



**MUMS Program Opening Workshop  
August 20-24, 2018**

**SPEAKER TITLES/ABSTRACTS**

**Edward George**

Wharton, University of Pennsylvania

**“Quantifying Nonparametric Modeling Uncertainty with BART”**

For the canonical regression setup where one wants to discover the relationship between  $Y$  and a  $p$ -dimensional vector  $x$ , BART (Bayesian Additive Regression Trees) approximates the conditional mean  $E[Y|x]$  with a sum of regression trees model, where each tree is constrained by a regularization prior to be a weak learner. Fitting and inference are accomplished via a scalable iterative Bayesian backfitting MCMC algorithm that generates samples from a posterior. Effectively, BART is a nonparametric Bayesian regression approach which uses dimensionally adaptive random basis elements. Motivated by ensemble methods in general, and boosting algorithms in particular, BART is defined by a statistical model: a prior and a likelihood. This approach enables full posterior inference including point and interval estimates of the unknown regression function as well as the marginal effects of potential predictors. By keeping track of predictor inclusion frequencies, BART can also be used for model-free variable selection. To further illustrate the modeling flexibility of BART, we introduce two elaborations, MBART and HBART. Exploiting the potential monotonicity of  $E[Y|x]$  in components of  $x$ , MBART incorporates such monotonicity with a multivariate basis of monotone trees. To allow for the possibility of heteroscedasticity, HBART incorporates an additional product of regression trees model component for the conditional