Bayesian Spatial Modeling in Functional Magnetic Resonance Imaging

Yongkai Qiu, Rongqian Zhang, Wenjing Zhou

April 2020

1 Project Outline

Functional magnetic resonance imaging (fMRI), a noninvasive neuroimaging method that provides an indirect measure of neuronal activity by detecting blood flow changes, has experienced an explosive growth in the past years. Statistical methods play a crucial role in understanding and analyzing fMRI data. Bayesian approaches, in particular, have shown great promise in applications. One of the most important purposes for fMRI analysis is to detect the Primary Motor Area (PMA) of a brain in a certain experiment. Early approaches to the analysis of such data would calculate voxel-wise t-test or ANOVA statistics and/or fit a linear model (GLM) at each voxel after a series of preprocessing steps(smoothing, spatial clustering, etc.). However, spatial correlation is expected in a voxel-level analysis of fMRI data because the response at a particular voxel is likely to be similar to the responses of neighboring voxels. Without applying the tools of Bayesian analysis, main approaches based on the time-series data on each voxel (ARIMA model, ARIMA model with BOLD and linear signals trend) would only focus on finding the pattern of each voxel but ignore the correlations among neighboring voxels (or even voxels with some distance.). Our project plans to focus on the spatial correlation problems related to fMRI data analysis. Under the assumption of a linear model is sufficient for the analysis of the data of each voxel, we would introduce several popular spatial priors to construct the posterior distribution of the coefficient vectors, then apply MCMC sampling techniques to simulate samples of the coefficients of each voxel. After applying a threshold vector for the value of those coefficients (aiming to rule out unactivated brain area), we would then construct a Posterior Probability Maps (PPM) of the brain scan image to detect the potential motor area.

Keywords: Functional magnetic resonance imaging (fMRI), generalized linear model(GLM), Markov chain Monte Carlo (MCMC) sampling, Spatial Priors, Posterior Probability Maps (PPM)

2 Data and Experiment Description

2.1 Experiment Description

Functional magnetic resonance imaging (fMRI) is one of the most widely used tools to study the neural underpinnings of human cognition. A typical fMRI data is a 4 dimension time-series data consists of 3 spatial dimensions (X, Y, and Z axes) representing the brain area. Each X, Y, and Z coordinate represents a cuboid element (a brain area) of the fMRI scan result, we call each of them as a "voxel". The last dimenson is a temporal dimension representing the time when the MRI machine conducts a brain scan on every voxel, which makes each voxel contain a time-series data which mainly captures the changes in blood flow and oxygenation associated with neural activity (the hemodynamic response), and on the differing magnetic properties of oxygenated and deoxygenated blood of that certain voxel. In particular, we call the signal captured by the time-series data in each voxel as the blood oxygen level-dependent (BOLD) signal.

For any certain given biain scan experimental design, we would ask our subjects to conduct that certain experiment with a certain time, we call all MRI scans conducted in those periods as "ON epochs". By contrast, there are also certain time that our subjects are asked to rest without any certain movement. Those periods would act as the control group with the "ON epochs", which we call as "OFF epochs". Moreover, for a given experiment, we would expect that only a few certain brain areas would be activated, especially for a given experiment related to the movement of a certain human body, we would expect activation of one or several certain areas in the Primary Motor Area (PMA) of the brain.

2.2 Data Description

In this report, the data we use comes from a sensitive source. It is pre-processed(combining the real and imaginary part of the data in each voxel and only leave the magnitdue data behind, which is the exact data of the brain scan
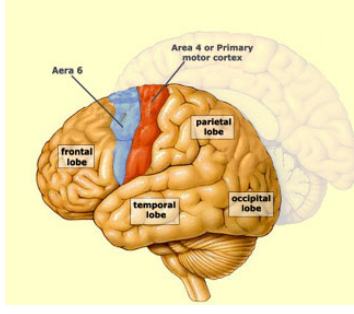
Figure 1: Primary Motor Area (PMA) of the brain

image) stored in the file "MOD1.RData". It contains the fMRI data which comes from the following finger tapping experiment:
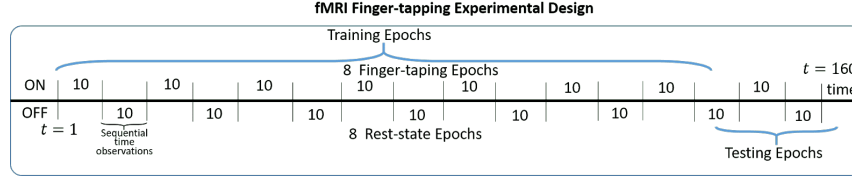

Figure 2: Finger-Tapping Experiment

Our data is a 4D fMRI data with $64 * 64 * 40$ voxel locations (i.e. Brain scan area: X axis: 64, Y axis: 64, Z axis: 40). Each voxel location has 160 scans arranged by time. There are 3 seconds pause between each scan. Thus each experiment would last for 8 minutes. Just like what is illustrated in the graph above, during those 160 scans there are alternant ON and OFF epochs, each ON and OFF epochs last for 30 seconds, thus give us 8 ON and OFF alternant periods for the whole experiment.

To make the computation easier, here we only take one brain slice location from z axis 33 (i.e z=33) to conduct the following analysis. With 160 scans at different time, our data could be viewed as 160 2D brain scan images, each with 64*64 pixels. Below are some examples of the scan image on the z axis 33. Notice that the X axis represents the brain area from left to right and Y axis represents the brain area from front to back.
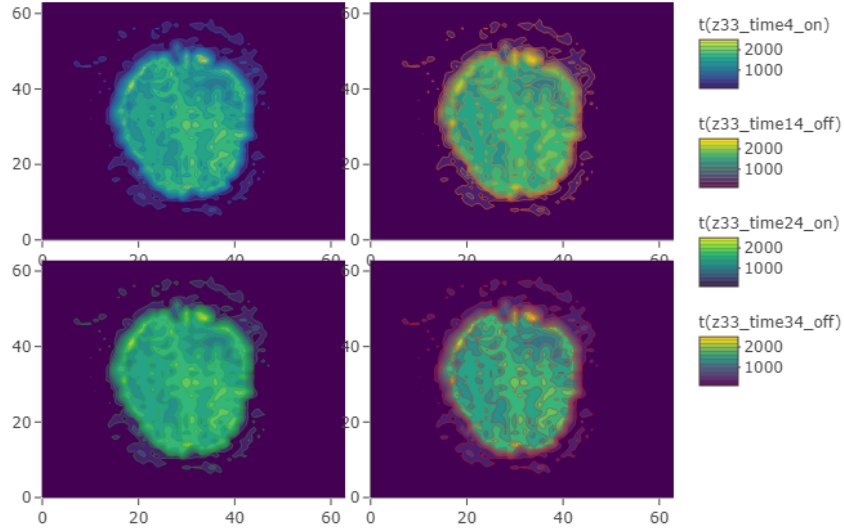

Figure 3: Z axis examples: Time: 4(ON), 14(OFF), 24(ON), 34(OFF)

All analysis in this report is based on this fMRI data. But this report aims to introduce a simple and general method that could be applied in any similiar MRI scan experiment.

3 Two-Stage Single-subject Modeling and Spatial Modeling

Ignoring the correlations between different times, a easy way to conduct Bayesian analysis on fmri data is only to consider the correlations between neighboring pixels (spatial information of pixels). However, the simulation requires an initial value of all coefficients. Thus, to fully conduct this model, we need initial value of coefficients from different pixels. Basically it is computed by a GLM model using all time-series data from each pixel. Typicall, we could choose from 3 different methods.

1. Regression models with baseline, linear signals, and BOLD signals.
2. ARIMA time series model with trends based on linear signals and BOLD signals.
3. Method discribed in the *Single-subject Model* part.

A good reason for us to apply the third method is that, the first two methods could only give us a single initial value for all coefficients. By the *Single-subject Model*, all coefficients are computed by simulation, and we could try multiple initial values on the Spatial model discribed here.

Thus in order to apply the spatial information of different pixels, here we developed a two stage modeling process, with the first stage simulating the coefficients for each single voxel, and the second stage apply the simulated single voxel coefficients data for spatial modeling.

3.1 Single-subject Modeling

3.1.1 Observation Model

For each pixel i, $i = 1, ..., I$, the time series $\{Y_{it}, t = 1, ..., T\}$ of magnetic resonance signals is related to the stimulus $\{X_{it}, t = 1, ..., T\}$ through a pixelwise linear model: $y_{it} = \alpha_i + z_t b_i + \epsilon_{it}$, $I = 1, 2, .., I, t = 1, 2, \ldots, T$. Here, we already transformed our original stimulus x_{it} to BOLD signal z_{it} by applying hemodynamic response function (HRF) h, such as $z_t = \int_0^t x(s)h(t-s)ds$ $(\ldots)$ In this model, we choose Poisson function as our HRF. Adding the HRF with the ON-OFF signals for this experiment, we could get the following BOLD signals.
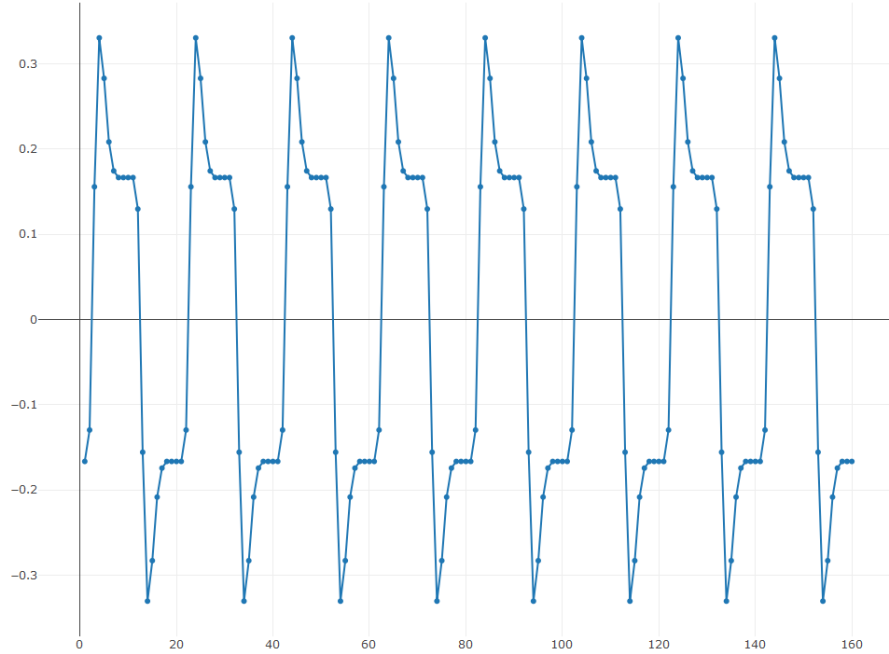

Figure 4: BOLD signal

For parameter b_i, we consider it as the effect of activation at pixel i. For error term ϵ_{it}, it captures random noise and various nuisance effects due to the machine as well as subject-related physiological noise.

According to our knowledge, a Gaussian assumption for the observations, conditional upon parameters, is suitable for this model. That is, the observation model for one pixel, i, i=1,2,...,I, is

$$y_{it}|\alpha_i, z_t, b_i, \sigma_i^2 \sim N(\alpha_i + z_t b_i, \sigma_i^2), \ t = 1, 2, \ldots, T$$

3.1.2   Prior Distribution

For the parameters b_i and a_i, we choose a highly dispersed diffuse Gaussian distribution to get a conjugate posterior distribution. For a_i, we have

$$a_i \sim N(\mu_a, 1/\lambda_a)$$

Where mu_a and λ_a are hyperparameters for a_i. The diffusion of this prior is controlled by λ, and when $\lambda \to 0$, the prior is diffuse and the Bayes estimator is in line with with the least squares estimator.

For b_i, we have

$$b_i \sim N(\mu_b, 1/\lambda_b)$$

The form for prior of b_i is the same as a_i, we only need to specify different hyperparameters.

For the parameter σ_i^2, we choose an Inverse-Gamma distribution.

$$\sigma_i^2 \sim Inv - Gamma(\gamma_a, \gamma_b)$$

Result for the Single-Subject Model

We apply the single-subject model to the dataset described in the Introduction section. After using 'fmri' in R, we convert original stimulus to BOLD signal. Then we fit the linear model with data in STAN. Our results contain 2000 iterations, with the first 1000 being discarded as burn-in. Hence, inference is performed with 1000 samples for each parameter. Parameters of the gamma hyperpriors were set to $\gamma_a = \gamma_b = 1$. Hyperparameters for a_i's prior were set to $\mu_a=0$, $\lambda_a=1$. Hyperparameters for b_i's prior were set to $\mu_b=0$, $\lambda_b=\frac{1}{10}$.
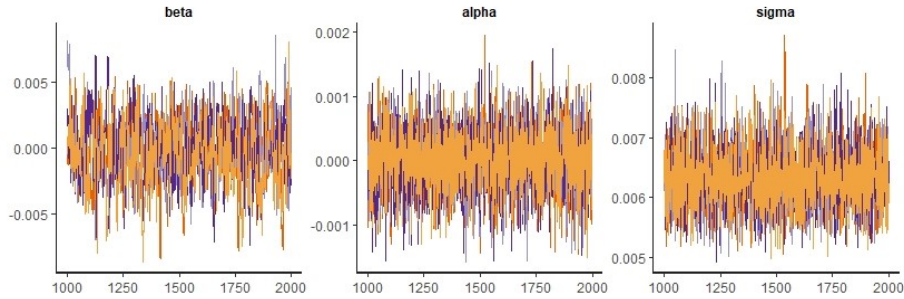

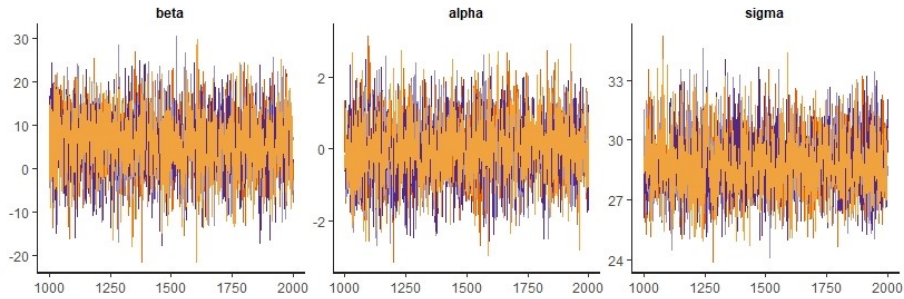Figure 5: STAN output for a non-active pixel (1,5) for b, a, $sigma^2$

Figure 6: STAN output for an active pixel (33,33) for b, a, $sigma^2$

Two plots above show posterior sampling results for b, a, $sigma^2$ in a non-active pixel(1,5,33) and an active pixel(33,33,33). We observe how 4 chains forget the initial value after some iterations and that the four traces are mixing well. We also can check the effective sample sizes and R hat. For non-active pixel (1,5,33), the effective

sample sizes of b, a and $sigma^2$ are 760, 4502 and 4017 respectively. All R hat equal to 1. For active pixel (33,33,33), the effective sample sizes are 4149,4377,3817,2122 respectively. All R hat also equal to 1.
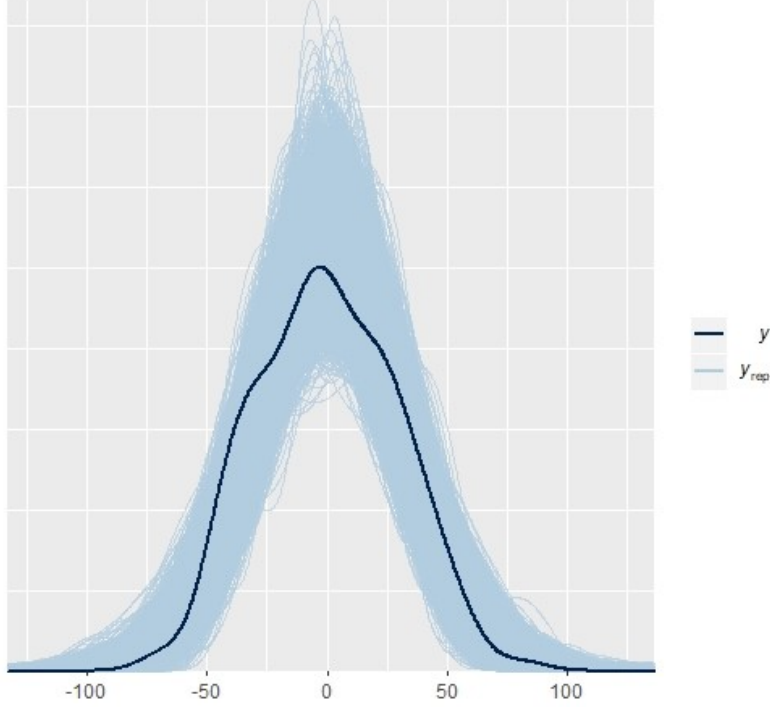

Figure 7: Posterior predictive density estimates for a pixel

In the plot above, the dark line is the distribution of the observed outcomes y and each of the 1000 lighter lines is the kernel density estimate of one of the replications of y from the posterior predictive distribution. This plot makes it easy to see that this model succeed to account for the bell-shaped of y.

Finally, we can save our results for parameters b_i to use in the next stage —- spatial model inference.

3.2 Spatial Model

Traditional generalized linear model(GLM) and time-series model focus on the analysis of the time series data based on each voxel (in our case, each pixel). However, it is reasonable to expect that there exist spatial correlation between different pixels because the response at a particular pixel is likely to be similar to the responses of neighboring pixels. Referring to the description of Zhang's paper, spatial dependence between brain pixels could be captured by imposing spatial priors on the model parameters. And a very commonly used approach is based on applying Gaussian Markov random field (GMRF) priors on the coefficients vector. Here, our original GLM model contains three signals, the baseline (intercept), linear signals, and BOLD signals. Thus GMRF priors are applied on this coefficients vector. And in it, we are mainly interested in the simulation for the coefficient of the BOLD signal as it represents if the signal in that certain pixel could follow the experimental design.

Structure for the GMRF priors

$$p(\beta_{(j)}|\lambda) \propto exp(-\frac{1}{2}\lambda\beta_{(j)}^T Q\beta_{(j)})$$

The precision matrix Q is defined as

$$Q_{v,k} = \begin{cases} n_v, & v = k \\ -1, & v \sim k \\ 0, & otherwise \end{cases}$$

In the above form, the symbol $v \sim k$ represents that pixel v and k are neighbors. The simplest and less computational cost assumption applied here indicate that v and k are defined as neighbors when they have only one pixel difference on the x axis or y axis.

Simulation iteration of β based on the GMRF priors

Based on the prior information and the penalty parameter λ, we could get the conditional distribution of beta based on the prior information. Thus our simulation iteration is based on the following distribution.

$$\beta_{v,j}|\beta_{-v,j}, \lambda \sim N(\frac{1}{n_v}\sum_{k\sim v}\beta_{k,j}, \frac{1}{n_v\lambda})$$

where $\beta_{-v,j} = \{\beta_{l,j}; l \neq v\}$ and $k \sim v$ represents point k and v are neighboring voxels.

Model discription and explaination

Essentially, the Spatial model is a smooting method. If there are few scattered pixels with very significant coefficients on the BOLD signal, by applying the neighboring information of other pixels, their coefficients significance woould be "smoothed" and become less significant. Thus this method is useful to rule out some single pixels with unusually high significance level. What we should expect is that after applying the spatial model, it would be easier for us to observe several areas with high significance levels, but no single pixels with high significance level.

Simulation results

Due to the heavy computational cost for the algorithm (each iteration needs to go through all 64*64 pixels), we could only conduct 200 iterations for each voxel. However, based on the conditional distribution given above, the Spatial model is all based on the normal distribution. Thus in theory, there is no need to warm-up the simulation procedure. Some simulation result is shown below:

Example 1: Pixel X=33,Y=12

This pixel is a very typical one, it has a very significant coefficient on the BOLD signal but its neighbors all have low significant level. What we could find is that after applying the spatial model, its coefficient is being "smoothed" to a value close to 0. Its variance is also controlled.

Visualization for Pixel X=33,Y=12
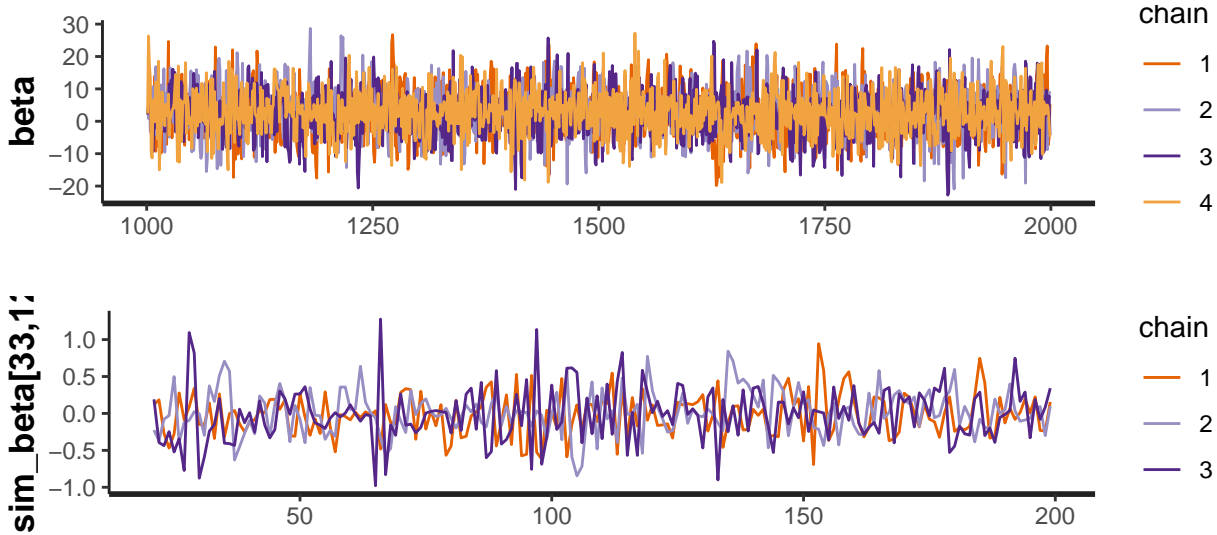
4 Other Spatial Model Variation

4.1 Spatial Model with Weighted Neighbors

The definition of neighbors in the previous model is only taken account of pixels with one distance from the original one. However, since each activated area could be an area with multiple pixels, we could think about expanding the neighboring area and put different weight to different neighbors based on some criteria. A very straightforward thought is to put weight based on the distance with the original pixel. However, correlation among voxels does not necessarily decay with distance. Thus weights for different pixels might also need to be determined by prior knowledge.

Below is the conditional distribution for a model like this

$$\beta_{v,j}|\beta_{-v,j}, \lambda \sim N(w_v \sum_{k \sim v} \beta_{k,j}, \frac{1}{n_v \lambda})$$

in which w_v represents the weight and we have $\sum w_v = 1$.

In our report, one weighted neighboring model is given taking 12 neighboring pixels with the weight determined by distance. Detailed result is shown in the conclusion.

4.2 Multiple Smooting Based on the Spatial Model

According to the conditional distribution of the spatial model, we could find that the essence for this model is taking the neighboring information and smooth a certain pixel. Thus a reasonable procedure to control the smoothness of our result is to conduct the same(or different weighted) spatial model for multiple times. One example of applying the same spatial model two times is shown in the conclusion part of this report. Notice that applying the spatial model too many times would only drag all coeffifients to the average level.

5 Posterior Probability Maps

The major goal for posterior inference is to create a spatial mapping of the activated brain regions. Based on the estimated regression parameters β, We construct posterior probability maps (PPMs). PPMs represent a complementary alternative to statistical parametric maps that are used to make classical inferences.The main idea of PPMs is to detect activations by mapping the estimates of β at each voxel of a single slice and then thresholding the conditional posterior probabilities at a specific confidence level.

At each voxel, the conditional posterior probability that a particular effect exceeds some threshold κ is calculted as

$$p = 1 - \Phi\left[\frac{\kappa - w^T M_{\beta_v|Y}}{(w^T C_{\beta_v|Y} w)^{1/2}}\right],$$

where $M_{\beta_v|Y}$ is the posterior mean, $C_{\beta_v|Y}$ is the posterior covariance of the parameter β_v, w is a contrast weight vector, $\Phi(.)$ is the cumulative density function of the standard normal distribution. As we have generated the postrior samples, we can easily obtain the posterior means and covariances, and the optimal threshold is formulated by minimizing a loss function. Specifically, we established a function to calculate the p value at each voxel and organized them into a matrix, then we drew the contour maps.

6 Spatial Model Posterior Probability Maps(PPM) Visualization

With a masking tools(BrainSuite19), we are able to narrow all p-values inside a reasonable brain area. Then based on the formula of PPM, we could construct a plot based on p-values.
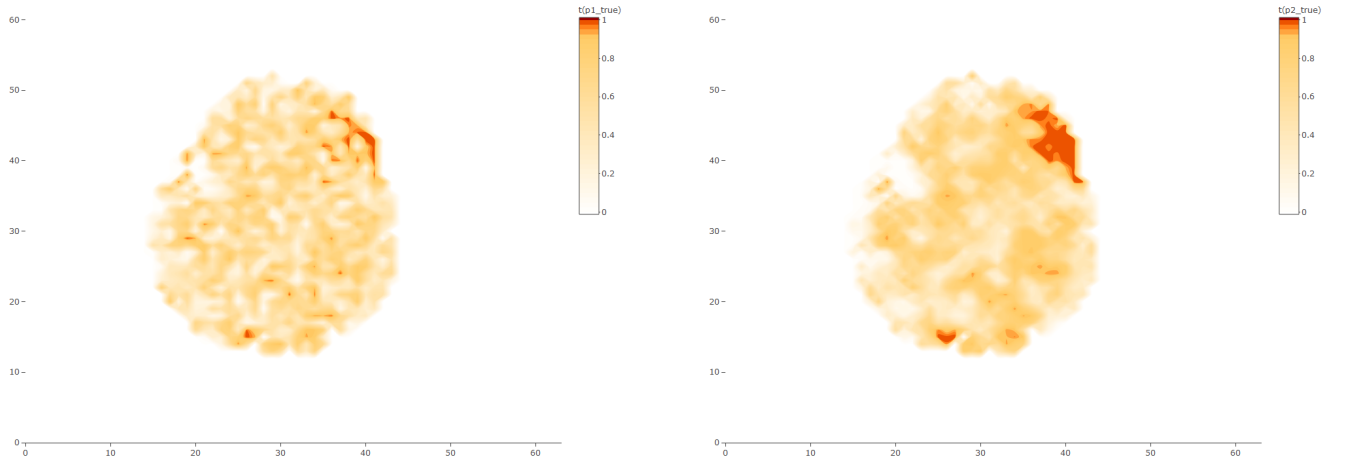

Fig. 8. First stage Observation model vs Two stage Spatial model

6.1 PPM for Coefficients of BOLD Signal from Single-subject Model

From the left graph of *Fig. 8.* on last page, we could see that based on the formula of PPM, though we are able to detect several areas with high p-values, all those are are pretty scattered, and it is hard for us to determine which area is the Primary Motor Area that we are looking for. Thus, without taking the spatial correlation into consideration, it is hard to find one or several particular area that could be detected as the Primary Motor Area.

6.2 PPM for Coefficients of BOLD Signal from Spatial Model

Now based on the right graph of *Fig. 8.* on last page, after applying the two stage spatial model, we could see that many scattered pixels with high p-value in the previous plot is now ruled out. And the scattered area with high p-values in the upper right of the brain is now connected into a clear area with high p-values. Thus we courd see, after applyig two stage spatial model, it is easier for us to determine which area is the Primary Motor Area.

However, we still see several small scattered area with high p-values. This could be reasonable, or this is caused by the limitation of the Two stage Spatial method. We might think about weighted spatial model or the multiple smoothing model discribed before. Or better, try more advanced bayesian model like the Nonadditive Spatial-temporal model.
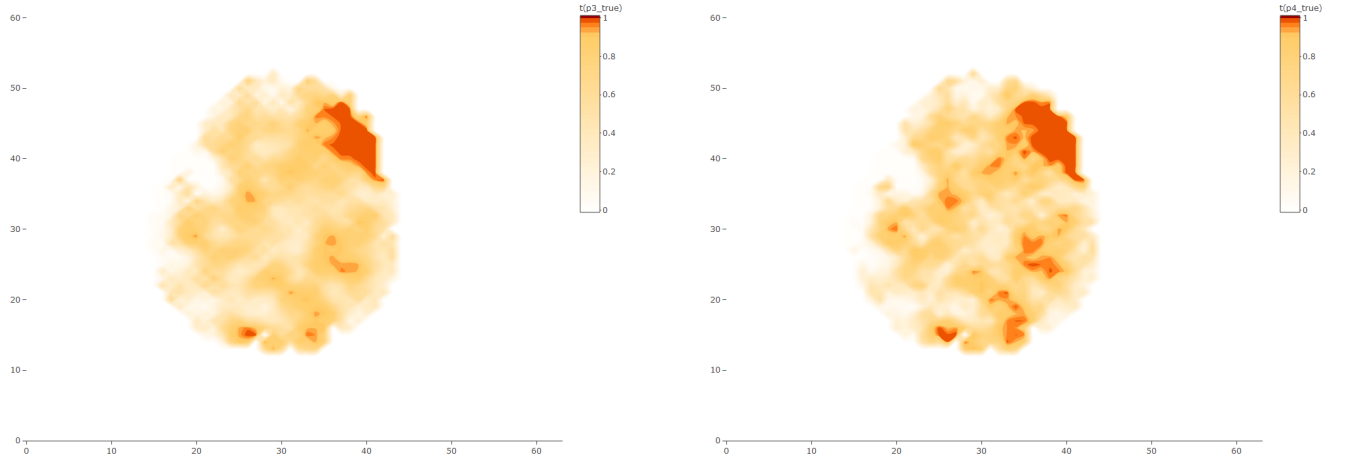

Fig. 9. Double Smoothing model vs Weighted Spatial model

6.3 PPM for Coefficients of BOLD Signal from Double Smoothing Model

The left graph of *Fig. 9.* comes from the double smoothing model. the result comes from applying the same spatial model on the result given by the previous spatial model. In this way, the data would be smoothed twice. Based on the result, we see that comparing to the two stage Spatial model, p-value given by the double smoothing model is clearly more smoothed. But there is not a huge difference than the previous one. Thus we should favor the simplier spatial smoothing model given before. However, we might able to get a better model if we combine more spatial model together.

6.4 PPM for coefficients of BOLD Signal from Weighted Spatial Model

The right graph of *Fig. 9.* comesfrom tne weighted spatial model, in which we applied not the nearest 4 pixels, but the nearest 12 pixels, and give them weight based on the distance of those pixels with the pixel we are computing. We can clearly see that, in our case, this model put more pixels into significant pixels under the same criteria of PPM. This suggests that it might not be very reasonable to assign weight only based on the distance of pixels. More detailed prior knowledge might need to be applied to get a reasonable weighted spatial model. Also, this might suggest that we should think about applying some FDR correction method on the p-value to get more reasonable conclusion. Like the benjamini hochberg correction.

7 Conclusion and Discussion

This project describes a Bayesian 2-Stage Spatial Modeling for detecting brain activity in fMRI studies. In the first stage, we use hemodynamic response function (HRF) to transform original stimulus to BOLD signal. Then we build pixelwise linear model with Gaussian prior. In the second stage, we use Gaussian Markov random fields (GMRF)

prior to capture the spatial dependence between brain pixels, since we have prior knowledge that evoked responses are spatially contiguous and locally homogeneous.

In this work, we use STAN to fit our fMRI data and draw samples. We check the effective sample sizes, trace plots and R hat of our quantities of interest and do Posterior Predictive Checking. By doing this, we show that our model is working in the right way.

Moreover, we construct posterior probability maps (PPMs) based on our estimated posterior parameters. By doing this, we produce a spatial mapping of the activated brain regions.

For further improvement, we consider introducing temporal effect to this model. By choosing appropriate priors for time-varying effect on brain activity, we can extend our model to a spatiotemporal model. Besides, we can also try other algorithms, such as Gibbs Sampling, to avoid some limitations of STAN and improve our computation efficiency.

8   Sources and References

1. Zhang, L., Guindani, M., & Vannucci, M. (2015). Bayesian models for functional magnetic resonance imaging data analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 7(1), 21-41.
2. Flandin, G., & Penny, W. D. (2007). Bayesian fMRI data analysis with sparse spatial basis function priors. NeuroImage, 34(3), 1108-1125.
3. Gössl, C., Auer, D. P., & Fahrmeir, L. (2001). Bayesian spatiotemporal inference in functional magnetic resonance imaging. Biometrics, 57(2), 554-562.
4. Friston, K. J., & Penny, W. (2003). Posterior probability maps and SPMs. Neuroimage, 19(3), 1240-1249.