BM&Ms: Bayesian Modular & Multiscale Regression

- Michele Peruzzi *Università Bocconi and Duke University*
- David Dunson *Duke University*

Introduction

Data at very high resolutions may be poorly manageable and/or interpretable. But low-resolution data possibly not as informative. ***BM&Ms*** work when the goals are:

1. using data at pre-specified resolutions in the same model →

2. using high-resolution information only if/when necessary →

Multiscale regression model

X_j is an *n x* p_j matrix collecting data at scale *j*, *y* an *n x 1* vector of scalar responses. Multiscale regression:

$$y = X_1\theta_1 + \dots + X_K\theta_K + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2 I_n)$$

If we assume $X_j = X_{j+1}L_j$, then we can write

$$\begin{aligned} y &= X_K(\mathcal{L}_1\theta_1 + \dots + \mathcal{L}_{K-1}\theta_{K-1} + \theta_K) + \varepsilon \\ &= X_1\theta_1 + \dots + X_K\theta_K + \varepsilon \\ &= X_K\beta + \varepsilon \end{aligned}$$

i.e. the multiscale regression model decomposes β into resolution contributions. But $\theta_1, \dots, \theta_K$ not identifiable.

Bayesian, Modular

We assign priors on each θ_j, but cut their dependence on higher resolutions via modularity. In linear regression:

Module **j**: Use prior $\theta_j \sim p(\theta_j)$ on model $e_{j-1} = X_j\theta_j + \varepsilon$

Sample $\tilde{\theta}_j \sim \pi_j(\theta_j | e_{j-1}, X_j)$ to get $e_j = e_{j-1} - X_j\tilde{\theta}_j$

The **Modular posterior** collects the modules' posteriors:

$$p_M(\theta_1, \dots, \theta_K | y, X_{1:K}) = \pi_1(\theta_1 | y, X_1) \cdots \pi_K(\theta_K | e_{K-1}, X_K)$$

If *K=2* and Normal priors, posterior is $N(\mu_{1:2}, \Sigma_{1:2})$. Use $Q_1 = \Sigma_2 X_2' X_1$ and define μ_{β_j}, Σ_j $j \in \{1, 2\}$ as posterior params from models of the form $y = X_j\beta_j + \varepsilon_j$ then we get

$$\mu_{1:2} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_{\beta_1} \\ \mu_{\beta_2} - Q_1\mu_{\beta_1} \end{bmatrix}$$

In large samples, this is equivalent to using first-stage residuals as responses for a second-stage model.

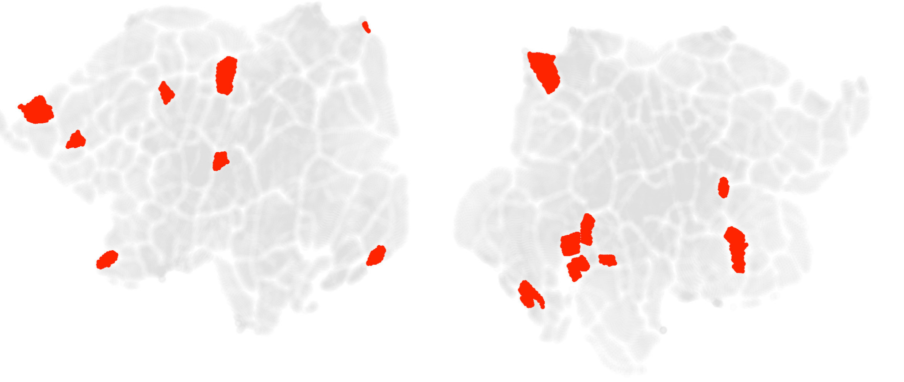
Gender classification with tfMRI data

Measure brain activity of *n=100* subjects on the 333 regions of Gordon (2016). Predictions on remaining 385 subjects. ***BM&Ms*** can be used in two ways:

1. with low-res lobes data alongside higher-res region data

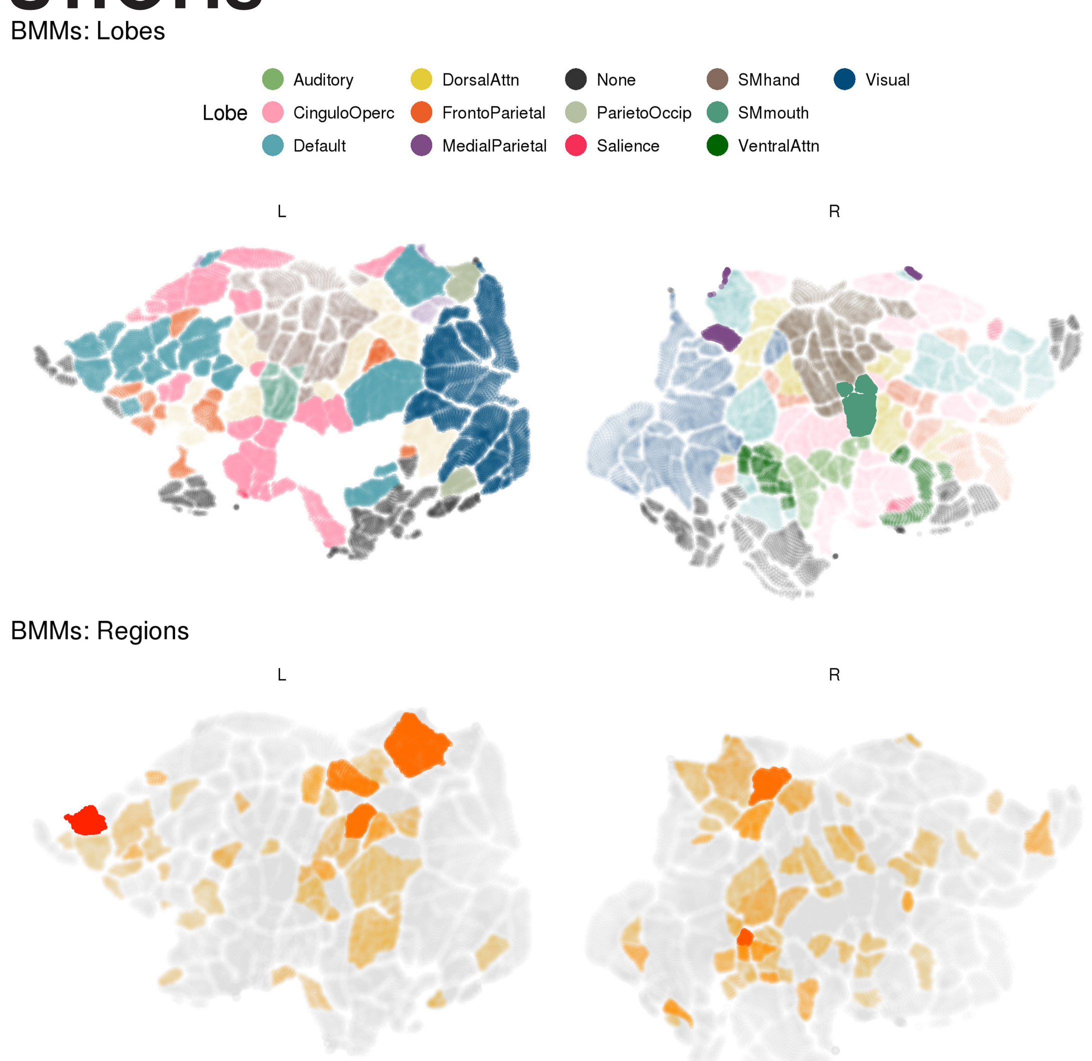
2. for multiscale interpretation using region location info

Quick comparison with Lasso



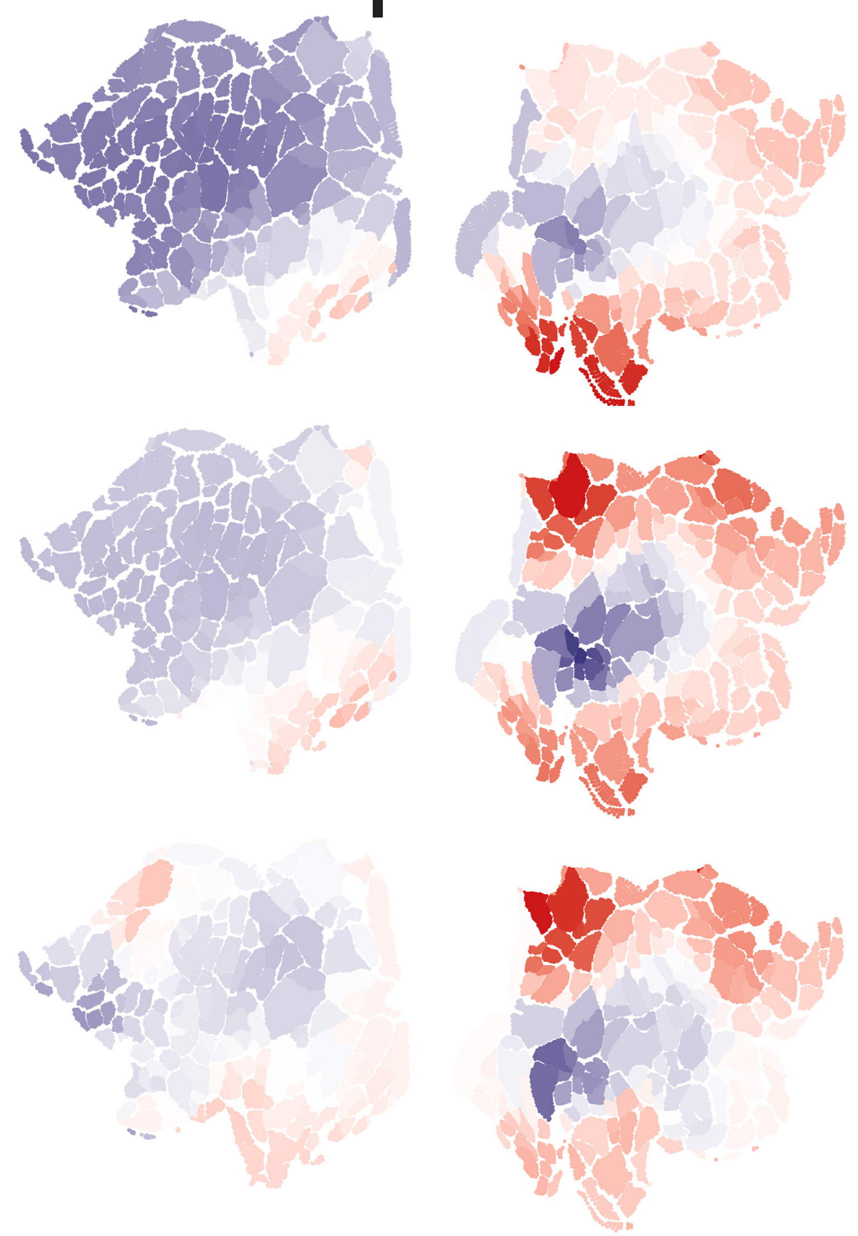
Selected regions in red. Difficult to understand pattern, if any. Average out-of-sample correct classification rate (accuracy): 74.5%. Uncertainty?

Pre-specified resolutions



In this case, matrices X_1 to X_K are pre-specified. We have *K=2*, with X_1 corresponding to brain lobes and X_2 to the 333 regions. Using variable selection priors, lobes alone would average 68% correct classification rate. Netting their effect, ***BM&Ms*** estimate which regions provide *additional* info. Posterior selection probability of lobes (top) and regions (bottom). Stronger colors correspond to more frequently selected lobes and regions. Accuracy 79%. *K=1* and no lobes? Lower sparsity.

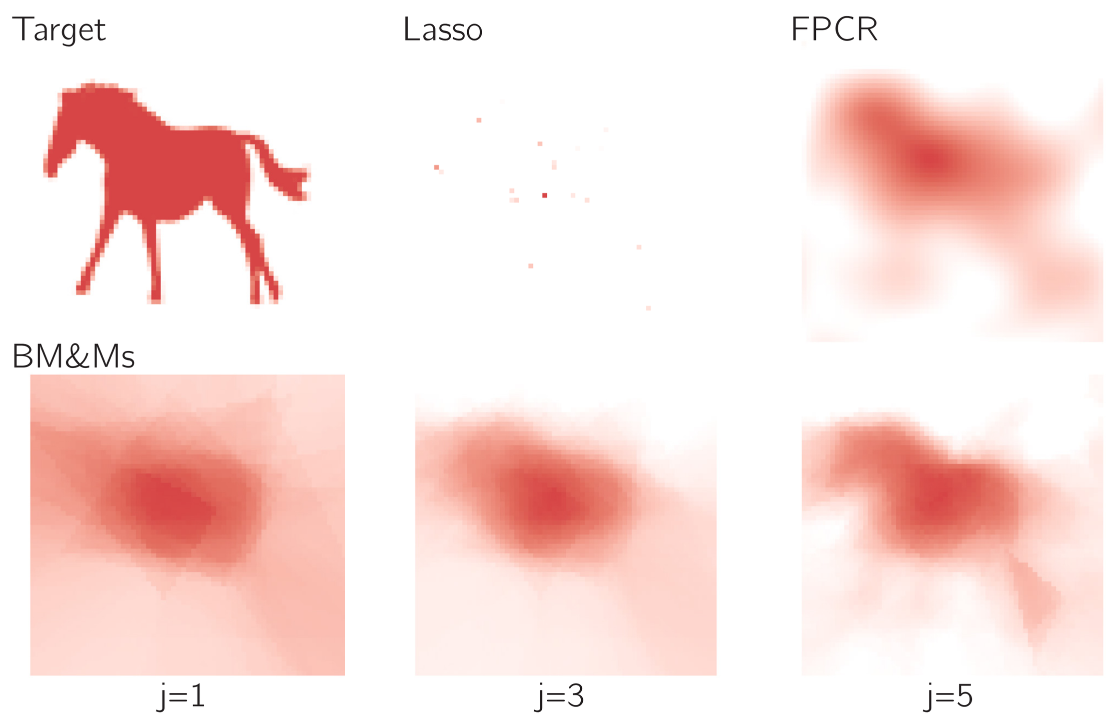
One-scale data, multiscale interpretation



We use regions' centroid locations to create predictor images of size *23x57* for each subject, then run MCMC with a hierarchical Voronoi prior on a multiresolution structure fixing *K=5*. This allows to obtain coarse-to-fine interpretation of the regression coefficients (from top to bottom, estimates for *j=2,3,5*). Can also obtain the *contributions* of each resolution level (not shown here). ***BM&Ms*** in this context average 81.5% accuracy out-of-sample. Possible extension to scalar-on-tensor regression using analogous algorithm.

Simulated data

For each of n=200 simulated observations, we generate a 2D predictor array and a continuous response, using the model $y_i = vec(X_i)vec(B) + \varepsilon_i$ with *B* corresponding to the *64x64* target image (top left). 3 stages of the ***BM&Ms*** model-averaged multiscale decomposition (using a total *of K=5*) are shown (bottom row) along with the single-scale estimates of Lasso (top center) and FPCR (top right).



Parameters for Lasso and FPCR were chosen via cross-validation.

Mean Square Error:

Lasso: 4.163

FPCR: 1.083

BM&Ms: 0.899