

Methods of analysis for production of indices of abundance and estimation of productivity

Abundance

Abundance data comprised counts from within the Seabird Monitoring Programme (between 1986 and the present) of whole colonies and sample plots (the latter for common guillemot, European shag, black-legged kittiwake, northern fulmar, razorbill, great skua and Arctic skua). The values of abundance for missing years in each colony (where these existed) were estimated using an ‘imputation’ method (Thomas, 1993). The method was implemented using bespoke code written in the R statistical programming language (Ihaka and Gentleman, 1996).

Specifically, Thomas (1993) calculate the value of the missing count Y_{ij} using a weighted sum of all the non-missing counts $\{y_{iq} : q \in T_i\}$ for colony i (T_i denotes the set of years for which data are available i.e. non-missing). Hence,

$$Y_{ij} = \sum_{q \in T_i} w_{jq} r_{jq} y_{iq} . \quad (3)$$

where T_i denotes the set of years for which data are available at site i . The weights w_{jq} determine the degree of temporal smoothing; we use equal weights ($w_{jq} = 1 / |T_i|$) – a standard, but not necessarily optimal, choice. An estimate of *total* abundance for year j is then found by summing both across the observed data (where available) and across the imputed counts (where the data are missing), so that

$$Y_j = \sum_{i \in S_j} y_{ij} + \sum_{i \notin S_j} Y_{ij} . \quad (4)$$

where S_j represents the set of sites that were recorded in year j .

Indices of abundance were produced by scaling total abundance to 100% in the base year (1986, the first year in the time series) and representing subsequent years as a percentage relative to 1986.

The imputation process introduces uncertainty. We quantified this uncertainty using a form of bootstrapping (Marchant *et al.*, 2004), in which we resample with replacement across the set of sites. This procedure allows us to generate confidence intervals for the estimated total abundance Y_j in each year j .

The accuracy of trend obtained using the SMP sample was assessed by comparing them with data from complete censuses of all breeding seabirds in the UK, during 1985-88 (Seabird Colony Register Census) and in 1998-2002 (Seabird 2000). An index was produced for each species in each of the census periods and the proportional change in index between censuses was compared with those obtained from the SMP sample. Where census counts were spread over a number of years, the year used to represent each census was taken as the year in which the greatest number of birds was counted. A Thomas trend was rejected as inaccurate where a discrepancy of more than 15% occurred between the Thomas estimate and the census figure.

Productivity

Data from the Seabird Monitoring Programme 1986 to present comprised samples of numbers of chicks fledged from a given number of nesting attempts at sample colonies. These were modelled so that estimates of productivity for each species in each year could be developed. The number of chicks fledged was used as the response variable in all analyses. Modelling was performed using a Generalised Linear Mixed Model, implemented in the Genstat statistical software package. Two classes of models were employed (binomial and Poisson), and applied to species depending on their maximum clutch size. For those species that lay a single egg (northern fulmar, Manx shearwater, northern gannet, common guillemot, razorbill, Atlantic puffin) the sample size was declared as a binomial denominator and the modelling proceeded with a binomial error distribution and logit link, which constrains the predicted values to lie between zero and one. For species that lay more than one egg (great cormorant, European shag, both skua spp., all tern and gull spp., and black guillemot), the sample size was declared as an offset and modelling proceeded with a Poisson error distribution and log link, which constrains predicted values to be greater than zero. In both types of model, overdispersion was corrected by scaling the deviance by the residual Pearson Chi-square divided by the residual degrees of freedom.

Five models, for each of the two classes were tested: a full interactive model of year and region (ie subdivisions of the UK, formerly adopted as reporting regions in the SMP, e.g. Mavor *et al.* 2006) effects, additive effects of year and region, year only, region only and constant productivity. Colony was included as a random effect in all models to allow for the fact that repeated measures of productivity from individual sites were not independent. Model fit was tested using F-ratio statistics and a backward elimination approach was used to arrive at the minimum adequate model. The parameter estimates were extracted from the minimum adequate model and back transformed to produce estimates of productivity. Confidence levels could not be ascertained using this procedure in Genstat.

Seabird abundance varies greatly among the regional groupings and so, for species where productivity was region-dependent *post hoc* weightings were used to produce UK productivity estimates. Abundance within the required regional groupings was extracted from the Seabird 2000 database. Productivity for each year was then estimated using analysis of variance, with the productivity value for each year being weighted by the number of pairs in the region. The least-square means from the analysis were then used to calculate the national year-dependent trends.

References

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314.

Marchant, J.H., Freeman, S.N., Crick, H.Q.P. & Beaven, L.P. (2004) The BTO Heronries Census of England and Wales 1928-2000: new indices and a comparison of analytical methods. *Ibis*, 146, 323-334.

Mavor, R.A, Parsons, M., Heubeck, M. and Schmitt, S. 2006. Seabird numbers and breeding success in Britain and Ireland, 2006. JNCC, Peterborough (UK Nature Conservation, No.30.)

Thomas, G.E. (1993) Estimating annual total heron population counts. *Appl. Statistics*, 42, 473-486.