

International Society for Bayesian Analysis, 9<sup>th</sup> World Meeting,  
Hamilton Island, Australia, 2008.

---

## FULLY NON-PARAMETRIC BAYESIAN ENSEMBLE MODELLING

Hugh A. Chipman<sup>1</sup>, Edward I. George<sup>2</sup> and Robert E. McCulloch<sup>3\*</sup>

<sup>1</sup> Acadia University, Nova Scotia, Canada

<sup>2</sup> University of Pennsylvania, Philadelphia, USA

<sup>3</sup> University of Chicago, Chicago, USA

\* `robert.mcculloch@chicagogsb.edu`

Suppose we would like to learn the relationship between  $y$  and a high dimensional vector  $x$  based on a limited number of observations. In “BART: Bayesian Additive Regression Trees” (2006), Chipman, George and McCulloch develop a fully Bayesian approach for discovering and drawing inference about an unknown function  $f$  based only on assuming  $y = f(x) + \epsilon$  with iid normal errors. In the spirit of “ensemble models”, BART approximates  $f$  by a sum of many simple regression tree models, each of which are kept small with a strong regularization prior. In terms of out-of-sample prediction, BART’s performance compares favorably with competing methods. Posterior evaluation by a well-mixing MCMC algorithm allows for the natural Bayesian quantification of uncertainty about  $f$ . Further, the modular nature of BART facilitates its embedding within larger hierarchical models (for example, see Zhang, Shih and Mueller 2006).

In this work, we further extend the flexibility of the BART approach by relaxing the simple iid normal error specification and replacing it with a Dirichlet process model for the errors. Various specification and prior choices are explored. The costs as well as the benefits of this more flexible approach are illustrated.