How to build (and compare) better roads to generality

shane.blowes@idiv.de

iDiv
Martin-Luther Universität Halle-Wittenberg
Funded by the European Union
European Research Council
Established by the European Commission

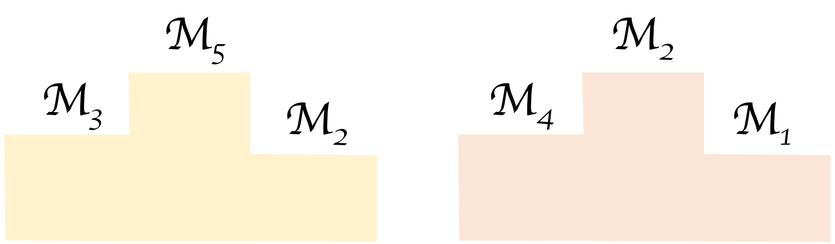
Background

Quantitative evidence synthesis, via meta-analysis and analyses of purpose-built data compilations, represents a prominent road to ***generality*** for many scientific fields[1].

The most common statistical models used in evidence synthesis allow effects to varying among studies, but often assume unexplained variation comes from a single (statistical) population (i.e., homoscedasticity). Consequences, and opportunities subsequent to relaxing this assumption are largely unexplored.

Model comparisons in evidence synthesis typically use the proportion of variation explained (R^2) or conditional predictions (e.g., AIC). However, ***transferability***, i.e., knowledge from one question or system that is informative about another[2], might be better assessed by comparing the ability of models to make predictions outside of the data they were parameterised with.

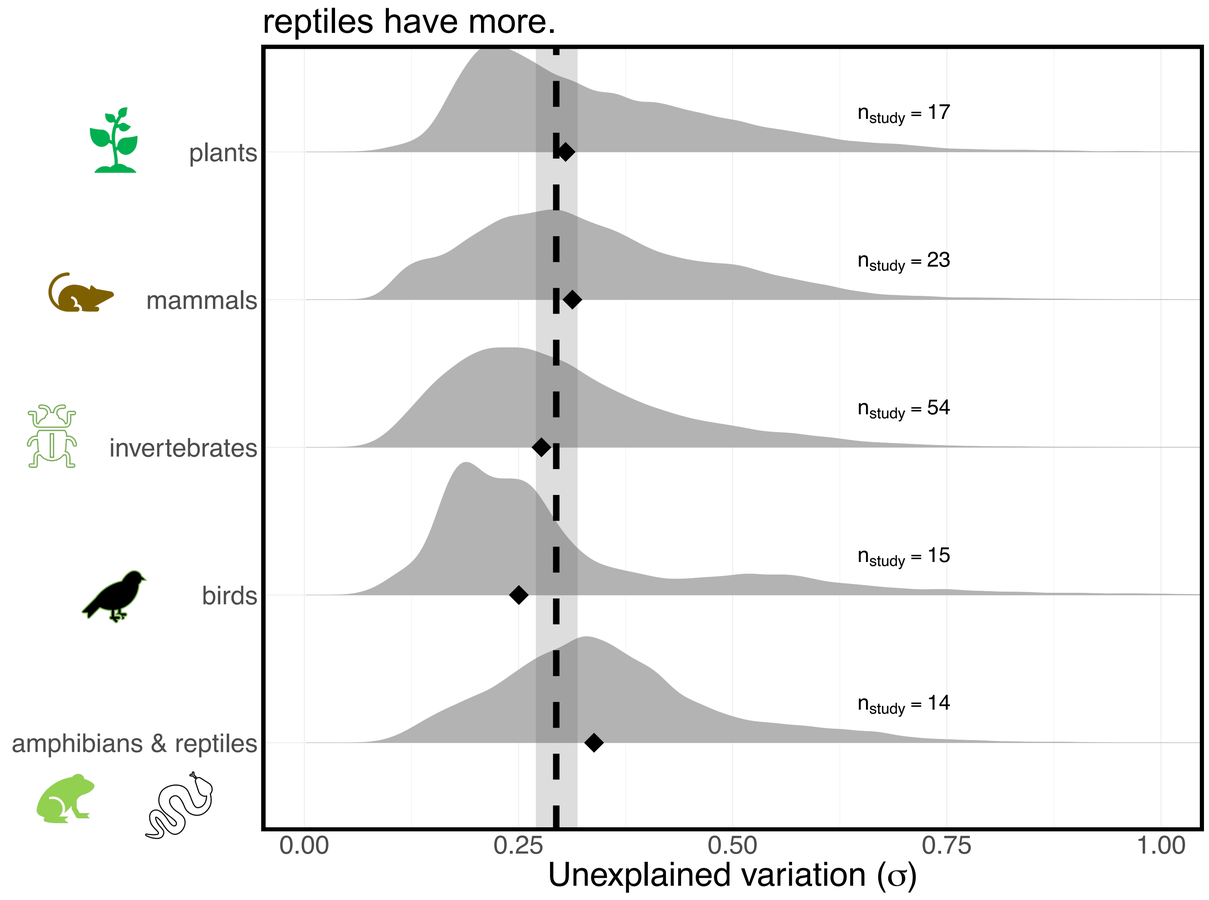
References
[1]Gurevitch *et al.* (2018) *Nature*
[2]Fox (2019) *Philosophical Topics*
[3]Chase *et al.* (2020) *Nature*
[4]Hastie *et al.* (2009) *The Elements of Statistical Learning*
[5]Yates *et al.* (2022) *Ecological Monographs*



Quantifying known unknowns

Statistical models that include predictors for unexplained variation can identify directions for future empirical and theoretical work. Quantifying systematic residual variation should also help describe known limits to transferability.

For example, unexplained variation in the response of species richness to patch size[3] varies among taxon groups. Birds have less residual variation than average, whereas amphibians and reptiles have more.

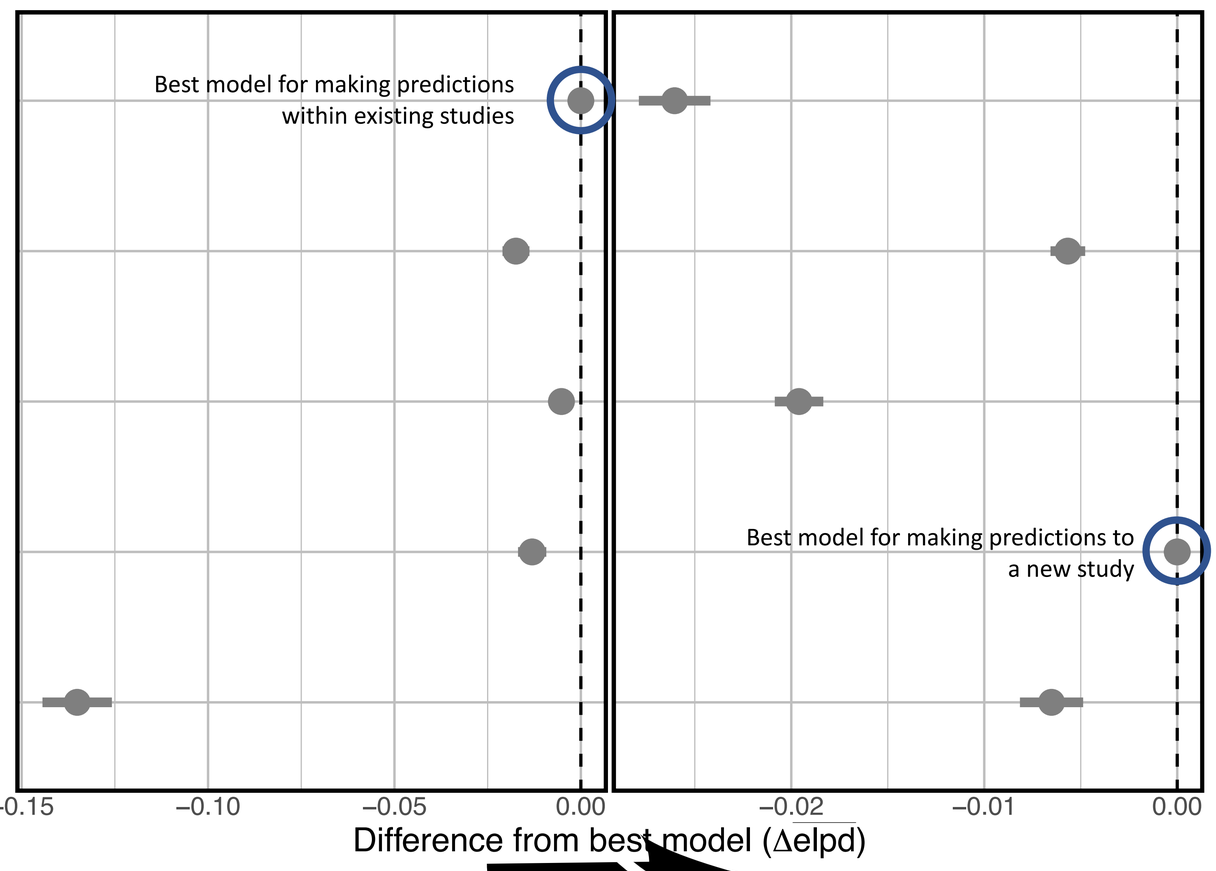


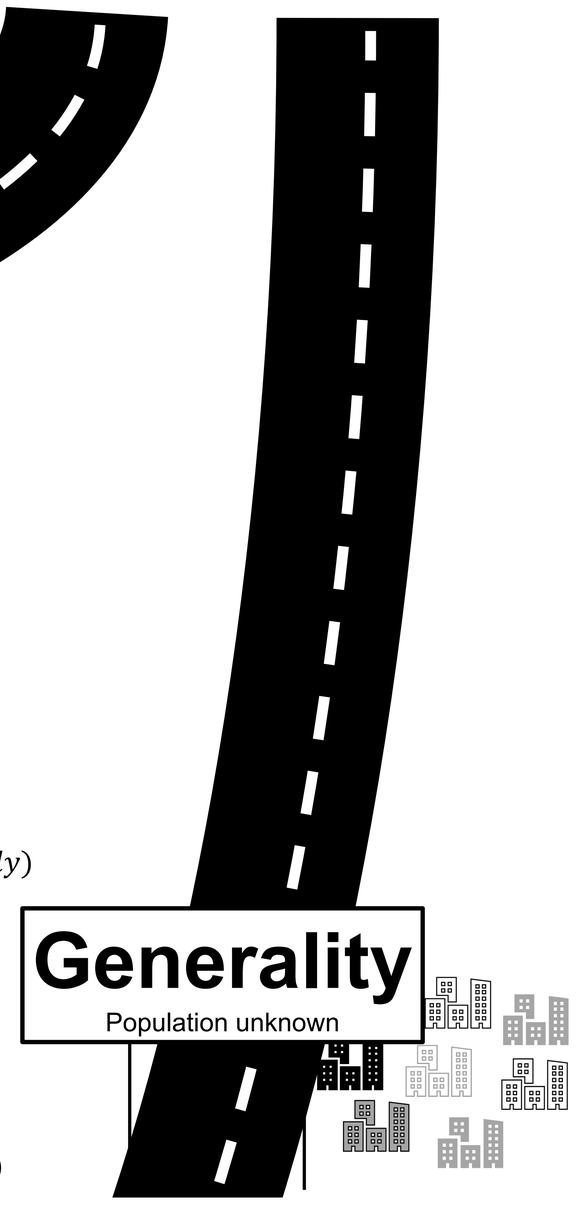
Cross validation for model comparison

Cross validation uses data splitting techniques to test model predictive performance. Models are fit to a 'training' data set and assessed on their ability to predict the 'test' data set[4].

Here, I evaluate two different types of predictions[5]: (i) conditional predictions to new data within existing studies; and, (ii) marginal predictions to data from a new study.

These different types of predictions can require different models. Fragmentation effects on species richness within studies are best predicted with a model where residuals are a function of patch size ($\mathcal{M}_5$). In contrast, predictions to new studies are best made using a simpler model with study-level variation in residuals ($\mathcal{M}_2$).





Generality in quantitative evidence synthesis is most often discussed in terms of central tendencies. The role of heterogeneity for advancing understanding is not as well developed.

Explicit models for potential predictors of unexplained variation can identify directions for future research and limits to transferability.

Cross validation can provide stronger, more flexible tests when assessing the generality and transferability goals of models used in quantitative evidence synthesis.

Acknowledgements
Thanks to members of the Biodiversity Synthesis Group for productive discussions and feedback at various stages of this work. Thanks to C. Krause, and the UFZ administrative and support staff of the High-Performance Computing Cluster EVE, a joint effort of the Helmholtz Centre for Environmental Research (UFZ) and iDiv.