124

A framework for predicting species abundances from distributional patterns of presences and absences: simulation and empirical tests

Understanding how species abundances vary across populations and species is important for theoretical, empirical, and applied contexts. Species Abundance Models (SAMs) are often used to model species abundances in space (e.g., across sites, lakes) as a function of environmental variables. While environmental variables are easily accessible, they do not always provide all the required sources of variation to predict abundances. Species interactions and missing environmental predictors are important factors that reduce SAMs predictive performances. Moreover, precise estimates of species abundances to calibrate SAMs are usually available for a much smaller fractions of sites for which the model is needed for prediction. Taken these challenges together, SAMs occasionally use presence-absence data across multiple species, which are easier to gather, to infer species interactions and missing environmental predictors. Because these models are known to have mixed performances, we investigate potential factors that may lead to lower SAMs predictive abilities. We focus on the generalised linear latent variable models (GLLVM) to model community composition that are then included in SAMs. We use simulations and a large empirical fish dataset for which there is accurate abundance estimates for 700 lakes and 84 species in Canada to assess the performance of this latent-based SAM approach. Our simulations consist of generating species abundances across a very large number of sites (universe); and then sample a much smaller number of sites, estimate our latent-based SAMs, and predict the known abundance for the remaining sites in the universe. We use different scenarios to generate random and systematic perturbations to emulate errors in abundance estimates used to fit the proposed latent-based SAM. By identifying the conditions in which model predictability is affected, we can determine strategies to minimize these errors. The empirical lake-fish abundances are used to evaluate when and how different strategies designed to reduce random and systematic errors may improve the performance of SAMs.