

Using remote-sensing variables to predict fine-scale arthropod distributions across a forested landscape

The US Forest Service is charged with achieving ‘sustainable multiple-use management’ to promote “the health, productivity, diversity, and beauty of forests and associated lands.” In short, US forests should produce timber, promote recreation, and protect biodiversity. This multiple-use mandate requires, as an input to decision-making, distribution maps of large numbers of species, to identify areas of high conservation value that should be protected. To generate these maps efficiently, we have designed a pipeline to predict the distribution of arthropod communities from DNA-based data.

We collected 121 Malaise-trap samples from 89 sample sites in and around the HJ Andrews Experimental Forest, Oregon. We shotgun-sequenced each sample and used Kelpie software to carry out in-silico PCR of the COI DNA barcode gene (BF3BR2 primer) to extract 889 Operational Taxonomic Units (OTU), which we filtered to 190 OTUs with ≥ 6 incidences. The resulting sample by species table was paired with Landsat, Lidar, and human-activity environmental covariates in a joint species distribution model. We used the sjSDM R package to apply a deep neural network (DNN) on the environmental covariates, paired with a linear model on spatial position and a species-species correlation matrix to direct model fitting via regularization. Model tuning was carried out using cross-validation, and we analyzed variable importance with explainable AI (xAI). The fitted model has a mean explanatory AUC of 0.82, and a mean predictive AUC of 0.66. xAI found that elevation and roads are most frequently the most important covariate, although 28 covariates in total were the most important for at least one species. We interpolated the model predictions for the 87 best-performing species (AUC_{predictive} > 0.7) across the 384 km² study area to generate a continuous biodiversity map, subdivided the landscape, and calculated each site’s irreplaceability value, a measure of conservation value. Site irreplaceability is negatively correlated with elevation and topographic position index.

We show that it is possible to efficiently fit a joint species distribution model with reasonable predictive performance on large numbers of data-poor species. Our pipeline can generate informative inputs to land-use decision-making and can be scaled up to larger areas.