220

A unifying modelling framework for metabarcoding data

Metabarcoding has been a breakthrough in the field of wildlife population monitoring, because it allows detection of substantially more species per sample compared to traditional sampling methods. The data obtained from metabarcoding protocols are in the form of counts, with each count representing the number of times the DNA of a particular species has been detected in the sample. Modelling metabarcoding data presents additional challenges compared to traditional sampling methods. For example, in addition to false negative observation error, it is important to account for false positive observation error due to contamination at the site or at the lab and to distinguish at which stage the contamination has occurred. An additional challenge is that parameters at different levels of the sampling process are unidentifiable or very close to being unidentifiable. For example, it is usually impossible to distinguish whether a low count is the result of lower biomass or of low amplification rate in the lab. We propose a new modelling framework for studying species communities using metabarcoding data, while taking into account false positive and false negative error at each stage of the sampling process. Compared to previous modelling approaches, we are able to infer changes in species distribution using the information from the number of reads and from the occupancy process, while accounting for errors at each stage and for PCR-specific biases. We tackle the identifiability issue in two ways. We propose an efficient MCMC strategy based on interweaving strategies that allows optimal mixing and we account for spike-ins, which are artificial controls introduced during the bioinformatic process to improve estimates of noise in the process. Finally, we account for the effect of covariates at all stage, and for species correlations at the species distribution stage, incorporating a joint species distribution model approach. We apply our approach to ingested DNA, collected by analyzing leeches, and DNA collected through malaise traps.