FMRI Analysis Methods: Classics and New Trends

Robert W Cox robertcox@mail.nih.gov
National Institute of Mental Health, Bethesda MD USA

Take Home Messages (for the *really* impatient)

- The classic linear regression model approach for analysis of task-based FMRI time series is robust and has been made easy-ish to use in various widely distributed software packages.
- Brain connectivity/network analyses (task- *and* resting-state) have become very widespread. There are many approaches, and all must be applied with care and caution.
- New types of data require new types of analysis—fusion of EEG and FMRI, fusion of DTI and FMRI, and very short TR whole brain FMRI are examples where the "best" ways to analyze such datasets are still not apparent.

Prolegomenon

FMRI data analysis is usually divided into 2 distinct stages: analysis of each individual subject's time series of images, followed by group (inter-subject) analysis of the results from the individuals. It is primarily the time series analysis that provides the greatest scope for flexibility and imagination. The goal is to wring some neurally-relevant information from the artifact- and noise-laden image sequence. However, innovations are also being made in group-level analyses, to get the most power from the limited number of subjects in a typical FMRI study, and to allow the results of more complicated first-level analyses to be combined and contrasted.

In what follows, I will elide virtually all details and exceptions (which as usual are matters for endless argument) and will paint with the broadest of brushes. No doubt some of the "trendy" methods will evolve to become beloved classics, others will become niche tools, and some will fade away to obscurity or be subsumed by the results of new insights.

The Classic Age (which has not yet passed away utterly)

In the beginning, the *t*-test and the correlation methods created the FMRI world. The *t*-test method simply took the EPI signal values from the "task on" phases as one sample, the EPI signal values from the "task off" phases as the second sample, and then computed a 2-sample *t*-statistic to test the null hypothesis that the mean of the "on" and "off" signals were equal. The correlation method that quickly followed introduced the "ideal" function, which is a model time series that allows for hemodynamic delay and smoothing, and simply computed the Pearson correlation coefficient between this ideal time series and each voxel's EPI signal time series.

Within a few years, the correlation method was generalized to allow for multiple ideal functions, corresponding to multiple task/stimulus conditions, and correlation was replaced with linear regression to compute fit coefficients for each component of the model—thus the "GLM" (general linear model) sprung into being. The GLM soon became the workhorse for task-based FMRI analyses, and is still a dominant force in the FMRI literature. The existence of several (relatively) easy-to-use software packages that implement GLM analyses makes using this type of first-level analysis fairly painless.

At the group level, the problem of combining brain maps from different subjects was solved in two ways. The first technique was simply to extract the average first-level results from a variety of anatomically-defined regions of interest (ROIs), and then use those few values in some standard statistical package (often just a spreadsheet). The second technique was to approximately align each subject's anatomical reference volume to a template volume, apply the same spatial transformation to the first-level results, and then carry out an appropriate voxel-wise statistical analysis on these "spatially normalized" datasets. For this purpose, a 12-parameter affine transformation was almost always used in the "classic" software packages, and is still the default transformation. (Nonlinear warping to a template has become popular, also.) At first, the target template was usually based on the Talairach-Tournoux atlas (inherited from the PET literature), but that was mostly superseded by the MNI-152 template created in Montreal.

Most often, the "appropriate" second-level analysis was the 1- or 2-sample *t*-test, as applicable—basically, testing if the mean "activation" (or activation difference) is nonzero at each brain voxel. Testing these results for significance was complicated by the number of voxels being analyzed—usually in the neighborhood of 100,000. The usual lenitive for this "curse of multiple comparisons" was a combination of spatial correlation and contiguity. Spatial correlation in the noise (induced largely by spatial smoothing of the EPI data) implies that there many fewer than 10^5 *independent* comparisons to be made. We also expect spatial contiguity in the results—colloquially speaking, we are hunting for "blobs" of activation, not scattered random-looking voxel patterns. Tools are available to "correct" the statistical significance of a *t*-statistic (say) from a per-voxel level to a per-contiguous-blob level.

The Trendy Age

The diversity of experiments, imaging methods, and analysis ideas multiplied in the second decade of FMRI. New types of imaging data and new types of neuropsychological experiment designs often require new analysis tools. In addition, new ways of thinking about the brain have prompted researchers to devise novel techniques to glean information from FMRI datasets.

For all of these novel techniques, there are serious issues of reliability, repeatability, and sensitivity to assumptions. These issues are often exacerbated by reluctance to share datasets which required a great deal of effort to acquire, and by complex software codes that are difficult to reproduce or verify, or require finesse to use well. "Extraordinary claims require extraordinary evidence"—and so very strong claims about the power of a new method require very strong evidence.

Connectivity: Most evident in this *novus ordo seclorum* is the upsurge of analyses to illuminate issues related to inter-regional brain "connectivity" or "networks". A partial list of such methods shows the flowering of possibilities that have blossomed in these attempts:

- Task-based connectivity analyses, such as Psycho-Physiological Interaction (PPI), Granger causality, and Stuctural Vector AutoRegression (SVAR); these methods are usually built around some sort of fitting to a linear model.
- Nonlinear decoding of "neural" signals from EPI time series datasets to then infer brain network information, including methods such as Dynamic Causal Modeling and Cubature Kalman filters.
- Resting-state connectivity analyses, included seed-based methods, "all-with-all" correlations, and spatio-temporal factorization of the data into components; the number of methods applied to resting-state FMRI datasets is mind-boggling (at least to my mind).
- Examination of the temporal fluctuations of connectivity patterns, sometimes in association with tasks and measurements of task performance.
- Graph theory concepts and metrics applied to brain networks (however derived), such as "small worlds" and "rich clubs".
- Combining genetics and connectivity measures in disorders such as autism and schizophrenia.
- Fusion of anatomical (DTI-based tractography) and functional connectivity results, including highly interactive visualization tools (cf. Fig. 1), and incorporating white matter "quality" as measured by DWI.

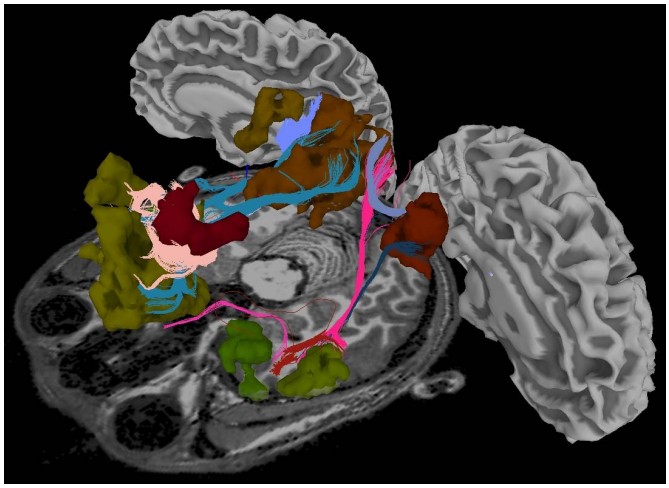

Figure 1. Screen snapshot of interactive visualization of DTI-base white matter tracts with functionally defined blobs as the tractography seeds and targets.

[*Strong personal opinions in this paragaph*] An important issue in all connectivity and network studies based on FMRI data is that the analyses are largely or completely self-referential—that is, the results are often only loosely tied to external (non-FMRI) measurements. In task-based FMRI, the activation map results are strongly tied to the task timings and task categories. At the other end of the spectrum, the results in resting-state FMRI often have no external associations. (Task-based connectivity results lie in between, with a combination of external and internal references.) The upshot is that task-based FMRI activation maps are robust to a number of processing strategy choices, but resting-state FMRI connectivity/network maps can be very sensitive to such choices. The recent back-and-forth in papers and discussions about whether/when it is appropriate to use the global (whole brain) mean signal as a nuisance regressor in resting-state FMRI is a salient example—to be blunt, it is my opinion that this practice is a bad idea. The future of FMRI-based connectivity experiments and analyses must include a large effort to tie the results to other data, either at the temporal level (e.g., eye-tracking, simultaneous EEG) or at least at the subject level (e.g., genetics, disease severity).

Trends prompted by new types of data: New types (or combinations) of imaging data require novel analysis methods. Some salient examples:

- Fusion of EEG and FMRI—technically quite difficult, this combination has long been desired as a way to combine the fine temporal resolution of EEG with the fine spatial resolution of FMRI. As the technical challenges become resolved, what comes to the fore are the analysis challenges of fusing, in a profound way, two very different types of data, each with its own artifacts (plus the combined artifacts).
- Correcting for extraneous physiology-induced signal fluctuations ("noise")—here, the new types of data are multi-echo EPI and multi-band EPI.
  - Multi-echo EPI yields 2 or 3 images with each RF shot, each image at a different TE. The BOLD effect should depend strongly on TE, but other physiological effect (respiration, heartbeat) should not. This difference makes it possible to filter out the physiological noise without directly assuming a regression model based on external monitoring of the subject (e.g., RETRICOR)—such models often do not work well.
  - Multi-band EPI also yields multiple images with each RF shot, each image being at different slice location. With this technology, it is possible to cover the whole cerebrum fast enough so that the heartbeat is not aliased into the much lower BOLD effect frequency band; thus, the physiological artifacts can largely be bandpassed out.
  - Another issue that arises with very rapid TR is the adequacy of the current simple methods for generating the ideal functions for task-based FMRI analysis. When TR is 2+ seconds, the rise and fall times of the BOLD effect (4-5 seconds) can be crudely modeled, since the data doesn't support complex shape fitting. As the TR drops to under 1 s and potentially down to perhaps 200 ms, the fundamentals of FMRI processing will require re-thinking.

Other Trends: Here, I limit myself to trends in *processing methods*, versus changes in analysis prompted by changes in image acquisition. Some prominent examples:

- Mining neuroimaging literature—there are many many functional neuroimaging results in the scientific literature. Extracting results and fusing them in "data mining" (better called "data wildcatting" IMHO) or in statistical meta-analyses is a growing field.
- "Mind Reading" FMRI—aka Multi Voxel Pattern Analysis (MVPA)—trying to get information out of the FMRI times series about what is happening in the subject's brain at any moment. There are many such pattern analysis algorithms, some of which have demonstrated remarkable discriminations. Of particular interest is potential application to clinical populations, as in discriminating between the various degrees of "coma-like" states, from vegetative to locked-in.

- Whole brain activation from *long* FMRI time series—recent work has shown that acquiring much more data than usual in a single subject (over 8 hours at TR=2 s) reveals that virtually the entire cortical gray matter is "activated"; that is, shows signal changes in lockstep with the task timing. Besides the obvious difficulties this result raises for neuroscientific interpretation, it also brings up some issues in FMRI data acquisition and analysis:
  - Can FMRI data acquisition and/or pre-processing be improved enough to see this effect in shorter datasets?
  - Would this result show up in very short TR datasets over more reasonable time spans?
- Surface and "grayordinate" datasets—the cerebral cortex can be approximated as a pair of folded up 2D manifolds; hence, the popularity of software that allow the visualization of FMRI data and results projected onto these surfaces. The Human Connectome Project has taken this one step farther, to include the more solid gray matter structures (such as basal ganglia) in a combined 2D (surface nodes) plus 3D (solid voxels) "grayordinate" system. This approach will undoubtedly become more popular, as it provides a convenient way to coordinate-ize the entire cerebrum.
- Multivariate and Meta group analyses—classic FMRI group analysis is ANOVA-like (sometimes with subject-level covariates), wherein the variance/covariance structure model of the noise in the data is assumed to be very simple, and the goal of the analysis is just to estimate the mean difference between groups (e.g.).
  - More complex "mixed effect" models also attempt to estimate the structure of the noise as well, which requires solving a nonlinear optimization problem at each voxel.
  - Using the *statistics* as well as the model fit parameters from the first-level analysis, as a measure of the reliability of the first-level results in each subject, can be done by incorporating the techniques of meta-analysis. Second-level group analysis is really a form of statistical meta-analysis, in the sense that it involves takes as the "data" the results of a prior statistical analysis—just as do meta-analyses combining the results of several clinical trials.
  - More complex experiments often give multiple functionally relevant parameters per voxel, per subject. Analyzing one such parameter (or one linear combination of parameters) at a time is potentially overlooking more complex inter-group differences. Multivariate analyses take all the first-level estimates into account, and can deal with these complicated scenarios. Such methods also have the potential for fusing results from MEG and FMRI, for example.