

Cajo J.F. ter Braak

Dance with statistics – quick, quick, slow with simple and complex figures

Simple models are the corner-stone of (applied) statistics and specified quickly. But even simple models need not be quick to fit to data and may need complex algorithms. I discuss shortly the Gaussian (constrained ordination) model used in ecological research. A quick simple approximate algorithm is still what is used in practice, despite the availability of a slow complex exact algorithm (that does not always work). On the other hand, mechanistic models may be complex to specify and slow to solve (and have often a hard time to outcompete simple models), but simple Bayesian algorithms can help to fit such models to data with due attention to prior information. I discuss an adaptive form of MCMC that is simple and looks sexy, but is it quick enough?

In Trondheim we are at the hard of approximate Bayesian computing using integrated nested Laplace approximation (INLA). INLA can fit models, from simple to (not too) complex, using a complex algorithm based on simple ideas. I discuss fitting Bayesian P-splines and Generalized Additive models using INLA for histogram smoothing, grouped regression and gams with ecological examples. What can and what cannot (yet) be done quickly?

Brezger, A. and S. Lang. 2006. Generalized structured additive regression based on Bayesian P-splines. *Computational Statistics & Data Analysis* 50:967-991.

Jamil, T. and C. J. F. ter Braak. 2012. Selection properties of type II maximum likelihood (empirical Bayes) in linear models with individual variance components for predictors. *Pattern Recognition Letters* 33:1205-1212.

Jamil, T. and C. J. F. ter Braak. 2013. Generalized linear mixed models can detect unimodal species-environment relationships. *PeerJ* 1:e95.

Martins, T. G., D. Simpson, F. Lindgren, and H. Rue. 2013. Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis* 67:68-83.

Rodríguez-Álvarez, M., D.-J. Lee, T. Kneib, M. Durbán, and P. Eilers. 2014. Fast smoothing parameter separation in multidimensional generalized P-splines: the SAP algorithm.1-17.

ter Braak, C. J. F. 2014. History of canonical correspondence analysis. Pages 61-75 in J. Blasius and M. Greenacre, editors. *Visualization and verbalization of Data*. Chapman and Hall/CRC, London.

ter Braak, C. J. F. and J. A. Vrugt. 2008. Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing* 18:435-446.

van Mourik, S., C. ter Braak, H. Stigter, and J. Molenaar. 2014. Prediction uncertainty assessment of a systems biology model requires a sample of the full probability distribution of its parameters. *PeerJ* 2:e433.

Yee, T. W. 2004. A new technique for maximum-likelihood canonical Gaussian ordination. *Ecological Monographs* 74:685-701.