

Tensions in ecological data collection: Integrating data sources in Bayesian design for species distribution modelling

Max Savery¹ and Stijn Luca¹

¹Ghent University, Department of Data Analysis and Mathematical Modelling

We present an analysis of Bayesian design of surveys for ecological data collection when integrating Presence-Only (PO) and Presence-Absence (PA) data in a site-occupancy model. Working in a simulated virtual environment, we analyse the effect of PA and PO data balance and contribution in varying data availability environments—both scarce and abundant. We consider the optimization of both structured survey sites and the unstructured locations for citizen scientists, conditional on previously collected data. Additionally, we consider a number of important issues regarding the behaviour of the integrated model behaviour in the designs. This includes correlation between detection probability and sampling bias, the use of single-visit PA surveys, the prior availability of each data source, and the tension between the corresponding likelihood components. Furthermore, we compare design-based and model-based Bayesian design strategies, highlighting how the balance of PA and PO data effects the outcome of different strategies in unexpected ways, particularly when considering prior data availability. Finally, we evaluate the ability of the designs to detect changes in species distribution over time, an essential requirement of designs for biodiversity monitoring for a quickly changing climate.