SPM99 Introduction

Summary:
SPM is a software package designed to analyze functional neuroimaging data. Its primary goal is to produce a statistically meaningful comparison between groups of images. However, every step of the analysis path is important to realize this goal, so SPM has developed into a comprehensive image analysis package, with modules to help guide the flow of data.

The software is written in Matlab, and is distributed freely by its authors, the Wellcome Department of Cognitive Neurology in London. Since they are a research group, SPM99 is officially unsupported, although a surprising amount of support is in fact available.

The major steps of data analysis include:
I. Data Acquisition
II. Pre-processing
III. Model Estimation
IV. Analysis of Results
V. Article sent to *Nature*

"SPM" stands for Statistical Parametric Mapping, which is the main output of the software. SPM99 seperately examines every pixel location across all images, and computes a parametric map containing a parameterized value at each pixel location. The parametric map is a form of data reduction, condensing information from a number of individual scans into a single image volume which can be more easily viewed and interpreted. The parameterized value is generally some form of a Student's t-test estimating the likelihood that a comparison of two image groups matches a given model that explains their possible differences.

Getting started, Getting help:
Most of the SPM resources are available online. There are links to some of the more helpful sources at:
http://psyphz.psych.wisc.edu/~oakes/spm/spm_resources.html

Images and Statistics:
The main issue that SPM99 addresses is the thorny problem of comparing images, or groups of images, in a statistically meaningful way. This is a problem because the underlying *distribution of variance* is unknown in most neuroimaging techniques. Without knowing this distribution, one cannot select any of the popular statistical distribution models (e.g. Poisson, normal, Gaussian), since these generally assume independent data points.

The individual data points (pixels or voxels) in most neuroimaging modalities (PET, fMRI, EEG, MEG, etc.) are heavily correlated with neighboring pixels. This means that a correction for multiple comparisons is needed. The traditional way of doing this is to use some version of a *Bonferroni correction*. However, due to the large number of pixels involved, a straightforward implementation would severely reduce the estimated number of degrees of freedom, yielding an overly conservative estimate of the t-value needed to achieve a significant result. As a matter of fact, the t-value resulting from a Bonferroni correction of neuroimaging data is generally so conservative that you would *never* get a significant result.

The major contribution of the SPM99 authors was to figure out a statistically valid approach that was not hobbled by the overly conservative assumptions of a standard Bonferroni correction. To the extent that the image data approximate a Random Gaussian Field (RGF), the theories behind RGFs can be used to yield valid statistical comparisons of images. This assumption is assured by applying a Gaussian smoothing filter to the image data in the Pre-processing stages.

Data Analysis Path:
Most image data analysis will follow the same general path. Within each step there is room for a large degree of customization.

I. Data Acquisition
This has been included as an item in the data analysis path because it is crucial that the data are acquired in a way that the experimental hypothesis can be addressed. This includes the experimental design as well as technical questions of modality, acquisition parameters, and reconstruction. The data fed into SPM99 should be as high-quality as possible.

II. Pre-processing
This stage includes several steps, all of which are aimed at massaging the data so it is suitable to be statistically analyzed by SPM99. The first several steps put each image volume into a standardized spatial reference frame. The last pre-processing step applies a Gaussian spatial filter. There are some special considerations for different imaging modalities, particularly for fMRI.

A. Realignment (motion correction)

This step is designed to coregister a series of image volumes of the *same brain* to a single representative volume. This is most frequently used to correct for small movements in a fMRI acquisition series.

B. Spatial normalization (a.k.a. Talairaching)
This step attempts to put all image volumes into the same spatial coordinate system. This provides a way to compare data from similar locations in different brains. This is a crucial step and generates more problems (and subsequent questions) than any other phase of SPM analysis. Most problems can be resolved by starting with well-behaved data, oriented in the axial format expected by SPM.

All image data are coregistered to a standard template. This was formerly based on the so-called Talairach brain, but has since been superceded by the MNI brain, which is a composite of approximately 305 individual brains (and counting!). The idea is to match each data volume to this template, so that specific anatomic structures (e.g. amygdala) will occupy the same voxels and can be compared across image volumes.

You must select a template that is as similar to your data as possible. There are separate templates provided with SPM99 for T1-weighted MRI, BOLD-signal fMRI, and [^{15}O]-H_2O PET. There are additional templates at the LfAN for [^{18}F]-FDG and for our own BOLD-fMRI data, which has a dropout artifact not present in the SPM99 template. Your success in achieving a good spatial normalization will be directly related to how well your data match the template you select. If your data do not match very well, you should consider creating your own template.

C. Gaussian smoothing

This step applies a Gaussian smoothing filter to the data. There are two purposes for this. The first is to make sure that the image data have the characteristics of a Random Gaussian Field in order that the statistical assumptions of SPM99 are valid. The second purpose is to recognize that the spatial normalization step is not perfect, and to correct for modest imprecisions in this process by smoothing the data so nearby voxels share more information.

The rule of thumb is that the Gaussian filter should be twice the resolution of your data. There is some confusion about this point, since true resolution is not the same as the size of your pixels. The size of your reconstructed pixels is fairly arbitrary, but the resolution is not. However, it is difficult to determine the precise resolution in most cases, since it is related to the modality, the scanner characteristics, the object being scanned, and the Signal:Noise ratio of the data. Hence, it is usually assumed that the data are reconstructed with pixel sizes similar to the dimensions of the resolution. This provides a good starting point, but some additional thought should go into determing the size of the Gaussian filter used on your data.

III. Model Estimation

This is the stage where you tell SPM99 what pattern you think activated areas should show. Usually this pattern corresponds to the groups designated for your images, e.g. normal/depressed, pre-treatment/post-treatment, stimulation/no stimulation, happy pictures/sad pictures. There are a myriad of options available, and this is where most people need help in wading through the menus and selecting appropriate analysis strategies.

This is the main analysis step, yielding the Student's t-test parametric maps and other goodies.

IV. Analysis of Results

This is where you see what happened. There are many ways to view your results, both from within SPM99 as well as using other programs. The two major issues are (i) figuring out what results are truly significant, and (ii) determining the precise anatomic location of a purported activation.

Several factors are involved in determining if a pixel or group of pixels reported as being above the threshold (activated) actually means anything. The first factor is the significance level you select; different values can be appropriate for different situations. Another factor has to do with the believability of an activated region; for instance, if your fMRI data show a lot of activations near the edge of the brain, some of these are probably motion artifacts.

Determing the precise anatomic location is not as easy it sounds, especially considering that most data are supposedly in a normalized coordinate system. There is a fair amount of variability between individual brains, making most spatial normalization procedures an estimate at best. It can be difficult to assign an activation to a specific anatomic structure. The most precise way to specify an activation is to see where it falls on *every subject's* high-resolution MRI, and to make a statement like "the activation occurred in the amygdala of 9 out of 12 of the subjects". However, this is not very satisfying, as we all want an unambiguous location. Preferably in a structure we hypothesized about prior to collecting the data…

V. Article sent to *Nature*

SPM99 can email your results directly to *Nature*, *Science*, and the *Journal of Erroneous Results* for publication. Do **not** hit this button until you are sure you like your results!