

Improving the estimation of the Boyce index using statistical smoothing methods for evaluating species distribution models with presence-only data

Canran Liu¹, Graeme Newell¹, Matt White¹, and Josephine Machunter¹

¹Victoria Department of Energy Environment and Climate Action

November 8, 2023

Abstract

Species distribution models (SDMs) underpin a wide range of decisions concerning biodiversity. Although SDMs can be built using presence-only data, rigorous evaluation of these models remains challenging. One evaluation method is the Boyce index (BI), which uses the relative frequencies between presence points and background points within a series of bins spanning the entire range of predicted values from the SDM (the original version), or a modified version using moving windows. Obtaining accurate estimates of the index using this approach relies upon having large number of presences which is often not feasible, particularly for rarer or restricted species that are often the focus of modelling. Wider application of BI requires a method that can accurately and reliably estimate the index using small numbers of presence records. In this study we investigated the effectiveness of five statistical smoothing methods and the mean of these five methods (denoted as ‘mean’) to estimate the BI. We simulated 800 species with varying levels of prevalence, built distribution models using random forest and Maxent methods with two levels of training presences (NPT: 20 and 500) together with 2×NPT and 10000 random points for the two modelling methods respectively. We used four levels of presences (NP: 1000, 200, 50 and 10) and 5000 random points to calculate the BI. Our results indicate that both the original and the modified versions of the calculation of the BI are severely affected by the decrease of NP, but one smoothing method (i.e., the thin plate spline) and the ‘mean’ were almost not affected by the decrease of NP for most realistic situations. Hence, these methods are recommended for estimating BI for evaluating species distribution models when verified absence data are unavailable.

Hosted file

main_text_file.docx available at <https://authorea.com/users/697081/articles/685378-improving-the-estimation-of-the-boyce-index-using-statistical-smoothing-methods-for-evaluating-species-distribution-models-with-presence-only-data>