Detecting Model Mis-specification in fMRI using Scan Statistics

Ji Meng Loh[1], Tor D. Wager[2] and Martin A. Lindquist[1]

[1]Department of Statistics, Columbia University

[2]Department of Psychology, Columbia University

INTRODUCTION

The voxel-wise general linear model (GLM) approach has arguably become the dominant way to analyze fMRI data. It is well-suited for testing how much of the variability in a voxel's time course can be explained by a set of *a priori* specified regressors. However, even a relatively small amount of mis-modeling can result in severe power loss (Fig. 1), and inflate the false positive rate beyond the nominal value. Due to the massive amount of data, performing model diagnostics is challenging, and only limited attention has been given to this problem (e.g. [1]).

In this work we:

- provide expressions for power loss and bias, given a statistical model with some degree of mis-modeling and a hypothetical true model.
- develop a procedure based on the use of scan statistics [2] for identifying voxels or regions where model misfit may be present.
- apply the procedure to detect systematic mis-modeling or artefacts in fMRI time courses.

METHODS

Mis-modeling in the GLM

In the GLM approach, the fMRI time series, $\mathbf{Y}$, is modeled as a linear combination of a number of different signal components summarized in a design matrix $\mathbf{X}$. The model can be written:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{V})$$

Suppose we erroneously mis-model the design matrix, i.e. we use $\mathbf{X} = \boldsymbol{\Pi} + \boldsymbol{\Gamma}$, where $\boldsymbol{\Pi}$ is the 'true' design matrix and $\boldsymbol{\Gamma}$ is a matrix describing the error due to mis-modeling. The bias in the estimate of $\boldsymbol{\beta}$ can be expressed as:

$$\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\boldsymbol{\Gamma}\mathbf{Y}$$

The power-loss attributable to model mis-specification is the difference between the power to detect activation when mis-modeling is present and when it is absent. Specifically this depends on the difference in the distribution of the regression variance under the two models. Under the correct model, $\boldsymbol{\Gamma}$=0, it follows a χ^2 distribution, otherwise, it follows a non-central χ^2 distribution, with non-centrality parameter

$$\delta = \boldsymbol{\beta}^T\boldsymbol{\Gamma}^T\left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\right)\boldsymbol{\Gamma}\boldsymbol{\beta}$$

Details can be found in [3].

Scan Statistic

If there is model misfit (e.g. mis-specification of onset, duration, or response shape) residuals will be systematically larger in mis-modeled segments of the time series. To detect whether mis-modeling is present, we can study the residuals from the GLM analysis using scan statistics. The scan statistic is often used to detect clusters in time courses or images.

Suppose *r(i)*, i=1,…T are the whitened residuals obtained from a voxel-wise GLM analysis. Let,

$$Y_w(t) = \sum_{i=t}^{t+w-1} r(i)K(t-i)$$

be the moving average of *w* consecutive observations starting at time *t*, where *K(t)* is a weighting function. Note that it is possible to use any kernel function here (e.g. Uniform or Gaussian) that sums to 1. Under H_0: $\boldsymbol{\Gamma}$=0, the statistic Y_w follows a normal distribution with mean 0 for all *w*, *t*. The window that yields the largest value, gives the strongest evidence of model misfit (See Fig. 2). This value is compared to the maximum that would be obtained if the residuals were *iid* Normal, as expected under the null hypothesis.

Calculating p-values

The distribution of the maximum statistic under the null hypothesis, and p-values for this statistic, can be found using a Monte Carlo simulation. Alternatively, an upper bound for the p-values can be approximated using Sidak's inequality. This latter approach provides effective results at a fraction of the computational cost.

If the weighting function is chosen to be a Gaussian, Gaussian random field theory can be used to determine the appropriate threshold and p-value. In addition, a 4D Gaussian random field can be used to correct for multiple comparisons over both space and time. This may ultimately be the most attractive approach.

Effect of mis-specification on power

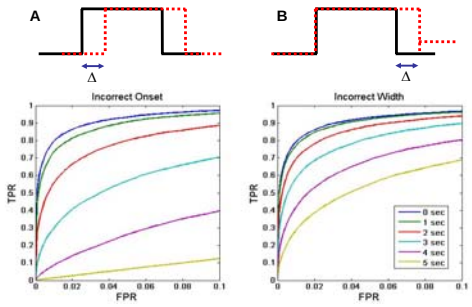

Figure 1. (A-B) Results of a simulation study. The solid and dashed lines in (A) and (B) show respectively the modeled and true activation. The difference between truth and model (delta) is allowed to vary from 0 to 5 seconds. The true activation paradigm is repeated 4 times and convolved with a canonical HRF. Noise is then added corresponding to a Cohen's *d* of 0.5. The GLM is fit using the modeled activation pattern (solid lines) convolved with the canonical HRF. This procedure is repeated 1000 times for both delayed onset (A) and prolonged width (B). Receiver operating characteristic (ROC) curves in the bottom panels show the false positive rate (FPR) vs. the true positive rate (TPR) across statistical significance thresholds. The curves show a substantial decrease in power as a function of model mis-specification.

Scan Statistic method at work

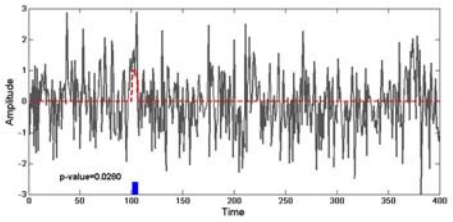

Figure 2. The bold line shows a sample time course of independent, identically distributed (*iid*) residuals with mean 0 (except in a cluster of length 5 time units where it is equal to 1 – see red dashed line) and standard deviation 1. Using the scan statistic approach outlined in this work, we found a cluster of 5 points with mean significantly different from 0 which we denote using a blue box to indicate its location and width (Note its exact correspondence with the boxcar function). The p-value of 0.0280 indicates that the data inside this cluster is not consistent with the null hypothesis that all the data points have mean 0. Hence, the scan statistic approach can effectively discover systematic variation in the residual time-course.

Activation, Bias and Power-loss maps

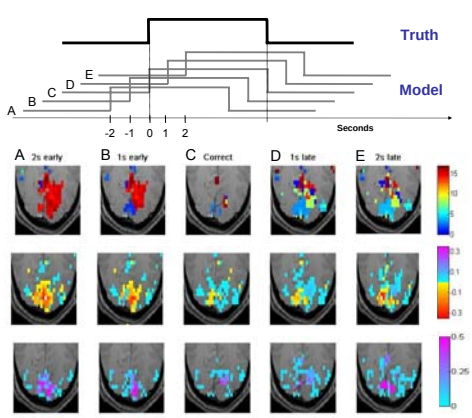

Figure 3. (Top) The experimental data is analyzed using a standard GLM where the onset of activation is purposefully mis-modeled so the difference in modeled and true onset time took the values -2, -1, 0, 1 and 2 seconds. (First row of images A-E) The location in time of the scan statistic is shown for voxels where the scan statistic indicates significant mis-modeling. Red indicates voxels that show a cluster of mis-modeled points toward the end of the visual stimuli, while blue indicates voxels that show a cluster in the beginning of the stimuli. The results are consistent with the erroneous model formulation. (Second row A-E) Bias maps show an increase in bias as the amount of mis-modeling increases. (Third row A-E) Power-loss maps show a decrease in power as the amount of mis-modeling increase.

Follow-up Analysis

Once we have detected evidence of mis-modeling in a voxel, we can estimate the mis-modeling parameter $\boldsymbol{\Gamma}$ and then do the following:

- Identify the presence of systematic mis-modeling as a function of the experimental stimulus. This can be done by calculating the average residual value relative to the onset of activation using an FIR basis set.
- Make maps that show the bias and power loss attributable to the mis-specification (see Fig. 3).

Experimental Design

The experimental data consisted of a blocked alternation of 11 s of full-field contrast-reversing checkerboards (16 Hz) with 30 s of open-eye fixation baseline. Blocks of stimulation were presented on an in-scanner LCD screen. Spiral-out gradient echo images were collected on a GE 3T fMRI scanner. Seven oblique slices were collected through visual and motor cortex, 3.12 x 3.12 x 5 mm voxels, TR = 0.5 s, TE = 25 ms, flip angle = 90, FOV = 20 cm, 410 images.

The data was analyzed using a standard GLM procedure, where the design matrix consisted of three regressors associated with a quadratic drift term and one regressor corresponding to the expected BOLD response. This regressor was calculated by convolving a boxcar function corresponding to the experimental design with SPMs canonical HRF. We performed this analysis five times, each time the onset of activation in the boxcar design was purposefully mis-modeled so the difference in modeled and true onset time took the values -2, -1, 0, 1 and 2 seconds. (Figure 3 top). For each of the five cases, the scan statistic approach was applied to the residuals to detect whether there was evidence of significant mis-modeling. Maps of the estimated bias and power-loss due to mis-modeling were computed for each case (Fig. 3 second and third row).

RESULTS

Figures 3A-E (top row) show the voxels found by the scan statistic to have significant mis-modeling. Red indicates voxels that show a cluster of mis-modeled points toward the end of the visual stimuli, while blue indicates voxels that show a cluster in the beginning of the stimuli. In each case the results are consistent with the error. Also, the number of significant voxels increase as the amount of mis-modeling increases. When the correct model is used, few significant voxels are present.

We have also included bias and power-loss maps (Figures 3 A-E, second and third rows). These maps tell a similar story and allow us to determine regions where mis-modeling has the greatest impact on statistical inference. These plots provide us with a means to judge the validity of the statistical parametric maps that are typically used to summarize the results of a GLM analysis. They indicate regions that we should study closer to check the model assumptions.

CONCLUSIONS

- Regression diagnostics are rarely performed when analysing fMRI data using the GLM.
- Mis-modeling can result in a severe decrease in statistical power when using the GLM. It may also inflate the false positive rate beyond the nominal value
- In this work we suggest the use of scan statistics to detect the presence of mis-modeling.
- Bias and power-loss maps can be constructed to indicate regions that are particularly influenced by mis-modeling.

REFERENCES

[1] Luo, W-L. Nichols, T.E. (2003). Diagnosis and exploration of massively univariate neuroimaging models, NeuroImage, 19, 1014-1032.

[2] Glaz, J., Balakrishnan, N. (Eds.) Scan Statistics and Applications, Birkhäuser, Boston, 1999.

[3] Loh, J-M., Lindquist, M.A. and Wager, T.D. (2007). Detecting Mis-modeling in fMRI using Scan Statistics. *In submission.*