

the broad range of people to involve in this work when eliciting patients or caregivers considerations with respect to effectiveness and efficacy, safety, means of therapy implementation, duration of use and effect, other characteristics that may inform assessments, and patients' willingness (and unwillingness) to tolerate therapies. We will discuss mixed methods approaches to integrate qualitative and quantitative patient input and experience about the relevance of clinical endpoints to describe how patients feel, function, and survive.

**EMAIL:** [laura.johnson@fda.hhs.gov](mailto:laura.johnson@fda.hhs.gov)

## 82. Variable Selection in Mixture of Regression

### ► A Hierarchical Hidden Markov Model Framework for Predicting Differential Methylation Profiles

Mayetri Gupta\*, University of Glasgow  
Tushar Ghosh, University of Glasgow  
Neil Robertson, University of Glasgow  
Peter Adams, University of Glasgow

DNA Methylation is an important epigenetic mechanism for controlling gene expression, silencing or genomic imprinting in living cells. High-throughput sequencing methods to study DNA methylation include sequencing of sodium bisulfite-treated DNA (BS-Seq). Several software tools for pre-processing and alignment of BS-seq data analysis have been published—however, methods for analyzing the methylation profiles, and detecting differentially methylated regions (DMRs) are relatively primitive, not taking specific dependence features of the data into account. Most current methods to detect DMRs rely on smoothing techniques that have high false discovery rates or are biased by experimental artefacts. We propose a novel method for predicting DMRs within a hierarchical Bayesian hidden Markov model framework, incorporating several levels of dependence between observations. Our method efficiently deals with nuisance parameters without leading to overwhelming analytical complexity and allows a principled way of building prior distributions based on partially known information, improving estimation of novel features. Our methods are illustrated through a study on human aging.

**EMAIL:** [mayetri.gupta@glasgow.ac.uk](mailto:mayetri.gupta@glasgow.ac.uk)

### ► Mixture of Generalized Linear Regression Models for Species-Rich Ecosystems

Frederic Mortier\*, CIRAD, France

Understanding how climate change could impact population dynamics is of primary importance for species conservation. In species-rich ecosystems with many rare species, the small population sizes hinder a good fit of species-specific models. We propose a mixture of regression models with variable selection allowing the simultaneous clustering of species into groups according to vital rate information (recruitment, growth, and mortality) and the identification of group-specific explicative environmental variables. We illustrate the effectiveness of the method on data from a tropical rain forest in the Central African Republic and demonstrate the accuracy of the model in successfully reproducing stand dynamics and classifying tree species into well-differentiated groups with clear ecological interpretations.

**EMAIL:** [fmortier@cirad.fr](mailto:fmortier@cirad.fr)

### ► Mixture of Regression/Classification Trees

Emanuele Mazzola\*, Dana-Farber Cancer Institute  
Mahlet Tadesse, Georgetown University  
Giovanni Parmigiani, Dana-Farber Cancer Institute

When analyzing high-dimensional datasets, it is often relevant to uncover homogeneous subgroups of individuals (cluster structures), and identify subgroup-specific variables associated with the responses. Assuming a latent or unobserved clustering structure, we propose a general method based on mixtures of regression or classification trees to identify relevant predictors associated to normally distributed (regression tree) or binary (classification tree) outcomes. We formulate the model in a Bayesian framework, allowing identification of homogeneous subgroups while searching for covariates that may have nonlinear association with the outcome. This is accomplished by alternating between the update of cluster allocations and the update of tree structures within the clusters at each MCMC iteration. We will illustrate the performance of the method using simulated and real datasets.

**EMAIL:** [mazzola@jimmy.harvard.edu](mailto:mazzola@jimmy.harvard.edu)