

Entropy-2988663 Multimodel Approaches Are Not the Best Way to Understand Multifactorial Systems – Bolker

An interesting paper with some useful insights. I share a few thoughts and critiques below.

Lines 11-12: Perhaps some rewording required here – (1) aren't all models artificial? Are you trying to say models of little interest?

Lines 13-16: Isn't all shrinkage going to be dependent on sample size relative to number of parameters in the model? Larger sample sizes relative to number of parameters resulting in less shrinkage.

Lines 31-44: It seems as if you are equating multiple processes with multiple variables in a model. Do your arguments still apply in a situation where multiple variables are used to characterize one encompassing process? Or at least where there is not a one-one mapping of processes and variables?

Line 66: Insert word "it", "unequivocally that it has".

Lines 72-77: One issue with only fitting the full model with all factors simultaneously even with very large sample sizes, is that the magnitude of the parameter estimates for each factor (variable) will reflect the multicollinearity among all the factors (variables) which alters their interpretation compared to a model with no collinearity among the factors (variables). See Cade (2015). There almost always is substantial multicollinearity among factors (predictor variables) in ecological models. One approach to help understand this better might be to not only consider the full model with all factors that has parameter estimates that reflect multicollinearity among factors (predictor variables), but to consider single factor (predictor) models where there is no collinearity possible. This could be used as a device to help understand how much rates of change associated with factor parameters have been altered by multicollinearity with other factors in the full model. It is easy to forget that regression coefficients estimated in a multiple regression model only provide an interpretation for the part of a factor (predictor variable) that is not linearly related to the other factors (predictor variables) as Cade (2015) points out. If we really want to know about changes in outcomes associated with the linearly correlated parts of multiple factors, then it seems like something else must be done. It seems to me that the crucial exploration of models with multiple factors that have some degree of collinearity (like precipitation and temperature in the example on lines 104-105), is to understand changes in outcomes with changes in factors that occur simultaneously. This will usually require more than just simple interpretation of individual parameter estimates in a multifactorial model when there is some degree of multicollinearity among the factors.

Line 88: Seems like Cade (2015) critique about MMA of coefficients is principally about issues associated with collinear predictor variables. The critique of Walker (2017) relies on fraught causal interpretations of regression models that is inconsistent with conventional statistical theory associated with regression models. I would eliminate this citation.

Lines 89-90: Seems like this needs to be reorganized a bit. The issue with variable collinearity (multicollinearity) is directly related to issues with model averaging estimates of predictor parameters (regression coefficients).

Lines 90-91: Again, Cade (2015) critique also applies to problems with summing model weights to assess relative importance of predictors.

Line 109: We could fit all five of these models and NOT average their parameters.

Line 149-150: There are several built in approaches to obtaining shrinkage estimates for quantile regression. The lasso is an option in Roger Koenker's quantreg package and other shrinkage type estimators are available in Fasiolo's qqam package.

Lines 155-161: I think I recall reading somewhere that with lasso shrinkage estimates that it may be appropriate to reestimate a model that excludes all the parameters that were shrunk to zero to obtain better standard errors. I would guess this might have something to do with correctly reflecting the multicollinearity among predictors. Can you offer any comments or insights on this model reestimation approach after shrinkage?

Line 199: Disagree. Some things we can imagine have no influence on the outcomes we observe. They may sometimes be weakly correlated with other relevant processes.