Assessing uncertainty in environmental sampling

Jennifer Brown[1] and Timothy Robinson[2]

[1]Biomathematics Research Centre, University of Canterbury, Private Bag 4800, Christchurch
Jennifer.Brown@canterbury.ac.nz

[2]Statistics Department of, University of Wyoming, 1000 E. University Avenue
Laramie, WY 82071, USA

Abstract

Uncertainty in estimating a true population parameter from survey data is a result of imperfect detection, imperfect observation, spatial and temporal variation, and sampling error. Some level of uncertainty is inevitable in all surveys. We discuss the sources of uncertainty in surveys of bird counts and show how we have used Monte Carlo simulations to illustrate the effect of uncertainty.

Keywords: Survey design, Monte Carlo simulation, variance.

1. Introduction

One of the common objectives in conservation management is assessment of change in status of an ecological system. Once change has been measured in some way management decides on how to mitigate the change, in the case of an undesirable negative change, or to enhance the change, in the case of desirable positive change.

Assessment of change in status of an ecological system is usually made by some inference based on data collected from field-based surveys (e.g., Thompson 1992, Thompson et al 1998, Borchers et al 2002, Thompson 2004). In all surveys, even those designed using some optimization criterion there will be a degree of uncertainty in how well the survey results reflect the true population. Survey-uncertainty is inevitable in conservation biology because environmental systems are variable, complex in multiple underlying population processes, and exceedingly hard to measure with high accuracy. This is especially true for mobile animals, inconspicuous plants and illusive and rare species (Thompson 2004, p. 1).

Understanding the sources of uncertainty and the implications of uncertainty to conservation management decisions can be difficult. We discuss here a method that we used to illustrate uncertainty. We used a Monte Carlo approach with repeat simulations to illustrate the cumulative effect of uncertainty in sampling situations.

We use two case studies. The first case study estimates uncertainty associated with counts from braided river bird surveys in New Zealand (Brown and Robinson 2009). The second case study estimates uncertainty associated with counts from penguins in Antarctica.

Understanding survey uncertainty

There are usually four main factors that contribute to uncertainty from surveys of animal counts; imperfect detection, imperfect observation, spatial and temporal variation, and sampling error.

Understanding survey uncertainty: Imperfect detection

For birds on braided river beds there are a number of reasons why it is not possible to detect all birds that are potentially available for counting. The main reasons are:

- Hidden birds—birds are hidden by rocks, dips in slopes, vegetation and other landscape features,
- Adverse weather—poor weather conditions make birds less visible,
- Diurnal behaviour patterns—birds may be more or less detectable at different times of the day,
- Group density—large groups of birds are more easily seen than small groups or solitary birds.

(Brown and Robinson 2009).

With the second case study the reasons why it was not possible to detect all penguins in the population at any one time included adverse weather (Antarctic weather can be extreme) and reasons associated with the observation platform.

Observing penguins on land by walking round the colony may have better detection rates than observing from a plane, or worst still, from a boat. Tight and large clusters of penguins often results in imprecise counts because not all individual birds can be easily seen.

Understanding survey uncertainty: Imperfect observation

The ability to detect birds varies between observers. The ability of each observer may also change through time; for example, it can decline as a result of fatigue as the day progresses, and improve as a result of increasing experience over a longer time frame. The use of more than one observation, where observations are pooled, can improve observation rates.

Understanding survey uncertainty: Spatial and temporal variation

When birds are mobile between seasons (migrating between distant places) and within seasons (moving around a particular area and onto and off adjacent lands and waters), the proportion of the total population available to be counted varies. This spatial and temporal variation reflects the change in the number of birds that could be counted given perfect detection and observation and with no change in the total bird population.

Penguin counts within a season have huge variation depending on what phase of the breeding cycle the population is at. At times the penguins are on land, and hence the population available to be counted will be both male and female from a breeding pair. At other times there will be only one partner on land while the other is feeding. The counts can double or half during the cycle depending on whether there is one or two of the pair on land. In our case study the exact phase of the breeding cycle was not able to be identified for each survey resulting in large uncertainty.

Understanding survey uncertainty: Statistical sampling error

Given only a fraction of a braided river, or Antarctica study area, is surveyed at any one time, and surveys can only occur in discrete sections of time, any count of birds is only a 'sample' both spatially and temporally, rather than a total count of the population. When this count is used in some way (such as deriving an index) to infer some biological state in the total bird population (e.g. the population is increasing or decreasing), the estimated uncertainty needs to include some measure to account for the fact that not all the potential habitat was surveyed, and surveys were not done at all points in (infinite) time. Instead, only a fraction of the area was surveyed in only a fraction of time, and there is no information on what bird counts would be in other parts of the area or at other times. However, with appropriate statistical survey design, counts for the areas and for other times that were not surveyed can be inferred from the survey results on hand. This uncertainty associated with counting only a fraction of the total population is referred to as statistical sampling error.

2. Methods

We used Monte Carlo simulations to illustrate uncertainty in estimating population trend from counts of the bird species. Data from past surveys of the river birds and of the penguins were used for defining suitable distributions of counts, and for estimating distribution parameters. Monte Carlo simulations were written in R (The R Foundation for Statistical Computing, 2007). This R code is available from the authors.

The Monte Carlo simulations for the river birds used the following steps:

1. Uncertainty as a result of spatial variation in bird counts along a river was estimated by drawing a random variate from a distribution that reflected the variation among counts from a survey of multiple sections of the river.
2. Uncertainty as a result of temporal variation between days in bird counts for a given section of river was estimated by drawing a random variate from a distribution that reflected the variation between counts taken over separate days (spaced 1 or 2 days apart). The mean of this distribution was the random variate drawn in step 1.
3. Uncertainty as a result of temporal variation within days in bird counts for a given section of river was estimated by drawing a random variate from a distribution that reflected the variation between counts in repeat surveys within the same day. The mean of this distribution was the random variate drawn in step 2.

See Brown and Robinson (2009) for details of the distributions and parameters. The other sources of uncertainty (imperfect detection and observation) are implicitly included because the estimates of spatial and temporal variation were derived from the provided bird count data.

The Monte Carlo bird count index was the random variate from step 3. This simulation method can be used for any survey design. Surveys with repeat counts within a day can be simulated with the index taken as the average of the random variates in step 3. To simulate a design for surveys with multiple sections, step 1 would be repeated for each section; and to simulate a design for surveys with repeat days, step 2 would be repeated for each day.

The Monte Carlo simulation was repeated 999 times for each survey design creating a distribution of 1000 values of each count index that mimicked variation in realistic bird counts. The distribution generated by this process was then 'grown-on' each year by the annual change (2%, 5% and 10% increase and decrease in counts) for up to 10 years to create a synthetic population of bird count indices changing over time for each species. This was done by returning to step 1 and increasing or decreasing the parameter for the mean, and any other related parameter of the distribution.

To simulate bird surveys, each synthetic bird count population produced using the process just described was sampled, and a regression line fitted. The slope of the line, converted to a more interpretable quantity, percent annual change in the mean, was a measure of the population trend. This re-sampling of the synthetic population was repeated 999 times, creating 1000 estimates of trend. Because the true trend in the synthetic population is known (it was defined in the simulation), percentile measures of uncertainty could be calculated; 25^{th}, 50^{th} and 75^{th} percentiles.

Uncertainty in penguin surveys was estimated using the same sequential Monte Carlo process with distributions for detection error from the different survey platforms, for temporal variation from different phases of the breeding cycle and for spatial variation among different colony sites. The variation in counts from different survey platforms was specifically included because of the relative size of this source of uncertainty.

3. Results

The results of the analyses were presented as a series of boxplots that showed the distribution of trends that could be reported from the synthetic population of survey counts. The effect of reducing, or increasing uncertainty by changing survey designs can be seen easily in these displays. Figure 1 shows one such boxplot display. See Brown and Robinson (2009) for full results.

The most disconcerting part of the results is the sheer size of uncertainty. Interpreting the box plots can suggest these survey results are uninformative. For example: 'given a 2% annual decline, there is a 25% chance that the reported trend will be greater than +4.6%, and only a 57% chance that the reported trend will, in fact, be negative'. Reporting that a population was increasing when in fact it was decreasing can have serious management consequences. However, interpretations such as these must be viewed in context. There is always uncertainty with any survey which does not involve a full census. Further, for environmental surveys, where populations are transitory and changing through time and in geographic space, there can be huge variation. Interpretation of survey results in environmental science with this level of uncertainty must be done cautiously and any decisions made to change an environment management practice should be well supported.

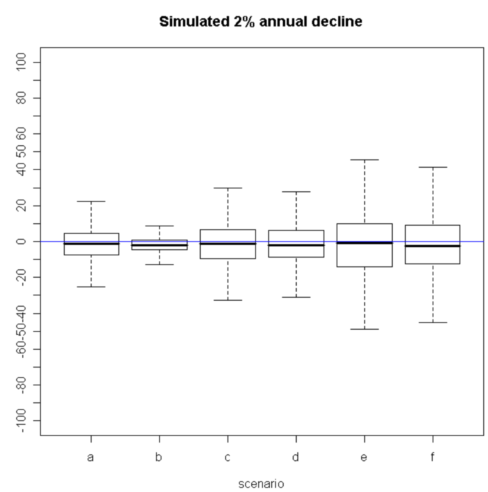

Figure 1 Distribution of estimated trends from Monte Carlo simulations for the river bird, with a simulated trend of 2% decline per year. The y-axis is the estimated change in the population per year (%), with a reference line for 0 (no change). Six different survey designs are compared, scenario a – f (from Brown and Robinson 2009).

4. Conclusions

The Monte Carlo method was used to illustrate uncertainty from survey counts to conservation biologists. The method was successful for this purpose and allowed the biologists to better understand sources of uncertainty. The simulations have been used by these biologists to explore the effect of changing survey designs.

The results have also been used in a more political environment to act as a warning for direct interpretation of estimates of trend where no estimates of uncertainty have been attached.

Often media reports of the devastating effect of an environmental impact have emotive statements like, "5% annual decline in species counts", to which the best reply is to ask what is the likely upper and lower limit of trend if this change were to be reported with some certainty. In the river birds case study we used a range between the 25^{th} and 75^{th} percentile because the concept of the middle 50 percentile can be understood (e.g., "there is a 50% chance"). We also shied away from using 95% intervals because the range would be so wide that the managers were unlikely to want to report any uncertainty.

Overall it is clear, as with any other survey, some basic principles apply, for survey design. One way to reduce uncertainty is to have clear survey objectives, where the population is defined both temporally and spatially, along with the type of change that is to be detected (linear trend, or the maximum change in the population etc).

Survey uncertainty can be reduced by using a common survey protocol in terms of observer training, the route observers walk, and their speed, and whether single or multiple observers are used. This will help with direct comparison among surveys.

Some general comments on allocating effort in the survey design can be made. Uncertainty will decrease with additional survey effort, but the marginal gains in reduction of uncertainty depend on where that extra effort is allocated. There was not sufficient variation in datasets to allow detailed exploration of alternative within-year survey-designs. Our personal observations and experience suggest multiple efforts among days is preferable to multiple effort within days, and spatial replication (e.g. multiple survey transects, multiple colonies) is very important if uncertainty is to be reduced.

The simulations show the gains in reducing uncertainty by surveying in multiple years through the survey period. If surveys were conducted at less than annual frequency then the desirable spacing between survey years depends on whether change is to be reported as a simple change (e.g., a 2% decline over 10 years), or in a more complex way to describe a non-linear trend. If, for reporting, a simple measure only was required then for non-annual surveys effort should be concentrated at the beginning and end of the time-period. If a more complex measure were required then surveys should be spaced more evenly. In the absence of common definition on how to report trend, and to allow for changes in reporting requirements, the most sensible approach would be to conduct annual surveys to ensure there is flexibility for any reporting framework.

5. Acknowledgements

Thank you to Andy Grant, Department of Conservation, Ken Hughey, Lincoln University for work on the river bird surveys. Thank you to Lyman McDonald for his work with me on the penguin surveys. Lyman was lead researcher on the penguin project so he must take all credit for any penguin comments. Also thank you to Ian Westbrooke, Department of Conservation for his statistical advice.

6. References

Borchers, D.L.; Buckland, S.T.; Zucchini, W. 2002: Estimating animal abundance. Springer-Verlag, London. 314 p.

Brown, J.A.; Robinson, T.J. 2009: Addressing uncertainty in braided river bird counts, DOC Research & Development Series 311, Published by Department of Conservation, Wellington.

Thompson, S. K. 2002: Sampling. 2nd Edition, Wiley, New York. 343 p.

Thompson, W.L. 2004: Introduction. Pp. 1-10 in Thompson, W.L.; (Ed.): Sampling Rare and Elusive Species: Concepts, Designs and Techniques for Estimating Population Parameters. Island Press, Washington D. C

Thompson, W.L.; (Ed.): Sampling Rare and Elusive Species: Concepts, Designs and Techniques for Estimating Population Parameters. Island Press, Washington D. C

Thompson, W.L.; White, G.C.; Gowan, C. 1998: Monitoring vertebrate populations Academic Press, London. 365 p.

White, G.C.; Bennetts, R.E. 1996: Analysis of Frequency Count Data Using the Negative Binomial Distribution Ecology 77: 2549-2557.