# Bayesian ultrahigh dimensional variable selection for mixed-type multivariate responses and Bayesian regression models for spatiotemporal data

Hsin-Hsiung Bill Huang[1]

[1]*Department of Statistics and Data Science, University of Central Florida, U.S.A.*

## Abstract

Inspired by our investigation on spatiotemporal data analysis for the NSF ATD challenges, we've investigated Bayesian clustering, variable selection for mixed-type multivariate responses and Gaussian process priors for spatiotemporal data. The proposed Bayesian approaches effectively and efficiently fit high-dimensional data with spatial and temporal features. We further propose a two-stage Gibbs sampler which leads a consistent estimator with a much faster posterior contraction rate than a one-step Gibbs sampler. For Bayesian ultrahigh dimensional variable selection, we have developed Bayesian sparse multivariate regression for mixed responses (BS-MRMR) with shrinkage priors model for mixed-type response generalized linear models. We consider a latent multivariate linear regression model associated with the observable mixed-type response vector through its link function. Under our proposed BS-MRMR model, multiple responses belonging to the exponential family are simultaneously modeled and mixed-type responses are allowed. We show that the MBSP-GLM model achieves posterior consistency and quantifies the posterior contraction rate. Additionally, we incorporate Gaussian processes into zero-inflated negative binomial regression. To conquer the computation bottleneck that GPs may suffer when the sample size is large, we adopt the nearest-neighbor GP approach that approximates the covariance matrix using local experts. We provide simulation studies and real-world gene data examples.