Bayesian Analysis (2016) **11**, Number 3, pp. 913–919

Comment on Article by Pratola*

Robert B. Gramacy†

I'd like to offer my congratulations to Pratola for engaging in timely study on a high-impact topic, namely the efficient exploration of the space of reasonable partition-based representations of the input–output relationships in data. Tree-based partitioning schemes for regression and classification have proliferated in machine learning, spatial statistics, and computer experiments. However, the Bayesian approach has long been limited by expensive Markov chain Monte Carlo (MCMC) and poor mixing therein.

Pratola said that the MCMC mixing "problem [with trees] has been recognized since such models were established ... and little progress has been made." That's true, but why? MCMC is falling out of fashion a bit, so that may be one explanation. Referees routinely ask authors to remove MCMC details from papers, or at best move them to an appendix, which discourages authors from embarking on the kind of very valuable study that Pratola has taken on in this work. But I think the main reason is that trees are a difficult data structure to deal with. The intersection of talented coders (particularly C data structures), and thoughtful experienced Bayesians, is unfortunately quite small. Not many people are qualified for the job.

My aim over the next several pages is to emphasize, primarily through a series of worked-code illustrations, the value of the contribution Pratola has made. Pratola has provided many of his own illustrations within the Bayesian Additive Regression Tree (BART, Chipman et al., 2010) framework, involving sums of trees, whereas mine will complement those by looking at single-tree models. Following that, I will mention a small potential downside, which I think could be addressed although it may involve a substantial undertaking. Finally, I will conclude with some comments on tree priors, a topic which has been similarly overlooked in the almost two decades since the first swarm of Bayesian tree methods arrived on the scene.

1 An illustration

Pratola talked about rotations, extending an idea from my PhD work (Gramacy, 2005). Whereas my version of rotations worked only on adjacent splits on identical input variables, Pratola's are far more general. Here my aim is to illustrate the value of rotations, and for ease of visualization I shall limit myself to a simple 1-dimensional regression problem.

The data generating mechanism is given by the R code below. This data was used to illustrate treed Gaussian processes in the original methods paper (Gramacy and Lee, 2008) and in the software paper (Gramacy, 2007) for the `tgp` package in R. It is part sinusoid and part linear; a visual will be provided shortly.

*Main article DOI: 10.1214/16-BA999.
†Department of Statistics, Virginia Tech, Blacksburg, VA, rbg@vt.edu

 DOI: 10.1214/16-BA999A