



Agreement on the Conservation of Albatrosses and Petrels

Second Meeting of Advisory Committee

Brasilia, Brazil, 5 – 8 June 2006

**Review of Trends Monitoring Methods as applied to
Seabird Populations**

Author: **Secretariat**

Review of Trends Monitoring Methods as applied to Seabird Populations

(This paper has been provided under contract to ACAP)

Chris Wilcox
Pelagic Fisheries and Ecosystems
CSIRO Marine and Atmospheric Research

Introduction and Scope

Estimation of abundance and distribution is a central theme in Ecology, and application of statistical methods to this question has been a thread throughout the history of the science. There is a substantial body of work on this topic, which has expanded rapidly with the development of the conservation biology literature. The focus of this document is to give a brief review of the philosophies that underlie the available methods, to discuss the issues that any analyses of the available data will have to accommodate, and finally suggest a method that may be useful in the analysis.

Analysis Philosophy

There are at least two possible approaches to take in analyzing the bird count data, and while either method will produce useful results it is important to make an informed choice between methods, as there are trade-offs. The basic choice is between modeling the counts as a phenomena of themselves, or trying to model the underlying population dynamics, out of which one can extract the counts. Past analyses of seabird trends have taken the approach of modeling the counts themselves (e.g. Woehler, Cooper et al. 2001), and have been able to produce both reasonable fits to the data and useful results. In general, this approach starts from the presumption that the count data is either Negative Binomially or Poisson distributed, and then uses an expression which is related to the either the shape or scale parameters (Negative Binomial) or mean of the distribution (Poisson) incorporating time and other potential covariates of interest. There are a variety of methods for analyzing the data in this context, largely dependent on the functional form chosen for the explanatory variables. If linear predictors are desired, one might use a generalized linear model or GLM (Venables and Ripley 1999), and for situations where linearity cannot be assumed one might use a generalized additive model or GAM (Venables and Ripley 1999). Subsequent to exploration with a GAM, a polynomial of suitable order could be substituted in a GLM. Further extensions are available using generalized linear mixed models, or GLMMs, in which some of the factors are fixed effects and others are random terms (Pinheiro and Bates 2000). These are particularly useful in cases where observation error occurs, and can be used to model the process inducing error thus removing one source of error from the time trend of interest.

An alternative approach to phenomenological modeling of the counts is to start explicitly from the biological process, modeling the counts as an outcome of the growth or decline

in the population under study. In this case the model is specified based on the life history and dynamical processes of the population, its parameters are estimated based on the count data, and then the trend in population size, i.e. the population growth rate, is inferred. As in the count modeling approach, these models range from very simple to quite complex. Dennis (1991) proposed a method for estimating the parameters in a very simple population model from count data. A very lucid discussion of this model and its extensions, along with instructions for estimating the parameters can be found in Morris and Doak (2002). The model is based on a diffusion approximation, and while there are clearly simplifying assumptions underlying this approach, it is well accepted in the biological literature and has been widely applied. At the opposite end of this scale are integrated assessment models, in which the underlying biological model might include survival, growth, and fecundity rates and variances for each age or size class, which are estimated across all available data including population counts, demographic data such as nest surveys, and values available in the literature. Tuck et al. (2001) applied a method along these lines to estimate the impacts of fishery bycatch mortality on albatross, and they have been widely applied elsewhere in fisheries management.

There is substantial overlap between the count modeling and the population modeling approaches. In fact, in the simplest versions of either model one is basically fitting a linear expression to the log of the count data. The essential difference is that in one case the model is concerned with the outcomes of the process (i.e. the counts themselves), while in the other it is concerned with the process of going from one count to the next (i.e. the biological process driving the counts). This difference in the analysis method affects its flexibility in matching the process, and the potential applications of parameters estimated in the model. Two examples will make these distinctions readily clear. A Poisson random variable can take integer values between 0 and infinity. If one is using a Poisson-based model to represent counts for a seabird population where there are a limited number of nesting sites available, the true count cannot exceed the number of nesting sites, although the statistical model will attach a nonzero probability to values above this ceiling. This lack of flexibility in the phenomenological approach is a fundamental result of assuming a random process (the Poisson process) that is phenomenologically similar, but mechanistically dissimilar to the actual biological process (births and deaths of individuals). The second, and potentially more important reason for using a biological process model, even a simple one, is that its parameters have meaningful interpretations. Thus in the case of a model fitted to a declining population, since the parameters can be linked to quantities that can be modified in the field, they may be informative about what actions might be taken to ameliorate the decline. For instance, in the limited nesting site example one might be able to distinguish between a loss of suitable habitat (i.e. a lower number of nesting sites) and a change in the population growth rate as causes for a change in the population counts over time. Depending on the methodology used for estimating the parameters in the biological model, there are also a number of other useful aspects which deserve mention. In biological models estimates for similar parameters (e.g. per capita fecundity) can be compared across models of differing complexity, and information can be transferred from data-rich to data-poor species. This requires estimating the models in a Bayesian framework, the discussion of which is beyond the scope of this paper. However, these

methods are well developed in fisheries and would be reasonably straightforward to apply to the types of models suggested here.

A further consideration in deciding between analysis methods is exactly what is meant by status, and how that will be quantified. There are at least two approaches to this issue, the assessment of status can be done based on the historic trend in counts or can be done based on forward projection of the counts. Typically, the historic analysis would involve testing whether there is a negative slope in the temporal trend in counts, in the case of Poisson regression, or whether the estimated population growth rate is negative in the case of a process model. An alternative, and preferable model is to compare the fit to the data of model with a trend versus the fit of a model assuming constant abundance or counts using a metric like AIC, as this avoids the problems of non-detection and power that have been widely discussed with monitoring based on null hypothesis test statistics. The forward projection approach has recently been suggested as a preferable to the use of historic trend for assessing status (Staples, Taper et al. 2005). The idea is to utilize the model that has been parameterized from the count data to predict the future abundance or count. A threshold is then set for the acceptable level of abundance, and the trend monitored is the likelihood of crossing that threshold over time. The approach has been championed for several reasons: 1) it projects forward, and is likely to indicate problems earlier than methods that look at historic trends; 2) status has a clear interpretation in the context of biodiversity protection policy; 3) relative risk (i.e. is risk increasing between years) has been shown to be robust to model misspecifications, e.g. not including density dependence; 4) the method requires relatively little data, and thus is not dependent on long time series (Staples, Taper et al. 2005). Staples et al. (Staples, Taper et al. 2005) provide a cogent discussion of this approach, and apply a diffusion approximation as suggested in the next section, demonstrating that it performs well for this purpose in a variety of situations.

Analysis Methodology

There are a variety of statistical packages that can be used to implement the analysis of the count data in a phenomenological approach, and the methods are well documented and generally accepted in the scientific literature. While there are certainly issues with assumptions underlying the analysis, such as how faithfully the distribution of the observations follows a Poisson, these can generally be accommodated through modifications to either the statistical model or transformation of the data. The TRIM package (<http://firmy.publikuj.cz/EBCC/index.php?ID=13>) is designed for analysis of count data, and incorporates a number of very useful features and would be a satisfactory package for doing a phenomenological analysis.

There are also well developed methods for estimating the parameters in the biological process models, although there is not a widely accepted standard package for estimation along the lines of TRIM. However, the computational methods are well known and easily implemented in a simple statistical language or in a spreadsheet. The attached spreadsheet estimates a diffusion approximation-based population model for a simple set

of count data. However, to get the error estimates around the rate of population change and the variance in the population change process it is easier to implement the analysis in a statistical language. R is a free statistical language and software package that readily runs on any platform. The following page contains instructions and code for implementing the analysis discussed in Morris and Doak (2002) for a density independent model (see Chapter 3, p. 51). Morris and Doak (2002) discuss various applications of the parameter estimates and diagnostic measures that can be used, along with model extensions. The R code provided in the next section can be readily modified for these purposes.

There are a few methodological differences that are important to point out in comparing the two analysis approaches. The TRIM package includes an alternative formulation of the likelihood for the Poisson model, to account for the non-independence of successive counts (a violation of the assumptions of standard GLMs). In comparison, the population dynamics models discussed here are a class of first order Markov processes, that is the value at the next time depends on the preceding value. In contrast to the Poisson GLM approach, the Markov models of the biological process assume that successive observations are correlated – and in fact are modeling the process that creates that correlation. However, these simple biological models assume that the rate of change in abundance is constant. That is, although the counts might change between years, the rate of that change is constant. There are methods for looking for discontinuities in the rate of change, as discussed by Morris and Doak (2002), and simple extensions are not difficult to implement. However, complex patterns in the rate of change across locations, through time, or with other conditions require more complex analyses. In this respect the TRIM package is more flexible in that it can accommodate covariates readily.

Literature Cited

- Dennis, B., P. L. Munholland, et al. (1991). "Estimation of Growth and Extinction Parameters for Endangered Species." *Ecological Monographs* **61**(2): 115-143.
- Morris, W. F. and D. F. Doak (2002). *Quantitative Conservation Biology Theory and Practice of Population Viability Analysis*. Sunderland, MA, Sinauer Associates.
- Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. New York, Springer.
- Staples, D. F., M. L. Taper, et al. (2005). "Risk-based viable population monitoring." *Conservation Biology* **19**(6): 1908-1916.
- Tuck, G. N., T. Polacheck, et al. (2001). "Modelling the impact of fishery by-catches on albatross populations." *Journal of Applied Ecology* **38**(6): 1182-1196.
- Venables, W. N. and B. D. Ripley (1999). *Modern applied statistics with S-PLUS*. New York, Springer.
- Woehler, E. J., J. Cooper, et al. (2001). A statistical assessment of the status and trends of Antarctic and Subantarctic seabirds. Hobart, SCAR Bird Biology Subcommittee, Scientific Committee of CCAMLR: 43.

Sample R code

This code will estimate the population growth rate “mu”, the 95% confidence interval around the mean, and the variance “sigma.sq” for a density independent diffusion approximation model as per Morris and Doak 2002.

To run the R code, the file “data for R.csv” needs to be created in the attached excel spreadsheet by replacing the data in the “sample data” sheet of the “sample analysis” workbook. Change to the sheet “export to R” and save this as a .csv file to the directory you want to work from. Be sure to paste the formulae down so all of the data is included. Then start R and set the working directory to the one where “data for R.csv” is saved. Paste the code below into the R window. Type “mu”, “mu.confidence.interval” or “sigma.sq” and the respective values will be printed to the screen.

```
raw.data <- read.csv("data for R.csv",header=T)
as.data.frame(raw.data)
attach(raw.data)
DA.est <- lm(trans.ln.change~0 + trans.interval)
Mu <- coefficients(DA.est)
names(Mu) = NULL
sigma.sq <- anova(DA.est)$"Mean Sq"[2]
Mu.confidence.interval <- confint(DA.est,1)
```