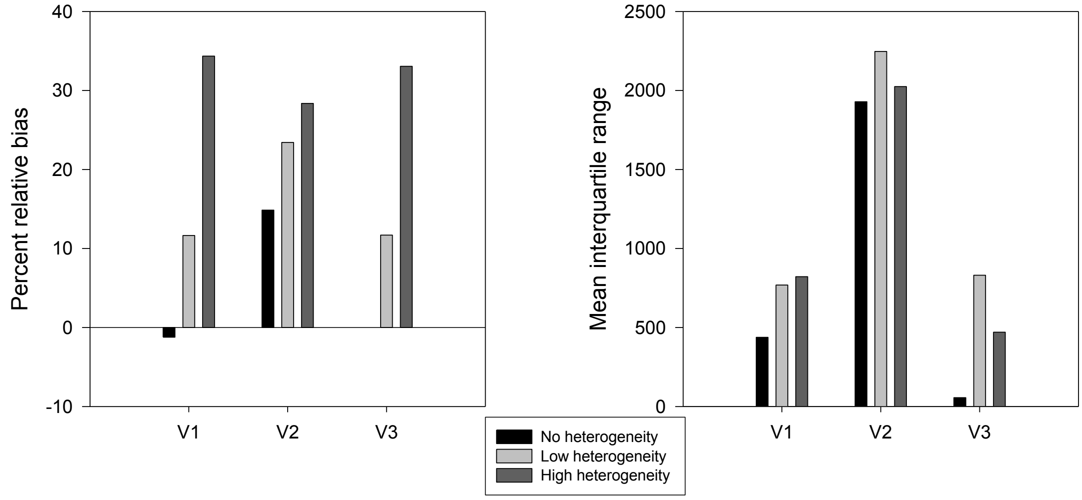
3.1.1 Scenario I - Number of individuals marked and daily resighting rate

We tested the effects of proportion of individuals marked and daily resighting probabilities on the accuracy and precision of volume estimates. Four sub-scenarios were tested for each of these two effects: 3, 6, 12, and 25% marked, and daily resighting probabilities of 0.25, 0.5, 0 75 and 1. The results showed that V2 was generally biased high, with a bias decreasing from approximately 25% to 10% with increasing values of both effects. In contrast, V1 and V3 were practically unbiased, except at the lowest daily resighting probability (0.25) when a positive bias of 3-5% was observed. V1 tended to very slightly underestimate volume in most other cases (<2%). Mean interquartile range decreased with increasing values of both effects, and the order of the three estimators was consistent: V2 > V1 > V3.



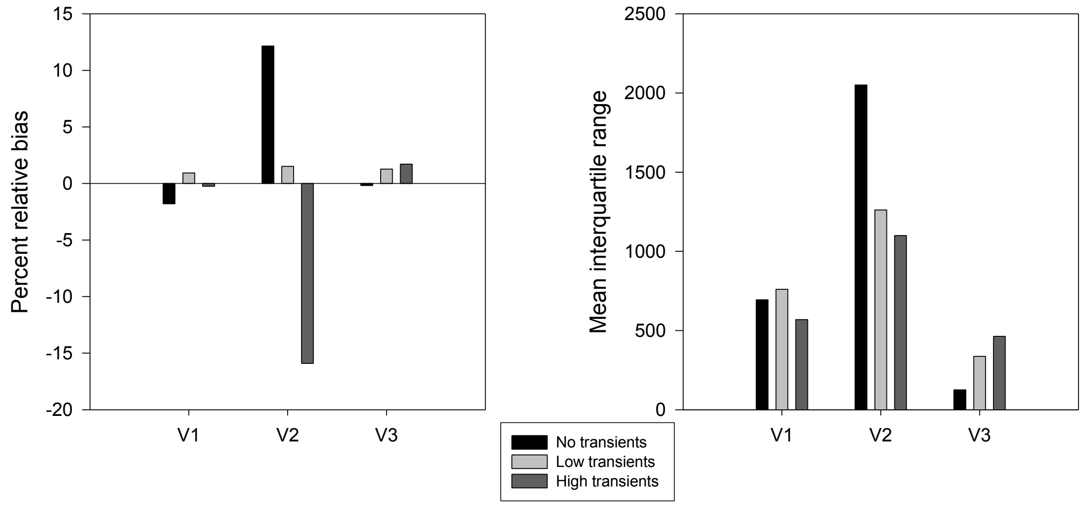
3.1.2 Scenario II - Resighting heterogeneity

The effect of resighting heterogeneity was investigated by including a variable proportion (0, 20% or 40%) of individuals with a low daily resighting probability. Other variables were held constant (12% marked, daily resighting probability = 1). For all estimators, a positive bias occurred, increasing with the proportion of less observable individuals to ~30%. Interquartile range also increased dramatically for V3, and less so for the other estimators. Resighting heterogeneity is thus a major source of bias in volume estimation, which needs to be accounted for. Preliminary results indicate that fitting trap-dependent models to data with low heterogeneity leads to a small negative bias in V3 (~2%), a larger negative bias in V1 (~15%), while V2 remains positively biased (~13%). The potential for using these models to compensate for resighting heterogeneity should be explored further.



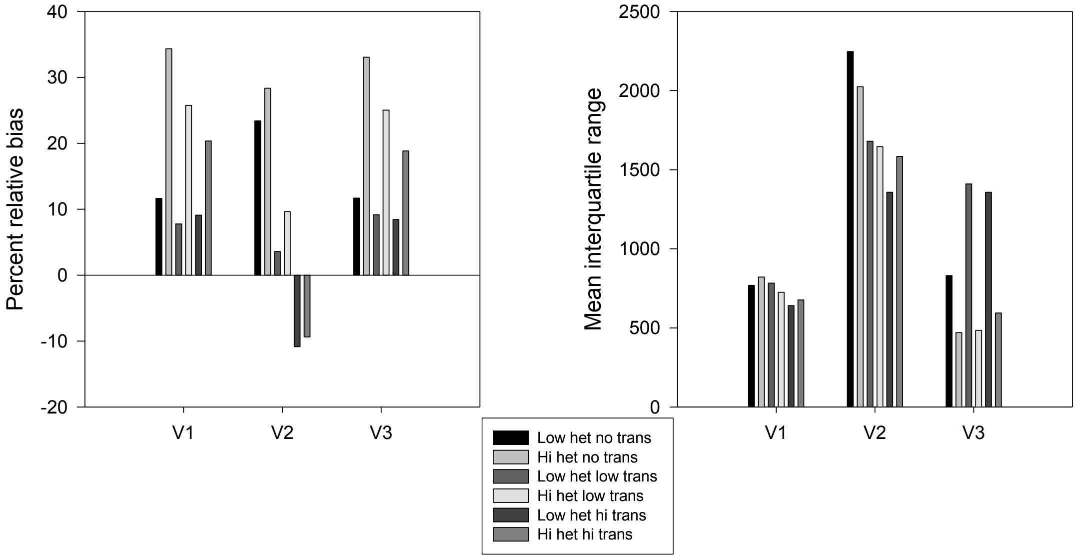
3.1.3 Scenario III - Presence of transients

The effect of transience was investigated by including a variable proportion (0, 20% or 50%) of individuals with a shorter stopover (mean = 2 days). Other variables were held constant (12% marked, daily resighting probability = 0.75). Transients did not cause any extra bias in V1 and V3, whereas V2 changed from positively biased without transience to negatively biased at high levels of transience. Interquartile range increased with the proportion of transients for V3, but decreased for V2.



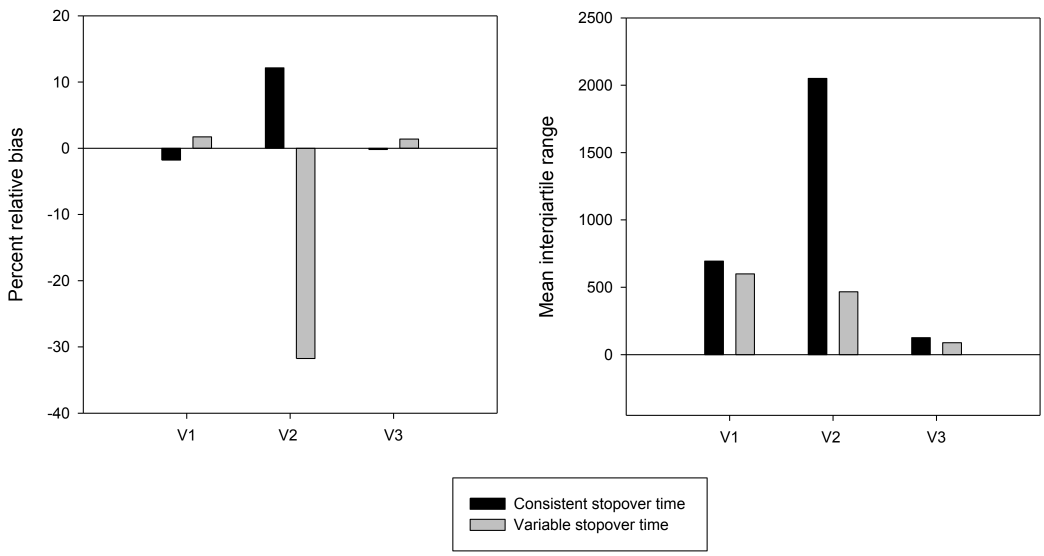
3.1.4 Scenario IV - Resighting heterogeneity & presence of transients

The combined effect of transience and resighting heterogeneity was investigated by varying the two effects (20% or 40% of individuals with a low daily resighting probability and 0, 20% or 50% of individuals with a shorter stopover. Other variables were held constant (12% marked, daily resighting probability = 1). For all estimators, the positive bias caused by heterogeneity decreased with increasing proportion of transients, although this effect was only pronounced for V2. Mean interquartile range of V3 was high for some low heterogeneity scenarios, reflecting a large number of outlying estimates.



3.1.5 Scenario V - Variation in arrival time

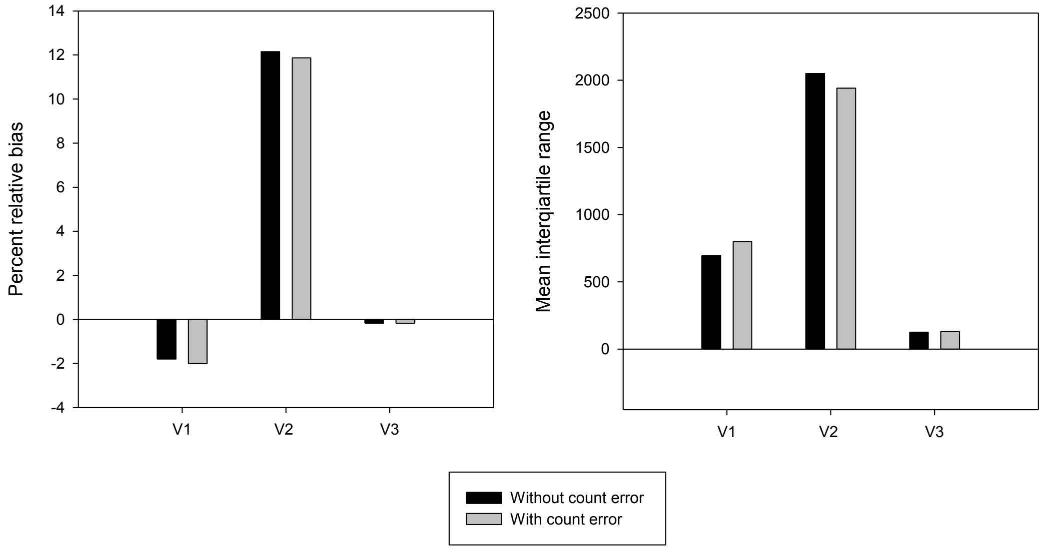
This scenario tested how the three volume estimates coped with two arrivals spread out such that no more than approximately half the 20000 birds were present on any one day. V3 was unbiased and exhibited a low interquartile range. V1 was also unbiased but with an increased inter-quartile range. V2 performed poorly, showing a strong negative bias and, perhaps surprisingly, a low interquartile range. Many of the V2 models did not converge.



3.1.6 Scenario VI - Count error

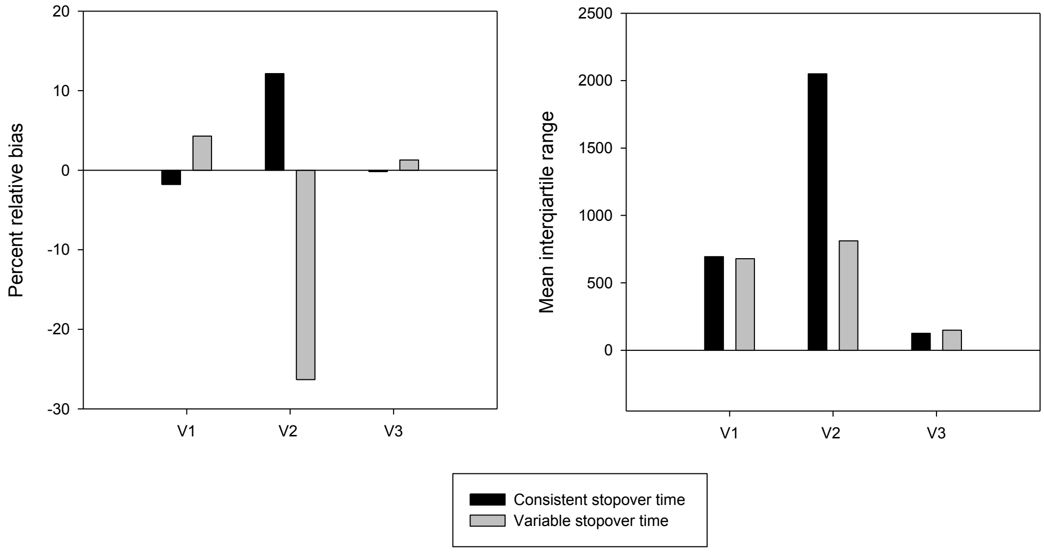
Count error was implemented by associating an error to each count. Errors associated with counts tend to be Poisson distributed. In this distribution the variance equals the mean and so the standard deviation for each count was calculated as being the square root of the count for each individual day (i.e. variance = SD^2).

The results of the simulation with count error were very similar to those without count error. V3 performed best being unbiased, whereas V2 was strongly positively biased and V1 showed a small negative bias. The inter quartile range was small for V3 but increasingly larger for V1 and V2.



3.1.7 Scenario VII - Variable length of stay

Both V1 and V3showed low bias in a situation when birds exhibited large variation in stopover times (mean = 14 days, variance = 3). The interquartile range varied such that V3<V1<V2.

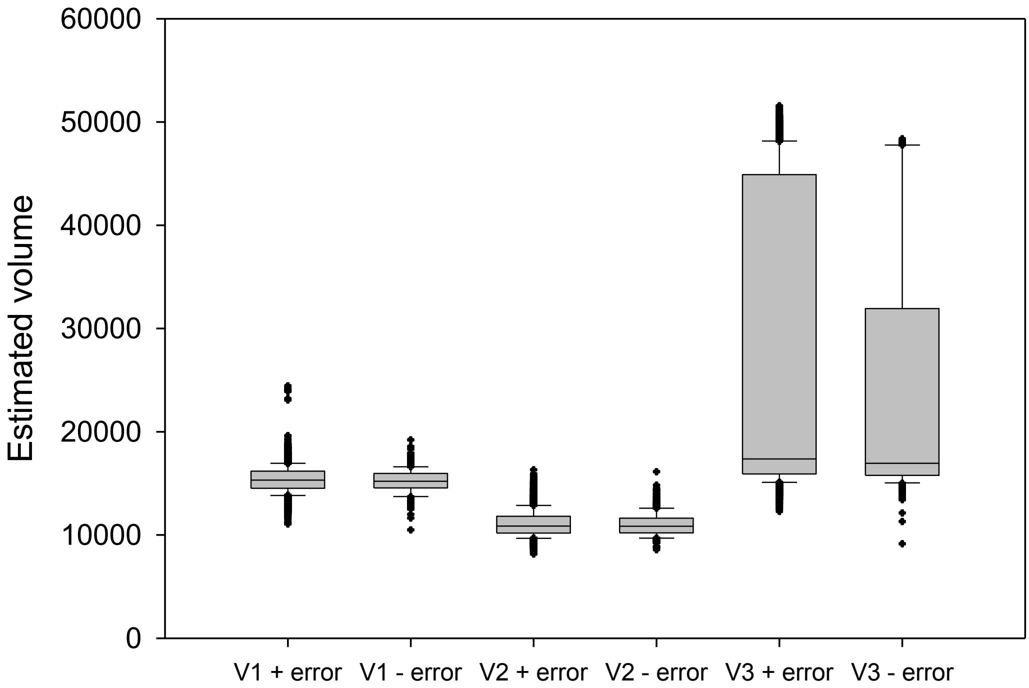


3.3 Pink-footed Goose

In 2001, a total of 195 marked individuals were seen 409 times from 5 to 26 May. The peak count was 7571. The goodness of fit was acceptable, although there was a slight trap-happiness effect (overall test: χ^2 = 43.5, *df* = 50, *P* = 0.73; test for trap-dependence: *z* = -2.29, *P* = 0.022). Model selection in

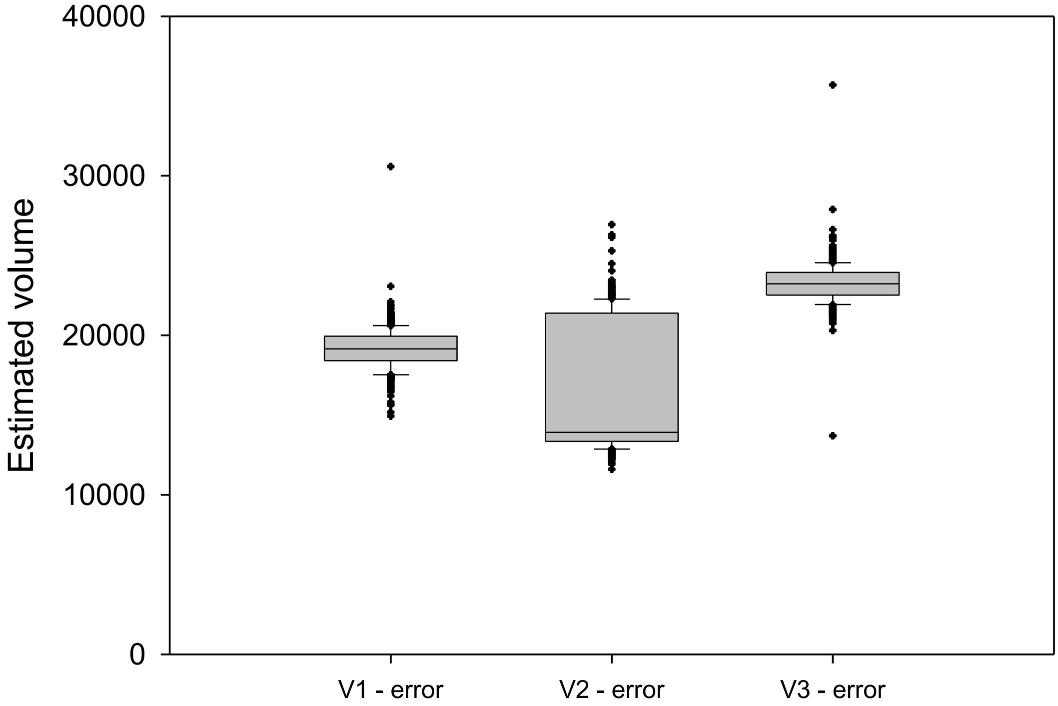
MARK indicated that a model with time-dependent survival and recruitment and constant resighting probabilities was preferable, and this model was run in SODA.

The addition of count error had very little effect on median volume estimates, and only affected precision for V3. Median volume was lowest for V2 (10840), and considerably higher for V1 (15200) and V3 (16950). Precision was much lower for V3, largely due to a bimodal distribution of volume estimates. Simulations showed that V2 can be biased low when transients are presents, and although the directional test for transience in U-CARE was not significant for this data set, V1 and V3 estimates are likely to be more reliable. In the figure: the box shows the 25th and 75th percentiles, the whiskers show the 10th and 90th percentiles, and the symbols show the outliers.



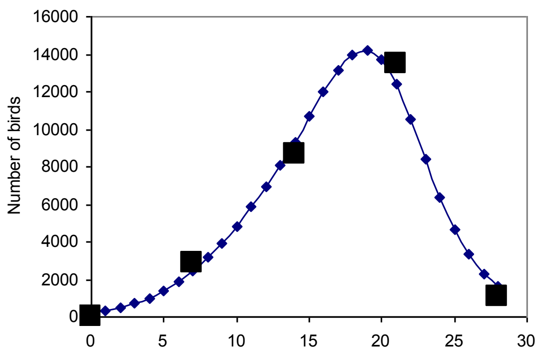
In 2002, a total of 481 marked individuals were seen 1246 times from 27 April to 22 May. The peak count was 11516. The goodness of fit was poor, although none of the specific tests were significant (overall test: $\chi^2 = 165.9$, $df = 89$, $P = 1.4*10^{-6}$). Model selection in MARK indicated that a model with time-dependent survival, recruitment and resighting probabilities was preferable, but due to convergence problems, a model with constant resighting probabilities was run in SODA.

SODA had some problems running the 2002 data. Without count error, 433 replicates were produced, and with count error 420 (rather than the 5000 expected). Results are therefore only presented without count error. Again, median estimated volume was lowest for V2 (13920), and considerably higher for V1 (19150) and V3 (23230). Precision was seemingly high, and there was no overlap between the approximate 80% confidence intervals of V1 and V3 (minimal overlap at 95% level). It is thus very difficult to evaluate which estimate is most realistic. (Note that two extreme outliers for V3 are not shown – values around 110,000).

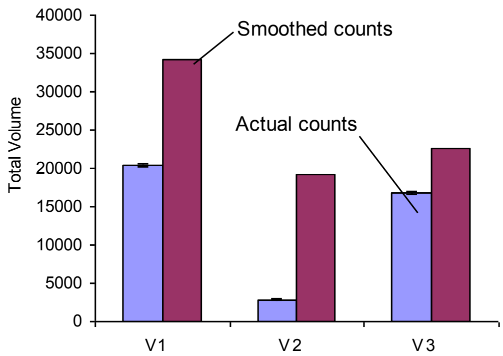


3.4 Red Knot in Delaware – does count interval impact on the volume estimates obtained?

We fitted a fully time-dependent model in SODA. The staging period for the population was relatively short – most birds arrive and depart within a 3 week window in late May and early June, the average stopover being 8-10 days. As such there is normally a rapid increase in numbers during the early part of the staging period. However, complete bay counts of shorebirds are only carried out every week and so the ratio of stopover duration to count period is rather high. To test whether this was likely to be a problem in estimating volume we tested two scenarios. The first included just the actual weekly counts and the second interpolated counts between the actual count times by fitting a smooth curve through the weekly data and estimating the numbers present on a daily basis.



Days from first count
(large squares – real counts, diamonds – smoothed counts)



Method of estimating volume (V1, V2, V3)

The two graphs above show the real and smoothed count data and the resulting volume estimates. Whether or not the smoothed line is a realistic interpretation, the key result from this is that choosing an appropriate count interval is crucial to obtaining a realistic estimate. The model underlying SODA assumes that the number of individuals between count periods does not change, which is unlikely to be realistic in this case and so the optimum count period does depend on individual's length of stay. In this case counts that are a week apart will not give a good estimate as there are only 2 counts on days 14 and 21 which have significant numbers of birds and counts every say 2-4 days would be much better.

The volume estimates obtained showed a very large variation. The smoothed count data resulted in similar V2 and V3 estimates but there was a large difference between these two methods when the actual counts were used. V3 performed best when simulated data were used and the difference between the smoothed and actual data, although different, were smaller than for V2 and V1.

4. DISCUSSION

4.1 Summary of the Simulations

There were clear differences in how the three methods of estimating volume dealt with the biases introduced into the data (Table 3). V2 was the poorest performer and was consistently biased, even when the original simulated data were not biased. V1 performed better but the variation associated with the volume estimates was generally in between V2 and V3. V3 was the best performer and was generally showed little or a slight positive bias in most situations.

The one form of bias in the simulated data that caused major problems was introducing heterogeneity in the daily resighting probability. The simulation was rather extreme with proportions of the population having resighting probabilities of 1 and 0.25. Whether these extreme values would be commonplace in the field is debatable but on a staging area birds with different migration strategies may well exhibit behaviour that comes close to this scenario. Although not presented in detail in this report, using models that take into account trap-dependence showed some promise. The V3 estimates for a low heterogeneity situation were reduced from a positive bias of c. 12% to -2% which was very similar to a dataset without resighting heterogeneity. V1 became negatively biased and V2 remained positively biased.

Out of these three methods, therefore V3 is the model of choice. Despite its ability to produce relatively unbiased estimates under different scenarios of data availability, arrival and stopover time and biases within the simulated data, it must be stressed that attention must be paid to reducing bias when considering any study.

	V1		V2		V3	
	Bias	Variation	Bias	Variation	Bias	Variation
Proportion marked	+	Moderate	+++	High	+	Low
Daily resighting probability	>0.25 + <0.25 ++	Moderate	+++	High	+	Low
Resighting heterogeneity	++ to +++	Moderate to High	++ to +++	Moderate to High	++ to +++	Moderate to High
Presence of transients	0	Moderate	-- to ++	High	0	Low
Variation in arrival times	0	Moderate	---	High	0	Low
Count error	0	Moderate	++	High	0	Low
Variable length of stay	0	Moderate	---	High	0	Low

Table 4.1 Summary of the effect of biases in the simulated data on the volume estimates for V1, V2 and V3. Bias: --- Low positive bias (<-20%); -- moderate negative bias (-20 to -10%); - low negative bias (-10 to -5%); 0 little or no bias; (-5 to 5%) + Low positive bias (5-10%); ++ moderate positive bias (10-20%); +++ high positive bias (>20%). Variation: Low (IQ range <500), Moderate (IQ range 500-1000), High (IQ > 1000).

4.2 Practical Ways to Reduce Biases

The recommendations in this section are aimed at providing guidance to schemes setting out to measure turnover. However, they must only be considered as guidelines for a successful study.

4.2.1 Census at the time of reading colour marks

Making counts at the time of reading colour rings prevents the issue of having to interpolate count data between resighting periods.

4.2.2 Make sure inter count periods are suitable

When surveying a staging site some prior knowledge of the approximate length of stay is needed so that an appropriate survey design can be initiated. For example, if individuals remain on site for only 2-3 days and censuses take place at ten-day intervals then many birds may be missed. However at a wintering site, where the rate of arrival and departure may be relatively slow, fortnightly or monthly censuses may be more appropriate. It is appropriate to sample across shorter periods as some can always be amalgamated at a later date. A rule of thumb is to choose an interval of the length of stay divided by 3.

This also raises the issue of mortality during the sighting period. In staging areas this is not usually a problem as length of stay is short and mortality is likely to negligible compared with remainder of the year. However, across a winter birds may die and the separation of emigration and mortality is likely to be difficult.

4.2.3 Are the marked individuals representative of the whole population?

One of the greatest causes of heterogeneity in the data is that individuals are not all alike. Variation can come from the fact that individual birds may adopt different strategies. However there may be more structured variation in resightings as in the case of male Icelandic Black-tailed Godwits spending tending to spend less time on spring staging areas in Iceland than females. Another example is where birds from one wintering area may pass through the staging area before birds from another as is the case in spring migration of Knots and Bar-tailed Godwits *Limosa lapponica* through the Wadden Sea. Birds wintering in northwest Europe stage and fatten in the Wadden Sea in April before migrating to their northern breeding areas, whereas birds that winter in western Africa pass through the same sites in May though do overlap to some extent. Marking a population in one or other wintering areas would bias estimates of LOS and total numbers to that population, which wouldn't be a problem if the different populations use a staging area at different times. In this specific case, it is known that the two populations are mostly separated temporally but, for populations that stage together, such as Red Knot staging in Delaware Bay in the USA (Atkinson *et al.* 2005), there would be the issue that the length of stay was being calculated for one population and, hence, applied to the other population as well. The length of stay of the different populations may not actually be the same and so other ways of separating the populations, such as stable isotopes, would be needed to determine the number of birds from each population passing through.

As long as the birds marked are representative of the whole population then bias should not be an issue but if one part of the population is marked more frequently than others then this bias will influence the mean length of stay and total number estimates. By their nature, birds that remain at one site have a greater likelihood of being caught and marked. It is also important to note that it is only possible to estimate volume for two mixed populations if they can be distinguished in the field so that counts can be split accordingly. If only marked individuals can be distinguished, LOS can be estimated separately, but not volume.

4.2.4 Should birds be marked at the staging site or in other areas (e.g. wintering areas) and how many birds should be marked?

The issue of bias from marking birds in specific wintering areas and monitoring turnover in staging areas has been highlighted above. This suggests that in cases where the issue is site- rather than population-specific, it is preferable to mark birds on the staging site. Given that the staging populations may be structured in some way, to help remove any biases it is preferable to catch birds in a larger number of small catches rather than one or two large catches. It is therefore important to collect as much information as possible on population structure by collecting data on age, sex and from intrinsic markers (stable isotopes or DNA), and individual quality determinants such as parasite loadings, for these will allow testing of whether stratification helps explain the variation in the resightings data.

Sample size is important. The error associated with the survival and recruitment estimates is likely to be lower with a greater absolute number of individuals resighted during each observation period and will probably also reduce heterogeneity if a higher proportion of individuals are resighted. Further simulations will be needed to determine this but studies where an average of >30 individuals have been resighted per time period have given relatively tight confidence intervals, but this will obviously vary according to the number ringed and ideally an individual would be resighted several times during its stay. So for example if you have 2000 individually marked birds then resighting 30 birds per time period would mean there is little chance of a bird being resighted more then once and maybe 100 would give a more realistic estimate of turnover rates.

4.2.5 Should individuals be marked during the period of observation?

Birds marked on the study site have the great advantage of being representative of the staging population as a whole, while birds marked elsewhere may only be representative of a component of this population. It is thus desirable to have such birds in the data set. For analyses that estimate recruitment, it is not possible to use the birds newly marked on site because marked and unmarked birds have different detection rates at first capture. Therefore, it will be necessary to use birds marked in a previous season for estimating recruitment. However, this is not an issue for estimating a survival probability, i.e. length of time a bird is in the area after capture.

4.2.6 How many 'sites' should be monitored?

In the majority of cases birds do not use single discrete sites and are often spread across a wider area that is too large to monitor simultaneously. There may also be movements of birds within the larger area. This is especially true for large estuarine areas, e.g. the Wash, Wadden Sea or Delaware Bay, where daily flights of over 10 km are not unusual for individual birds (Rehfisch *et al*. 1996, 2003).

Resighting individually-marked birds is often a specialist and labour-intensive job and so it may be that only one site within a larger area can be monitored whereas count information may be obtained for the whole site through, for example, aerial or coordinated ground counts. If movements were not an issue and birds behaved in a similar manner across the wider area (not a safe assumption) then the survival/recruitment parameters from the smaller site could be used in conjunction with the wider area count data to get an estimate of 'total' numbers of birds using that wider area.

However if there are significant movements of individuals across the wider area during the staging period, there will be severe bias in the estimate of total numbers. Without the development of multi-strata models that estimate transitions between states (in this case, sites within the wider area), the issue of individuals moving restricts studies to individual discrete sites without reference to the wider area.

4.2.7 Should data on the proportion of colour-marked birds be collected?

Although not used much at present, such information could be used in the future to estimate the number of birds present on a site. Detailed observations (how many birds are unmarked until first marking event; how many are unmarked until second marking event, etc) would enable the derivation of an estimate of the precision of this measure. We believe that such information should continue to be collected.

4.2.8 Do observation periods need to be equal?

Although equal time intervals may make interpretation easier, it is possible to census and look for colour marked birds at different time intervals.

Acknowledgements

Especial thanks are due to Natural England and Allan Drewitt in particular for helping to fund this study. Dr Jesper Madsen of the National Environmental Research Institute (NERI), Denmark, and the British Trust for Ornithology Delaware team generously provided the Pink-footed Goose and Red Knot data, respectively, used to test the new model. We would also like to thank Dr Niall Burton for proof-reading the report and Ms Nicola Read for formatting it.

References

Atkinson, P.W., Baker, A.J., Bevan, R.M., Clark, N.A., Cole, K.B., Gonzalez, P.M., Newton, J. & Robinson, R.A. 2005. Unravelling the migratory strategies of a long-distance migrant using stable isotopes: Red Knot *Calidris canutus* movements in the Americas. *Ibis*, **147**, 738-749.

Burnham, K.P. & Anderson, D.R. 1998. *Model Selection and Inference: A Practical Information-Theoretic Approach.* Springer-Verlag, New York.

Choquet, R., Reboulet, A-M., Lebreton, J-D., Olivier, G. & Pradel, R. 2005. *U-CARE 2.2 user's manual (Utilities-CApture-REcapture)*. CEFE, Montpellier, France.

Choquet, R. & Pradel, R. 2007. *SODA 2.0 User's Manual*. CEFE U.M.R, Montpellier, France.

Clark, N.A., Gillings, S., Baker, A.J., González, P.M. & Porter, R. 2005. The production and use of permanently inscribed leg flags for waders. *Wader Study Group Bulletin*, **108**, 38-41.

Farmer, A., & Durbian, F. 2006. Estimating shorebird numbers at a migration stopover site. *The Condor,* **108**, 792-807.

Frederiksen, M., Fox, A.D., Madsen, J. & Colhoun, K. 2001. Estimating the total number of birds using a staging site. *Journal of Wildlife Management*, **65**, 282-289.

Fuller, R.J. & Langslow, D.R. 1986. Ornithological evaluation for wildlife conservation. Pages 248-269 in M.B. Usher (ed.). *Wildlife conservation evaluation*. Chapman & Hall, London.

Gunnarsson, T.G., Gill, J.A., Atkinson, P.W., Gélinaud, G., Potts, P.M., Croger, R.E., Gudmundsson, G.A., Appleton, G.F. & Sutherland, W.J. 2006. Population-scale drivers of individual arrival times in migratory birds. *Journal of Animal Ecology*, **75,** 1119-1127.

Jolly, G.M. 1965. Explicit estimates from capture-recapture data with both death and immigration – Stochastic model. *Biometrika*, **52**, 225-247.

Kersten, M., Rappoldt, C. & Smit, C. 1981. Over de nauwkeurigheid van wadvogeltellingen. *Limosa*, **54**, 37-46. [On the accuracy of shorebird counts.]

Pradel, R. 1996. Utilization of Capture-Mark-Recapture for the study of recruitment and population growth rate. *Biometrics*, **52**, 703-709.

Pradel, R., Rioux, N., Tamisier, A. & Lebreton, J-D. 1997. Individual turnover among wintering teal in Camargue: a mark-recapture study. *Journal of Wildlife Management*, **61**, 816-821.

Ramsar. 1988. *Proceedings of the Third Meeting of the Conference of the Contracting Parties. Regina, Saskatchewan, Canada, 27 May - 5 June 1987.* Ramsar Convention Bureau, Gland, Switzerland.

Rappoldt, C., Kersten, M. & Smit, C. 1985. Errors in large-scale shorebird counts. *Ardea*, **73**, 13-24.

Rehfisch, M.M., Clark, N.A., Langston, R.H.W. & Greenwood, J.J.D. 1996. A guide to the provision of refuges for waders: an analysis of thirty years of ringing data from the Wash, England. *Journal of Applied Ecology*, **33**, 673-687.

Rehfisch, M.M., Insley, H. & Swann, B. 2003. Fidelity of overwintering shorebirds to roosts on the Moray Basin, Scotland: implications for predicting impacts of habitat loss. *Ardea*, **91**, 53-70.

Robinson, R.A., Clark, N.A., Lanctot, R., Nebel, S., Harrington, B., Clark, J.A., Gill, J.A., Meltofte, H., Rogers, D.I., Rogers, K.G., Ens, B.J., Reynolds, C.M., Ward, R.M., Piersma, T. & Atkinson, P.W. 2005. Long term demographic monitoring of wader populations in non-breeding areas. *Wader Study Group Bulletin*, **106**,17-29.

Schaub, M., Pradel, R., Jenni, L. & Lebreton, J-D. 2001. Migrating birds stop over longer than usually thought: an improved capture-recapture analysis. *Ecology*, **82**, 852-859.

Seber, G.A.F. 1965. A note on the multiple recapture census. *Biometrika*, **52**, 249-259.

White, G.C. & Burnham, K.P. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46 Supplement**, 120-138.