

JASA A&CS Reproducibility Initiative - Author Contributions Checklist Form

The purpose of the Author Contributions Checklist (ACC) Form is to document the code and data supporting a manuscript, and describe how to reproduce its main results.

As of Sept. 1, 2016, the ACC Form must be included with all new submissions to JASA A&CS.

This document is the initial version of the template that will be provided to authors. The JASA Associate Editors for Reproducibility will update this document with more detailed instructions and information about best practices for many of the listed requirements over time.

Data

Abstract (Mandatory)

Study subjects included 20 randomly sampled participants from the Human Connectome Project (HCP) 500-subject release. For each subject, one run of the motor task and gambling task studies were each analyzed. We utilized the fMRI data included in the HCP data release that was minimally preprocessed and projected to the cortical surface using the HCP fMRISurface pipeline. Several additional files were also utilized, listed below.

Availability (Mandatory)

The HCP dataset is publicly available.

Description (Mandatory if data available)

The HCP dataset is available for download or ordering at <http://www.humanconnectome.org/>. For each subject, a variety of imaging and demographic data are available. For this analysis, the following files in each subject's **MNINonLinear/Results** directory were utilized:

- For the motor task analysis, within the **tfMRI_MOTOR_RL** directory:
 - Task fMRI data (CIFTI format): **tfMRI_MOTOR_RL_Atlas.dtseries.nii**
 - Task timing (text format): **EVs/[task].txt**, where task = { 'cue', 'lf', 'lh', 'rf', 'rh', 't' } are the tasks performed in the motor task study.
 - Head motion parameters (text format): **Movement_Regressors.txt**
- For the gambling task analysis, within the **tfMRI_GAMBLING_RL** directory:
 - Task fMRI data (CIFTI format): **tfMRI_GAMBLING_RL_Atlas.dtseries.nii**
 - Task timing (text format): **EVs/[task].txt**, where task = { 'loss_event', 'neut_event', and 'win_event' } are the tasks performed in the gambling task study.
 - Head motion parameters (text format): **Movement_Regressors.txt**

Code (fMRI Data Analysis)

Abstract: The code to perform the analysis consists of several files, to be executed in the following order:

1. **BayGLM0.sh**, in which Connectome Workbench bash commands are used to manipulate image files.
2. **BayGLM1.m**, a MATLAB script that creates the task design matrix for each subject and converts CIFTI files to CSV. *Note: The use of MATLAB here is not strictly required, as BayGLM1.m could be rewritten in R using the `cifti` and `gifti` R packages to read in CIFTI and GIFTI-format files.*
3. **BayGLM2.R**, in which several setup steps are performed.
4. **BayGLM3_classical.R** or **BayGLM3_bayesian.R**, the main analysis files in which the classical and Bayesian GLMs are fit, respectively. *Note: Both call the `writeCIFTIs` function defined in **BayGLMfun.R**, which calls MATLAB to create CIFTI-format output files. Currently there is no R package capable of writing out CIFTI files.*

The file **BayGLMfun.R** includes several functions utilized in BayGLM2.R, BayGLM3_classical.R and BayGLM3_bayesian.R.

Description: The code consists of the following bash, MATLAB and R files. Computation times listed below are based on a high-memory cluster node (100GB RAM except for the fully Bayesian group model, which requires approximately 270GB). For more information on memory requirements and computation times, see Appendix E.

To achieve these computation times, the user should obtain a free PARDISO sparse matrix library license at <https://pardiso-project.org/r-inla/>, save the license key in a text file titled `pardiso.lic`, then activate it within the R session by running `inla.setOption(pardiso.license = "/path/to/pardiso.lic")`. The code can also be run without the PARDISO library installed, but computation times for the INLA models will be slower.

1. **BayGLM0.sh** – Create files needed for resampling fMRI data from 32K to 6K resolution (and vice-versa).
Computation time: approximately 30 seconds
2. **BayGLM1.m** – Select subjects, get coordinates of surface vertices, and for each subject:
 - a. Set up design matrix (task & nuisance)
 - b. Smooth surface fMRI data (for classical GLM)
 - c. Resample fMRI data from 32K to 6K resolution (for Bayesian GLM)
 - d. Write smoothed and resampled fMRI data to CSV

Relies on the following **toolboxes** and *functions*:

- **cifti-matlab** (*ft_read_cifti*) – <https://github.com/Washington-University/cifti-matlab>
- **cifti** (*ciftiopen*, **gifti-1.6** toolbox) – <https://github.com/Washington-University/HCPpipelines/tree/master/global/matlab>
- **spm8** – <https://www.fil.ion.ucl.ac.uk/spm/software/spm8/>
- *CanonicalBasisSet.m* – depends on **spm8**

Computation time: approximately 20-25 minutes per subject and task.

3. **BayGLM2.R** – Perform setup steps for spatial analysis, prewhitening, and permutation testing. Relies on the following packages: INLA, Matrix, ggplot2, reshape2

Computation time: approximately 1.5 hours per task

4. **BayGLM3_classical.R** – Prewhiten fMRI data (32K smoothed and unsmoothed); fit subject-level and group-level classical GLM, correct for multiple comparisons with permutation test for FWER control and Benjamini-Yekutieli for FDR control. Relies on the following packages: matrixStats, gifti

Computation time: 4-5 hours per subject (can be parallelized) plus 3-4 hours for the group level model.

5. **BayGLM3_bayesian.R** – Fit Bayesian GLM at subject level; estimate Bayesian group-level model using fully Bayesian, joint and two-level approaches; identify areas of activation using joint PPM approach at group level; resample results from 6K to 32K; write results as CIFTI files for visualization. Rely on the following packages: matrixStats, gifti, INLA, excursions, fields, expm, ggplot2, RColorBrewer, reshape2, MASS, parallel.

Computation time: Approximately 1-1.5 hours per subject per task for subject-level models using the INLA PARDISO library with 4 parallel workers. Subjects can be parallelized to speed computation. Approximately 12-15 hours for group models (10-13 hours for the fully Bayesian group model, 1 hour for the joint group approach, and 1 hour for the two-level approach). Note that to obtain group-level activation estimates, only one approach is needed.

6. **BayGLMfun.R** – Functions utilized in R code files (3-6).

In addition, the Connectome Workbench software package was used to perform surface smoothing and resampling of task fMRI data and to visualize images. This software is publicly available at <http://www.humanconnectome.org/software/get-connectome-workbench>.

Code (Simulation)

Abstract: The code to perform simulation studies 1 (single subject) and 2 (multi-subject) consists of several files, to be executed in the following order. The files **BayGLMfun.R** and **functions_sim.R** include several functions utilized in the R scripts.

1. **BayGLMSim0.m**, in which the fMRI data for both studies is simulated.
2. **BayGLMSim1.R**, the main analysis file for simulation 1.
3. **BayGLMSim2.R**, the main analysis file for simulation 2.

Description: The code consists of the following MATLAB and R files. Computation times listed below are based on a cluster node with 50GB RAM. To achieve these computation times, the user should obtain and activate a free PARDISO sparse matrix library license as described in the previous section. The code can also be run without the PARDISO library installed, but computation times for the INLA models will be slower.

1. **BayGLMSim0.m** – Simulate true activation profiles, areas of activation, and fMRI data for both simulation studies. For simulation 1 (single subject), fMRI data with AR(1) errors are simulated, and a smoothed version of the data is created for the classical GLM. For simulation 2 (multi-subject), fMRI data for 10 subjects with different areas of activation are simulated with independent errors. Relies on the spm12 toolbox (spm8 could also be used), available at <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>.
Computation time: less than 1 minute
2. **BayGLMSim1.R** – For single-subject simulation study, visualizes true activation amplitudes and areas of activation; performs classical GLM on smoothed and unsmoothed data, identifying areas of activation with FWER through a permutation test and FDR correction through Benjamini-Hochberg; performs Bayesian GLM, identifying areas of activation through marginal and joint PPM; and visualizes all activation estimates, areas of activation and ROC curves. Relies on the following packages: INLA, RColorBrewer, fields, matrixStats, pROC and excursions.
Computation time: approximately 1 hour
3. **BayGLMSim2.R** – For multi-subject simulation study, visualizes true subject- and group-level activations; estimates subject-level Bayesian GLMs; estimates group-level Bayesian GLM and identifies areas of activation using each proposed approach (GE model, RE model, FE model, joint approach and two-level approach); visualizes activation estimates and areas of activation. Relies on the following packages: INLA, RColorBrewer, fields, matrixStats, excursions, MASS and parallel.
Computation time: approximately 30 minutes for setup and subject models, followed by group models: 30 minutes for GE model, 5 hours for RE model, 2 hours for FE model, 15 minutes for joint approach, and 6 minutes for two-level approach
4. **BayGLMfun.R** and **functions_sim.R** – Functions utilized in R code files (2 and 3).

Instructions for Use

Reproducibility

fMRI Data analysis: Figures 12-16 and all figures in Appendices C and D can be reproduced by accessing the HCP data for the 20 randomly sampled subjects listed below, downloading the Connectome Workbench, and executing the scripts BayGLM0.sh, BayGLM1.m, BayGLM2.R, BayGLM3_classical.R and BayGLM3_bayesian.R. To create surface visualizations from the final CIFTI output files (ending in .dscalar.nii), the Connectome Workbench wb_view application is required.

To download HCP data, you must register with ConnectomeDB at <https://db.humanconnectome.org>. To perform Bayesian computation for single-subject models and fully Bayesian group models using INLA, a large memory cluster node is recommended. See Appendix E for more information on memory requirements.

HCP Subjects: 100307, 120111, 126628, 131924, 154936, 182840, 198451, 211417, 334635, 371843, 620434, 638049, 677968, 833148, 845458, 849971, 901442, 904044, 912447, 922854

Simulation: Figures 3-11 and Appendix B figures can be reproduced by executing the scripts BayGLMSim0.m, BayGLMSim1.R and BayGLMSim2.R.