Bayesian Bootstrap Uncertainty Quantification for Spatial Lesion Regression Modelling

Anna Menacher[*1], Thomas E. Nichols[2], Chris Holmes[1], and Habib Ganjgahi[1,2]

[1] Department of Statistics, University of Oxford, [2] Big Data Institute, University of Oxford

Motivation

Goal: Perform 3D brain lesion mapping by identifying spatial locations in the brain where lesion incidence is associated with different covariates (e.g. age, disease duration and severity) to enable a better understanding of the aging brain and multiple sclerosis.

Existing lesion mapping approaches:

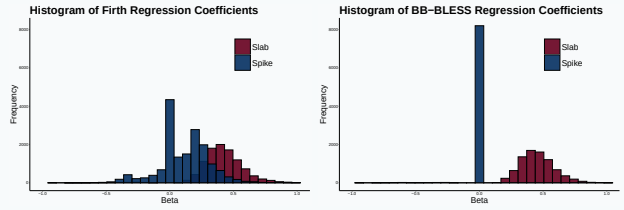
- **Mass-univariate method:** Firth Regression [2]
  - → Fit a model at every voxel location independently.
  - → Ignore any form of spatial dependence.
- **Bayesian spatial method:** BSGLMM [3]
  - → Accounts for shared information between neighboring voxels.
  - → High computational cost of MCMC methods.

BLESS

Bayesian **L**esion **E**stimation with **S**tructured **S**pike-and-Slab (**BLESS**) Prior

- Handles many covariates using **shrinkage priors** to do **variable selection**.
- **Scales to thousands of subjects** and accounts for **spatial dependency** in 3D lesion mapping studies containing over 50,000 voxel locations.
- Relies on **optimization** rather than MCMC for faster parameter estimation and inference.
- Offers **uncertainty estimates** of any spatial statistics, such as **cluster size**, providing credible intervals of cluster size and measures of reliability of cluster occurrence.

Spike-and-Slab Prior



- Firth finds many negligible effects that overlap with a set of large ones, motivating the spike-and-slab prior.
- Firth regression coefficients (top left) suggest a mixture of two normal distributions with different variances.
- BB-BLESS coefficients (top right) show how the spike-and-slab prior pulls negligible effects to 0 while leaving large coefficients unaffected in a slab distribution.

Model

(1) Probit Model

$$\begin{cases} [y_i(s_j)|p_i(s_j)] \sim \text{Bernoulli}[p_i(s_j)] \\ \Phi^{-1}\{\mathbb{E}[y_i(s_j)|p_i(s_j)]\} = \eta_i(s_j) = \boldsymbol{x}_i^\top \boldsymbol{\beta}(s_j) + \beta_0(s_j) \end{cases}$$

(2) Latent Model

$$\begin{cases} \Pr[y_i(s_j)|z_i(s_j)] = \begin{cases} 1, & z_i(s_j) > 0, \\ 0, & z_i(s_j) \le 0, \end{cases} \\ z_i(s_j) \sim \mathcal{N}(\boldsymbol{x}_i^\top \boldsymbol{\beta}(s_j) + \beta_0(s_j),\ 1) \end{cases}$$

(3) Spike-and-Slab Prior

$$\begin{cases} \beta_p(s_j) \sim \mathcal{N}(0,\ \nu_0(1-\gamma_p(s_j)) + \nu_1\gamma_p(s_j)) \\ \gamma_p(s_j) \sim \text{Bernoulli}(\sigma(\theta_p(s_j))) \end{cases}$$

(4) MCAR Prior

$$\begin{cases} [\boldsymbol{\theta}(s_j) \mid \boldsymbol{\theta}(-s_j), \boldsymbol{\Sigma}] \sim \mathcal{N}\left(\frac{\sum_{s_r \in \partial s_j} \boldsymbol{\theta}(s_r)}{n(s_j)}, \frac{\boldsymbol{\Sigma}}{n(s_j)}\right) \\ \boldsymbol{\Sigma}^{-1} \sim \text{Wishart}(\nu,\ \boldsymbol{I}) \end{cases}$$

(1) Probit regression modeling lesion presence $y_i(s_j)$ for every subject i at every voxel s_j.

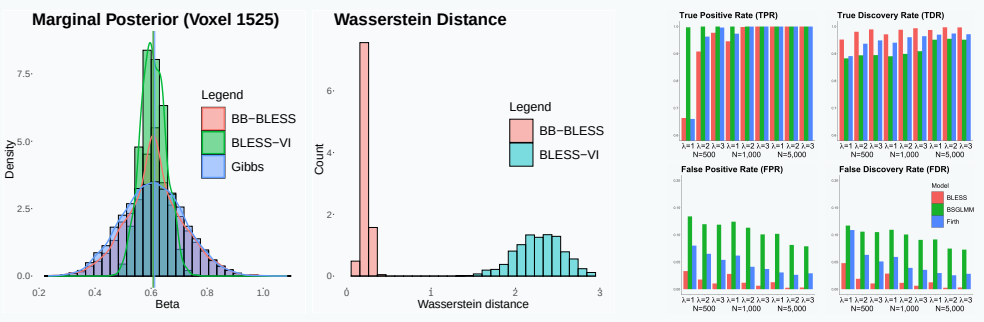
(2) Latent model using data augmentation approach assuming that the binary outcomes $y_i(s_j)$ have an underlying normal regression structure on continuous latents $\boldsymbol{z}_i(s_j)$.

(3) Spatially-varying, continuous version of the spike-and-slab prior on the parameters $\boldsymbol{\beta}(s_j)$ in the form of a mixture of normal distributions where $0 < \nu_0 < \nu_1$.

(4) Sparsity parameter $\boldsymbol{\theta}(s_j)$ introduces the spatial structure within the probability of inclusion $\sigma(\boldsymbol{\theta}(s_j))$, where $\sigma(\cdot)$ is the logistic function, with a multivariate conditional autoregressive (MCAR) prior.

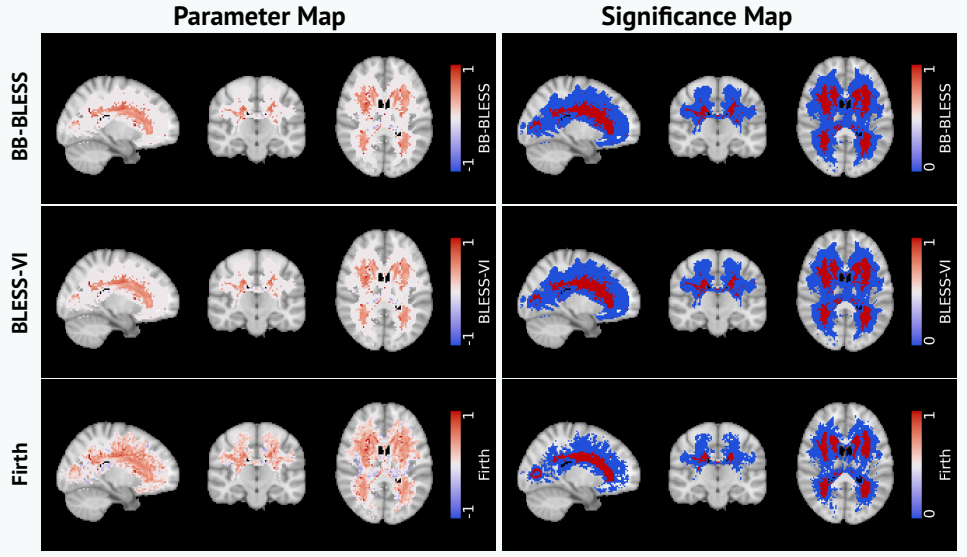
→ For inference, we use approximate posterior sampling based on Bayesian bootstrap (BB-BLESS) [4] which consists of parallelizable variational optimizations (BLESS-VI) [5].

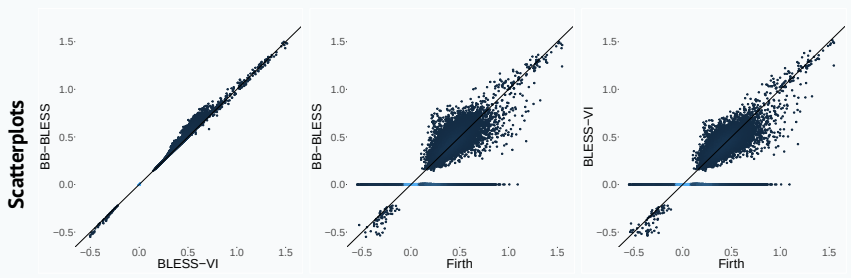
Simulation Study



- Evaluation of simulation studies with varying sample sizes N, base rate intensities λ, and sizes of effect which compare the performance of BLESS to BSGLMM and Firth.
- BLESS outperforms BSGLMM and Firth with a lower number of false positives. Sensitivity and specificity for BLESS remains high for all sample sizes and base rate intensities.
- The marginal posterior of a single voxel approximated via BB-BLESS matches the Gibbs sampled posterior (gold standard MCMC method) more closely than BLESS-VI.
- Wasserstein distance summarizes the performance of BB-BLESS across all voxels in an image where a lower distance equals a higher alignment to the Gibbs posterior.

UK Biobank Application





Data & Analysis:

- 2,000 subjects from the UK Biobank with lesion masks generated via segmentation tool BIANCA. (Model 3D and scalable, and has been fit on ∼40,000 subjects.)
- 3D regression model of lesion incidence on covariates, age, sex, headsize scaling factor, and age by sex [1], with an image mask of 54,728 voxels.

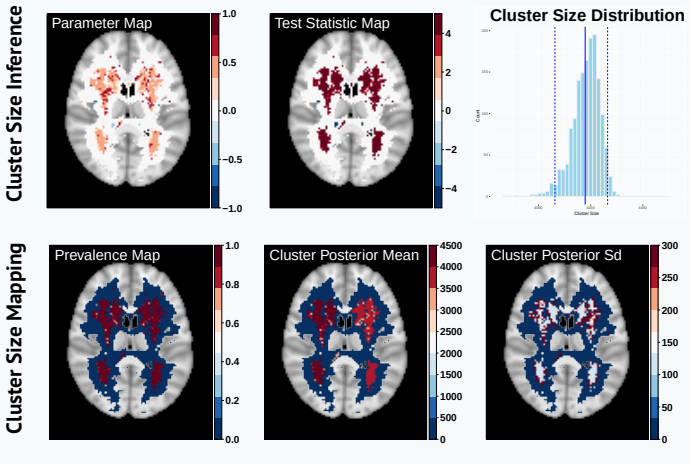
Results:

- Parameter maps: Negligible age coefficients are shrunk to nearly zero via BB-BLESS and BLESS-VI compared to Firth regression. The point estimates for BB-BLESS and BLESS-VI are almost identical.
- Thresholded significance maps: Inference results for BLESS-VI (thresholded based on posterior inclusion probabilities) and BB-BLESS (thresholded based on test statistics) are similar with 8,385 and 8,350 voxels detected. For Firth, only 6,278 voxel pass the FDR adjusted threshold (Benjamini-Hochberg FDR threshold at a 5% level.)
- Scatterplots: BB-BLESS and BLESS-VI age coefficients show the regularization effect with small values being pulled towards zero relative to Firth regression.

Conclusions:

- Key advantages of BLESS are the explicit modelling of the spatial dependence structure in the lesion data, the identification of active predictors based on a fixed thresholding rule via BLESS-VI and the equivalent detection of effects via test statistics via BB-BLESS.

Cluster Size Based Imaging Statistics



Cluster Size Inference:

- Estimate bootstrap samples of parameters, calculate test statistic maps and acquire cluster size maps (via cluster defining threshold of 2.3).
- The histogram shows the cluster size distribution of the largest detected cluster and the credible interval captures the uncertainty of cluster size.

Cluster Size Mapping:

- Prevalence map is determined by cluster occurrence across resampled bootstrap maps where locations exceeding a probability of 50% indicate a reliably large effect. Both clusters have reliably large effects with values close to 1.
- The posterior mean and standard deviation map of cluster size are acquired by thresholding cluster size quantities via the prevalence map.

Contact Details

Please scan the QR code to find out more about BLESS!

References

[1] F. Alfaro-Almagro et al. Confound Modelling in UK Biobank Brain Imaging. *NeuroImage*, 224, 2021.

[2] D. Firth. 'Bias Reduction of Maximum Likelihood Estimates'. *Biometrika*, 80(1):27–38, 1993.

[3] T. Ge et al. Analysis of Multiple Sclerosis Lesions via Spatially Varying Coefficients. *Ann. Appl. Stat.*, 8(2):1095–1118, 06 2014.

[4] L. Nie and V. Ročková. Bayesian Bootstrap Spike-and-slab LASSO. *JASA*, 0(0):1–16, 2022.

[5] V. Ročková et al. EMVS: The EM Approach to Bayesian Variable Selection. *JASA*, 109(506):828–846, 2014.