International Society for Bayesian Analysis, 9^{th} World Meeting,
Hamilton Island, Australia, 2008.

Bayesian Synthesis

Q. Yu[1], S.N. MacEachern[2] and M. Peruggia[2⋆]

[1] Lousiana State University, New Orleans, USA
[2] The Ohio State University, Columbus, USA
⋆ peruggia@stat.osu.edu

Bayesian model averaging enables one to combine the disparate predictions of a number of models in a coherent fashion, leading to superior predictive performance. The improvement in performance arises from averaging models that make different predictions. In this work, we tap into perhaps the biggest driver of different predictions—different analysts—in order to gain the full benefits of model averaging. In a standard implementation of our method, several data analysts work independently on portions of a data set, eliciting separate models which are eventually updated and combined through Bayesian synthesis. The methodology helps to alleviate concerns about the sizeable gap between the foundational underpinnings of the Bayesian paradigm and the practice of Bayesian statistics.

We provide theoretical results that characterize general conditions under which data-splitting results in improved estimation which, in turn, carries over to improved prediction. These results suggest general principles of good modeling practice. In experimental work we show that the method has predictive performance superior to that of many automatic modeling techniques, including AIC, BIC, Smoothing Splines, CART, Bagged CART, Bayes CART, BMA, BART and LARS. Compared to competing modeling methods, the data-splitting approach 1) exhibits superior predictive performance for real data sets and simulations; 2) makes more efficient use of human knowledge; 3) selects sparser models with better explanatory ability and 4) avoids multiple uses of the data in the Bayesian framework.