

Modelling species communities through space and time using crowdsourced datasets

Joint species distribution models (JSDMs) are becoming increasingly popular for the analysis of species community data. Their rising popularity coincides with the increasing availability of crowdsourced biodiversity data through citizen science initiatives and the digitization of natural history collections. Most JSDMs, however, cannot deal with the challenges intrinsic to these opportunistic data sources, such as imperfect and heterogeneous detection probabilities. We developed a spatio-temporal joint species distribution model that flexibly acknowledges imperfect detection. Specifically, we combined a spatio-temporal Factorisation Machine approach with a comprehensive site-occupancy approach to model occupancy and detection patterns across the considered species. We performed Bayesian inference through the probabilistic programming language Stan and applied the developed model to large datasets on invertebrate occurrences in Belgium, encompassing historical records as well as contemporary sightings by naturalists. By doing so, we were able to gain insights at both the species and the community level, ranging from species-specific distribution maps, distributional trends and phenological patterns, to species-to-species associations and shifts in the composition of natural communities over time. We argue that extending joint species distribution models to flexibly accommodate imperfect detection enables the study of species communities at an unprecedented scale due to their ability to harness a wider variety of datasets.