

# Locally Regularized Spatiotemporal Modeling and Model Comparison for Functional MRI

Patrick L. Purdon,<sup>\*†</sup> Victor Solo,<sup>\*‡</sup> Robert M. Weisskoff,<sup>\*</sup> and Emery N. Brown<sup>†§</sup>

<sup>\*</sup>Massachusetts General Hospital NMR Center, Charlestown, Massachusetts; <sup>‡</sup>School of Electrical Engineering, University of New South Wales, Sydney NSW, Australia; <sup>§</sup>Department of Anesthesia and Critical Care, Neuroscience Statistics Research Laboratory, Massachusetts General Hospital, Boston, Massachusetts; and <sup>†</sup>Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts

Received December 22, 2000

**In this work we treat fMRI data analysis as a spatiotemporal system identification problem and address issues of model formulation, estimation, and model comparison. We present a new model that includes a physiologically based hemodynamic response and an empirically derived low-frequency noise model. We introduce an estimation method employing spatial regularization that improves the precision of spatially varying noise estimates. We call the algorithm locally regularized spatiotemporal (LRST) modeling. We develop a new model selection criterion and compare our model to the SPM-GLM method. Our findings suggest that our method offers a better approach to identifying appropriate statistical models for fMRI studies.**

© 2001 Academic Press

**Key Words:** functional MRI; regularization; spatiotemporal modeling; system identification; model comparison.

## 1. INTRODUCTION

The fMRI data analysis problem is a problem of system identification. It involves the spatiotemporal relationship between an input such as a sensory stimulus or cognitive task and the cerebral response to that task measured with fMRI. This problem of system identification involves a number of essential steps: (1) model formulation, (2) model fitting, (3) model comparison, and (4) inference (Box *et al.*, 1994). In the model formulation step, a set of models is chosen or formulated to represent possible alternative descriptions for the relationship between the input and output. Once the models have been formulated, model fitting methods are used to estimate parameters in the models. The choice of a fitting method can depend largely upon the details of the model as well as the nature of the data being fitted. In the model comparison step, a “best” model is chosen from among the fitted models based on

estimates of the proximity between the models and the “true” process generating the data. This model comparison step considers not only the goodness-of-fit of the model, but also the complexity of the model in relation to the data, providing a mechanism to avoid overfitting. Only after the best model has been chosen can we do inference—the model that best represents the underlying process generating the data should be used to answer questions about that process.

In this paper, we develop a unique approach to addressing the model formulation, model fitting, and model comparison steps in the system identification framework. In the fMRI literature, a number of different models relating stimuli and response have been developed and applied (e.g., Worsley and Friston, 1995; Bullmore *et al.*, 1996; Lange and Zeger, 1997; Locascio *et al.*, 1997; Friston *et al.*, 1998). In this work, we develop a parametric, physiologically inspired model based on animal and simulation studies of the blood oxygen level-dependent (BOLD) response (Mandeville *et al.*, 1996, 1998; Marota *et al.*, 1996). This model incorporates spatially-varying first-order autoregressive noise (AR(1)) to account for low-frequency physiological fluctuations and white noise to account for scanner noise. Proper noise estimation is essential because the accuracy of the confidence intervals used in the inference stage depends upon how well the noise is estimated (e.g., Purdon and Weisskoff, 1998). Because fMRI time-series are short, on the order of 100 time points, and because much of the noise power comes from low-frequency physiological fluctuations (Weisskoff *et al.*, 1993; Mitras and Pesaran, 1999), estimating the noise on a pixel-wise basis can be difficult.

We introduce a novel model fitting method that uses spatial regularization on the model parameters. For model comparison, we develop a novel method based on nearly unbiased risk estimation (NURE; Linhart and Zucchini, 1986; Solo, 1998a) that overcomes many of the problems associated with hypothesis testing methods, and that allows for comparison between models

with different structures and fitting methods. We then use this method to compare the results of our model and fitting method to the widely-used Generalized Linear model-Statistical Parametric Map (GLM-SPM) method (Friston, 1997).

## 2. STATISTICAL MODEL

### 2.1. Model Formulation

We use the known properties of the physiological changes underlying the fMRI hemodynamic response along with the observed structure in fMRI signals and noise to develop our statistical model of the fMRI signal. Our model of the fMRI signal consists of three parts: the hemodynamic response signal, a colored noise, and a drift term.

**2.1.1. fMRI BOLD signal.** The physiological studies on which we base our model come from using both animal models and Monte Carlo simulations of BOLD, susceptibility-based contrast (Mandeville *et al.*, 1996, 1998; Marota *et al.*, 1996). We assume that the changes in the T2\*-based fMRI signal are primarily due to the net average deoxyhemoglobin in the voxel, which is the product of the blood volume and the blood oxygenation. That is, we assume blood has an average deoxygenation level, and that the BOLD signal will behave like the product

$$V(t) \times Hb(t) = (V_0 + \Delta V(t)) \times (Hb_0 + \Delta Hb(t)), \quad (2.1)$$

where  $V_0$  is the initial blood volume,  $\Delta V(t)$  is the change in blood volume,  $Hb_0$  is the baseline deoxyhemoglobin level, and  $\Delta Hb(t)$  is the change in deoxygenated hemoglobin in a time interval  $\Delta t$  due to underlying neuronal activation. We also assume that local blood flow changes and local tissue oxygen consumption changes are tightly coupled in time, though not necessarily proportionally coupled.<sup>1</sup> The deoxygenation term is dictated in terms of oxygen consumption and blood flow by Fick's principle (conservation of oxygen). The time-course of the deoxygenation and blood volume changes are difficult to separate in studies of normal human volunteers. In animal studies using intravascular contrast agents (Mandeville *et al.*, 1996, 1998; Marota *et al.*, 1996), it has been shown that the blood volume changes are slower than the changes in deoxygenation. The form of our model is also guided by the empirical observation that the hemodynamic component of the fMRI response for block stimuli has four

<sup>1</sup> These assumptions, are consistent with accepted properties of both PET and MRI data, yet ignore small, transient ( $\leq 2$  s) changes in blood  $O_2$  that have been detected in optical recordings and the "dip" reported in high-field fMRI data. While interesting and scientifically important, this transient dip is small enough that it may be neglected in the current formulation of the model.

characteristics: (a) delay, (b) peak response, (c) droop, and (d) poststimulus undershoot.

By combining the basic properties of the BOLD signal change with these empirical observations, we can create a model of fMRI signal change in response to a stimulus input. In a given pixel, we make the assumption that the blood hemoglobin changes are driven by the stimulus through a linear coupling (Boynton *et al.*, 1996), approximating the deoxyhemoglobin term as

$$\begin{aligned} Hb(t) &= k_1 + k_2(g_a * c) \\ g_a(t) &= (1 - e^{-t/d_a})^2(t + 1)e^{-t/d_a}, \end{aligned} \quad (2.2)$$

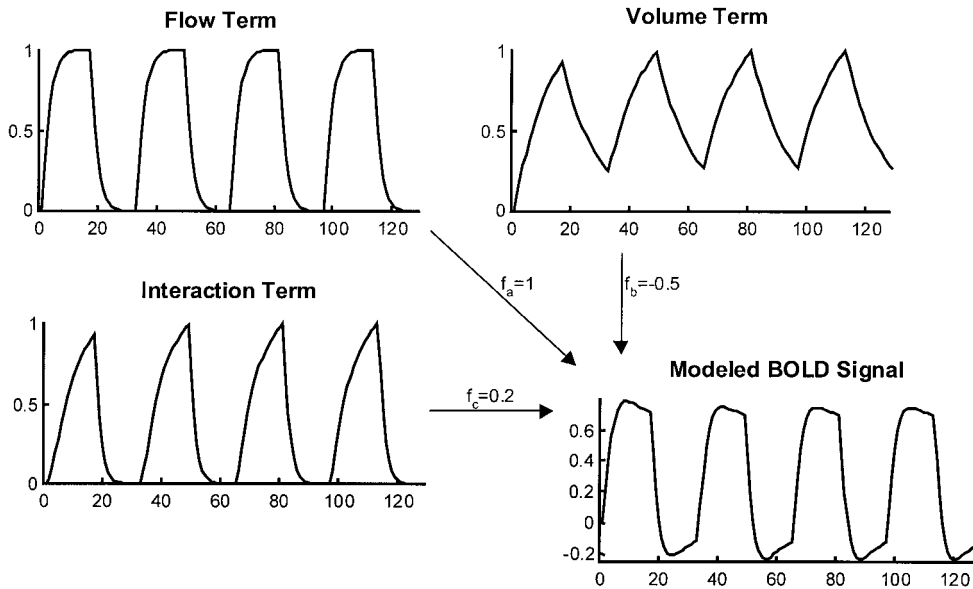
where  $k_1$  and  $k_2$  are constants,  $c(t)$  is the stimulus,  $g_a(t)$  is a hemodynamic impulse response, chosen to be a discrete gamma function with exponential time constant  $d_a = 1.5$  s, and  $*$  denotes convolution [ $g_a * c = \sum_0^\infty g_a(u)c(t - u)$ ]. The hemodynamic impulse response  $g_a(t)$  satisfies a normalization  $\sum_0^\infty g_a(t) = 1$  to avoid scaling problems and its maximum occurs at  $t = d_a - 1$ . The response for the blood volume is

$$\begin{aligned} V(t) &= k_3 + k_4 g_b * c \\ g_b(t) &= (1 - e^{-t/d_b})e^{-t/d_b}, \end{aligned} \quad (2.3)$$

where  $c(t)$  is the stimulus, and  $k_4 g_b * c$  reflects the blood volume response to this stimulus, and  $g_b(t)$  is an exponential impulse response with time constant  $d_b = 12$  s. Several authors have suggested the convolution model used here (Boynton *et al.*, 1996; Cohen, 1997). In the block design experiment we analyze, the stimulus is not sufficiently rich in excitation energy to estimate too many parameters. For this reason, values have been set for the  $d_a$ ,  $d_b$  time constants, based on empirical observations. Both responses can be shifted by a time delay,  $D$ , constrained to be positive, to represent a causal system. Combining these elements, we have a pixel by pixel model for the BOLD signal which after multiplying Eq. 2.2 by Eq. 2.3 and collecting constants is equal to a baseline  $m$  plus three response terms

$$\begin{aligned} s_t &= f_a(g_a * c)_{t-D} + f_b(g_b * c)_{t-D} \\ &\quad + f_c(g_a * c)_{t-D}(g_b * c)_{t-D} \end{aligned} \quad (2.4)$$

where  $f_a$ ,  $f_b$ , and  $f_c$  are amplitudes that parameterize the activation. From a physiological point of view, the  $f_a$  term corresponds to the flow response, the  $f_b$  term corresponds to the volume response, and  $f_c$  term represents their interaction. Figure 1 illustrates the relative contribution of these terms in generating a synthetic block-paradigm response time-course.



**FIG. 1.** Illustration of BOLD signal model for block-paradigm stimulus. The contribution of flow ( $f_a$ ), volume ( $f_b$ ), and interaction ( $f_c$ ) terms combine to form an overall signal.

**2.1.2. fMRI noise model.** The noise in fMRI data can be thought to come from two sources: (1) The background noise of the scanner and (2) physiological noise from low-frequency hemodynamic fluctuations (Weisskoff *et al.*, 1993; Mitra and Pesaran, 1999). We model the scanner noise as a white noise and the low-frequency physiological noise as a first-order autoregression (AR(1)), leading to the model:

$$\begin{aligned} v_t &= w_t + \eta_t \\ w_t &= \rho w_{t-1} + \epsilon_t \end{aligned} \quad (2.5)$$

where  $v_t$  is the noise,  $\eta_t$  is a zero-mean white Gaussian noise with variance  $\sigma_\eta^2$  used to represent the scanner noise,  $w_t$  is the AR(1) noise with correlation coefficient  $\rho$ , and  $\epsilon_t$  is a zero-mean white Gaussian noise of variance  $\sigma_\epsilon^2$ . This model can be thought of equivalently as a first-order autoregressive moving average model (ARMA(1,1); Box *et al.*, 1994).

**2.1.3. Drift terms.** To complete the model, we include a term to allow for slow drifts in the static magnetic field and residual motion not accounted for by prior motion-correction. Although we represent this drift with a linear term, it can be easily generalized to more complicated representations (Genovese, 1999). Adding on an additional term  $m$  to account for the mean signal, we then have

$$\text{drift} = m + bt. \quad (2.6)$$

**2.1.4. Complete model.** Combining the BOLD signal, noise, and drift components, we have the overall fMRI signal

$$\begin{aligned} \text{fMRI signal} &= \text{drift} + \text{hemodynamic response} \\ &+ \text{noise} \end{aligned} \quad (2.7)$$

$$x_{t,p} = (m_p + b_p t) + s_{t,p} + v_{t,p}$$

where  $t = 1 \dots N$  represents time and  $p = 1 \dots P$  represents the pixel.

## 2.2. Model Fitting

We introduce here a fitting procedure employing spatial regularization that is able to make use of local spatial continuity in the model parameters to improve estimation efficiency. The parameters to be estimated at each pixel are the signal parameters  $\beta = (m, b, f_a, f_b, f_c, D)$  and the noise parameters  $\alpha = (\sigma_\epsilon^2, \sigma_\eta^2, \rho)$ , which we collect into one parameter vector  $\theta = (\beta, \alpha)$ . To estimate the parameters we maximize a local, spatially weighted Gaussian log-likelihood. The local spatial weighting is a regularization procedure that captures spatial continuity. Let  $\theta_p = (\beta_p, \alpha_p)$  be the parameters at pixel  $p$ . The overall fitting criterion  $\mathcal{J}(\theta)$  is pixel-wise separable,

$$\mathcal{J}(\theta) = \mathcal{J}(\theta_1 \dots \theta_p) = \sum_p \mathcal{J}_p(\theta_p), \quad (2.8)$$

and the criterion at pixel  $p$  is a spatially locally weighted log-likelihood given by

$$J_p(\theta_p) = \sum_q K_q^h L_q(\theta_p), \quad (2.9)$$

where  $L_q(\theta_p)$  is the Gaussian log-likelihood based only on the time series at pixel  $q$ ,  $K_q^h$  (see the Appendix) is a positive spatial weighting function or kernel that sums to unity,  $h$  is a tuning parameter that controls the neighborhood size on which  $K_q^h$  is concentrated, and  $q$  is the spatial summation variable.

We implement this procedure with a cyclic descent algorithm that iterates between estimating signal parameters and noise parameters. This allows the noise and signal parameter subsets to be regularized separately. Here, we focus on regularizing the AR(1) noise parameters only, exploiting local spatial continuity to improve their estimation, while allowing the signal parameters to remain at the native spatial resolution of the imaging experiment. In general, the cyclic descent algorithm can be partitioned arbitrarily, allowing different parameter subsets to be regularized with different degrees of spatial smoothing. Within the cyclic descent algorithm, the signal parameters are estimated using weighted least-squares and the noise parameters are estimated using an Expectation–Maximization algorithm. Further details are given in Solo *et al.* (1997) and Solo *et al.* (2001). To reduce computational cost, we estimated the white noise variance  $\sigma_\eta^2$  from background pixels outside the brain, with a scaling factor to account for the fact that the background pixels are Rayleigh distributed, becoming approximately Gaussian distributed in regions with nonzero image intensity (Papoulis, 1991). We call our overall method locally regularized spatiotemporal modeling (LRST).

### 2.3. Model Comparison

We have developed a particular model based on physiological and empirical observations, but there are a variety of other models that have been proposed for fMRI data analysis. Most new models arise from advances in BOLD physiology and changes in experimental design. With the numerous model choices available, each of which represents an approximation to the true underlying process being observed, we must choose the model that best approximates the underlying process before we can make reliable inferences about the experiment. This notion of how well the model approximates the behavior of the underlying process is different from the notion of how well the model fits the data. For instance, arbitrarily increasing the number of parameters in a model will necessarily improve the fit, but for a fixed amount of data, the accuracy of the parameter estimates will decrease as the number of parameters increases. Somehow, a balance must be struck between model complexity and goodness of fit.

Furthermore, the measure of how well a model approximates the underlying process should allow for comparison between models that are formulated or fitted differently.

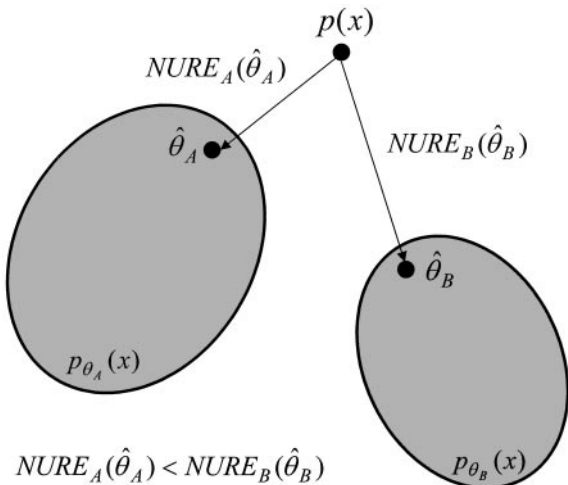
In this section we present a new model comparison technique based on a nearly unbiased risk estimator (NURE) and use it to compare our method with the GLM technique (Friston, 1997). The approach we take is an extension of the work of Linhart and Zucchini (1986) and Solo (1998a) and has connections to the Akaike information criterion (AIC). But unlike AIC, NURE is not restricted to cases where model complexity is measured by model dimension. In our case model complexity is measured by a neighborhood size. Details of the derivation are given in Solo (1998a) and Solo *et al.* (2001). A very readable introduction to this methodology, focused on the related AIC technique, is contained in (Linhart and Zucchini, 1986; Burnham and Anderson, 1998).

The basic problem is one of estimating how “close” a set of models is to a true but unknown process, based on the observed data. Proximity is measured by the Kullback–Leibler (KL) distance between the true probability density of the data,  $p(x)$ , and the probability density given by the model,  $p_\theta(x)$ :

$$R(\theta) = \int p(x) \ln \left( \frac{p(x)}{p_\theta(x)} \right) dx. \quad (2.10)$$

Because we do not know the true probability density, we cannot directly compute the KL distance. Furthermore, we have only an estimate of  $\theta$ ,  $\hat{\theta}$ , rather than the “true” value  $\theta_0$  that provides the best representation for the given model structure. Consequently, we compare models based on an estimate of the expected KL distance,  $E[R(\hat{\theta})]$ , which we compute using a Taylor series expansion (Solo, 1998; Solo *et al.*, 2001). This Taylor series expansion produces a model complexity term that corrects for the bias introduced by using  $\hat{\theta}$  in place of  $\theta_0$  and quantifies the trade-off between model complexity and goodness of fit. Model selection criteria such as AIC are special cases of this NURE approach. The estimate can be computed without having to know the “true” probability density by using the data as a surrogate for the true density. In many cases, including fMRI data analysis, the set of competing candidate models can have very different structures and may be estimated with different fitting procedures, yielding different NURE formulas for each model or method. As long as the same discrepancy measure is used to derive these NURE formulas, they can be used to compare different models.

To illustrate how we apply the NURE technique, consider the situation depicted in Fig. 2, where there are two classes of models, A and B, with different



**FIG. 2.** Illustration of model comparison using NURE. Models are compared based on the “information distance” between the models (with probability density functions  $p_{\theta_A}(x)$  and  $p_{\theta_B}(x)$ ) and the true but unknown process (with probability density function  $p(x)$ ). The NURE provides an estimate of this distance. In this case, since  $NURE_A(\hat{\theta}_A) < NURE_B(\hat{\theta}_B)$ , we choose the model represented by  $\hat{\theta}_A$  and use that model for inferences.

model structures parameterized respectively by  $\theta_A$  and  $\theta_B$ . For each model class A and B, we derive a NURE formula based on the KL distance (2.10). Since the model structures for A and B are different, the functional form for each NURE should be different as well. We then estimate the best-fitting models for each class, giving estimated parameters  $\hat{\theta}_A$  and  $\hat{\theta}_B$ , and then use those estimated parameters to compute a NURE for each model. In this example, since  $NURE_A(\hat{\theta}_A) < NURE_B(\hat{\theta}_B)$ , we choose the model corresponding to  $\hat{\theta}_A$  and base our inferences upon that model.

In the fMRI models we consider, the data fitting criterion are pixel-wise separable so that  $J(\theta) = \sum_p J_p(\theta_p)$ . Again using a Taylor series argument, it can then be shown (Solo *et al.*, 2001) that the NURE is given by an expression of the form

$$\text{NURE} = \text{data-model discrepancy} + \text{model complexity.} \quad (2.11)$$

The intuition behind this form, analogous to the AIC case, is that there is a trade-off between how well the model fits the data and the model complexity. As the model complexity increases, the data-model discrepancy decreases since more parameters can be fit to the data. However, if the model is “too complicated” (over-fitted), representing the possibility that inferences could be widely different if the experiment were repeated, the overall NURE will be larger than for a simpler model.

For the LRST model, using the frequency-domain form of the Gaussian log likelihood function (described in the Appendix), we have:

$$\begin{aligned} \text{NURE}_{LRST} &= \sum_p \hat{R}_{LRST,p}(\hat{\theta}_p) + P \frac{\dim \beta}{h} + 2P \frac{\dim \alpha}{h} \\ \hat{R}_{LRST,p}(\hat{\theta}_p) &= \frac{1}{2} \sum_k \frac{|\tilde{x}_{k,p} - \tilde{\mu}_{k,p}(\beta_p)|^2}{N^2 F_k(\alpha_p)} \\ &\quad + \frac{1}{2} \sum_k \ln F_k(\alpha_p), \\ \mu_{t,p} &= m + bt + s_{t,p} \end{aligned} \quad (2.12)$$

where  $\hat{R}_{LRST,p}(\hat{\theta}_p)$  is the data-model discrepancy at pixel  $p$ ,  $\beta_p$  and  $\alpha_p$  are the signal and noise parameters, respectively, as defined in section 2.2,  $\tilde{x}_{k,p}$  ( $\tilde{\mu}_{k,p}(\beta_p)$ ) is the discrete Fourier transform of the data (signal and drift terms) at temporal frequencies  $w_k = 2\pi k/N$  and pixel  $P$ ,  $F_k(\alpha_p)$  is the spectrum according to the model,  $\dim \beta$  is the number of parameters in  $\beta$ ,  $P$  is the number of pixels, and  $h$  is the regularization neighborhood parameter (described further in the Appendix).

For the GLM (Worsley and Friston, 1995), the signal model is

$$x_{t,p}^{GLM} = \mu_{t,p}^{GLM}(\beta_p^{GLM}) + \epsilon_{t,p}^{GLM}, \quad (2.13)$$

where  $\epsilon_{t,p}^{GLM}$  is a white noise and  $\mu_{t,p}^{GLM}(\beta_p^{GLM})$  is a linear combination of covariates with coefficients given by  $\beta_p^{GLM}$ . In the GLM method the data is smoothed and  $\beta_p^{GLM}$  is estimated from the smoothed data by ordinary least squares. If we denote their smoothing filter by  $\phi_t$  and its DFT by  $\hat{\phi}_k$  then  $|\hat{\phi}_k|^2$  is essentially a whitening filter and takes the place of  $1/F_k(\alpha_p)$  in the expression for the data-model complexity for the LRST model. The NURE for the GLM model is then

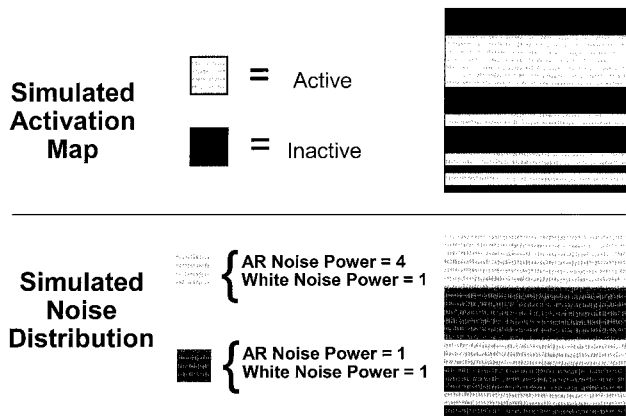
$$\begin{aligned} \text{NURE}_{GLM} &= \sum_p \hat{R}_{GLM,p}(\hat{\theta}_p) + P \dim \beta_p^{GLM} \\ \hat{R}_{GLM,p}(\hat{\theta}_p) &= \frac{1}{2} \sum_k |\tilde{x}_{k,p} - \tilde{\mu}_{k,p}^{GLM}(\hat{\beta}_p^{GLM})|^2 |\hat{\phi}_k|^2 \\ &\quad + \frac{1}{2} \sum_k \ln \frac{1}{|\hat{\phi}_k|^2}. \end{aligned} \quad (2.14)$$

Further details for the derivation of these expressions are given in Solo *et al.* (2001).

### 3. RESULTS

#### 3.1. Simulated Data

We constructed a simulated data set to investigate the performance of the regularization and cyclic de-



**FIG. 3.** Spatial distribution of activation signal and noise for simulated data set. In order to test the efficacy of the cyclic descent regularization algorithm, spatially varying activation signals were superimposed over spatially varying noise. The activation signals were aligned spatially over regions of transition from high to low AR noise power. Ideally, the regularization algorithm should be able to separate the spatial patterns, making use of spatial continuity to estimate the noise parameters without blurring the activation pattern.

scant procedures. This simulation set consisted of a 28 by 28 pixel image of 256-length time-series with spatial variations in both simulated BOLD signal and noise. The main reason for employing the regularization technique is to take advantage of local spatial smoothness in order to estimate more precisely the low-frequency physiological fMRI noise. However, it is also important that the method preserve spatial variations in this low-frequency noise without over-smoothing transitions from regions of high physiological noise power to regions with lower physiological noise power. To investigate this, we simulated spatial variations in AR(1) noise, with alternating stripes of high and low AR noise power, as illustrated in Fig. 3. The AR noise power is the integral of the power spectrum of the AR portion of the noise model (AR noise power =  $\sigma_\eta^2 / (1 - \rho^2)$ ). The noise varies abruptly across space from regions of high AR noise power (AR noise power = 4) to regions of lower AR noise power (AR noise power = 1) with increasing spatial frequency as one proceeds from the top of the image to the bottom of the image. The AR correlation parameter  $\rho$  and white noise variance  $\sigma_\eta^2$  are constant across the image ( $\rho = 0.75$ ;  $\sigma_\eta^2 = 1$ ) and the noise time-series are pixel-wise independent.

On top of the spatially varying noise, we superimposed a spatially varying simulated activation signal consisting of stripes of “active” and “inactive” regions, as illustrated in Fig. 3. In active regions, the time series contain a synthesized BOLD signal with  $f_a = 2$ ,  $f_b = f_c = 0$ , produced from a square-wave stimulus with eight OFF-ON periods. In inactive regions, there is only noise. To test the efficacy of the spatial regular-

ization and cyclic-descent fitting methods in separating the signal from the noise, we placed the active regions of the noise over the transitions from high to low AR(1) noise.

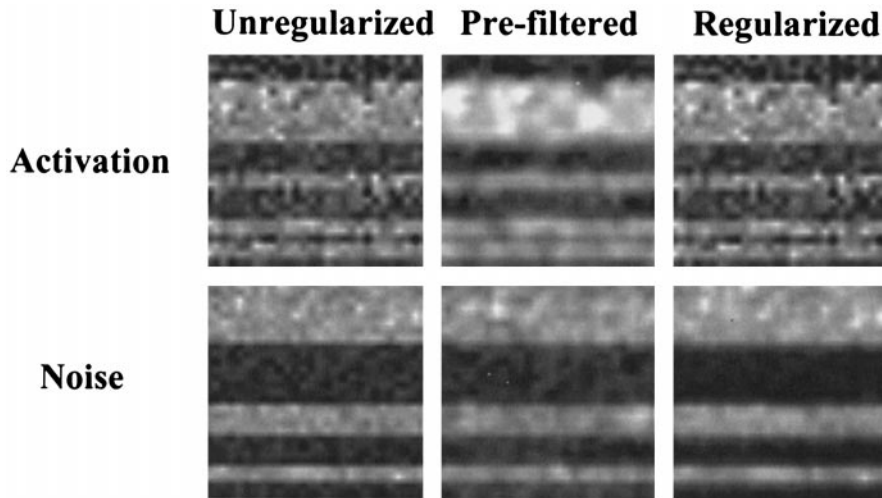
To compare the regularization technique to other techniques, this data set was analyzed in three different ways: (1) With no regularization, (2) With spatial smoothing prior to analysis, but with no regularization, and (3) With regularization on the AR parameters. The spatial presmoothing was accomplished by spatially filtering the data set with a 3-by-3 separable Hanning kernel with relative weights given by  $[1 \ 2 \ 1] [1 \ 2 \ 1]^T$ . A comparably smooth kernel was used for regularization. We gauge the performance of each method by examining: (1) the smoothness of the estimates in the regions where the activation or noise are constant across space; and (2) the sharpness of the transition between “active” and “inactive” regions or “high” and “low” noise.

Figure 4 shows the results of the data analysis from the simulated data set. The data processed with regularization showed a smoother estimate of the noise in regions of constant noise power (as seen through a map of the AR power) than either of the other two methods. The contrast is particularly clear in the “high-noise” regions. The sharpness of transition in the AR noise power under regularization was comparable to that under spatial prefiltering. The activation map shows substantial blurring for the spatially prefiltered case, particularly at the lower portion of the image where the activation varies rapidly across space, whereas the regularized data show identical spatial resolution to the untreated case. These data demonstrate that the regularization scheme is effective in improving the noise estimates while retaining the native spatial resolution of the activation map.

### 3.2. Human Data

Visual stimulus data from a previous study were analyzed to assess the performance of the LRST method on real data. These data were collected using an EPI at 1.5T (GE Signa scanner modified by Advanced NMR Systems). A subject was presented with a visual stimulus consisting of a full-field flickering checkerboard, presented in a 12.8 s OFF, 12.8 s ON pattern. The experiment included eight such periods, imaged at a TR of 800 ms, with a single slice chosen to transect the primary visual cortex. These data were analyzed in a manner identical to the simulated data set described above: (1) no regularization, (2) spatial smoothing prior to analysis, but no regularization, and (3) regularization on the AR parameters.

The results of this analysis are shown in Fig. 5 in terms of an activation map and noise map. The activation maps are quantified in terms of the square root of the weighted 2-norm of the estimated activation signal,



**FIG. 4.** Activation and noise plots for simulated data. Activation plots show the value for the parameter  $f_a$ , while the noise is displayed as the power in the AR portion of the noise spectrum. Pre-filtering blurs both the activation and noise estimates, while regularization is able to produce smooth estimates of the noise without blurring the activation estimates.

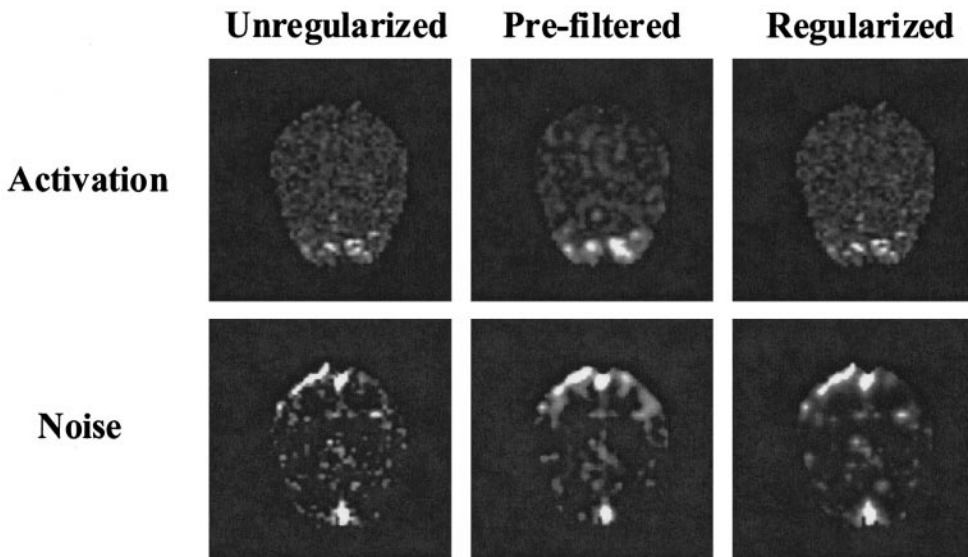
where the weighting is given by the local noise power. Therefore, at pixel  $p$ , we have

$$(\text{Activation})_p = \sqrt{\sum_k \frac{|\tilde{s}_{k,p}|^2}{F_{k,p}}},$$

where  $\sim$  denotes a discrete Fourier transform,  $s$  is the estimated BOLD signal defined in Eq. 2.4,  $F$  is the modeled noise spectrum, and the summation is over all the Fourier frequencies. An activation metric of this kind ensures that truly “active” regions of the brain

can be distinguished from regions that simply have large values of  $f_a$ ,  $f_b$ , and  $f_c$  due to a preponderance of noise. The noise images given here are displayed as the square root of the power in the AR portion of the noise model.

In all cases, the visual cortex was found to be active, in agreement with the experimental paradigm. Because we are only regularizing over the AR parameters, the activation maps under regularization have the same high spatial resolution as the unregularized map, whereas the spatially prefiltered map shows substantial blurring of the activation signal. We see that



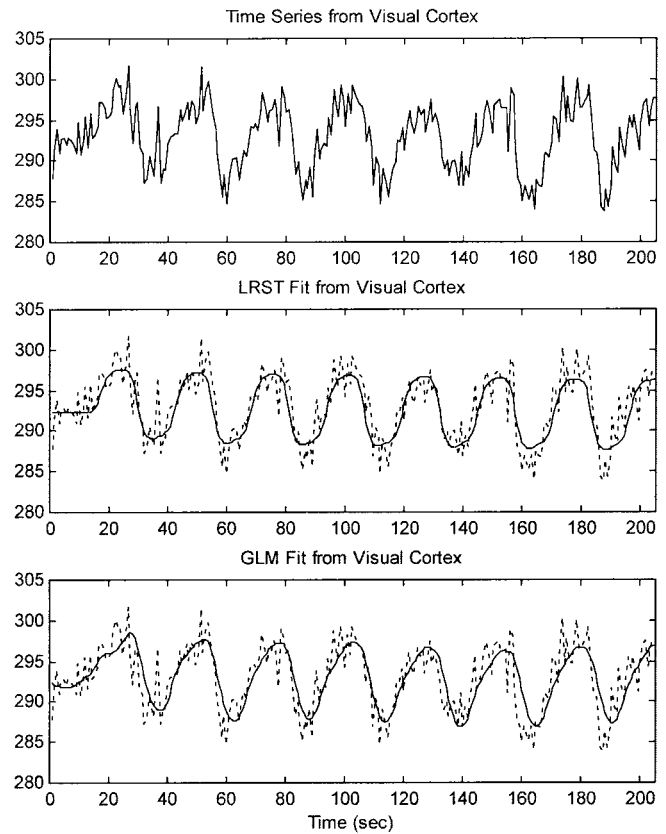
**FIG. 5.** Activation and noise plots for visual stimulus data. Activation is represented in terms of the square root of the power in the estimated activation signal, while the noise is displayed as the square root of the AR noise power. As with the simulated data, prefiltering blurs both the activation and noise estimates, while regularization is able to produce smooth estimates of the noise without blurring the activation estimates.

the noise estimates for the regularized case are smoother than the unregularized estimates and comparably smooth to the prefiltered estimates. The noise for the regularized and prefiltered estimates show some degree of contrast between cortical gray matter regions and white matter, coinciding with the notion that the noise stems from hemodynamic fluctuations. In the unregularized map, this contrast is less clear.

### 3.3. Model Comparison

To provide a comparison with existing analysis methods, the visual stimulus data were also analyzed with the GLM method (Worsley and Friston, 1995). This method was implemented using a second-order Volterra basis set as described in Friston *et al.* (1998) and seven low-frequency cosinusoids to act as the “high-pass filter,” accounting for slow drifts in the data (Friston, 1997). In order to explore on an intuitive level how well each model represents the dynamical features of the data, time-series of the model fits for the LRST and GLM were plotted and compared over several regions, as shown in Figs. 6 and 7. The time-series of the model fit were drawn from ROIs over the primary visual cortex (Fig. 6), superior sagittal sinus, white matter from the parietal region, and gray matter from the frontal region (Fig. 7). The pixel-wise NURE for each analysis method was then plotted and compared for an oblique axial slice superior to the ACPC line (Fig. 8).

The basic dynamical features of the BOLD signal—the delay, undershoot, and drift—are well accounted for by both the LRST and GLM methods, as seen in Fig. 6. In areas that should not be active under this paradigm, the estimated signal for both the LRST and GLM methods are very small (Fig. 7). Figures 8A and 8B show plots of the pixel-wise NURE for the LRST and the GLM. Uniformly across the image, the NURE for the LRST is nearly three times smaller than that for the GLM, suggesting that the LRST is closer to the true model, despite the similarity in time-series fit. Given the similarity in the fit, and since the model complexity terms are small compared to the overall NURE value for both models, the difference in the computed NURE values is due largely to the different noise specifications in the two models. To illustrate this, the autocovariance functions of the residuals from the visual cortex were computed after whitening the residuals with a whitening filter based on the noise model for both the LRST and GLM methods (Fig. 9). In the LRST case, this corresponds to the whitening filter implied by the AR(1) + white noise model (Purdon and Weisskoff, 1998), while in the GLM case this corresponds to the smoothing kernel used prior to fitting with ordinary least-squares. The autocovariance function from the LRST method shows little correlation at nonzero lags, suggesting that the noise model provides

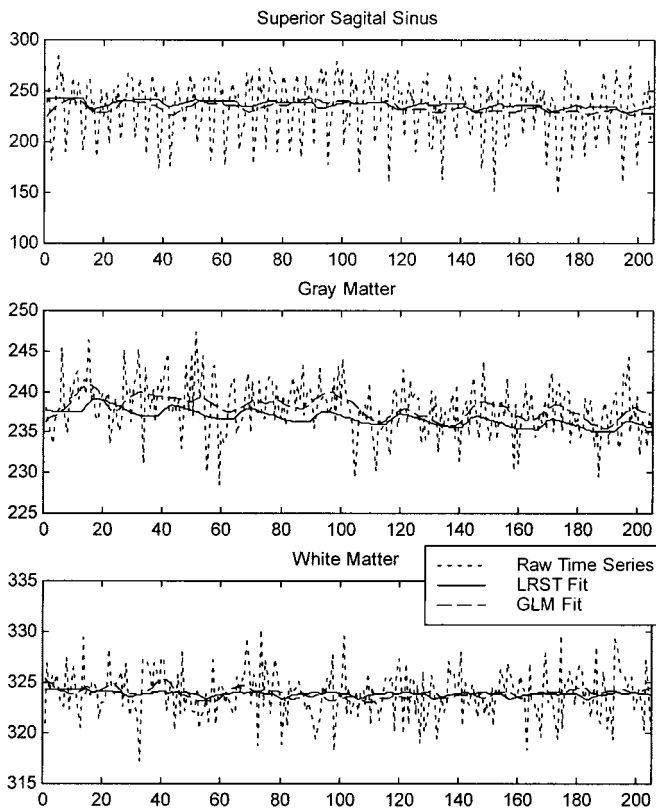


**FIG. 6.** Time-series and model fits for LRST and GLM from the visual cortex. The displayed plots are for time series and model fits averaged over an ROI in the visual cortex. The top panel shows the average time series data over the ROI, while the lower two panels show the model fits (solid line) superimposed over the data (dashed line) for the LRST and GLM models. Both the LRST and GLM methods provide a good temporal fit to the BOLD response.

an adequate representation of the temporal correlations in the time series. The autocovariance function from the GLM method shows correlation over several nonzero lags, suggesting that the noise model implied by the smoothing kernel improperly represents the temporal correlation structure.

### 3.4. Discussion

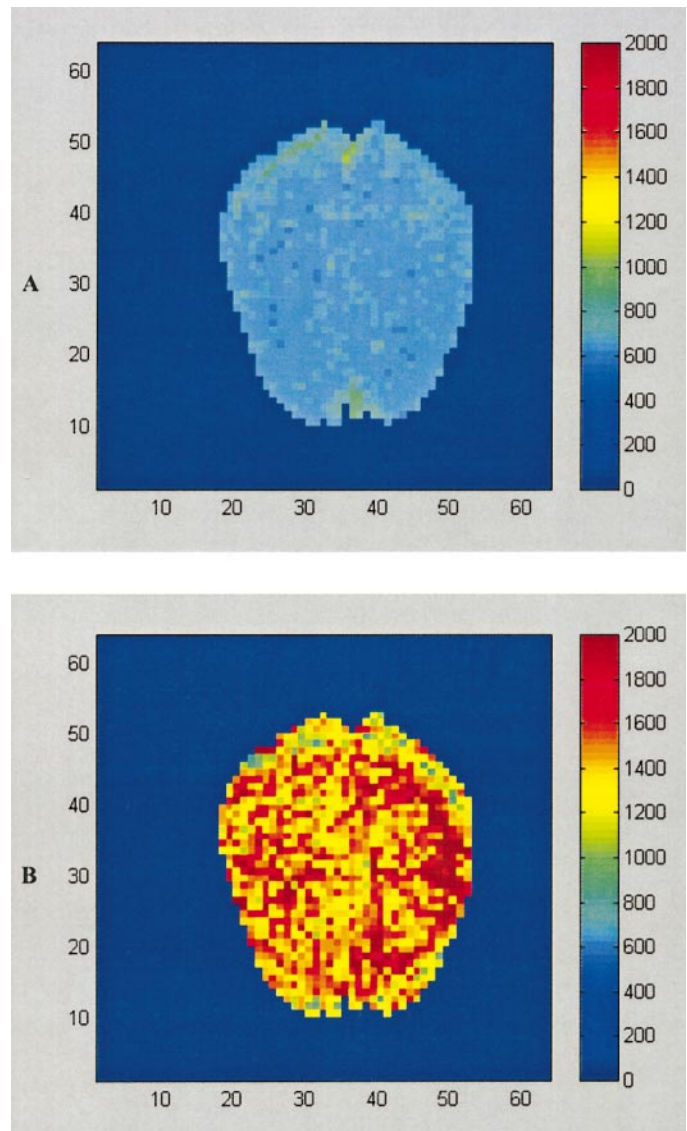
**3.4.1. Related work.** Friston and colleagues (Friston *et al.*, 1994, 1995; Worsley and Friston, 1995) have laid out a framework in their general linear model-statistical parametric map (GLM-SPM) that encompasses most correlation-type (e.g., Bandettini *et al.*, 1992) and *t*- and *F*-test strategies. They have used a spatially invariant hemodynamic response function both to smooth the time data and to make an approximate correction for the number of degrees of freedom in inferences based on the fMRI time series. Because this hemodynamic function goes to zero at high (0.5 Hz) frequencies, it is unstable under a weighted least



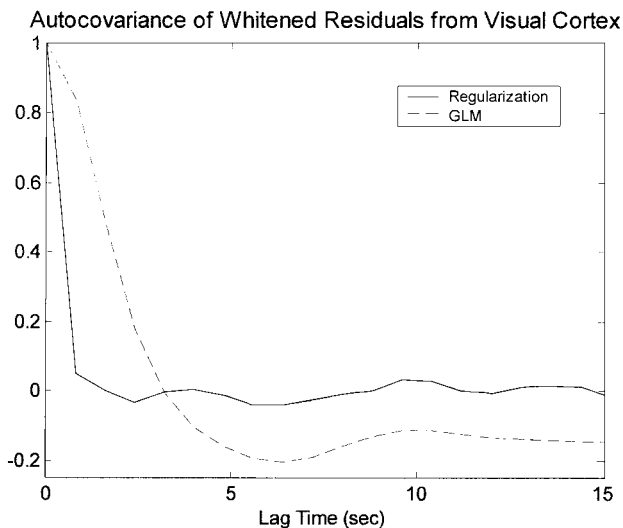
**FIG. 7.** Time-series and model fits for LRST and GLM from superior sagittal sinus, gray matter, and white matter. None of these areas should show activation with a visual stimulus, and both the LRST and GLM model fits reflect this.

squares fit. For this reason, they have chosen a strategy of an unweighted fitting of the covariates. Recent work has shown (Weisskoff *et al.*, 1993; Bullmore *et al.*, 1996) that the actual noise in the fMRI data sets is driven by both low frequency physiologic noise and white noise from the scanner. The situation is further complicated by the fact that the amount of noise coloring varies dramatically between cortical, subcortical, and white matter regions (Purdon *et al.*, 1998). The results of the model comparison suggest that the spatially varying AR(1) plus white noise model provides a representation of the noise that is superior to that used in the GLM. The GLM approach while suboptimal, is nonetheless computationally straight forward. However, using the hemodynamic response as the smoothing kernel assumes that the primary source of fluctuations in fMRI data are real activations and can induce fluctuations that appear like response activations. Use of the actual shape of the fluctuations in the fMRI noise leads to more reliable inferences by properly accounting for the true time correlation in the fMRI samples. While noise models can be difficult to estimate given the short time-series in fMRI studies, the use of spatial regularization can improve estimates of the noise parameters.

By choosing a physiologically motivated impulse response, we use four parameters (activation ( $f_a$ ), droop ( $f_b$ ,  $f_c$ ), poststimulus undershoot ( $f_d$ ), and time delay ( $D$ )) to model the activation and fit two AR parameters (amplitude and correlation) to the noise. While this model may not fully characterize the signal and noise, we believe that it captures the most important factors needed for inference and confidence interval generation, providing a reasonable and efficient starting point. While the functional form of the LRST model is similar to the Volterra expansion employed in Friston *et al.* (1998), with the “interaction” term resembling the second-order terms in the Volterra expansion of Fris-



**FIG. 8.** Pixel-wise NURE plots for LRST (A) and GLM (B) for an oblique axial slice superior to the ACPC line. The NURE values estimated for the LRST model are substantially lower than those for the GLM method, suggesting that the LRST model provides a better representation of the underlying process than does the GLM.



**FIG. 9.** Autocovariance functions for whitenes residuals under LRST and GLM methods. The autocovariance function from the LRST method shows little correlation at non-zero lags, suggesting that the noise model provides an adequate representation of the temporal correlations in the time series. The autocovariance function from the GLM method shows correlation over several nonzero lags, suggesting that the noise model implied by the smoothing kernel improperly represents the temporal correlation structure.

ton *et al.* (1998), the LRST model was formulated from a physiologically inspired standpoint. The resulting LRST model is much simpler than the Volterra expansion, requiring only three terms to achieve the BOLD signal fit in Fig. 5, compared to the 12 terms used in the Volterra expansion to achieve a similar fit.

There have been other approaches for incorporating more realistic signal and noise models in the fMRI analysis that parallel the spirit of our work. Bullmore *et al.* (1996) used a harmonic regression model for the signal and a first-order autoregressive (AR(1)) noise model for the noise. They used resampling methods to derive standard errors and to make inferences. We use a similar, though more physiologically based signal model whose parameters are more readily interpretable. Locascio *et al.* (1997) used a Box-Jenkins time series approach to model the noise assuming a boxcar type stimulus input, fitting an autoregressive moving average (ARMA) model in order to whiten the residuals from the fit. They have reported that the ARMA model is in many cases superior to the AR(1) model. The AR(1) plus white noise model used in the present study is equivalent to an ARMA(1,1) model. Burock and Dale (2000) applied a spatially varying AR(1) plus white noise model in the context of event-related fMRI and discovered that it maximized statistical efficiency while maintaining accurate control of type I error, relative to a global noise model of the same form or a white noise model. Studies by Bullmore *et al.* (2001) have shown that an AR(3) model, when used in conjunction with resampling methods, could provide for

accurate inferences. Furthermore, they introduce a wavelet-based resampling method that may be useful for cases of nonstationary noise. In a strictly periodic paradigm, Marchini and Ripley (2000) used a pixel-wise nonparametric spectral estimate in conjunction with nonlinear trend removal. This method produced very accurate inferences for periodic designs, but extensions to other types of designs have not yet been studied. Signal and noise modeling in fMRI will continue to evolve as statistical methods, computational methods, and knowledge of BOLD physiology improve. The model comparison method presented in this paper provides a framework for choosing between these various models for a given experiment.

Lange and Zeger (1997) dealt with the problem of spatial correlation in the frequency-domain. They used a fixed spatial smoothing to derive frequency-domain noise estimates, and then used them in standard frequency-domain tests of significance (Priestly, 1981). This approach has the advantage that it can account naturally for periodic fluctuations such as cardiac and respiratory-linked noises (Guimaraes *et al.*, 1995; Le *et al.*, 1996). Our proposed method differs in that: (a) model of the noises we propose; (b) the form of our signal model; and (c) way we accomplish the regularization. Ruttiman *et al.* (1996, 1998) used wavelets to test differences between ON-OFF responses in block design experiments. They do not estimate an impulse response and the thresholding technique they use is sensitive to colored noise (Silverman and Johnstone, 1997; Solo, 1998b). Descombes *et al.* (1998) take a Markov random field (MRF) approach. The same MRF prior is applied in both space and time but this does not seem wise since the spatial variation will be much more inhomogeneous than the temporal variation. Also an impulse response is estimated only after an initial MRF fit. Thus bias is introduced by applying the stimulus response relation late in the model fitting.

**3.4.2. Regularization.** A fundamental difference between current approaches and ours is our use of spatial regularization in the estimation process. Regularization, while familiar to applied mathematicians, engineers, and statistical modelers for solving ill-posed inverse problems (e.g., Groetsch, 1993) is less well-known to the functional MRI community. In GLM, the noise shape was assumed to be constant across the whole brain, requiring only a local estimate of the noise variance. While fMRI data sets typically have enough independent images to estimate this variance without pooling pixels, estimating pixel-by-pixel the more physiologically-based noise parameters, critical for producing accurate standard errors and inference, is more challenging. Since the noise varies dramatically over the brain, pooling across the entire brain to estimate the noise is not a physiologically sound approach. Empirically, the noise variation appears smoother

than the variation of local activation and thus its estimation is reasonably approached by spatial regularization. The regularization scheme does handle other signal parameters as well. It would be more usual to achieve spatial regularization by penalizing the global Gaussian log-likelihood  $\sum_p L_p(\theta_p)$  with a Tikhonov penalty (Natterer, 1986), but this leads to a prohibitively large global optimization problem. Our procedure requires only local computations and the actual optimization of our criterion function is carried out by cyclic descent (Solo *et al.*, 2001). Our methods are not intended to be a rigid solution for all types of functional experiments, though they naturally lead to generalizations for many fMRI-based explorations.

**3.4.3. Model comparison.** The common approach to selecting and/or comparing statistical models is based on hypothesis tests of model parameters. If there are a few parameters or a few well-formulated hypotheses, this works well, as for example in classical analysis of variance. But in many problems, and the fMRI problem in particular, the number of potential parameters is large compared to the number of actual data values. In such cases the hypothesis testing procedure is less appropriate because the number of hypotheses to be tested simultaneously can be intractably large. Problems like this have increasingly arisen as technological advances in all branches of engineering and science have produced challenging ill-conditioned inverse problems. Statistical methodology has moved away from hypothesis testing to comparison based on discrepancy measures, also called risk or distance measures. The risk measures compare models and evaluate model complexity by measuring for each model the expected “distance” between the true underlying process and the model. The theory behind this approach is further detailed in Linhart and Zucchini (1986) and Solo (1998a).

#### 4. CONCLUSIONS

We have developed a locally regularized spatiotemporal (LRST) system identification approach to fMRI data analysis and a systematic method for model comparison. The signal model was motivated by fMRI physiology, while the noise model was based on empirical studies. The model was constrained by specifying two time constants and a background noise variance. We have developed a new fitting procedure employing local spatial regularization that captures spatial continuity at low computational cost in order to improve estimates of noise parameters. We have also developed a model comparison procedure appropriate for fMRI data analysis based on Kullback–Leibler information distance.

These methods can be extended, both to include the much richer stimuli available in single-trial designs

and multiple-tasks, and to include methods, such as wavelets, that can better model spatial inhomogeneity.

#### APPENDIX

##### *Gaussian Log-Likelihood*

In this appendix we give expressions for the frequency-domain log-likelihood used in Eq. (2.8). The Fourier representation simplifies the log-likelihood function because the covariance matrix of  $v_{t,p}$ , the AR(1) plus white noise process, is approximately diagonal in the frequency domain. This log-likelihood function at pixel  $q$  is given by

$$L_q(\theta_p) = -\frac{1}{2} \sum_k \frac{|\tilde{x}_{k,q} - \tilde{\mu}_{k,p}(\beta_p)|^2}{N^2 F_k(\alpha_p)} - \frac{1}{2} \sum_k \log F_k(\alpha_p), \quad (\text{A.1})$$

where  $\tilde{x}_{k,q}$  is the discrete Fourier transform (DFT) of the data  $x_{t,q}$ ,  $F_k(\alpha_p)$  is the noise spectrum, and  $\tilde{\mu}_{k,p}(\beta_p)$  is the DFT of  $\mu_{t,p} = m_p + b_p t + s_{t,p}$ . The parameter vector  $\theta_p$  is indexed differently than the data  $\tilde{x}_{t,q}$  to allow for the spatially weighted regularization described by (2.8), which essentially models the spatial variation in the parameters as being locally constant. If we set  $p = q$ , (A.1) gives the expression for the log-likelihood at a single pixel.

##### *Weighting Kernel*

To keep the likelihood computations in (2.8) local, we use a separable, finite impulse-response Epanechnikov weighting kernel of the following form:

$$K_q^h = K\left(\frac{q_1}{M_1 h}\right) K\left(\frac{q_2}{M_2 h}\right) \frac{\kappa}{h^2 M_1 M_2}$$

$$K(u) = \frac{3}{4} (1 - u^2) \text{ for } |u| \leq 1$$

$$= 0 \text{ for } |u| \geq 1$$

$$\kappa = \left(1 - \frac{1}{4M_1^2}\right) \left(1 - \frac{1}{4M_2^2}\right), \quad (\text{A.2})$$

where  $M_1$  and  $M_2$  are the image dimensions,  $q_1$  and  $q_2$  are the pixel coordinates,  $h$  is the smoothness parameter controlling the width of the kernel, and  $\kappa$  is a correction factor for pixel discreteness.

#### ACKNOWLEDGMENTS

This research was partially supported by NIH grants NCRR P41 RR14075 to the Center for Functional Imaging Technologies at Massachusetts General Hospital, NIMH KO2 MH61637 to E.N.B., a

Whitaker Foundation graduate fellowship to P.P. and an HST Reed Scholars Award to P.P.

## REFERENCES

- Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., and Hyde, J. S. 1992. Time course EPI for human brain function during task activation. *Mag. Reson. Med.* **25**(2): 390–397.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. 1994. *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs.
- Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. 1996. Linear systems analysis of functional MRI in human V1. *J. Neurosci.* **16**: 4207–4221.
- Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., and Sham, P. 1996. Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Imag.* **35**: 261–277.
- Bullmore, E., Long, C., Suckling, J., Fadili, J., Calvert, J., Zelaya, F., Carpenter, T. A., and Brammer, M. 2001. Colored noise and computational inference in neurophysiological (fMRI) time series analysis: Resampling methods in time and wavelet domains. *Hum. Brain Mapp.* **12**: 61–78.
- Burnham, K. P., and Anderson, D. R. 1998. *Model Selection and Inference: A Practical Information Theoretic Approach*. Springer, New York.
- Burock, M. A., and Dale, A. M. 2000. Estimation and detection of event-related fMRI signals with temporally correlated noise: A statistically efficient and unbiased approach. *Hum. Brain Mapp.* **11**: 249–260.
- Cohen, M. S. 1997. Parametric analysis of fMRI using linear systems methods. *Neuroimage* **6**: 96–103.
- Cover, T. M., and Thomas, J. A. 1991. *Elements of Information Theory*. Wiley, New York.
- Descombes, X., Kruggel, F., and vonCramon, D. 1998. Spatio-temporal fMRI analysis using Markov random fields. *IEEE Trans. Med. Imag.* **17**: 1028–1039.
- Frank, L. R., Buxton, R. B., and Wong, E. C. 1998. Probabilistic analysis of functional magnetic resonance imaging data. *Magn. Res. Med.* **39**(1): 132–148.
- Friston, K. J., Jezzard, P., and Turner, R. 1994. Analysis of functional MRI time series. *Hum. Brain Mapp.* **1**: 153–171.
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., and Frackowiak, R. S. J. 1995. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* **2**: 173–181.
- Friston, K. J. 1997. SPM course notes. <http://www.fil.ion.ucl.ac.uk/spm/course/notes.html>.
- Friston, K. J., Josephs, O., Rees, G., and Turner, R. 1998. Nonlinear event-related responses in fMRI. *Magn. Res. Med.* **39**: 41–52.
- Genovese, C. R. 1999. Statistical inference in functional magnetic resonance imaging. Tech. Rep., Dept. of Statistics, Carnegie Mellon University.
- Groetsch, C. 1993. *Inverse Problems in the Mathematical Sciences*. Vieweg, Weisbaden.
- Guimaraes, A., Baker, J., and Weisskoff, R. 1995. Cardiac-gated functional MRI with T1 correction. *Soc. Magn. Reson.*, 3rd meeting, Nice, p. 798.
- Lange, N., and Zeger, S. L. 1997. Non-linear Fourier time-series analysis for human brain mapping by functional MRI. *Appl. Stat.* **46**: 1–29.
- Le, T., and Hu, X. 1996. Retrospective estimation and correction of physiological artifacts in fMRI by direct extraction of physiological activity from MR data. *Magn. Reson. Med.* **35**, 290–298.
- Linhart, H., and Zucchini, W. 1986. *Model Selection*. Wiley, New York.
- Locascio, J., Jennings, P., Moore, C., and Corkin, S. 1997. Time series analysis in the time domain and resampling methods for studies of functional Magnetic Resonance Brain Imaging. *Hum. Brain Mapp.* **5**, 168–193.
- Mandeville, J., Marota, J., Keltner, J., Kosofsky, B., Burke, J., Hyman, S., LaPointe, L., Reese, T., Kwong, K., Rosen, B., Weissleder, R., and Weisskoff, R. 1996. CBV functional imaging in rat brain using iron oxide agent at steady state concentration. *Int. Soc. Magn. Reson. Med.* 292.
- Mandeville, J., Marota, J., Kosofsky, B., Keltner, J., Weissleder, R., Rosen, B., and Weisskoff, R. 1998. Dynamic functional imaging of relative cerebral blood volume during rat forepaw stimulation. *Magn. Reson. Med.* **39**(4): 615–624.
- Marchini, J. L., and Ripley, B. D. 2000. A new statistical approach to detecting significant activation in functional MRI. *NeuroImage* **12**: 366–380.
- Marota, J., Mandeville, J., Kosofsky, B., Keltner, J., Burke, J., LaPointe, L., Weisslander, R., Rosen, B., and Weisskoff, R. 1996. Somatosensory stimulation mapping in rat brain: A functional MRI study. *Soc. Neurosci.*
- Mitra, P. P. and Pesaran, B. 1999. Analysis of dynamic brain imaging data. *Biophysical J.* **76**(2): 691–708.
- Natterer, F. 1986. *The Mathematics of Computerized Tomography*. Wiley, New York.
- Papoulis, A. 1991. *Probability, Random Variables, and Stochastic Processes*, 3rd ed. McGraw Hill, Boston.
- Priestley, M. 1981. *Spectral Analysis and Times Series*. Academic Press, Boston.
- Purdon, P. L., and Weisskoff, R. M. 1998. Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI. *Hum. Brain Mapp.* **6**(4): 239–249.
- Purdon, P. L., Solo, V., Rotte, M., Buckner, R. L., Brown, E. N., and Weisskoff, R. M. 1998. FMRI noise variability across subjects and trials: Insights for signal processing strategies. *Proc. ISMRM 6th Scientific Meeting*, Sydney, Australia, p. 253.
- Ruttimann, U. E., Unser, M., Thevenaz, P., Lee, C., Rio, D., and Hommer, D. W. 1996. Statistical analysis of differences by wavelet decomposition. In *Wavelets in Medicine and Biology* (A. Aldroubi and M. Unser, Eds.), pp. 115–124. CRC Press.
- Ruttiman, U. E., Unser, M., Rawlings, R. R., Rio, D., Ramsey, M. F., Mattay, V. S., Hommer, D. W., Frank, J. A., and Weinberger, D. R. 1998. Statistical analysis of functional MRI data in the wavelet domain. *IEEE Trans. Med. Imag.* **17**(2): 142–152.
- Silverman, B., and Johnstone, I. 1997. Wavelet threshold estimators for data with correlated noise. *J. R. Stat. Soc. Ser B* **59**: 319–352.
- Solo, V. 1998a. Transfer function order estimation with a  $H_\infty$  criterion. In *Proc. IEEE CDC98*.
- Solo, V. 1998b. Wavelet signal estimation in coloured noise with extension to transfer function estimation. In *Proc. IEEE CDC98*.
- Solo, V., Brown, E. N., and Weisskoff, R. 1997. A signal processing approach to functional MRI. In *Proc. IEEE ICIP97*, pp. 121–124, Vol. II.
- Solo, V., Purdon, P., Brown, E. N., and Weisskoff, R. 1998. Regularization for functional MRI models. In *Proc. IEEE ICIP98*.
- Solo, V., Purdon, P., Brown, E. N., and Weisskoff, R. 2001. A signal estimation approach to functional MRI. *IEEE Trans. Med. Imaging*, in press.
- Weisskoff, R. M., Baker, J., Belliveau, J., Davis, T. L., Kwong, K. K., Cohen, M. S., and Rosen, B. R. 1993. Power spectrum analysis of functionally-weighted MR data: What's in the noise. In *Proc. Soc. Magnetic Resonance in Medicine*, pp. 1–7.
- Worsley, K. J., and Friston, K. J. 1995. Analysis of fMRI time-series revisited—Again. *Neuroimage* **2**: 173–181.