Building robust abundance indices combining commercial data and scientific survey

Marie-Pierre Etienne[a], **Eric Parent**[a], **Jean-Baptiste Lecomte**[a] and **Robyn Forrest**[b]

[a] AgroParisTech
UMR 518 MIA
F-75005 Paris, France
marie.etienne@agroparistech.fr, eric.parent@agroparistech.fr
jean-baptiste.lecomte@agroparistech.fr

[b] Fisheries and Oceans Canada
Pacific Biological Station
Nanaimo, Canada
robyn.forrest@dfo-mpo.gc.ca

Keywords: abundance; big data; spatial ecology

Abstract: The quality of a stock assessment strongly depends on the quality of the underlying fitted population dynamic model used to investigate the consequence of different harvest rules. The estimation and even the selection of the appropriate model is mainly driven by the abundance indices considered.
Those indices are derived either from marine scientific surveys or obtained with ad-hoc normalisation of the commercial catch, called Catch Per Unit of Effort (CPUE). Data sampled during marine scientific surveys and CPUE often present a high proportion of zeros with, possibly skewed, positive continuous values. In addition for commercial fisheries, the sampling process and the latent biomass process are stochastically dependent since professionals - conversely to scientists - use to concentrate in areas that are thought likely to yield high fish abundance. The spatial locations of the commercial catch provide information on the spatial biomass repartition. Furthermore, because of the spatial structure of the population and of the sampling process, a naive abundance index produces biased estimates of the trends. The spatial aspect of the sampling process has to be accounted for. But zero inflated continuous data and preferential sampling prevent from using standard geostatistical methods. In addition, the fishermen data are cheap but massive and require to use model that can accommodate big datasets.
We develop a model that addresses the two limiting specific features, give a general expression for the likelihood function, and discuss how inference can be performed, at least approximately, using advanced Monte Carlo methods. We finally describe a possible application of such a model to bottom fish data from the continental shelf of the coast of British Columbia, Canada collected both by marine scientists (Ocean and Fisheries, Canada) and by fishermen.

References

R. Menezes, T. Su, P.J. Diggle (2010). Geostatistical inference under preferential sampling, The Annals of Statistics, Vol. 59, No. 2, 191 - 232.
C. Andrieu, G.O. Roberts (2009) The pseudo marginal approach for efficient Monte Carlo simulations, The Annals of Statistics, Vol. 37, No.2, 697-725.
D. Higdon. (1998) A process-convolution approach to modelling temperatures in the North Atlantic Ocean, Environmental and Ecological Statistics, Vol. 5, No. 2, 173-190.