

## Effects of including community-level or species-specific data on the performances of a Joint Species Distribution Model: a case study with a species-rich marine benthic communities

Understanding and predicting the variability of species communities remain a major challenge in ecology. Relative to single species distribution models, Joint Species Distribution Models (JSDM) offer an improvement to characterise community variability as all species abundances or occurrences are modelled jointly. In JSDM, species can share similar responses to environmental conditions based on their phylogenetic or functional traits similarities. In addition to environmental filtering, recent development in JSDM can also characterise species responses to missing abiotic and biotic predictors by estimating species correlation at community-level through latent variables. Thus, the explanatory and predictive power of JSDM can increase thanks to: (1) additional species-specific information such as phylogeny or functional trait data, or (2) inclusion of abundance or occurrence of non-target species.

Here, we tested the influence on JSDM performances of including species-specific information or data about non-target species. Focusing on a target assemblage of 99 bristle worm species out of total of 278 benthic species sampled across 21 sites along the French Atlantic coast over 9 years, we fitted alternative model structures using the HMSC framework. Each of 3 alternative structures (1) with phylogeny, (2) with phylogeny and functional traits and (3) with the 278 species were fitted using abundance and occurrence data. We assessed their performances relative to a baseline model that only considers the 99 bristle worm species without phylogeny nor functional traits.

Our results suggest that inclusion of phylogenetic or trait information does not improve model abilities to predict bristle worms occurrences and abundances. However, models explanatory and predictive abilities for bristle worms occurrences and abundance do improve when including information about the 179 non-target species. Thus, inclusion of non-target species in the model seems more informative than adding traits or phylogeny data. Moreover, this performance gain comes at a cost as the computational time and the RAM needed to fit the model exponentially increase with the number of species included in the model. Since these conclusions only rely on a single species-rich marine dataset, similar applications to other case studies will be needed to confirm these findings.