The Composite-Hybrid Matlab Toolbox

This is a step by step guide on how to generate all the results shown in the paper **Novel Bayesian method for simultaneous detection of activation signatures and background connectivity for task fMRI data** from authors Michelle F. Miranda and Jeffrey S. Morris. For the simulation part, you can run the default analysis and it will generate the plots and tables shown in the paper. All the scripts call for data in the folder `/Simulations/Datasets`. If you wish to generate new data, we provide the code as described in Section 3.1. For the real data application, you can download the original data from the Human Connectome Website and run the pre-process steps described in Section 4.1. All analyses and plots are done using the Matlab scripts described below. To visualize the images, we use the software MRICron that can be downloaded at https://www.nitrc.org/projects/mricron.

1 Installation and Setup

1. Download **CompositeHybrid_Toolbox.zip**, decompress the file, and move the folder to your preferred directory.
2. Open your Matlab and make sure that your current working directory is CompositeHybrid_toolbox. To confirm the current working directory, type `pwd` on your Matlab command window.
3. Open the file `ToolboxPath.m`. If you are using a macOS then keep the default (variable **Mac=1**). If your are using Windows, change the variable to **Mac=0**. Run the code to save the path of the Toolbox.

2 Data Generation for Simulation

- **This section is optional**. You can skip to next section and use the default data provided.
- **Beware**. Running the files described here will replace the default data provided in the Toolbox.

Before running each script, make sure your current directory is the ToolBox main directory or update the path to load your file `'ToolboxPath'` in the first line of each script.

```
load('ToolboxPath')
```

There are three scripts to generate the following data.

- `GeneratePartitionMatrix_Sim.m` will generate the matrix `Sim_PhiMatrix.mat`, a matrix of size *Number of Partitions* vs. *Number of Voxels*, containing an indicator of which voxels belong to which partitions. Therefore, if generating data, this script must be run before the next ones.

- GenerateDataReplications_LongTerm.m will generate 100 datasets for the task data with long term memory noise.
- GenerateDataReplications_ShortTerm.m will generate 100 datasets for the task data with short term memory noise.
- GenerateNullDataReplications_LongTerm.m will generate 100 datasets for the null data with long term memory noise.

The files will be stored in /Simulations/Datasets.

3 Simulation

This step-by-step guide will generate the following results displayed in Section 3 of the manuscript.

1. Single data set simulation with long-term noise. Results are in Figure 1, Figure 2, and Table 1.
2. One hundred replications- simulation with long-term noise. Results in Table 2.
3. One hundred replications- simulation with short-term noise. Results in the Supplementary Material.
4. Null simulation. Results reported in Section 3.2.

3.1 Single data task simulation - long-term noise

The goal is to compare the four models as described in Section 3. The results for each model is obtained by running their corresponding scripts. The files start with the prefix Sim1 followed by the method and the noise type (either long or short term, e.g. Sim1_CHSB_LongTerm.m.

These files will use data in the file DataNull_longterm_sim1.mat in the folder Simulations to generate the results presented in Section 3. Running Sim1_CHSB_LongTerm.m will also generate **figures 1 and 2** in the manuscript.

Each file above will generate a .mat file (e.g. ResultsSim1_CHSB.mat), which includes the variables the IntervalWidth, MSE, FP, RP, representing average joint bands interval width computed across all 25600 voxels, the mean squared error of the contrast estimator, the average rate of false-positive voxels, and average rate of real positive voxels, respectively.

3.2 Task data replications - long and short term noises

Instructions here will generate the results for the simulated task data with short and long term memory. This includes results in **Table 2** in the manuscript and **Table 1** in the supplement material. Script files are included in the folder Replications_100 and data is provided in /Simulations/Datasets.

The files start with the prefix Sim100 followed by the method and the noise type (either long or short term, e.g. Sim100_CHSB_LongTerm.m. Running this file will generate the results for the composite-hybrid spatial model with long term noise and save the variables IntervalWidthArray, MSEArray, FPArray, and RPArray. These variables are arrays containing, for each dataset, the average joint bands interval width computed across all 25600 voxels, the mean squared error of the contrast estimator, the average rate of false-positive voxels, and average rate of real positive voxels, respectively.

The script Generate_Table2.m will collect the results for all methods presented in Table 2 and generate the variable Table 2 with 12 rows and 5 columns presented in the same order as Table 2 in the manuscript.

3.3 Single data null simulation

The results for each model is obtained by running their corresponding scripts. The files start with the prefix SimNull1 followed by the method, e.g. SimNull1_CHSB_LongTerm.m.

These files will use data in the file DataNull_longterm_sim1.mat in the folder Simulations to generate the results presented in Section 3.2.

Each file above will generate a .mat file (e.g. ResultsSimNull1_CHSB.mat), which includes the variables the Howmany and MSE, representing the total number of false-positive voxel and the mean squared error of the contrast coefficient. For the null simulation, the true contrast is zero.

3.4 Null data replications

Instructions here will generate the results for the simulated null data with long term memory. These results are presented in Section 3.2. Script files are included in the folder Replications_100 and data is provided in /Simulations/Datasets.

The files start with the prefix SimNull100 followed by the corresponding method, e.g. SimNull100_CHSB_LongTerm. Running this file will generate the results for the composite-hybrid spatial model with long term noise and save the variables MSEArray and FPArray. These variables are arrays containing, for each dataset, the mean squared error of the contrast estimator and the average rate of false-positive voxels, respectively.

4 Real data

This step-by-step guide will generate the results displayed in Section 4 of the manuscript. This includes:

1. Individual panels in Figure 4.
2. Results in Table 3.

3. Panels (a) to (e) in Figure 5
4. Results in Figure 6.
5. Figure 2 in the Supplementary Material.
6. A list of labels indicating regions as in Table 4.
7. Instructions on how to generate Figure 7 based on the results.

Note: before running the scripts, make sure to follow the instructions from Section 1.

4.1 Imaging data and design matrix setup

For reviewers only: You can skip this section and use the data provided in the Toolbox.

Data used in the manuscript was downloaded from `https://db.humanconnectome.org/`. We considered the Working Memory volumes collected from the right-left phase of the example Subject 100307. After downloading the data, the path of the image file should be `/100307/MNINonLinear/Results/tfMRI_W` We used a digitalized version of the original Talairach structural labelling that was registered into the MNI152 space. The atlas can be obtained from the FSL atlas library with information found at `https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases`. We will run three scripts to generate the following data files.

- `GetImageData.m` will read the data and set up the imaging data matrix. The data will be store in `ExampleSubjectData.mat`.
- `GenerateDesignMatrix.m` will generate the design matrix using the EV files downloaded with the data. This will be done by creating the boxcar function and convoluting it with the double gamma HRF and its derivative function. It will store the design matrix in the file `DesignMatrix_100307.mat`.
- `GeneratePartitionMatrix.m` will generate the matrix `InverseTrans_ROI_part_Phi.mat`, a matrix of size *Number of Partitions* vs. *Number of Voxels*, containing an indicator of which voxels belong to which partitions.

4.2 Composite-hybrid model

For the composite-hybrid model results we will run three scripts, `Application_CHSB.m`, `ReorderConnectivity.R`, and `Connectivity.m`.

- `Application_CHSB.m`. Running this script will generate various files that will be stored in the folder `HCP_Application/Results`. This includes:
  - `Results_Table3_FacePlaces_task_CHSB.mat` contains 8 variables with sufix `_CHSB`. The prefix `NClusters` followed by a number, e.g. 64,indicates the number of clusters larger

than 64 as in Table 3 of the manuscript. The prefix NVoxels followed by a number, e.g. 64, indicates the combined number of voxels in all clusters larger than 64 as in Table 3 of the manuscript. The variable IntWidth is the average jointband interval width as in Table 3 of the manuscript. Finally, the variable BiggCl is the number of voxels in the largest cluster identified as in Table 3 of the manuscript.

- Contrast_FacePlaces followed by the suffix of the model _CHSB with extensions .img/.hdr, .mat is the image of posterior mean of the contrast Face versus Places as Figure 2 in the Supplementary Material.
- Cluster followed by a number and the suffix of the model _CHSB with extensions .img/.hdr, .mat is an image mask of clusters larger than 64 as shown in figures 4 and 5 of the manuscript. For visualization, one can load the ImageBackground.hdr in MRICron and click on Overlay<-Add to add all clusters. Visualizations using the Window Multislice (Slices: 30, 34, 47, 48, and 50) and Render will generate panels (a) in figures 4 and 5, respectively.
- RV2.txt contains the square root of the RV coefficient as explained in Section 2.5 of the manuscript. This file is used to generate the panels in Figure 6 of the manuscript.
- ConnectivityNotOrdered.png (Figure 1) is a heatmap of the connectivity before it is reordered using the R script ReorderConnectivity.R.

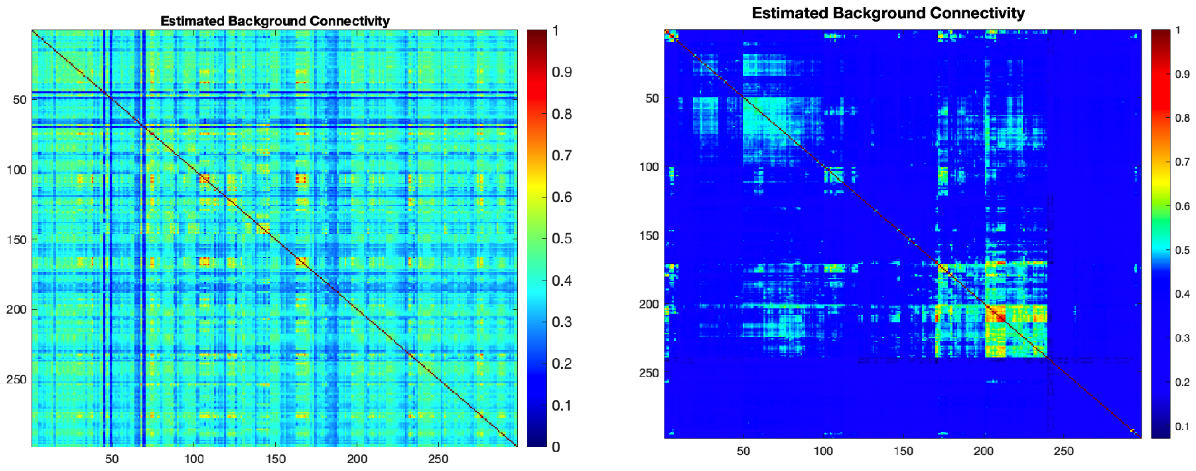


Figure 1: Estimated background connectivity. Values are the square root of the RV coefficients as defined in Section 2.5. Left panel displays the estimated connectivity values between the 298 ROIs before ordering. Right panel displays the same values ordered using hierarchical cluster.

Note: Additional outputs such as MCMC results and SimBas values (Section 2.4 of the manuscript) can also be saved by removing the comments in lines starting with save as bellow.

```
save(strcat(path2,‘Results_MCMC_WMtask_CHSB’),‘Results_MCMC’,‘Results_MCMC_psi’,
‘MCMC_scalenewpsi’,‘RV_PC’,‘V0’,‘p0’,‘p’,‘K’,‘tictoc_mcmc’,
‘tictoc_variancecomponents’)
save(strcat(path2,‘Results_Simbas_FacePlaces_task_CHSB’),‘SimbasResult_FacePlaces’,
‘upper_CI’, ‘lower_CI’)
```

- `ReorderConnectivity.R`. This R script will read the file `RV2.txt` described above and will use the hierarchical clustering algorithm in the R function `hclust` to reorganize the heatmap in Figure 1 for better visualization. Please adjust the path in the command `read.csv` accordingly and the file path in `write.table` to reflect where the reordered version of the heatmap and its regions' labels will be stored. Running this R script will generate two text files: (i) `ConnectivityOrdered.txt` contains the RV coefficients from the file RV2.txt after reordering; (ii) `ConnectivityROIorder.txt` contains the labels of the reordered matrix.

- `Connectivity.m`. This script will take the results from files `ConnectivityOrdered.txt` and `ConnectivityROIorder.txt` to generate the plots displayed in Figure 6 of the manuscript. The figures will be stored in the folder `HCP Application <- Results`. For better visualization we manually changed the colors in the colormap to show all values of connectivity below 0.5 in blue. Panels (a) and (b) in Figure 6 are saved as `ConnectivityOrdered.png` and `ConnectivityReordered_SelectedRegions.png` respectively. This script will also save the connectivity matrix for the selected regions to be displayed using `BrainNet Viewer` as in Figure 7 of the manuscript. This file is called `BrainViewRV.txt`.

Using BrainNet Viewer to generate Figure 7. To learn more about BrainNet Viewer please visit `https://www.nitrc.org/projects/bnv/`. For data visualization we need to provide 4 files: a surface file, a node file, a edge file, and a mapping file.

Surface File: we use the template available in BrainNet Viewer that includes the cerebellum, file BrainMesh_Ch2withCerebellum.nv.

Node File: the node file contains information about the coordinates of each ROI to be displayed. To generate this file, we obtained the center coordinates for regions 81, 238, 244, 473, 478, 667, 844, 902, and 987, which were estimated using the Tailarach Atlas. These regions are a subset of the regions presented in Table 4 of the manuscript. We load the file BrainViewNode.node, which is stored in the folder
`HCP_Application/BrainViewFiles`.

Edge File: the edge file is a matrix of connectivities between the ROIs defined in the Node file. We used the file BrainViewRV.edge, which is stored in the folder
`HCP_Application/BrainViewFiles`.

Mapping File: we use the Tailarach Atlas (file Talairach-labels-2mm.nii) for the mapping file, which is stored in the main directory of the ToolBox.

Specific settings for figure generation: to generate the plot in Figure 7 of the manuscript we chose surface opacity of 0.15 in the `Surface` tab, same color and size in the `Node` tab, we changed the color in the `Edge` tab of the edge to colormap *Jet* and values ranging from 0.5 to 0.85 to reflect the estimated connectivity values. Finally, we manually changed the colormap of the displayed volumes to *Gray* for better visualization.

4.3 Local Spatial and Global Spatial Models

To generate the results for the Local Spatial Model (LSB) and the Global Spatial Model (GSB) we run the scripts `Application_LSB.m` and `Application_GSB.m`. Running these scripts will generate files that will be stored in the folder `HCP_Application/Results`. This includes:

- `Results_Table3_FacePlaces_task` followed by either `_LSB.mat` or `_GSB.mat` containing 8 variables with sufixes `_LSB` and `_GSB`. The prefix `NClusters` followed by a number, e.g. 64,indicates the number of clusters larger than 64 as in Table 3 of the manuscript. The prefix `NVoxels` followed by a number, e.g. 64, indicates the combined number of voxels in all clusters larger than 64 as in Table 3 of the manuscript. The variable `IntWidth` is the average jointband interval width as in Table 3 of the manuscript. Finally, the variable `BiggCl` is the number of voxels in the largest cluster identified as in Table 3 of the manuscript.

- `Contrast_FacePlaces` followed by the suffix of the model with extensions `.img/.hdr, .mat` is the image of posterior mean of the contrast Face versus Places as Figure 2 in the Supplementary Material.

- `Cluster` followed by a number and the suffix of the model with extensions `.img/.hdr, .mat` is an image mask of clusters larger than 64 as shown in figures 4 and 5 of the manuscript (panels (b) and (c)). For visualization, one can load the BackgroundImage.hdr in MRICron and click on `Overlay<-Add` to add all clusters. Visualizations using the Window Multislice (Slices: 30, 34, 47, 48, and 50) and Render will generate panels (b) and (c) in figures 4 and 5, respectively.

4.4 No Spatial Basis Model

To obtain the results for the No Spatial Basis Model (NSB) we need approximately 300 hours. Because of its time, we use the following strategy.

- `Application_NSB.m`. This script applies the wavelet transform in the time domain of the data matrix. Running this script will generate 10 files, each containing a portion of the wavelet-transformed data. These files are named `DTransformed_1.mat`, with numbers ranging from 1 to 10. Since there is no spatial basis, the purpose of splitting the data is that the model can be run independently for each of these wavelet-transformed files.

- `ClusterNoBasis_1.m` with numbers ranging from 1 to 10. This script will take the corresponding `DTransformed` data and will run the wavelet based model with no spatial basis. The results will be stored at `Results_MCMC_NoBasis1.mat`, which are numbered according to which DTransform file was used. We ran the 10 scripts separately using the Cedar cluster from Compute Canada.

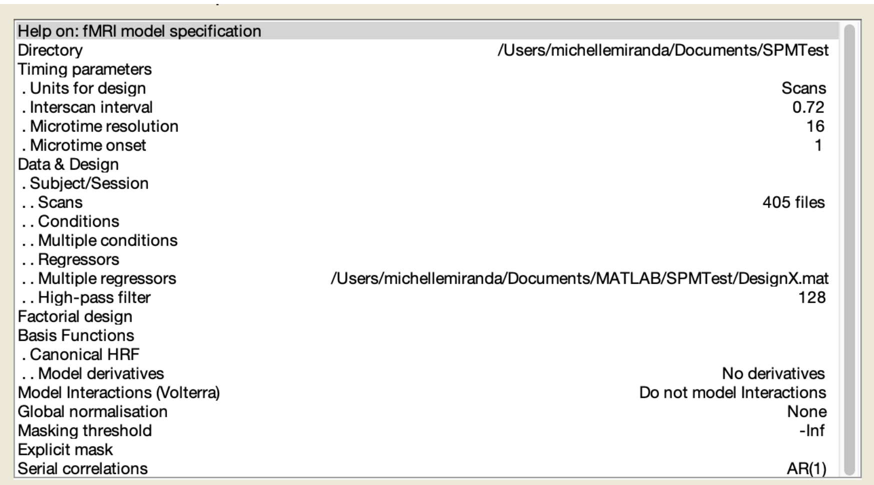
- `Application_NSB_GetResults.m`. This script combines the results stored in the files `Results_MCMC_NoBasis1` with numbers ranging from 1 to 10. Similarly to the previous models, it will generate the following results:

- Results_Table3_FacePlaces_task_NSB.mat containing 8 variables with suffix _NSB. The prefix NClusters followed by a number, e.g. 64,indicates the number of clusters larger than 64 as in Table 3 of the manuscript. The prefix NVoxels followed by a number, e.g. 64, indicates the combined number of voxels in all clusters larger than 64 as in Table 3 of the manuscript. The variable IntWidth is the average jointband interval width as in Table 3 of the manuscript. Finally, the variable BiggCl is the number of voxels in the largest cluster identified as in Table 3 of the manuscript.
- Contrast_FacePlaces_NSB.mat with extensions .img/.hdr, .mat is the image of posterior mean of the contrast Face versus Places as Figure 2 in the Supplementary Material.
- Cluster followed by a number and the suffix of the model with extensions .img/.hdr, .mat is an image mask of clusters larger than 64 as shown in figures 4 and 5 of the manuscript (panel (d)). For visualization, one can load the BackgroundImage.hdr in MRICron and click on Overlay<-Add to add all clusters. Visualizations using the Window Multislice (Slices: 30, 34, 47, 48, and 50) and Render will generate panel (d)) in figures 4 and 5, respectively.

4.5 Statistical Parametric Model - SPM

We ran the SPM sofware (https://www.fil.ion.ucl.ac.uk/spm/) with an auto-regressive time component (AR(1)) and design matrix as shown in Figure 3 of the manuscript. The HRF derivative is included in the design matrix. To specify the first level model in SPM click on fMRI <- Specify 1st- level tab. Most parameters are left to their defaults values with exceptions of Units for Design = Scans, TR=0.72, and Microtime Onset=1, Mask Threshold = -Inf. Figure 2 shows a screenshot of the defined parameters. After specification we click on the tab Estimate to estimate all parameters. After estimation, we generate the results by clicking on the tab Results and defining the contrast of *Face versus Places* and a family-wise error rate (FWE) of 0.05. SPM will generate the image and will display results which includes a *Height Threshold* value for the contrast map. This value is 5.14 and we use this number to determine activate voxels and estimate the clusters as in the previous models. After SPM estimation is done and the contrast image saved, we use the script described bellow to generate the plots presented in the manuscript. Notice that SPM stores the results in a folder created by the user (field Directory in Figure 2).

- Application_SPM.m. Please update the variable path_spm to reflect the path where the SPM results were stored. The script loads the images of the estimated coefficients to generate the contrast image showed in panel (e) of Figure 2 in the Supplementary Material. The variable *SpmI* in the script loads the contrast image for Face versus Places. If only one contrast was specified in the Results tab, the image name should be *spmT_ 0001.nii.* Running this script will produce the file:
  * Results_Table3_FacePlaces_task_SPM.mat containing 7 variables with suffix _SPM. The prefix NClusters followed by a number, e.g. 64,indicates the number of clusters larger than 64 as in Table 3 of the manuscript. The prefix NVoxels followed by a number, e.g. 64, indicates the combined number of voxels in all clusters larger than 64 as in Table 3 of the manuscript. Finally, the variable BiggCl is the number of voxels in the largest cluster identified as in Table 3 of the manuscript.



```
Help on: fMRI model specification
Directory                                      /Users/michellemiranda/Documents/SPMTest
Timing parameters
 . Units for design                                                               Scans
 . Interscan interval                                                              0.72
 . Microtime resolution                                                              16
 . Microtime onset                                                                    1
Data & Design
 . Subject/Session
 . . Scans                                                                    405 files
 . . Conditions
 . . Multiple conditions
 . . Regressors
 . . Multiple regressors   /Users/michellemiranda/Documents/MATLAB/SPMTest/DesignX.mat
 . . High-pass filter                                                               128
Factorial design
Basis Functions
 . Canonical HRF
 . . Model derivatives                                                   No derivatives
Model Interactions (Volterra)                                Do not model Interactions
Global normalisation                                                               None
Masking threshold                                                                  -Inf
Explicit mask
Serial correlations                                                               AR(1)
```

Figure 2: Screenshot of the parameters selected in SPM.