MODEL SELECTION IN DISTANCE SAMPLING

Russell Alpizar-Jara[1,2] and Fátima Morgado[2]

CIMA[1]/Department of Mathematics[2], University of Évora, Portugal

Model selection is a fundamental part of distance sampling methodology [1]. The Akaike's Information Criterion (AIC) has been the preferred framework to choose a detectability function that fits a given data set, consisting on perpendicular or radial distances from line or point transects [2]. Inferences about population density or abundance are then based on the selected and estimated detection probability density function. We use a Monte Carlo simulation study to assess the performance of the model selection routine implemented in program DISTANCE [3]. Two approaches are used for estimation purposes after generating simulated data from a half-normal (HN) detectability function. We first use the traditional approach (SM) of selecting the model that provides the minimum AIC. The second approach consists on model averaging (MA) a set of the best candidate models after fitting the four key functions with different combinations of series expansions. Since the negative exponential (NE) detectability function has been often strongly unadvised, we also redid our analysis excluding it for both approaches. The effect of truncating data was also analysed. Our results show that the suggested model selection criteria tends to be biased towards selecting other than a HN detectability function in about 65% of the cases when a NE is considered, and in about 55% of the cases when a NE detection function is not included as potential candidate. We also found that the MA approach provides less biased estimates than if we were to use a single model with minimum AIC in about 60% of the times. When estimating the detection probability density function evaluated at zero distance, f(0), the distribution of bias is positively skewed, and biases could be much more severe if the a NE detection function is included in the analyses. Conversely, we found that the modelling averaging approach performed slightly better if the NE was included as a candidate model. This last result is particularly intriguing as the NE detection function is often not recommended. We illustrate our analyses with a controlled field experiment.

References

[1] Buckland, S. T., Burnham, K.P., & Augustin, N.H. 1997. Model Selection: An integral part of inference. Biometrics 53, 603-618.

[2] Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. & Thomas, L. 2001. Introduction to distance sampling - Estimating abundance of biological populations . Oxford University Press. UK.

[3] Thomas, L., Laake, J.L., Strindberg, S., Marques, F.F.C., Buckland, S.T., Borchers, D.L., Anderson, D.R., Burnham, K.P., Hedley, S.L., Pollard, J.H., Bishop, J.R.B. & Marques, T.A. 2006. Distance 5.0. Release 2. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/