Comparing species distribution model accuracies for various species groups and divergent prevalences

T. Van Daele [a] and D. Maes [a]

[a] Research Institute for Nature and Forest
Brussels, Belgium
toon.vandaele@inbo.be

Abstract: Species distribution models (SDM) relate the occurrence of species to a set of explanatory environmental variables. Many different types of measures have been proposed to indicate the models accuracy (Liu et al., 2011). These measures are known to be dependent of the species prevalence for both the training and validation datasets. Most species datasets don't have explicit data on species absences. One way to overcome this is to select species absences from locations that have been thoroughly investigated. We investigate how different selection strategies from grid based species datasets influence the model results and accuracy measures. For this study we use a collection of datasets with species presences of five different species groups (butterflies, birds, plants, dragonflies and amphibians) in 1x1km grid cells in Flanders (Belgium). Within each species group, a range of rare and common species are selected. The selection of the data is based on different levels of intensity the grid cells have been investigated. We compared some of the most common accuracy measures (AUC, kappa, TSS and overall accuracy) for several model techniques (GAM, RF and Maxent) for all species in each group. The analysis allows determining the optimal selection for each species group.

References
Liu C.R., White M., Newell G. (2011). Measuring and comparing the accuracy of species distribution models with presence-absence data. Ecography 34(2):232-243.