ICES CM 2007/O:42

Random effects modeling in fisheries science using AD Model Builder

Hans J. Skaug (skaug@math.uib.no), University of Bergen, Norway
David A. Fournier, Otter Research Ltd., Canada

The need to model heterogeneity

Statistical modeling and estimation form an integral part of fisheries science. Heterogeneity in parameters arises in almost every study, along the axes of time, space and possibly other dimensions. Examples include: Mortality changing over time in mark-recapture studies (Maunder et al, 2007), abundance varying spatially (Skaug et al 2004).

Key question How much heterogeneity to allow? In practice this means estimating a variance parameter. Typically there exist no prior information about the amount of heterogeneity, so the variance must be estimated from data along with the other parameters of the model.

The Bayesian revolution

Classical statistical techniques have been very limited in their ability to account for heterogeneity. Bayesian statistics offers a solution: parameters are treated as random variables with a distribution, and can hence be allowed to vary in time or space. As a result, Bayesian statistics has gained enormous popularity during the last decade. This is also due to the invention of MCMC used as a computational tool to fit Bayesian models to data.

The frequentist alternative: random effects

Not all researchers are comfortable with the Bayesian paradigm, in particular the need to put prior distributions on all parameter. In cases where real prior information about a parameter exist, most researchers would adhere to the Bayesian philosophy. The problem arises when no prior information is available; the Bayesian must then resort to a "non-informative" prior. However, what is non-informative is not always obvious, and the choice made may influence the result in an undesirable way.

Random effects models provide an alternative, where parameter heterogeneity can be models via random variables, but other parameters are treated as non-random quantities that are estimated by maximum likelihood. This avoids the need to use non-informative priors. The variance of the heterogeneity (called a "variance component" in regression models) is among the parameters that can be estimated by maximum likelihood.

Correlation and over-dispersion

The effect of heterogeneity is that data become correlated (due to the influence of a common random factor). Random effects provide a simple and intuitive way of building correlation structures. A second application is modeling of over-dispersion in Poisson count data.

Non-linear statistical models

Linear and Gaussian models with random effects can be fit in most statistical software packages. However, many models in fisheries are either non-linear or non-Gaussian. Examples includes: population dynamics, mark-recapture models, generalized linear models. This models are not easily implemented in standards software packages.

The Bayesian software package WinBUGS allows a large range of non-linear models to be fit via MCMC. The nice things about WinBUGS are: 1) *Flexibility* (you have a full programming

language to your disposal) and 2) It is *automatic* (you do not have to worry about the nitty-gritty details about MCMC).

Computational aspects

A drawback with WinBUGS is that the underlying computational algorithms are slow, limiting the range of models that can be fitted in practice.

An alternative to doing MCMC on all parameters in the model: fit some parameters by maximum likelihood. Numerical optimization is a much faster procedure than MCMC! The random effects module of AD Model Builder (ADMB-RE) implements this approach, and joins AD Model Builder's functionality for fitting non-linear statistical models with random effects modelling.

The present study

The present study review studies from the published literature where ADMB-RE has been used to fit nonlinear statistical models with random effects (Cope, Punt, 2007; Maunder *et al.*, 2007; Skaug, Fournier, 2006; Skaug *et al.*, 2004; Trenkel, Skaug, 2005).

References

Cope JM, Punt AE (2007) Admitting ageing error when fitting growth curves: an example using the von Bertalanffy growth function with random effects. *Canadian Journal of Fisheries and Aquatic Sciences* **64**, 205-218.

Maunder M, Skaug H, Fournier D, Hoyle S (2007) Comparison of estimators for mark-recapture models: random effects, hierarchical Bayes, and AD Model Builder. In: *Euring 2007*, Dunedin, New Zealand.

Skaug HJ, Fournier DA (2006) Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Computational Statistics & Data Analysis* **51**, 699-709.

Skaug HJ, Oien N, Schweder T, Bothun G (2004) Abundance of minke whales (Balaenoptera acutorostrata) in the Northeast Atlantic: variability in time and space. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 870-886.

Trenkel VM, Skaug HJ (2005) Disentangling the effects of capture efficiency and population abundance on catch data using random effects models. *Ices Journal of Marine Science* **62**, 1543-1555.