BAYESIAN STATISTICS 6, pp. 000--000
J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.)

Functional Magnetic Resonance Imaging and Spatio-temporal Inference

CHRISTOPHER R. GENOVESE
Carnegie Mellon University

SUMMARY

With ever greater computational power at our fingertips, statisticians can now approach problems of unprecedented complexity, both in terms of the mechanics of modeling and the substantive content of inferences to be drawn from models. One class of such problems, of particular relevance in many current applications, is modeling and drawing inferences from data derived from a spatio-temporal process. This paper illustrates some of the new challenges that can arise in this context, using functional Magnetic Resonance Imaging (fMRI) as an exemplar. The paper begins with a detailed description of a hierarchical model for fMRI data that focuses on the temporal structure only. The model is applied to fMRI data from a recent neuroscientific study of the eye-movement system in the human brain. The discussion then extends to cover a selection of open problems that are of vital interest to neuroscientists. In fMRI, each of these problems can be approached by extending or applying the hierarchical model in different ways. However, these problems can also be abstracted beyond fMRI and indeed exemplify broader methods that can be useful in a wide range of applications.

Keywords: SPATIO-TEMPORAL INFERENCE, HIERARCHICAL MODELS, FUNCTIONAL NEUROIMAGING, OCULOMOTOR SYSTEM.

1. INTRODUCTION

As the available computational power has grown to the inexorable drum beat of Moore's Law, it has become possible for statisticians to attack applied problems of greater complexity than could be dreamt in the past. A general class of problems for which this new-found power can be put to good use arises when drawing inferences from data realized by a process that varies over both space and time. Instead of modeling a single time-series or random field, it is necessary to model a collection of related time-series or a time-varying random field. More than just the curse of dimensionality is at work here: potentially intricate interactions across space and time can be difficult to account for and hard to interpret in a substantive context.

My goal in this paper is to offer exemplars of this distinction that illustrate some of the new challenges involved in spatio-temporal inference. I will describe inferential problems that I am currently thinking about in the context of functional Magnetic Resonance Imaging (fMRI). Most of these ideas are still work in progress, if even that, but my hope here is that the range of new problems will provide a first step towards new thinking about spatio-temporal inference in other applications.

It is worth noting that spatio-temporal inference involves many practical problems as well as conceptual and theoretical ones. The data sets are large, and the methods used to make inferences are often computationally intensive. In fMRI, for example, the data from a single experiment on a single subject contain from 100,000 to 400,000 time series, along with various behavioral and physiological data acquired during the experiment. It is common for each such data set to require a day of processing on the fastest computers readily available. While these practical concerns are not to be overlooked, they have been discussed at length elsewhere (Eddy et al, 1996), and I will ignore them here. My focus is on the conceptual side of the problems with an eye towards transferring ideas from fMRI to other areas.

In the next section, I briefly describe the background on fMRI that will be helpful for understanding the problems discussed later. In section 3, I discuss the current frontier in drawing inferences from fMRI data, highlighting a hierarchical model for fMRI time-series that I have developed. In sections 4 through 7, I cross the frontier and describe some of the problems that need to be solved to advance the scientific objectives of researchers using fMRI. Section 8 closes with general musings on spatio-temporal inference.

2. BACKGROUND ON FMRI

Magnetic Resonance Imaging (MR Imaging or MRI) is a technique for acquiring detailed pictures of the inside of a body. A patient is placed in a strong magnetic field which effectively magnetizes the tissue. Carefully modulating the magnetic field affects a specific type of atomic nucleus, such as hydrogen, in the patient's tissue and produces a measurable signal in a nearby receiver coil. The greater the density of the given nucleus in the imaged volume, the stronger the signal. An MRI scanner is thus a very expensive tool for determining how much hydrogen is contained in a volume of tissue. By manipulating the magnetic fields, however, the scanner can effectively acquire such measurements of hydrogen content for many small volume elements, or *voxels*. The result is a high-resolution, three-dimensional image of the tissue volume with each voxel shaded by the strength of the measured signal. These are called structural images because they show the anatomical structure of the tissue in a way that is useful for clinical radiology.

In *functional* Magnetic Resonance Imaging (fMRI), this technology is applied in an attempt to infer how the brain works. While we use MR images of the brain, we are not interested in the images *per se* but rather in subtle changes in the measured signal across time. These changes are caused by physiological effects related to neural activity (Belliveau et al., 1992; Kwong et al., 1992; Ogawa et al., 1992).

During an fMRI experiment, a subject performs a carefully arranged sequence of behavioral tasks while MR images of the subject's brain are acquired at regular intervals. The resulting data set is a time-series of three-dimensional images. Considering a single voxel, the time-series of measurements associated with that voxel tells something about what is happening in that small volume of brain tissue during the experiment. Figure 1 shows one such time series with the experimental design indicated on the time axis. This design consists of two conditions: one in which the subject engages in some task of interest (in this case, tapping index finger to thumb while staring at a dot) and one in which the subject engages in a control task (in this case, no motion while staring at a dot). It is hoped that the control task involves all aspects of processing in the task of interest except those being studied. In the figure, the measured signal rises slightly during the task of interest (with some lag) and then drops during the control. This systematic change in the signal is thought to be a by-product of concentrated neural activity within the voxel, the result of a physiological blood-flow (hemodynamic) response.

The scientific problem of fMRI is to determine from a spatio-temporal pattern of task-related changes how the brain subserves various processes of interest, such as language, cognition, memory, motor control, and so forth. The statistical problem of fMRI is to infer the pattern of task-related changes (if any) from the data. The true challenge here lies at the interface between these two objectives.

Although a plethora of statistical methods have been applied to fMRI data, the standard (and still most common) methods are based on classification (Lange, 1996). Each voxel is classified as *active* or *inactive* via a statistical hypothesis test; the distinction lies in whether the voxel time-series exhibits systematic task related changes like those in Figure 1. For example, one might use the t-test to compare the average signal level in the two conditions. Genovese (1997) argues that although classification-based approaches can provide effective identifications of what ''lights up'' in the brain, they cannot address many questions of critical scientific interest. This argument motivates many of the issues covered in this paper.

3. A MODEL FOR THE VOXEL TIME SERIES IN FMRI DATA

Functional MRI has become an industry of sorts, with thousands of researchers in a wide range of disciplines either applying or developing imaging techniques. Over the past two years, acquisition and reconstruction methods have improved tremendously, with corresponding gains in the achievable resolution and signal-to-noise ratio of the data. For behavioral and neuro-scientists, fMRI has become an indispensable tool; indeed, fMRI data is being applied to study a vast range of scientific questions. The statistical methods for analyzing fMRI data have also improved during this time as researchers have accumulated experience with the data and applied an wide array of available techniques to the problem of identifying task-related signal changes. But the scientific scope of fMRI studies has out-paced the available statistical techniques, and researchers are beginning to ask new questions: about the operation of the brain as a complex system, about the connectivity and interaction of different processing nodes, about the spatial clustering and temporal dynamics of activity. These questions lie at the boundary of what the data can answer, but they are approachable. Most current statistical techniques focus on where activity occurs but offer few formal mechanisms for inferring the structure and dynamics of the underlying system. I believe that the long-term scientific potential of fMRI depends on being able to draw inferences about these new questions.

So how do we take the next step? One approach is to construct a detailed model of the data that describes the system about which inferences are to be made. This makes it possible to identify the location of activity better than current methods and also provides an inferential framework from which to attack new questions. This is the approach begun in (Genovese, 1997) which develops a hierarchical model for the voxel time-series, discussed further below. This model describes the temporal structure of the data in detail but treats the signals at every voxel as independent, which ignores spatial relationships in the data. The approach here was to begin by capturing the most important structure and refine from there. As such, this model takes a spatio-temporal approach to fMRI data, where we think of temporal structure varying over space, as opposed to what we might call a temporo-spatial approach (spatial structure varying over time) or a more general spatial-temporal approach (spatial and temporal structure intertwined). In fMRI, a spatio-temporal perspective seems most natural because the temporal pattern of responses to the experimental manipulation is what characterizes the brain's basic function. The model is discussed in detail in (Genovese, 1997) along with a thorough discussion of the objectives; a case study using the model can be found in (Genovese & Sweeney, 1998).

3.1. *Sources of Variation*

The data from an fMRI experiment are the results of a very complex process with both biological and technological sources of variation. Some of these sources can be identified, but many have not been. A critical source of variation is subject movement. If the subject's brain did not move during the experiment, then the voxel time-series would consist of measurements of the same volume of tissue. When the subject moves, this mapping between image voxel and tissue voxel breaks down; with voxels ranging from 1--5mm on a side, even a small movement can cause abrupt signal changes much larger in magnitude than task-related changes. Fortunately, by combining experimental set-up (e.g., head restraint), subject training, and post-processing, the impact of movement can be kept manageable. Beyond explicit head movement, the physiological processes that occur in living tissue can cause significant fluctuations in the measured MR signal. For example, the subject's cardiac and respiratory cycle cause systematic signal changes whose character varies across the brain. Other known sources of variation include mis-calibrations in the scanner and thermal noise in the data. But whether identified or not, we would like to account for these sources of variation as they manifest themselves in the data. A typical voxel time-series in fMRI data exhibits a variety of irregular changes, not all of which are attributable to a specific cause. The most salient of these is signal drift, large-magnitude signal changes over the course of the experiment. These drifts are generally smooth, although abrupt localized changes can occur, but they can be inhomogeneous over space and time. The magnitude of the drifts can be an order of magnitude larger than the task-related signal changes of interest. In addition, the data exhibit spatial and temporal correlation, often highly nonlocal, and outlying or even missing data are common.

The variation of interest in the MR signal is the impact of neural activity. Concentrated activity in the brain leads to a localized increase in the flow of oxygenated blood to the area of activity, called the hemodynamic response. Because of the different magnetic properties between oxygenated and de-oxygenated blood, the hemodynamic response manifests itself as a slight change in the measured signal. Where oxygenated blood-flow increases, there is a slight rise in the signal. The time-scale of the hemodynamic response is much slower than the underlying neural activity and is also susceptible to the presence of large blood-vessels and other artifacts. The typical hemodynamic signal change for a single period of task performance becomes detectable after a lag of approximately 500ms and begins a slow rise on the order of 6s. If task performance continues, the signal levels off to a plateau where it remains until the task ends. After some delay, the signal then begins to decay, dropping slowly back to baseline in on the order of 10-15s. An undershoot is frequently observed, where the signal drops below baseline before returning. (A smaller undershoot may also occur before signal rise, although that is still controversial.)

3.2. *The Hierarchy*

A fMRI voxel time-series consists of T measurements of the MR signal acquired over a regularly spaced grid of times $t_j = t_0 + j^\cdot$ for $j = 0, \ldots, T-1$. An fMRI experiment consists of a sequence of tasks, each assigned to a specific condition c and performed for some period of time during the imaging period. Each experimental condition c is replicated over several *epochs* labeled $k = 1, \ldots, K_c$. (For example, consider a very simple experimental design in which the subject repeats the following five times: finger tapping for 30s followed by rest for 30s. This design has two conditions and each condition consists of 5 epochs, 30 seconds in length.)

The voxel time-series combines the sources of variation described above. If not constrained, the contributions from these different sources of variation would not be identifiable. The model provides constraints on these features based on substantive information about the underlying processes. The design objectives for the model are as follow:

- Make the parameterization interpretable in terms of the underlying processes,
- Use substantively based priors built on current empirical information,

- Build a framework that can evolve as new information comes to light.

In the model, all voxels are taken as independent, structurally and stochastically. This is almost certainly an optimistic assumption but is a reasonable short-term approximation that captures most of the interesting structure. The model parameters are arranged in groups, where each group relates to a single source of variation:

μ	Baseline level of signal in the absence of activity
Drift	Coefficients of drift profile in current basis
Response	Amplitude of response within an epoch ($\theta^{\text{Response}}_{c,k}$)
	Average amplitude of response in a condition ($\theta^{\text{Response}}_{c}$)
Shape	Shape of response curve (2--8 parameters $\boldsymbol{\theta}^{\text{Shape}}$)
Noise	Noise Precision

Parameter groups at each level of the hierarchy are taken as independent. Below, I use ρ, appropriately subscripted, for fixed hyperparameters; I give sample values after describing the model. Also, in describing the conditional distributions, I use ''...'' to refer to all parameters farther down the hierarchy and $\boldsymbol{\theta}_{-i}$ to refer to all parameters in a group except θ_i.

The likelihood is an additive model

$$Y(t_j) \;=\; \mu \;+\; d(t_j;\boldsymbol{\theta}^{\text{Drift}}) \;+\; a(t_j;\theta^{\text{Response}}_{c_j,k_j},\boldsymbol{\theta}^{\text{Shape}},\mu) \;+\; \epsilon(t_j;\boldsymbol{\theta}^{\text{Noise}}), \tag{1}$$

where c_j and k_j denote respectively the condition and epoch associated with time t_j. In the fits in this paper, the noise model is simplistic, i.i.d. Normal(0, $1/\boldsymbol{\theta}^{\text{Noise}}$), but I have used more complicated forms and am working on further extensions.

The next level of the hierarchy, which is optional, accounts for variations in the response amplitudes across epochs of the *same* condition.

$$\boldsymbol{\theta}^{\text{Response}}_{c,k} \mid \boldsymbol{\theta}^{\text{Response}}_{-(c,k)},\,\ldots \;\sim\; \pi_{\text{Epoch}}(\boldsymbol{\theta}^{\text{Response}}_{c,k} \mid \boldsymbol{\theta}^{\text{Response}}_{c}), \tag{2}$$

where

$$\pi_{\text{Epoch}}\left(\theta^{\text{Response}}_{c,k} \mid \theta^{\text{Response}}_{c}\right) = \begin{cases} N_{+}(\theta^{\text{Response}}_{c},\rho_0^2) & \text{if } \theta^{\text{Response}}_{c} > 0 \\ \text{point mass at } 0 & \text{otherwise.} \end{cases} \tag{3}$$

Here, N_+ denotes a truncated Normal; the response amplitudes are currently constrained to be non-negative, and they have a positive probability of being exactly 0.

The remaining levels of the hierarchy account for voxelwise variations in the basic parameters. First, the drift is modeled effectively as a smoothing spline, with a large number of regularly spaced knots.

$$\boldsymbol{\theta}^{\text{Drift}} \mid \theta^{\text{Noise}},\lambda,\,\ldots \;\sim\; A\exp(-Q(d)/2\lambda)) \tag{4}$$

$$\lambda \mid \theta^{\text{Noise}},\,\ldots \;\sim\; \text{Exponential}(1/\theta^{\text{Noise}}\lambda_0). \tag{5}$$

The quadratic form Q penalizes the weighted Sobolev norm of the drift profile as follows

$$Q(d) \propto -\left[a_n \left\|d(t)\right\|^2_{2,w} + a_c \left\|d''(t)\right\|^2_{2,w}\right]. \tag{6}$$

The constants a_n and a_c here are fixed and indicate the relative weight of the norm and curvature of the profile. The drift profile is parameterized so that it always has mean value 0; hence the norm of the profile directly measures the drift magnitude. The hyper-parameter λ controls the Q-smoothness of the profile; large λ corresponds to little smoothness. Second, the condition-average response amplitudes are taken as independent with a fixed mixture of an atom at 0 and a positive distribution covering the expected range of values:

$$\theta_c^{\text{Response}} \mid \boldsymbol{\theta}_{-c}^{\text{Response}}, \ldots \sim \eta \text{ point mass at } 0 + (1-\eta)\,\text{Gamma}(\rho_{1c}, \rho_{2c}) \tag{7}$$

The shape parameters describe different features of the hemodynamic response function; all of these are modeled as independent. While this is not necessarily the optimal choice, the parameterization was chosen to separate out the distinct features, and there is currently little substantive information available to suggest a dependence structure. The form of the shape priors are Gamma with parameters chosen based on current empirical evidence about the structure of the hemodynamic response.

$$\boldsymbol{\theta}^{\text{Shape}} \mid \theta^{\text{Noise}} \sim \prod_s \text{Gamma}(\rho_{3s}, \rho_{4s}) \tag{8}$$

Finally, the noise precision is given a proper and diffuse prior

$$\theta^{\text{Noise}} \sim \text{Inverse Gamma}(\rho_5, \rho_6). \tag{9}$$

The specific prior can be chosen to match available diagnostic summaries of the system noise level that are typically recorded every time the scanner is used. The noise distribution depends on the image acquisition scheme, the strength of the main magnet, and many other factors.

All the ρ hyper-parameters used above are fixed at specific values, chosen on the basis of the imaging scheme, the scanner, and the experimental paradigm. For standard echo-planar imaging, currently the most commonly used scheme at the University of Pittsburgh MR Research Center, typical default values are as follows: $\rho_{1c} = 1$ and $\rho_{2c} = 50.0$ for all conditions; $\rho_3 = (2, 4, 2, 4)$ and $\rho_4 = (4.3, 1, 4.3, 0.43)$ for the four-parameter activation model; and $\rho_5 = 1.6$, $\rho_6 = 100$ for the noise model. Relative weighings for the drift penalty are set to $a_n = 0.01$ and $a_c = 1.0$ by default; these are both scaled by the tuning hyper-parameter λ.

3.3. *Model Averaging*

A critical benefit of Bayesian inference for fMRI is that the qualitative structure of the model can be varied as part of the inferential procedure. This is particularly relevant for estimating the response amplitude parameters because for every condition, the true amplitude of the task-related signal changes is zero or negligible for the vast majority of voxels. If the ''full'' model were fit with positive amplitudes associated with every condition in every voxel, this would encourage small but usually artifactual estimated effects. The model's priors for the amplitudes account for this constraint by including an atom at 0. This atom induces a set of possible sub-models over which every subset of the experimental conditions is constrained to have no effect. Model averaging is required here to account for the uncertainty in which conditions evoke a response. The meaning of the parameters is retained across all sub-models.

In practice, I use two computational strategies to fit the model, depending on the needs of the investigator and the time constraints. The first, and by far the fastest, is to compute maximum posterior estimates with standard errors derived from the observed Fisher Information matrix. In this case, Bayes' Factors are estimated from the individual sub-model fits using the Laplace approximation to the normalizing constant (I have tried other methods with good effect but higher computational cost; see DiCiccio et al., 1997) The estimated factors are combined with specified prior probabilities on the sub-models to obtain model posterior probabilities. The second computational strategy is to use a Markov Chain Monte Carlo Simulation to estimate the full posterior of the parameters. In this case, the sub-models are integrated into a single chain using reversible jumps across models (Green, 1995).

3.4. *Example: The Oculomotor System*

The oculomotor system in the brain controls eye-movements. These simple motions require surprisingly complex and distributed processing. Two types of eye movements are commonly studied to understand the system and its implications for various clinical disorders:

1. Saccades---Sudden, sharp repositioning movements, and
2. Pursuit---Smooth tracking of a moving target.

These systems appear to be differentiated in the brain, and the specific nature of the differences offer insights into open questions of pure and clinical neuroscience.

Genovese and Sweeney (1997) provide background on the oculomotor system and a case-study of fMRI experiments devoted to understanding it. In these initial fMRI studies, the goal was to contrast the locations of the pursuit and saccade systems. Earlier work by Sweeney and his group found suggestions of systematic differences but could not conclusively separate the two systems. In their initial studies, using the model described above for analysis, Genovese and Sweeney are able to detect differences between pursuit and saccade that were previously obscured. Figure 2 shows a summary of selected results. The two panels show the posterior probabilities across the brain of non-zero task-related activation in the saccade and pursuit systems respectively, each overlaied on a mean functional image for a single slice and single subject. In this example, the pursuit and saccade activations appear to follow different sulci (folds), which suggests a structural difference in the regions. The differences detected in this study appear robust across slices and make anatomic sense (Sweeney, Personal Communication). Although the questions here are based on location of brain areas subserving these two eye-movement types, the future development of these studies will be directed at more complex issues such as the functional connectivity of the oculomotor system.

3.5. *Extensions*

The model as stated above is being extended in several ways to more closely match the properties of the data and the required inferential procedures. The noise model needs to account for spectral features caused by respiratory and cardiac affects, and it may be desirable to allow for temporal autocorrelation. There are also several viable response characterizations that can be empirically compared, and ANOVA-style models of response amplitudes are needed for some current designs. All of these are work in progress, and some of these efforts are described briefly below.

4. TWO SPATIAL PROBLEMS IN THE BRAIN

Two primary challenges arise when incorporating spatial structure into a model for fMRI data. First, the noise process can exhibit highly non-local and non-isotropic patterns of correlation. Second, the activation response can have a very similar structure across complicated local neighborhoods. These two challenges relate directly to two corresponding lines of research in spatial statistics and image processing: the specification of covariance structures for spatio-temporal random fields and hierarchical image models. Functional MRI adds some new twists to both of these problems.

4.1. *Nonlocal Dependence*

Different parts of the brain interact in myriad ways, by vascular, chemical, fluid, and electrical connections. These connections make it possible for irregular and non-local correlations to arise in fMRI data. For example, the brain moves non-rigidly as the heart beats, and breathing changes the blood pressure throughout the brain, although differentially through the vascular network. Such physiological effects are already known to manifest themselves in fMRI data (Hu et al., 1995); they add a variation to the voxel time series but also induce dependence that is not necessarily related to physical proximity. There is mounting empirical evidence that these non-local correlations are non-trivial (Weiskopff, Personal Communication), although the magnitude of spatial correlation depends strongly on the image acquisition scheme (Fitzgerald, 1996). These correlations may also induce task-related signal changes in locations that are not directly active.

Our understanding of these effects is improving, but it is clear that irregular and non-local correlations will need to be handled. The question becomes how best to incorporate them into our inferences. Current models for spatial and spatio-temporal processes are based primarily on local correlations, and those with non-trivial correlation at large lags tend to have very regular correlation structure. While it is always possible to directly model the correlation structure, this is a very high dimensional problem with little substantive information (currently) to provide priors or constraints. As a first step, we can assume that the spatial and temporal correlation structures are separable and use a smoother iteratively to estimate the spatial correlation structure. Further experience with many data sets and acquisition schemes will allow us to validate the separability assumption and to assess the importance of spatial correlations.

4.2. *Borrowing Strength*

Of more direct relevance to our objective is the structural dependence among different voxels. The shape of the hemodynamic response function, the magnitude of the response, and the impact of physiological variations all tend to be consistent across localized groups of voxels while varying substantially on a larger scale. These localized groups presumably represent regions with some common physiologic and/or functional properties, but their boundaries are convoluted and piecemeal in the brain. This is because the cortex, a thin layer of cells, is folded intricately throughout the brain and appears and disappears in the images as it meanders through the slices. There are many cases in which tissue type, vasculature, and functional response change on a millimeter scale, so there can be sharp changes across the region boundaries.

From a modeling standpoint, these dependencies induce spatial relationships among the model parameters associated with different voxels. Modeling these relationships increases the precision of inferences about the parameters because multiple voxels contribute information about features they have in common. Such a model also facilitates more global inferences about spatial patterns of task-related activity. The challenge is to find the ‘‘dependence neighborhoods’’ over which we can borrow strength in estimating the model parameters.

One straightforward approach involves segmenting each image into grey-matter (cortex), white-matter, and blood-vessel voxels using the baseline (time-averaged) signal magnitude along with physical constraints. This is motivated by the different imaging properties of these tissue types. Such a categorization does not depend on the task-related signal changes, but it is still a somewhat computationally intensive procedure. A plausible dependence neighborhood for a given voxel could then be defined by taking nearby voxels in the same category.

A better approach is to let the data (particularly the temporal changes in the signal) determine what the neighborhoods should be. This requires an adaptive partitioning of the voxels that is tied to the temporal structure of the signal at each voxel. The partition consists of sets of contiguous voxels whose union is the entire image. Voxels in the same partition element would share the response amplitude and shape parameters. The partition is tied to the model by making the contents of each set and the number of sets parameters of the model at the next level of the hierarchy. By tying the partition parameter to the response amplitude and shape only, the partitioning is guided by the temporal structure of the signal that relates to potential activation. Features such as the drift and the baseline signal can vary even across adjacent voxels for many reasons that are independent of the hemodynamic response to activation.

The principal methodological question here is how to deal with the partition parameter. The partition elements are three-dimensional and may be highly irregular; moreover, the vast majority of voxels should lie in a ‘‘no response’’ partition because they exhibit a negligible response to the experimental tasks. Johnson (1994,1995) proposed a method for image partitioning that is well-suited to this problem. The method adds a level to the hierarchy and includes the number N and descriptors $R_1, \ldots, R_N$ of the dependence neighborhoods as additional model parameters. Given a specification of the neighborhoods, we take $\boldsymbol{\theta}^{\text{Shape}}$ and $\boldsymbol{\theta}^{\text{Response}}$ for voxels in each neighborhood to be either constant (the simplest case) or to vary slightly about a common mean (the more accurate case). Because most of the voxels exhibit no activation at all, to make this method work in the fMRI context, the prior for the partition parameter used in (Johnson, 1994; Johnson, 1995) needs to be adjusted to encourage a moderate number of small partition elements.

5. INFERENCES ACROSS SUBJECTS

Although vast amounts of data are collected for each subject in an fMRI experiment, these data are effectively a single observation when it comes to generalizing results about the brain to a broader population. Cognitive psychologists and others who study high-level brain functions want to test their theories with fMRI data. But to do this, the results need to be consistent across a large number of subjects. The big problem here is variation in the topography of subjects' brains: while the functional organization of the brain is basically the same for everyone, the geometry of how the cortex is placed and folded varies substantially. Thus, there is no effective way to map between ‘‘brain-coordinates’’ among individuals. How then to combine results for individual subjects into a picture of how the brain works in general?

One method that is currently in widespread use is based on the Talairach atlas (Talairach and Tournoux, 1988) derived from a detailed study of a few brains. Selected gross measurements of the subject's brain are used to map the subjects brain onto the atlas coordinate system. The summary images from an fMRI analysis are typically then averaged in these coordinates. Large individual differences remain after the Talairach mapping, squandering hard won resolution. This approach is popular for lack of other convenient options, but Talairach averaging is coarse at best. A variety of computationally intensive approaches based on segmenting, projecting, or flattening the cortex from the brain images are raising hopes, but these are very hard problems.

Here, I propose three alternatives to control individual differences:

1. Restrict inferences to regions defined by *a priori* anatomical or functional relations,
2. Model the individual variations directly.
3. Express questions in terms that are abstracted away from the anatomy,

Unlike the Talairach approach, none of these approaches apply in every situation; their applicability and efficacy depends on the questions being asked. But when they do apply, they offer a promising alternative.

The first approach is to restrict our attention to a region of the brain based on *a priori* anatomical (or functional) information. The idea is that the functionality of such regions is maintained across subjects even when the geometry of the region is not. For many experimental tasks, a great deal is known about which regions play an important role in processing, and there are often basic functional studies that can sharpen this information for a given subject. The role of anatomical expertise cannot be overemphasized here, for it is challenging to identify small anatomical structures from the images. Naturally, there is also a trade-off between size of the region and level of controversy associated with the identification; no one will argue with a claim to have identified the left hemisphere, but for very small structures, there is much greater uncertainty. Having identified a region and analyzing the fMRI data, results for that region can be compared across subjects. This need not (and usually should not) involve voxelwise averaging; rather quantitative summaries for the region can be compared. For example, using my model described earlier, one could compute the posterior distribution of the average signal change over the region and compare subjects via these distributions. The effectiveness of this method depends on how large the region is relative to the patterns of activity within the region. In some cases, it might even be possible to further partition the given region in a way that highlights spatial patterns of activity that are consistently observed across subjects. This would refine the regional mappings and provide more detailed information about the nature of the underlying neural processing.

The second approach is to explicitly model the across-subject variations in brain topography using the estimated patterns of activity in addition to detailed anatomical information. This approach treats the Talairach mapping as a coarse first-stage approximation to the individual differences but then adds a spatial component of variation to the mapping. For example, if two subjects exhibit clusters of activation that are close to each other in Talairach coordinates, it suggests that these represent a common functional response. From this point of view, the problem becomes two-fold: (i) determine the distance scale over which individuals vary in the location of activity for the same functional response, and (ii) find a correspondence among the observed patterns of activity that enables us to compare results across subjects. With Jonathan Raz, I have begun tackling this problem. Our initial efforts have focused on the locations of peak activation estimated from the data; a spatial point process model for these locations can account for the components of variation and solves both (i) and (ii) simultaneously. Further refinements of this approach will bring to bear more features of the activation pattern. Other similar approaches are based on more sophisticated mapping schemes.

The third approach applies to specific questions that are not directly tied to anatomical structures. While a question such as ''Is there significant activation in the middle temporal gyrus under the experimental paradigm?'' is fundamentally tied to the specifics of the anatomy, some questions can be abstracted away from the anatomy. For example, in the oculomotor studies described above, one question of interest is whether pursuit and saccades are spatially dissociated and moreover whether one is systematically anterior to the other (i.e., closer to the front of the brain). Having fit the model for the ith subject, we can compute the posterior expectation $E(D(\boldsymbol{\theta}^i) \mid \boldsymbol{Y}^i)$, where $D(\boldsymbol{\theta}^i)$ is the binary functional on the parameter space that indicates whether the underlying activation responses for pursuit and saccade are dissociated in such a way that saccade activity is anterior to pursuit. Here, $\boldsymbol{\theta}^i$ and $\boldsymbol{Y}^i$ are the model parameter vector (including all voxels) and the fMRI data for the ith subject. Computing D is non-trivial since it depends on the spatial distribution of activity, but the point is that each posterior probability is encapsulated within a single subject's anatomy. Notice that the parameter vectors for different subjects cannot be directly mapped; nonetheless, D means the same thing for every subject. The question at hand relates to the population distribution of D We can get this by an unorthodox combination of the posterior expectations across the N subjects:

$$E(D) \approx \frac{1}{N} \sum_{i=1}^{N} E(D(\boldsymbol{\theta}^i) \mid \boldsymbol{Y}^i) \tag{10}$$

$$\text{Var}(D) \approx E(D)\,(1 - E(D)) \tag{11}$$

This yields an estimate and a measure of uncertainty directly relevant to the question. (Clearly, this can be generalized for non-binary functionals.) Moreover, all of the techniques that can be applied to answer such a question at the single-subject level can be generalized to multiple subjects in this way. Although subjects' data are conditionally independent of each other given the parameters $\boldsymbol{\theta}$, there are invariants in the response across subjects (e.g., D) that induce unconditional dependence. Such invariants are unknown and relatively few in number. Thus, constructing a full joint specification of the model across subjects (from which might be derived $E(D \mid \boldsymbol{Y}^1, \ldots, \boldsymbol{Y}^N)$) would be tantamount to computing the full anatomical mapping from brain to brain---exactly the problem we need to solve. The approach described above can make across-subject inferences without requiring this mapping.

6. NETWORK SPECIFICATION

The high-level functioning of the brain relies on a complex and nonlinear flow of information, a chain of associations operating in parallel at many levels. Understanding such a complex system requires a multi-faceted approach. Basic neuroscience research has clarified the detailed workings of neurons and the function of specific and localized brain regions. At the same time, cognitive psychologists have developed computational models for learning and memory that make specific predictions about behavior. With the advent of functional neuroimaging, a synthesis of these two perspectives has become possible. Now, we can begin to study how *systems* in the brain operate through the interaction of their components.

This interaction is complex. Consider a relatively simple network underlying the oculomotor system described previously. If we restrict our attention to the most important components of this system, we have a network of approximately eight regions (in each hemisphere) that are distributed throughout the brain (see Genovese & Sweeney, 1997). Neuroimaging techniques such as fMRI make it possible for the first time to study the interaction of these components. Discerning the connectivity and relative timing is crucial to understanding how high-level behaviors can emerge from low-level circuitry. What can fMRI actually tell us about this question?

Consider a highly simplified model for the dynamics of information flow in such a network. Suppose we have R nodes, which we think of as specific functional regions in the brain, that interact by passing "activation" to other nodes in the network. Two nodes can interact by either facilitating or inhibiting the activation in the other node. For $r = 1, \ldots, R$, let $\mathcal{I}_r(t) = \alpha_r \mathcal{I}(t)$ be the driving input to the rth node, where $\alpha_r \geq 0$ and $\mathcal{I}(t)$ is the exogenous input to the system, which we take as a known, filtered version of the stimulus presentation schedule. Let $A_r(t)$ be the level of activation in region r at time t. I take $A_r(t)$ to be the activation profile estimated by fitting my model described above to the original fMRI data. The computed posterior on the model parameters induces a joint distribution for the A_r functions that will be used to derive a likelihood below. A model for the system specifies the strength of the connections among these different regions and the relative timings of activation flow. One plausible model is to take

$$A_r(t) = \gamma_r F(\mathcal{I}_r(t) + \sum_{1 \leq m \leq R} \beta_{rm} A_m(t - \tau_{rm})), \tag{12}$$

where F is a fixed sigmoidal function (e.g., standard normal cdf), β_{rm} determines the strength of facilitation or inhibition for activation flowing to region r from region m, $\tau_{rm} \geq 0$ is the time lag for this flow. The goal is to estimate the parameters $\boldsymbol{\beta}$, $\boldsymbol{\tau}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$ that describe the system. The relation in equation (12) is based on a classical neural network model. Several variants are possible that yield different temporal dynamics.

7. COMBINING DATA FROM DISPARATE SOURCES

Alone, fMRI data provide information on the blood-flow response to neural activity. This technique provides a superior combination of spatial and temporal resolution, but fMRI based on this hemodynamic mechanism offers an inherently limited view of the behavior of actual neurons. The hemodynamic processes are delayed and dispersed relative to the underlying activation, and the position and drainage of blood-vessels can cause various artifacts. The strengths of other imaging modalities complement fMRI quite well.

For example, Event-Related Potentials (ERP) and Positron Emission Tomography (PET) can provide detailed information on other aspects of brain function. ERP data consist of measures at the scalp of changes in the brain's electrical potential, which is caused directly by neural activity. ERP reconstruction is a seriously ill-posed problem, and little spatial resolution can be obtained from these data. However, temporal changes in ERP are on the order of milliseconds rather than seconds as in fMRI. In PET, a radioactive tracer attached to a neurochemical is injected into the subject's blood stream and concentrates in the brain where the neurochemical is used. Positrons emitted by radioactive decay of the tracer give rise to pairs of photons which can be counted, and the resulting counts can be reconstructed into a brain image. Because of the time-scale of decay and radioactive exposure, only a limited number of PET images can be collected in a study, and both the temporal and spatial resolution are lacking. The great strength of PET, however, is that the radioactive tracer can be selected to target one of many different metabolic processes. Although for functional experiments PET is dominated by fMRI, PET can image many other processes that can be traced in no other way.

The basic problem here is to combine data acquired simultaneously from several different modalities during the experiment. These data have different temporal and spatial resolutions, different sensitivities, and different practical issues that must be tackled. Although these disparate data are derived from distinct mechanisms, the underlying processes (hemodynamic response, electric potential changes, neurochemical concentration) still have strong relationships. By combining the data from disparate sources, we can exploit these relationships to get a better estimate of brain function.

Consider a simple version of the problem where data are acquired simultaneously from two modalities with different resolutions. Suppose these data sets are samples from mean-zero processes $Y(\boldsymbol{u}, s)$ and $Z(\boldsymbol{v}, t)$ taken on different spatio-temporal grids $(\boldsymbol{u}_j, s_k)$ and $(\boldsymbol{v}_m, t_n)$. Assume that the time points of the two sample grids are separated by α_1 and β_1 respectively, with $\alpha_1 > \beta_1$ and that the spatial points are separated by α_2 and β_2 respectively, with $\alpha_2 < \beta_2$. The assumption is that both of these data sets are related to the underlying neural activity, albeit in different ways:

$$Y(\boldsymbol{u}_j, s_k) = F\left(\int K(\boldsymbol{x}, t)\, a(\boldsymbol{u}_j - \boldsymbol{x}, s_k - t)\, d\boldsymbol{x}\, dt\right) + \delta(\boldsymbol{u}_j, s_k) \tag{13}$$

$$Z(\boldsymbol{v}_m, t_n) = G\left(\int L(\boldsymbol{x}, t)\, a(\boldsymbol{v}_m - \boldsymbol{x}, t_n - t)\, d\boldsymbol{x}\, dt\right) + \epsilon(\boldsymbol{v}_m, t_n). \tag{14}$$

where F and G are unknown, monotonic functions, K and L are known point-spread kernels, δ and ϵ are noise processes, and a is the underlying activation process common to both modalities. Strong constraints on $a(\boldsymbol{x}, t)$, F, and G are needed to make these terms identifiable. Some of these can be derived substantively because the experimental paradigm and various properties of neuronal firing are known. Other constraints are more practically motivated. For example, it makes sense to take $a(\boldsymbol{x}, t) \in [0, 1]$ since there is no natural scale for the activation. Moreover, a should be sparse, which drastically reduces the effective degrees of freedom required, although complicating the fitting. Finally, since the data contain no information about any component of $a(\boldsymbol{x}, t)$ that is annihilated by both kernels, $a()$ can be taken as smooth at the finest discernible resolutions.

This model specification and associated constraints give a joint likelihood on which estimates can be based, although it would be desirable to estimate F and G nonparametrically. One way to fit such a model is to iteratively alternate fitting the individual pieces subject to varying constraints; I suspect this will prove to be a practical computational procedure. The resulting model fit provides an estimate of the activation pattern that benefits from the resolution of both modalities. The fit also yields an estimate of the quantitative relationship between the modalities which may prove useful.

In practice, the simplified form of the model above will change somewhat to meet the requirements of specifics modalities, but the structure remains similar. For example, in combining ERP and fMRI, there is a further level of indirection: known functionals of Z, instead of Z itself, are actually measured. Similarly, the function $a()$ is vector valued, but the fMRI data in this case are sensitive only to the norm of this function.

8. DISCUSSION AND CONCLUSIONS

In this paper, I presented a model for fMRI data that captures the critical sources of variation in the voxel time-series. I then described four new challenges that arise in applying fMRI data to scientific questions. These include:

- Constructing models that borrow strength across regions with complex geometries,
- Comparing inferences across different coordinate systems,
- Characterizing nonlinear relations embedded in spatio-temporal fluctuations, and
- Combining data with different spatial and temporal resolutions.

Although I have described these problems in the context of fMRI, these issues arise more broadly in spatio-temporal applications. For example, consider remote sensing data obtained from different instruments and consequently acquired with different spatial and temporal resolutions. Methods for dealing with such problems are only beginning to be developed, so what we learn about these issues in fMRI will benefit other areas.

REFERENCES

Belliveau, J.W., Kennedy, D.N., McKinstry, R.C., Buchbinder, B.R., Weisskoff, R.M., Cohen, M.S., Vevea, J.M., Brady, T.J., and Rosen, B.R. (1992). Functional mapping of the human visual cortex by magnetic resonance imaging. *Science* **254**, 716--719.

DiCiccio, T.J., Kass, R.E., Raftery, A., and Wasserman, L. (1997). Computing Bayes Factors by Combining Simulation and Asymptotic Approximations. *J. Amer. Statist. Assoc.* **92**, 903--915.

Eddy, W.F., Fitzgerald, M., Genovese, C.R., Mockus, A., and Noll, D.C. (1996). *Functional image analysis software - computational olio.* Proceedings in Computational Statistics (ed. Prat, A.). Heidelberg: Physica-Verlag, **12**, pp. 39--49.

Fitzgerald, M. (1996). Registration and Estimation of Functional Magnetic Resonance Images. Ph.D. Thesis, Carnegie Mellon University, Department of Statistics.

Genovese, C.R. (1997) Statistical Inference in Functional Magnetic Resonance Imaging. Technical Report 674, Carnegie Mellon University, Department of Statistics.

Genovese, C.R. and Sweeney, J.A. (1998). *Functional Connectivity in the Cortical Circuits Subserving Eye Movements*. Case Studies in Bayesian Statistics (eds. Kass, R.E., Carlin, B.P., Carriquiry, A.L., Catsonis, C., Gelman, A., Verdinelli, I., and West, M.). Springer Verlag, to appear.

Green, P.J. (1995) Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika* **82**, 711--732.

Hu, X., Le, T.H., Parrish, T., and Erhard, P (1995). Retrospective estimation and correction of physiological fluctuation in functional MRI. *Magnetic Resonance in Medicine*, **34**, 201--212.

Johnson, V.E. (1994). A Model for Segmentation and Analysis of Noisy Images. *J. Amer. Statist. Assoc.* **89**, 230--241.

Johnson, V.E., Bowsher, J.E., Turkington, T.G., and Jaszczak, R.J. (1995). *Analysis and Reconstruction of Medical Images Using Prior Information*. Case Studies in Applied Bayesian Statistics II, 149-238, Springer-Verlag, New York.

Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E. and Cohen, M.S., Turner, R., Cheng, H., Brady, T.J., and Rosen, B.R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5675.

Lange, N. (1996). Statistical Approaches to Human Brain Mapping By Functional Magnetic Resonance Imaging. *Stat. in Med.* **15**, 389--428.

Ogawa, S., Tank, D.W., Menon, D.W., Ellermann, J.M., Kim, S., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping using MRI. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5951-5955

Talairach, J. and Tounoux, P. (1988). Coplanar Stereotaxic Atlas of the Human Brain. Three-dimensional Proportional System: An Approach to Cerebral Imaging. Paris:Thieme.

Figure 1. A single voxel time-course. The horizontal axis gives image index, with an image acquired every 3 seconds. The vertical axis gives the MR signal strength in arbitrary units. The hashed bar indicates the experimental design. The \\\-hashed blocks (red) correspond to the ‘‘Task’’ condition, finger tapping while staring at a dot. The ///-hashed blocks (blue) correspond to the ‘‘Control’’ condition, staring at a dot. The line style indicates the condition associated with each data point: solid (red) lines correspond to ‘‘Task’’ and dotted (blue) lines correspond to ‘‘Control’’. Note the slight lag in the perturbation of response relative to task performance.

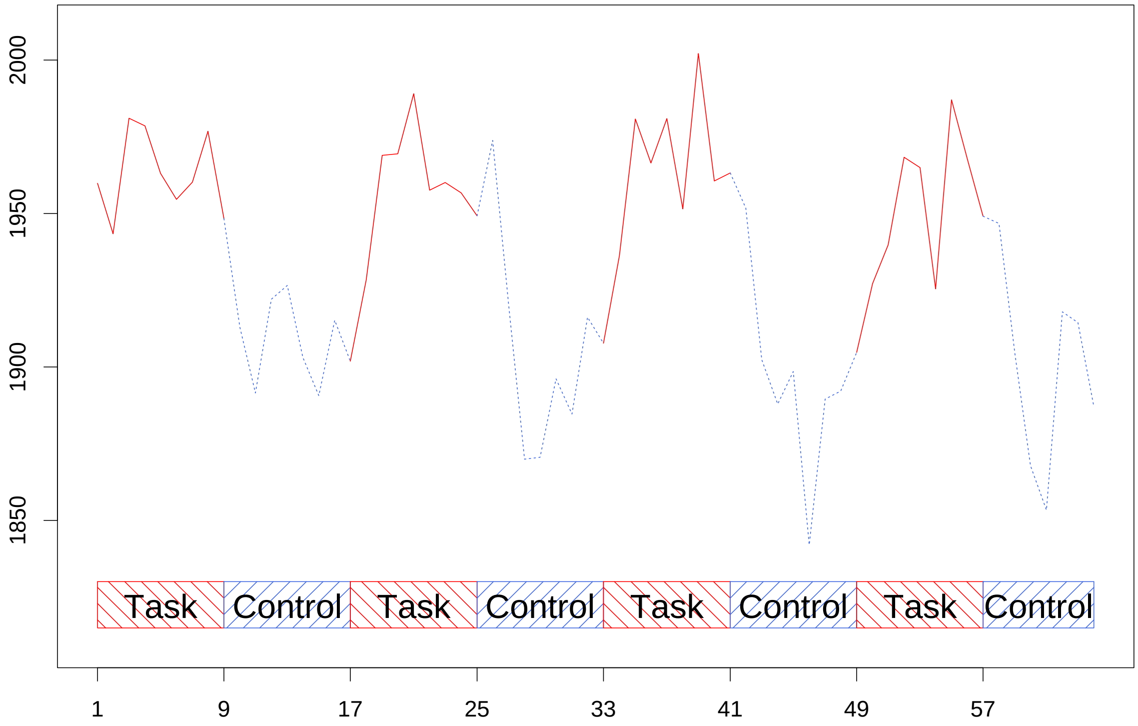

Figure 2. (a) The mean of the functional images for one slice from one subject with posterior probabilities of visual saccade activation superimposed. The probabilities are thresholded at 0.2 to enhance interpretation. (b) The mean of the functional images for one slice from one subject with posterior probabilities of pursuit activation superimposed. The probabilities are thresholded at 0.2 to enhance interpretation.