

## HST.583 - Lab 5: fMRI Data Analysis Tutorial

# Table of Contents

- I. [Introduction](#)
    1. [Data](#)
    2. [Paradigm](#)
  - II. [Understanding temporal and spatial correlation in fMRI data](#)
    1. [Temporal Autocorrelation](#)
    2. [Spatial Correlation](#)
  - III. [Understanding how to construct a statistical model for fMRI data](#)
    1. [Statistical Maps](#)
    2. [Design Matrix](#)
    3. [Gamma function HDR](#)
    4. [Thresholding](#)
    5. [FIR Model](#)
  - IV. [Identifying sources of noise and their contribution to fMRI signals](#)
  - V. [Understanding motion correction and spatial filtering for processing fMRI data](#)
    1. [Motion Correction](#)
    2. [Spatial Filtering](#)
- 

## I. Introduction

These laboratory exercises complement the lectures and assigned readings on statistical analysis of fMRI data.

The goals of this tutorial are to help you with the following:

1. Understanding temporal and spatial correlation in fMRI data;
2. Understanding how to construct a statistical model for fMRI data;
3. Identifying sources of noise and their contribution to fMRI signals; and
4. Understanding the effects of motion correction and spatial filtering on the outcome of statistical analysis of fMRI data.

### 1. Data

The data was acquired on a a Siemens 3 T scanner using a head coil and the following acquisition parameters:

- TR = 2 s
- TE = 30 ms
- alpha = 90°

- in-plane resolution = 4mm x 4mm (256 mm FOV on 64x64 matrix)
- 180 time points
- 21 slices, each 4mm thick with 1mm gap (acquired in interleaved order)

## 2. Paradigm

There were 20 epochs of 18 seconds (9 TRs) each. During each epoch, a visual pattern was shown during the first two seconds (1 TR). The pattern consisted of a flashing checkerboard annulus (ring) pattern against a gray background. For the rest of the epoch (16 seconds/8 TRs), only the gray background was present, along with a black dot for fixation.

## II. Understanding temporal/spatial correlation in fMRI data

**Question 1\*.** What do you currently know about autocorrelation in general, and in fMRI data specifically? This and other starred questions will not be graded, but will help focus your learning. (No fair looking ahead! :-)

We will first explore the characteristics of the temporal autocorrelation in functional data. Correlation, as you will learn in lecture, is the nonrandomness in data due to lack of independence between adjacent data points in time and space; autocorrelation is the correlation of a time course with itself (as opposed to between two different time courses). The demo allows you to display the temporal autocorrelation at each voxel of the functional data via Dview, so you can examine the heterogeneity in signal variance in different areas of the brain.

**Question 2.** Most statistical analyses assume independence among the sample data points. How is this assumption violated in fMRI data with respect to the distance between voxels (spatially) and on the time between image acquisitions (temporally)?

### 1. Temporal Autocorrelation

1. In the **Input Data** box, click on the drop-down menu and select the **Human Experimental** data set.
2. Click **Autocorr** in the **Autocorr** box. Dview will now show the temporal autocorrelation for the voxel under the yellow crosshair cursor.
3. Move the cursor by clicking on another part of the brain in any of the transverse, sagittal, or coronal views. You should see the autocorrelation change depending on which voxel you have selected.

4. Now examine the autocorrelation in air, gray matter, white matter, and CSF. For gray matter, pick three locations among the frontal, parietal, and temporal cortices, and one in the basal ganglia.

**Question 3. (a)** Plot the autocorrelation graph for each of the seven tissue regions that you examined and label each. Describe what you see; is it what you expected? **(b)** Given what you know about the experimental paradigm, find an area in the brain that has paradigm-related signal changes. Plot this separately and label with the correct anatomical region. How does the autocorrelation here differ from other regions?

5. Select the **Phantom** data set in the **Input Data** box, and click **Autocorr** to show the autocorrelation. Move around different areas of the phantom and the surrounding air.

**Question 4.** Plot the autocorrelation graph for 2 locations of your choice in the phantom. What does the autocorrelation look like in the phantom? How does it vary within the phantom? Is it what you expected?

Detrending is a way of removing signal drift and other low-frequency noise, which may result from slow head movements, gradient heating, etc. Here we look at how the use of detrending impacts the temporal autocorrelation.

1. In the **Input Data** box, again select the **Human Experimental** data set. Turn off **Spatial Filtering** if it is on.
2. Turn on **Detrending** and set the **polynomial order** to 3.
3. Now examine the autocorrelation in the same tissues (CSF, 4 gray matter areas, white matter, air) that you used above by clicking on **Autocorr** and moving the crosshair to the appropriate regions.

**Question 5.** What effect does detrending have on temporal autocorrelation? Is it the same in all 7 regions? Explain why the effects occurred.

**Aside: Polynomial Detrending.** In the demo, we remove trends in the data which are polynomial in nature. The order of the polynomial trend removed can be varied to yield different results. For instance, removing 0th-order polynomials simply subtracts the mean from the signal, and removing 1st order polynomials does that and also removes any linear

trend. As a thought exercise, consider the effects of removing higher-order polynomial trends. Why wouldn't it be a good idea to perform detrending with, say, 85th-order polynomials?

## 2. Spatial Correlation

We now turn to standard deviation maps of the temporal data at each voxel to help us investigate spatial correlation.

The standard deviation map allows you to look at how much the signal varies over time in each voxel, and how that parameter varies across structures. Along with spatial filtering (which you will learn about in a later section), this lets us examine spatial correlation indirectly.

1. Click on **Reset Parameters** to restore the parameters to their defaults, and select the **Human resting** data set in the **Input Data** box.
2. Click on **Std Dev Map** in the **Autocorr** box to show the standard deviation map.

**Question 6.** Note the differences in standard deviation among different regions of the brain (white matter, gray matter, CSF, air). Describe what you see. Are they what you expected?

3. Now turn on **Spatial Filtering** and set the **Filter Width** to 6 mm. Click on **Std Dev** again and again examine the standard deviation map.

**Question 7.** Where do you see reduction in standard deviation? You may toggle between the filtered and unfiltered maps by clicking on **Recall** in the **Autocorr** box and selecting the appropriate parameters.

4. Now examine the temporal autocorrelation in the spatially filtered data.

**Question 8.** What is the relation between changes in standard deviation and temporal autocorrelation after spatial filtering in CSF, white matter, and gray matter?

### III. Understanding how to construct a statistical model for fMRI data

Now that you've explored noise characteristics in data and thought about how it impacts the data, we will actually perform a GLM (general linear model) analysis. In the next few sections we will start with a barebones statistical model and add more components to it in order to increase the power of our analysis.

We can, as a first step, model the signal that we record from an fMRI session as a linear combination of the actual signal and noise; in subsequent sections we will further refine our model. With GLM analysis, we use the signal and noise as regressors, and solve for the coefficients for the linear combination. The mathematical equation for the model is  $y = \beta x + \epsilon$ , where  $y$  is the observed signal,  $x$  is the estimated signal,  $\beta$  is the parameter estimate for  $x$ , and  $\epsilon$  is error term that corresponds to our noise. You will learn more about it in the next few lectures; more information can also be found in Stephen Smith's chapter in *Functional MRI: An Introduction to Methods* (Ch. 11).

For now, let us simply perform the statistical analysis and see what happens.

1. Reset the parameters and select the **Human Experiment** data.
2. Turn on **Motion Correction** and **Detrending**, setting the polynomial order to 3.
3. Click **Fit** in the **Statistical Model** box.
4. You will now see a dialog box with the settings for GLM analysis; they allow changes to different aspects of the model. Note that the parameters (order of polynomial detrending, width of spatial filter, etc.) that you have set in the demo are reflected in the box. For the purposes of this lab, examine the settings and leave them as is; the following is the list of settings and their explanations:
  - **Time step**  
The TR (time between each volume acquisition) in our experiment.
  - **Base for output file names**  
The prefix of the output file names for the results of the analysis. Used only if they are saved to disk.
  - **Hemodynamic response function**  
Formula for HDR function if not using an FIR model (more on this soon).
  - **Contrasts**  
Vector specifying which explanatory variables to compare if there are more than two conditions. For the Gamma model, contrasts are ignored in this exercise since our paradigm has only two conditions; more on this for the FIR model soon.
  - **FIR time window**  
Length of time window to use in the FIR model (more on this soon).

- **FIR time offset**  
Offset of FIR model from event onset (more on this soon).
- **Frames to exclude**  
Time points to exclude from analysis. Ignored in this lab.
- **Intensity threshold for each volume**  
Threshold above which voxels are considered in brain. Ignored in this lab.
- **Order of polynomial to remove**  
Order of polynomial used in detrending.
- **FWHM used to smooth data**  
Size of the filter used in spatial filtering. Certain steps of the analysis require this information to interpret the data properly.
- **Estimated AR1 parameter**  
The first order autoregressive (AR(1)) model parameter (used in whitening).

5. After you have examined the settings, click **Ok**.
6. A dialog box with the relevant plots and maps to compute will pop up. Click **Ok** again to start the analysis; it will take a minute or so to complete.

## 1. Statistical Maps

Once the analysis is complete, you can view the results by selecting different statistical maps and plots in the **Inferences** box.

There are two types of activation (statistical) maps available to view: *t-maps* and *p-maps*. The *t-map* is a voxel-by-voxel map of the *t*-statistic comparing the two conditions for this paradigm (visual stimulus versus fixation); at each voxel, the value of the *t*-statistic is calculated for the coefficient in the model which compares the two test conditions. Corresponding to each statistic is a significance probability, which is plotted to form the *p-map*. The two maps are typically plotted on a color scale, with different colors along a spectrum indicating low and high significance.

Both maps are useful for determining what parts of the brain are being activated during the experiment, but they differ in two important ways. The *t-map* is based on a relative measure, the standardized difference between the means of the two conditions, whereas the *p-map* is based on the probability distribution of the *t*-statistic and thus takes on a value between 0 and 1. The *p-map* also takes into account the *degrees of freedom (DOF)*, or the amount of independent information, at each voxel; the more time points (observations) there are in our data, the higher our *DOF*.

1. Select either the **p-map** or **t-map** under **Maps** in the **Inference** box and click **View**.
2. Locate the activation region in the brain.

**Question 9.** Is the activation region where you would expect it to be? Why or why not?

## 2. Design Matrix

You can also see a representation of the regressors used in the analysis with the *design matrix* and *stimulus covariance matrix*. The design matrix is a visual representation of the regressors used in the analysis, and consists of one or more columns. The first column represents the paradigm convolved with the hemodynamic response function, and the subsequent columns represent polynomials used in detrending. The stimulus covariance matrix is necessary when testing contrasts involving multiple parameters under the general linear model (for example, if there are more than two conditions), and will not be used in this lab.

1. Select **matrices & regressors** under **Graphs** in the **Inference** box and click **Plot**.
2. Click on different vertical columns of the design matrix to view the corresponding regressor below.

## 3. Gamma function HDR

Recall that in our signal model, we assumed that the hemodynamic response function took a certain shape. In our statistical analysis, we assumed that the HDR took the shape of a *gamma function*; you can see the shape of the Gamma function by selecting the **Gamma/FIR Model HDR** graph and clicking **Plot**. The demo will show you the fitted gamma function corresponding to the selected location in the brain in Dview. Notice that it consists of a delayed peak, followed by a post-stimulus undershoot; this type of gamma function is known as a *dual gamma*, as opposed to a *single gamma* which just has the initial peak. More sophisticated models may also be used to add an undershoot before the peak, if the investigator hypothesizes that the data will be best fit using such a model.

**Question 10.** What is the physiological basis for using a single gamma HDR function? What about a dual gamma HDR function?

## 4. Thresholding

In addition to seeing the default statistical map that shows every voxel, you can also specify a *threshold* such that only the activation is visible.

1. Go back and select the **p-map** under **Maps**.
2. Click on the colorbar in the lower left-hand corner of the Dview window to change the threshold for the statistical map. A contour will surround the region(s) that are below the threshold (*i.e.* have a p-value less than the threshold).

**Question 11.** At what threshold value do you start seeing activation in the air? What does this tell you?

## 5. FIR Model HDR

Instead of fitting the signal to a predetermined model of the hemodynamic response, we can also estimate the timecourse of the response to the stimulus or condition from the signal. The *FIR (finite impulse response) model* averages the signal from the beginning of each epoch of stimulus onset to give an estimate of the underlying hemodynamic response. The name comes from signal processing and tells us that the estimated response is finite in time.

Note that in the GLM analysis settings that pop up when you perform the analysis, a time window and offset are specified for the FIR model. The time window specifies how long of a response you would like to estimate, and the offset specifies how long after each event onset you want the time window to begin. We have chosen appropriate default values for the purposes of this lab.

Also, we have set up the default contrast vector to pick out the TRs in which the HDR has its peak; this allows us to compare those timepoints with the baseline and to calculate the significance probability and t-statistic appropriately.

1. Change the **signal model** from **Gamma** to **FIR** and perform the analysis again.
2. Examine the **p-map** and **t-map**.

**Question 12.** How does the activation compare to what you observed earlier using the Gamma model? Are there places where the observed activation has greater statistical significance?

3. Now select the **HDR** plot and click **Plot**.

**Question X.** Describe the estimated hemodynamic response. Does it resemble the gamma function? Give an example of an experimental paradigm when the FIR model might be better suited to the analysis, and the gamma model might lead to erroneous results.

## IV. Identifying sources of noise and their contribution to fMRI signals

In Lab 1, Lab 2, and Lab 3, you have already seen examples of how noise in fMRI data affects your ability to detect neuronal activation. Here we examine how different methods of modeling noise improve the fit of our statistical model, and hence improve our ability to detect expected signal changes associated with our experimental paradigm.

We will investigate detrending and whitening as two methods of modeling the noise. You have seen detrending earlier in the lab when we examined the characteristics of temporal autocorrelation. With *whitening (or the first-order autoregressive model, AR(1))*, we assume that the error at each time point,  $\epsilon_i$ , is a combination of the error from the previous time point,  $\epsilon_{i-1}$ , and white noise,  $\chi_i$ . The equation that relates the errors is thus  $\epsilon_i = \rho\epsilon_{i-1} + \chi_i$ , where  $\rho$  is a parameter between -1 and 1. This takes into account correlation between the noise at different time points. (For more information, consult Worsley's chapter in your assigned reading.)

We will revisit the phantom and human resting data sets to gain insight into the properties of our noise. We will examine the *false positive rate*, the probability that we detect a significant response using our analysis when no stimulus was applied. By plotting the actual false positive rate versus the theoretical false positive rate, we can see how different methods of noise modeling can help us be more confident in our interpretation of the analysis results.

1. Reset the parameters and select the **Phantom** data set.
2. Make sure that **motion correction**, **spatial filtering**, and **detrending** are off, and that the **gamma** model and **white noise** are selected for the signal model and noise model, respectively.
3. Click **Fit** in the **Statistical Model** box to perform the analysis.
4. Output the false positive rate into a variable by selecting **False Positive Rate** in the list of plots and clicking **Plot**. If you now type `global fpr` in MATLAB®, you can see the false positive rates from 0 to 1 in the `fpr` vector. Copy the vector into another variable of your choice (e.g. `fpr1 = fpr`); you will need it for graphing later.
5. Now select the **Human Resting** data set, and repeat steps 3-5 four times, once with each set of statistical modeling parameters listed below. To specify whitening, select the **AR(1)** noise model.
  - Gamma model, no detrending, white noise (no whitening)
  - FIR model, no detrending, white noise
  - Gamma model, detrend (polynomial order 2), white noise
  - Gamma model, detrend (polynomial order 2), whitening

**Question 14.** Plot the false positive rate graphs for all five sets of parameters on one graph. How do detrending and whitening affect the false positive rate? What is the difference between gamma and FIR?

Now we will examine the human experimental data set, which contains both noise and the signal resulting from the experimental paradigm and.

1. Reset the parameters and select the **Human Experimental** data set.
2. Make sure that **motion correction**, **spatial filtering**, and **detrending** are off, and that the **gamma** model and **white noise** are selected for the signal model and noise model, respectively.
3. Click **Fit** in the **Statistical Model** box to perform the analysis.
4. Examine the statistical map (either the t-map or p-map).
5. Repeat steps 3-4 twice, once with each set of parameters listed below.
  1. Gamma, detrend (order 2), white noise
  2. Gamma, detrend (order 2), AR(1)

**Question 15.** How do detrending and whitening affect the activation map?

## V. Understanding motion correction and spatial filtering

### 1. Motion Correction

We have so far examined the effects of detrending and whitening. Another source of noise in fMRI data is head motion. To attempt to minimize the impact of motion on our ability to detect significant information, we add a preprocessing step, *motion correction*, that tries to correct such movement within a session.

In motion correction, one volume time point in the session is designated as the reference, and all of the other volumes are rotated and shifted to try and match the position of the reference. Such a rotation and translation (shift) is called a *rigid-body motion*, as it preserves the shape of the object. For each scan, a *cost function* is computed as a metric of how well the scan aligns with the reference timepoint. To align the timepoints, the magnitude and direction of rotation and translation are determined by optimizing the cost function.

Because each scan can be rotated three different ways (pitch, yaw, and roll) and translated in three dimensions, there are six total axes to consider in aligning two scans, which form a space where each point represents a different rigid-body

motion. Since the space of all possible rigid-body motions is quite large, we use *search algorithms* for finding the one that optimizes the cost function. One commonly used algorithm (which is used in this lab) is *gradient descent*.

Using the tools in this lab you can explore the effects of choosing different cost functions, algorithms, and reference points for motion correction on your own if you wish. We will now look at how motion correction affects our analysis results.

1. Reset the parameters and select the **Human Experimental** data set.
2. Click on **Std Dev** to bring up the standard deviation map.
3. Turn on **motion correction** and click on **Std Dev** again to bring up the standard deviation map after motion correction.

**Question 16.** How does the standard deviation differ between the motion-corrected and uncorrected data? You may toggle between the two by clicking on **Recall** in the **Autocorr** box and selecting the appropriate parameters.

**Question 17.** Repeat the same comparison with activation maps (either the t-map or the p-map made from motion-corrected and uncorrected data). How do the activation maps differ?

1. Reset the parameters and select the **Phantom** data set.
2. Turn on **motion correction** and click on **Plot** in the **Motion Correction** box to show the estimated movement.

**Question 18.** The phantom did not stay completely still, according to the graph of the estimated movement. What do you think caused the apparent motion?

## 2. Spatial Filtering

Now we turn to *spatial filtering*, another preprocessing step that can aid in our analysis. Spatial filtering basically blurs the image; the main reasons for doing it before analysis are that it increases signal-to-noise ratio, and that the validity of certain steps in statistical analyses may require the data to be smooth.

One can filter the image with different *kernels*, or matrices that are convolved spatially with the data. We have chosen as our default the *Gaussian filter*, a good general-purpose filter for smoothing. The width of the kernel at half its maximum

point (full width half max, or FWHM) is typically used to as a measurement of its size.

For more in-depth information about spatial filtering, consult Stephen Smith's chapter in your readings.

1. Reset the parameters and select the **Human Experimental** data set. Make sure **spatial filtering** is off.
2. Make sure **motion correction** and **detrending** are off, and the signal model and noise model are set to **gamma** and **white noise**, respectively.
3. Perform the analysis by clicking **Fit** in the **Statistical Model** box and examine the **t-map**.
4. Now turn on **spatial filtering** with the checkbox in the **Spatial Filtering** box and specify a **filter width** of 4 mm FWHM. Repeat steps 2 and 3.
5. Now change the **filter width** to 8 mm FWHM, and repeat steps 2 and 3.

**Question 19.** How do the activation maps differ with and without spatial filtering? What about with different filter widths?