213

Wrong but Useful? Identifiability Regimes of Species Distribution and Abundance Models Under Model Mis-specification

Many ecological questions require hard-won data, and resource constraints often push us to do more with less. There have been continued debates about which models can provide trustworthy information about species prevalence, occupancy, detection, and abundance, especially with the inevitability of model mis-specification. Competing findings about the identifiability of parameters of interest beg the question: why should we care if another ecologist's model isn't identifiable when ours is?

We reconcile these competing identifiability results by bringing terminology from the econometrics literature about a hierarchy of identifiability to the ecology community. Different levels of identifiability exist depending on the parametric assumptions imposed on the model. The relevant, and strongest, form of identifiability, non-parametric identifiability, exists when a model could, in theory, be estimated without parametric assumptions like common link functions that are at the center of many of the debates in the current ecology literature.

We also propose a way to investigate whether this stronger form of non-parametric identifiability holds in targeted parts of a species distribution or abundance model. We relax certain parametric assumptions, like the linearity of the mean response, approximate a non-parametric relationship with a flexible, unpenalized spline, and fit this to simulated data with increasing sample sizes to assess the potential to fit a non-parametric model. A successful approximation shows that the data is information rich enough to robustly inform quantities of interest.

We find that when a property of interest lacks non-parametric identifiability, the "best approximating" model is not "close" to the truth in the way that we might expect, increasing the flexibility within a mis-specified parametric model does not get us closer to the truth, and mis-specification of the model for the distribution of both observable and unobservable quantities is not diagnosable.

In this talk, I will argue that a useful approximation to the truth can only be found when the stronger non-parametric identifiability holds in key parts of the model. The consequence of continuing to rely on parametric assumptions is identifiability by fiat where the data cannot distinguish alternative models and inference about important ecological quantities is put into serious question.