HST 583

FUNCTIONAL MAGNETIC RESONANCE IMAGING DATA ANALYSIS AND ACQUISITION

A REVIEW OF STATISTICS FOR fMRI DATA ANALYSIS

EMERY N. BROWN AND CHRIS LONG

NEUROSCIENCE STATISTICS RESEARCH LABORATORY
DEPARTMENT OF ANESTHESIA AND CRITICAL CARE
MASSACHUSETTS GENERAL HOSPITAL

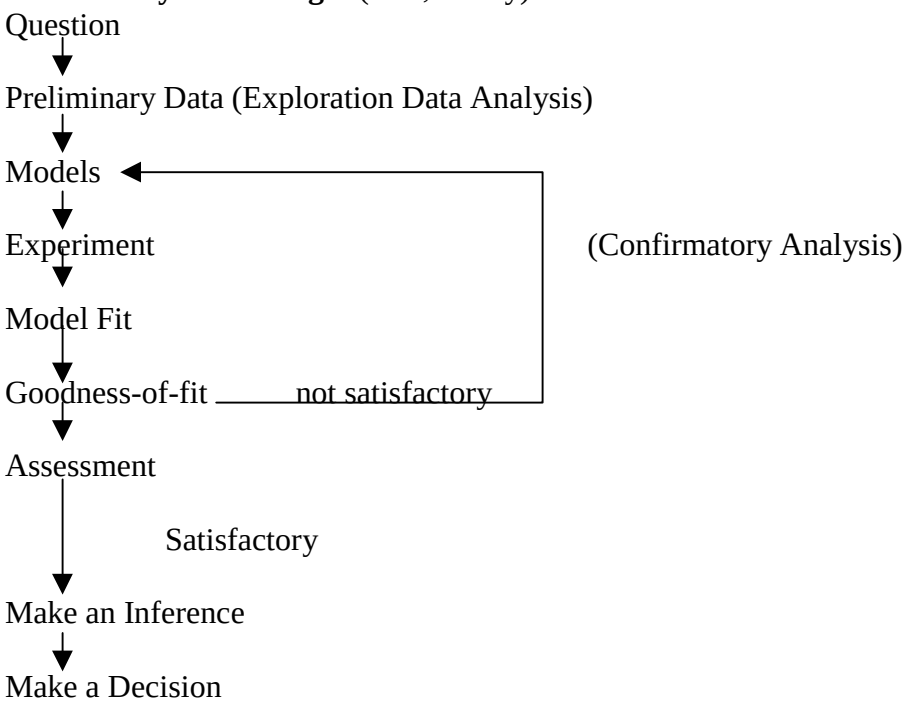
DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
HARVARD MEDICAL SCHOOL / MIT

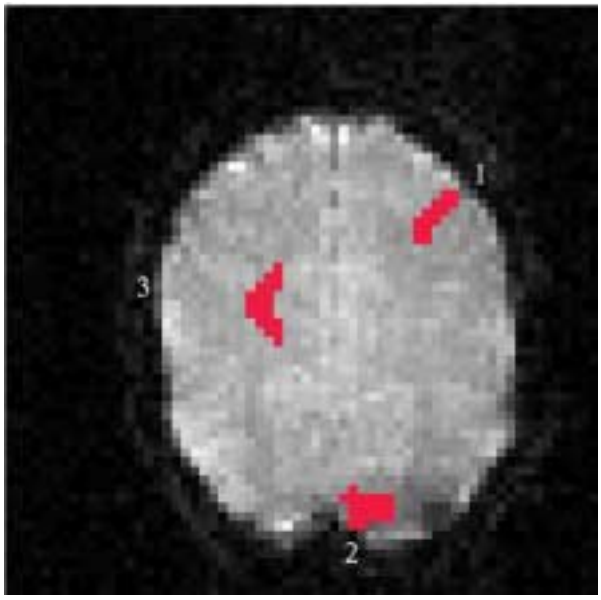
REVISED
THURSDAY, NOVEMBER 29, 2001

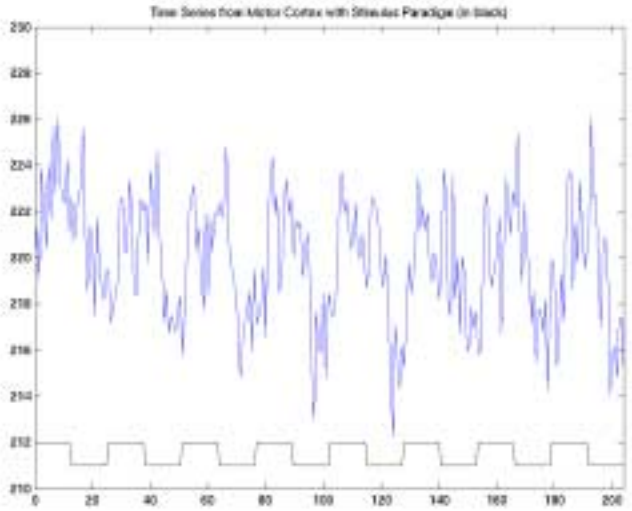
STATISTICS

The science of making decisions under uncertainty using mathematical models from probability theory.

1. **Statistical Analysis Paradigm** (Box, Tukey)

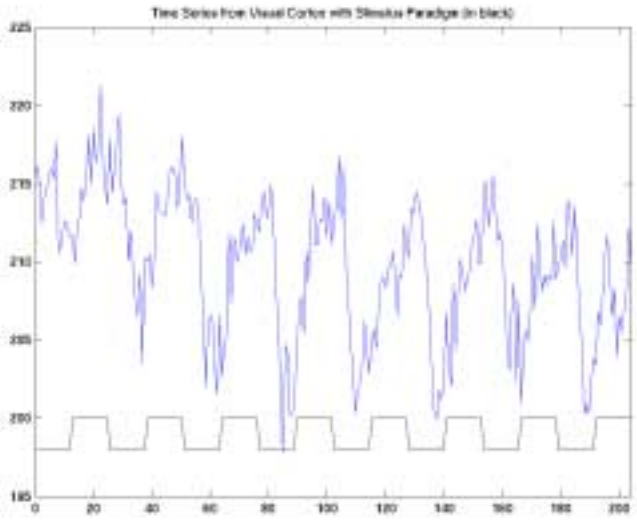






a)

b)



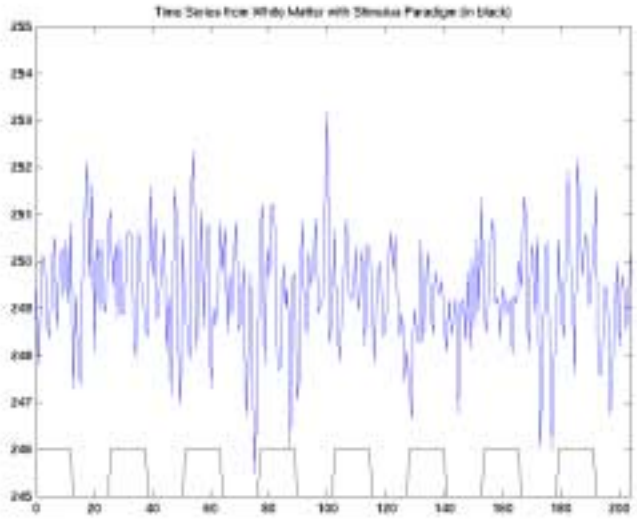


Figure 1: a) This panel shows a slice taken from a combined visual and motor fMRI experiment. The subject was presented with a full-field flickering checkerboard, in a 12.8-s OFF, 12.8-s ON pattern, repeated 8 times. The slice shown was chosen to transect both the visual and motor cortices, and was imaged once every 800ms for the duration of the experiment. Three regions of interest have been selected, corresponding to the motor cortex 1), the visual cortex 2), and the white matter 3). Figures b) – d) illustrate the raw timeseries taken from each of these regions, along with timing diagrams of the input stimulus.

Question: Is there significant activation in the visual and motor cortices during combined motor and visual tasks? Is the level of activation greater in the visual area than in the motor area?

2. What Makes Up An fMRI Signal

Hemodynamic Response/MR Physics

i) stimulus paradigm
   a) event-related
   b) block
ii) blood flow
iii) blood volume
iv) hemoglobin and deoxy hemoglobin content

Noise

Stochastic

i) physiologic
ii) scanner noise

Systematic

i) motion artifact
ii) drift
iii) [distortion]
iv) [registration]
v) [susceptibility]

3. Statistical Model

A fMRI Bold Signal (Neurovascular Coupling)

In a small time interval Δt we have on a given pixel

$$V(t) \times Hb(t) = (V_0 + \Delta V(t)) \times Hb_0 + \Delta Hb(t) . \quad (1)$$

V_0 initial blood volume
$\Delta V(t)$ change in blood volume
Hb_0 is the baseline deoxygenated hemoglobin
$\Delta Hb(t)$ change in deoxygenated hemoglobin

Assume that there is linear coupling of the stimulus to

$$Hb(t) = k_1 + k_2 (g_a * c) \quad (2)$$

$$g_a(t) = (1 - e^{-t/d_a})^2 (t+1) e^{-t/d_a} , \quad (3)$$

where k_1 and k_2 are constants, $c(t)$ is the stimulus input, $g_a(t)$ is a hemodynamic impulse response, chosen to be a discrete gamma function, d_a is a time constant

$$g_a * c = \sum_0^\infty g_a(u) c(t-u) . \quad (4)$$

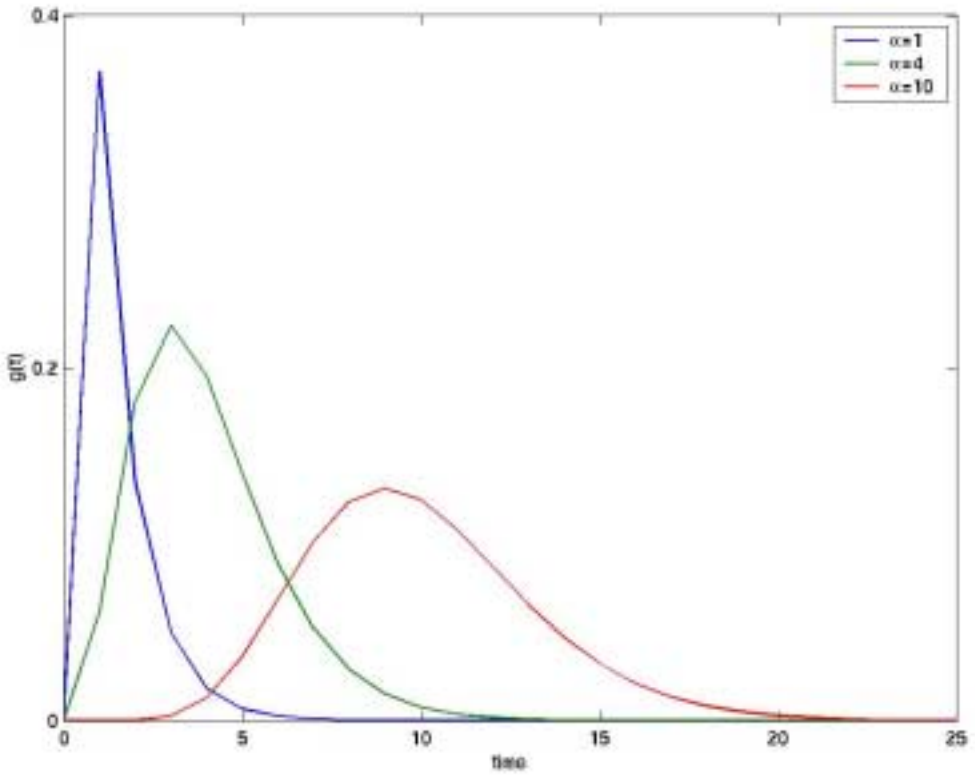


Figure 2: Examples of the gamma function when calculated with some different choices of α.

Now, the response to the blood volume is

$$V(t) = k_3 + k_4(g_b * c) \tag{5}$$

$$g_b(t) = (1 - e^{-t/d_a})e^{-t/d_b}, \tag{6}$$

where k_3 and k_4 are constants, $c(t)$ is the stimulus, $g_b(t)$ is an exponential impulse response function with time constant d_b.

The responses most likely have a delay D. Multiplying Eq. 5 by Eq. 6 and collecting terms, we obtain

$$s_t = f_a(g_a * c)_{t-D} + f_b(g_b * c)_{t-D} + f_c(g_a * c)_{t-D}(g_b * c)_{t-D},$$

where $f_a = k_2k_3$, $f_b = k_1k_4$ and $f_c = k_2k_4$. Physiologically f_a corresponds to the flow response, f_b the volume response, and f_c represents their interaction.

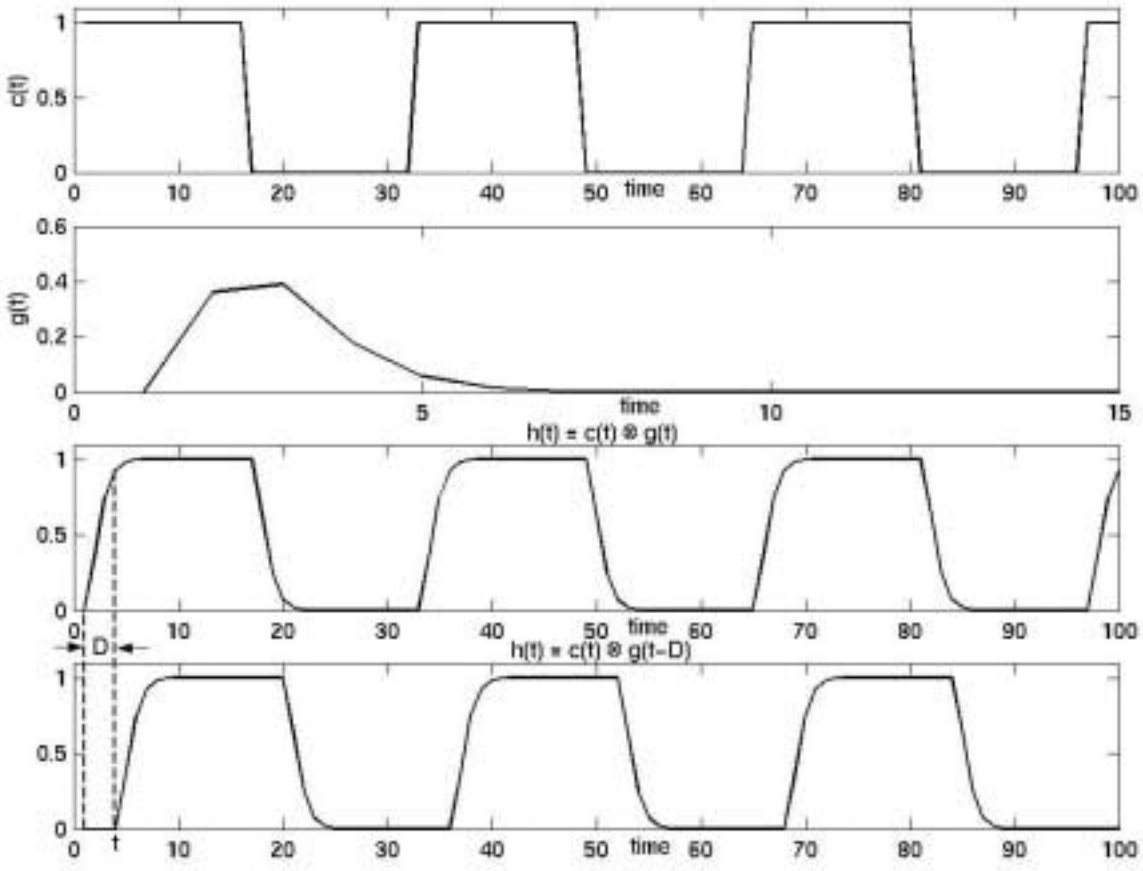


Figure 3 Sequence of steps followed when estimating a model for an fMRI experiment. We take advantage of the input function (top) and the physiology of the bold response (second down), to formulate a likely response for the brain (third and fourth panels). This is shown for the simplest case of convolution with one basis function.

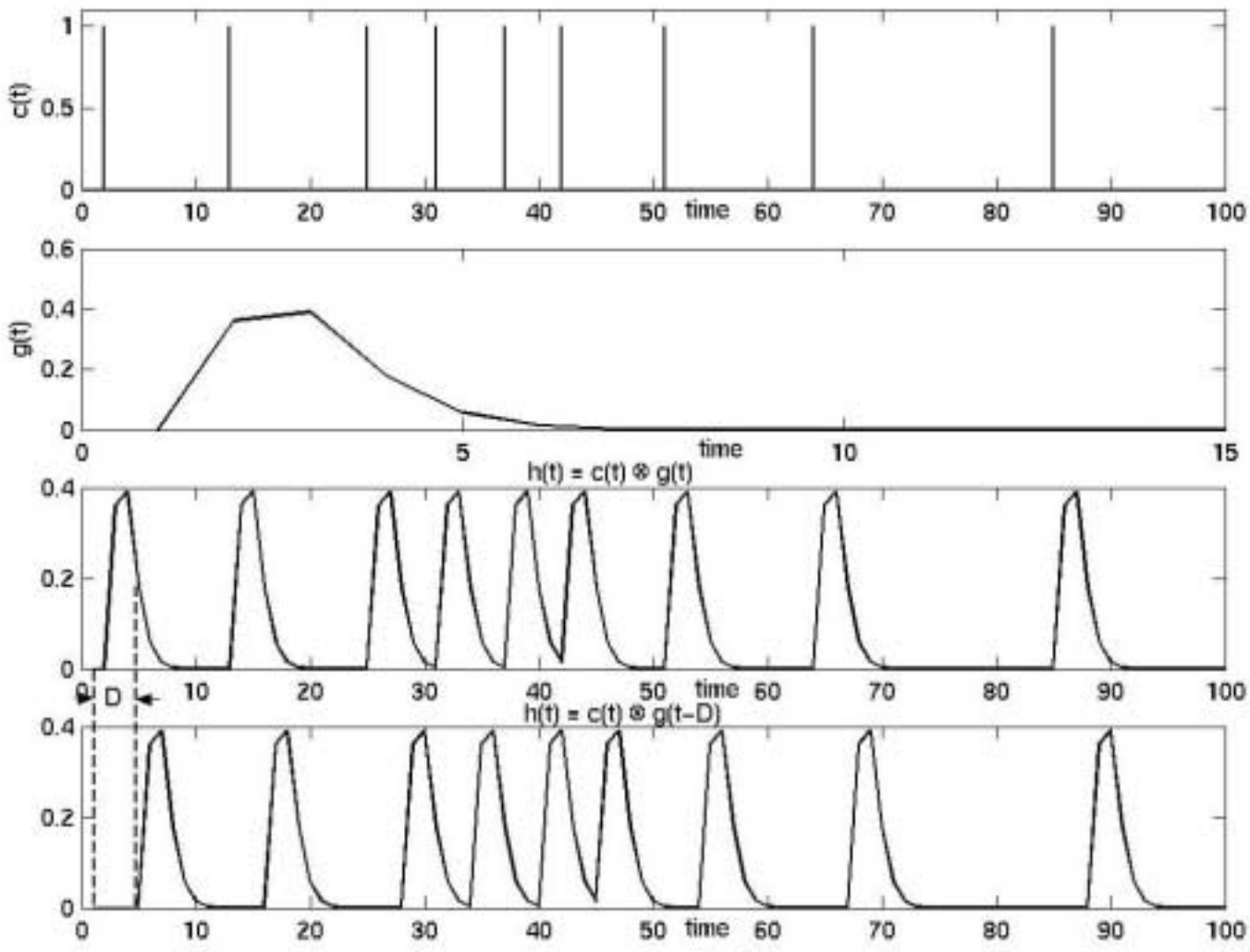


Figure 4 Same sequence of steps as for **Figure 3** except the input function is now event related rather than blocked-periodic. Furthermore the time intervals between consecutive responses is random.

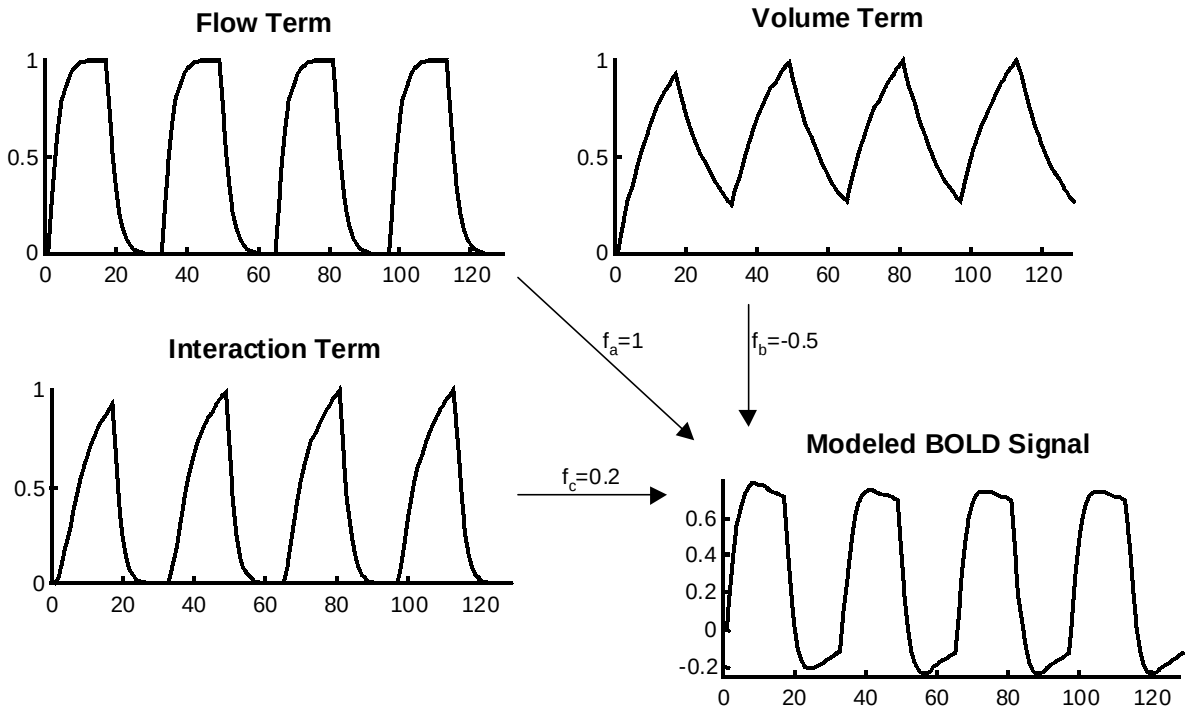


Figure 5 Illustration of BOLD signal model for a block-paradigm stimulus. The contribution of flow (f_a), volume (f_b), and interaction (f_c) terms combine to form an overall signal.

B. fMRI noise model (stochastic)

Physiologic low frequency noise and scanner noise is generally modeled as

$$v_t = w_t + n_t$$

$$w_t = \rho w_{t-1} + \varepsilon_t,$$

w_t first order Gaussian AR(1) process, ρ is the correlation coefficient, ε_t is zero mean Gaussian white noise with variance σ_t^2, n_t is zero mean Gaussian white noise with variance σ_η^2.

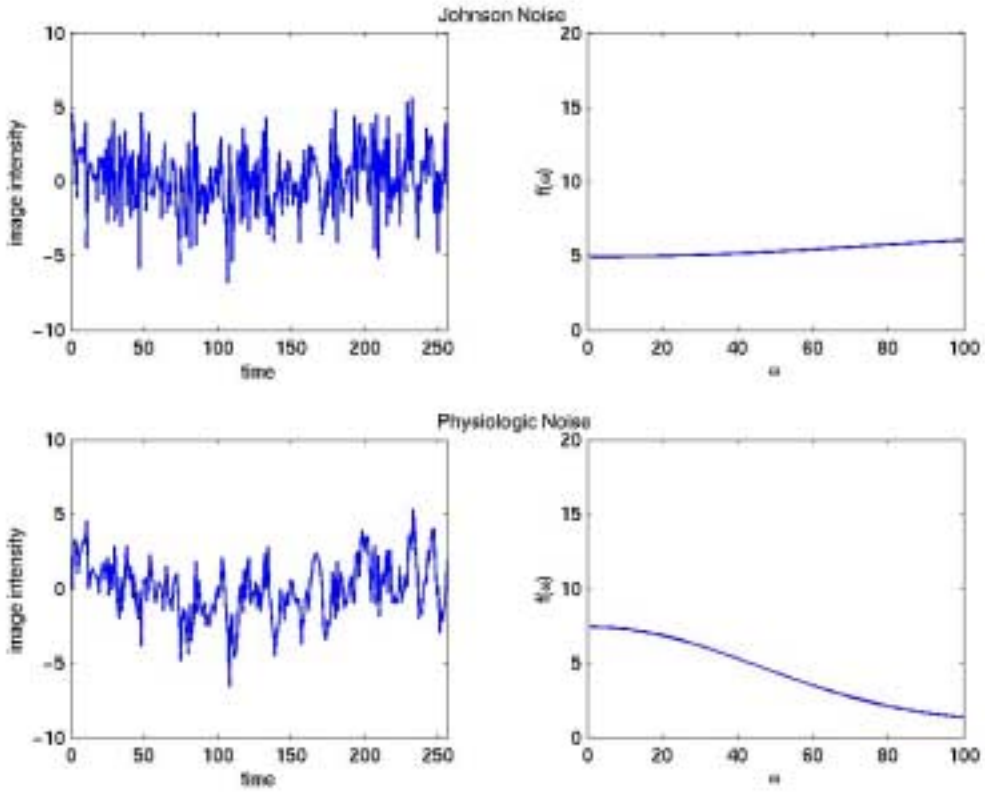


Figure 6 Illustration of two major noise classes encountered in fMRI. The left hand column summarizes the associated spectra.

C. Drift Term

Slow drifts of the static magnetic field and residual motion not accounted for by prior motion correction.

$$\text{drift} = m + bt$$

D. Complete Model is

fMRI signal = drift + hemodynamic response + noise

$$Y_t = m + bt + s_t + v_t$$

for $t = 1, \ldots, N$. In matrix notation

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} 1 & 1 & g_a * c_{1-D} & g_b * c_{1-D} & (g_a * c)_{1-D}(g_b * c)_{1-D} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & N & g_a * c_{N-D} & g_b * c_{N-D} & (g_a * c)_{N-D}(g_b * c)_{N-D} \end{bmatrix} \begin{bmatrix} m \\ b \\ f_a \\ f_b \\ f_c \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}$$

Or we have

$$Y = X(D, d_a, d_b)\beta + v(\rho, \sigma_\varepsilon^2, \sigma_\eta^2), \tag{7}$$

where $\beta = [m, b, f_a, f_b, f_c]$ is the set of linear parameters and D, d_a, d_b are the nonlinear parameters for the signal and $, \sigma_\varepsilon^2$ and σ_η^2 are the nonlinear parameters for the noise process.

Alternative fMRI Signal Model

Harmonic Regression Model

$$y_t = \mu + \beta t + \sum_{r=1}^{q} A_r \cos(\frac{2\pi}{\tau} rt) + B_r \sin(\frac{2\pi}{\tau} rt) + \varepsilon_t$$

In matrix notation becomes

$$\begin{bmatrix} y_1 \\ \\ \\ \\ \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cos(\frac{2\pi 1}{\tau}) & \sin(\frac{2\pi 1}{\tau}) & \cdots & \cos(\frac{2\pi q}{\tau}) & \sin(\frac{2\pi q}{\tau}) \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & N & \cos(\frac{2\pi N}{\tau}) & \sin(\frac{2\pi N}{\tau}) & & \cos(\frac{2\pi qN}{\tau}) & \sin(\frac{2\pi qN}{\tau}) \end{bmatrix} \begin{bmatrix} \mu \\ \beta \\ A_1 \\ b_1 \\ \vdots \\ A_q \\ B_q \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \\ \\ \\ \\ \varepsilon_N \end{bmatrix}$$

We can define the activation as $A_{c_t} = \sum_{r=1}^{q} (A_r^2 + B_r^2)^{\frac{1}{2}}$.

The joint probability of the signal plus noise model is the n-dimensional Gaussian density

$$f(v \mid \rho, \sigma_\varepsilon^2, \sigma_\eta^2) = \left(\frac{1}{2\pi}\right)^{N/2} \mid \Sigma_v \mid^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} v^T \Sigma_v^{-1} v\right\}, \tag{8}$$

where Σ_v is $n \times n$ the covariance matrix v. Now because $v = Y - X\beta$ and the Jacobian of the transformation is 1, we have the joint distribution of

$$f(Y \mid X, \beta, D, d_a, d_b, \rho, \sigma_\varepsilon^2, \sigma_\eta^2) = \left(\frac{1}{2\pi}\right)^{N/2} \mid \Sigma_v \mid^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(Y - X\beta)^T \Sigma_v^{-1} (Y - X\beta)\right\}. \tag{9}$$

Our objective is to estimate the parameters. This problem is complicated, we therefore study it as a set of simpler problems. Taking the log of both sides we obtain

$$\log f(Y \mid X\beta, D, d_a, d_b, \rho, \sigma_\varepsilon^2, \sigma_\eta^2) \propto -\frac{1}{2} \log \mid \Sigma_v \mid^{-\frac{1}{2}} - \frac{1}{2}(Y - X\beta)^T \Sigma_v^{-1} (Y - X\beta),$$

which is the log likelihood. The likelihood is the probability density of Y viewed now as a function of the parameters.

E. Maximum Likelihood Estimation for the fMRI Signal + Noise Model

Case I. D, d_a, d_b are known and $\Sigma_v = I\sigma_\varepsilon^2$. In this case the data are assumed to be independent identically distributed with common error ε_t's which have zero mean and common variance σ_ε^2. The ML estimates are just the least squares estimates. (See WAND 2001 Notes on Regression and the Generalized Linear Model).

Case II. D, d_a, d_b are known and Σ_v is known. The data are assumed to be correlated. The maximum likelihood estimates are the generalized least squares estimates (Homework, Problem 4).

Case III. D, d_a, d_b are unknown and Σ_v, i.e. $\sigma_\varepsilon^2, \sigma_\eta^2$ and ρ are unknown. We must do full maximum likelihood estimation. (Homework, Problem 5 illustrates with a simple example how the model parameters can be estimated simultaneously using cyclic descent).

Let $g(\theta) = \log f(X\beta, D, d_a, d_b, \rho, \sigma_\varepsilon^2, \sigma_\eta^2 Y)$ where $\theta = (\beta, D, d_a, \rho, \sigma_\varepsilon^2, \sigma_\eta^2)$ and suppose we knew D, d_a, d_b then compute $\hat{\beta}_{GLS}$ (Homework, Problem 4) and now $g(\theta)$ only depends $\theta' = (D, d_a, d_b, \rho, \sigma_\varepsilon^2, \sigma_\eta^2)$

$$g(\theta') = -\frac{1}{2}\log|\Sigma_v|^{-\frac{1}{2}} - \frac{1}{2}(Y - X\hat{\beta}_{GLS})^T \Sigma_v^{-1}(Y - X\beta_{GLS}) .$$

Now suppose θ' is close to the maximum likelihood estimate then by Taylor series expansion we get

$$\nabla g(\theta'_{ML}) = \nabla g(\theta') + \nabla^2 g(\theta)(\theta'_{ML} - \theta') + R .$$

At the maximum likelihood estimate $\nabla g = 0$ so we can write the following Newton's algorithm, by rearranging the Taylor series expansion

$$\theta'^{(\ell)} = \theta'^{(\ell-1)} - \nabla^2 g(\theta'^{(\ell-1)})^{-1} \nabla g(\theta'^{(\ell-1)} ,$$

where $\theta'^{(\infty)}$ is the maximum likelihood estimate. By our factoids

$$\hat{\beta}_{ML} = \hat{\beta}(\theta'_{ML}) .$$

Note that

$$I(\theta') = -E\left[\nabla^2 g(\theta')\right],$$

and hence an estimate of the Fisher information for θ' is

$$I(\theta'_{ML}) = -\nabla^2 g(\theta'_{ML}) ,$$

which is called the observed Fisher information.

Remarks

1. Starting values are important.

2. Computing $|\Sigma_v|$ and Σ_v^{-1} are in general hard. Note it is $N \times N$. Computation must be carried out so that parameter values are admissible, e.g. $0 < \rho < 1, \quad \sigma_\varepsilon^2 > 0, \sigma_\eta^2 > 0$.

3. We can use alternatives to Newton's method.

   i) Fisher scoring algorithm.
   ii) Cyclic descent (Homework Problem 5).

4. Once we have the maximum likelihood estimates computing confidence intervals is straightforward, because of our factoids.

G. Examples of Data Analyses

G.0 Model Selection Techniques: (Akaike's Information Criterion)

We saw in the least squares analyses that increasing the number of model parameters to equal the number of data points leads to a perfect model fit. Various criteria to measure the parasimony of statistical models have been proposed. Intuitively, the least squares or log likelihood function is penalized for each increase in the number of parameters in order to measure the trade-off between improving goodness of fit and increasing the number of parameters. The criterion we use to measure this trade-off is Akaike's Information Criterion (AIC) (Box, Jenkins and Reinsel, 1994). It is defined as

$$AIC = -2 \log f(\hat{\theta}_{ML}) + 2p$$

P is the number of the parameters in the model parameters and the remaining expression on the rhs is the log likelihood function. AIC can be used to compare different models as we illustrate below. The smaller the AIC the more parsimonious the model.

G.1 Harmonic Regression Hemodynamic Response Model

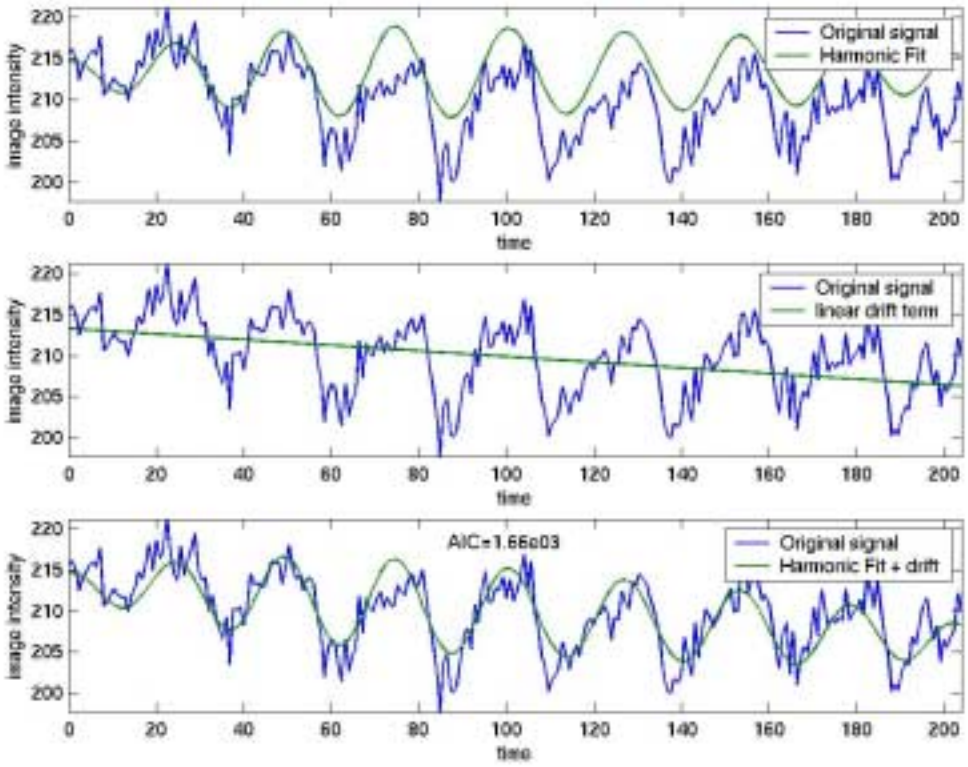


Figure 7 This figure shows the effect of fitting a simple harmonic model to the time series in the presence of an assumed white noise. The top panel exhibits the reconstruction of the model from the estimated regression parameters, the middle panel indicates the extent to which a linear drift was present in the data. The bottom panel merges these two effects into one graph. Note the Akaike Information criterion (AIC) for this method was 1.66e+03, *c.f.* **Figure 9**.

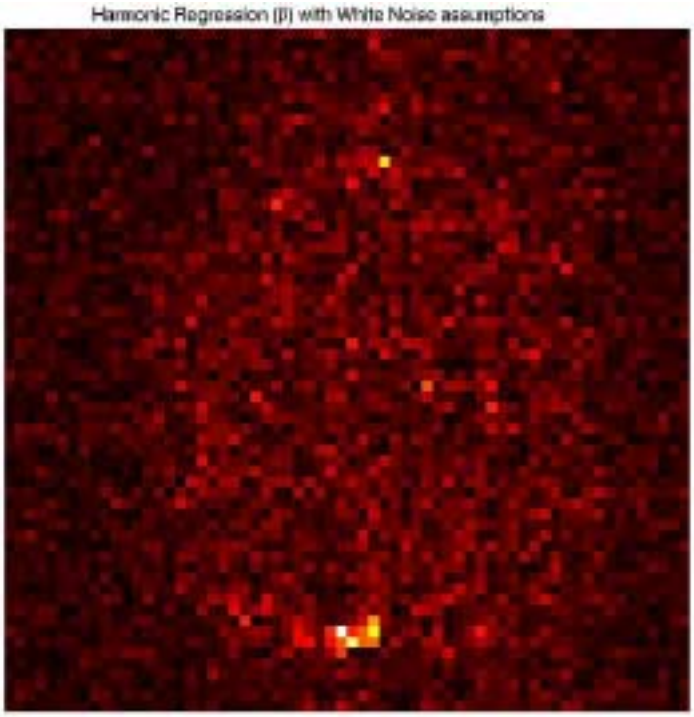


Figure 8 Raw activation map of the Visual experiment when analyzed using the harmonic regression method. Some strong activation is seen in the visual cortex (bottom center), however there is a large amount of background noise in the overall estimate.

G.2. Simple Convolution Plus Correlated Noise Model

We assume only a flow component in our model and that there is no volume or interaction between the flow and volume components of our model. We also assume that the scanner noise is zero, so that we have just the AR(1) noise model.

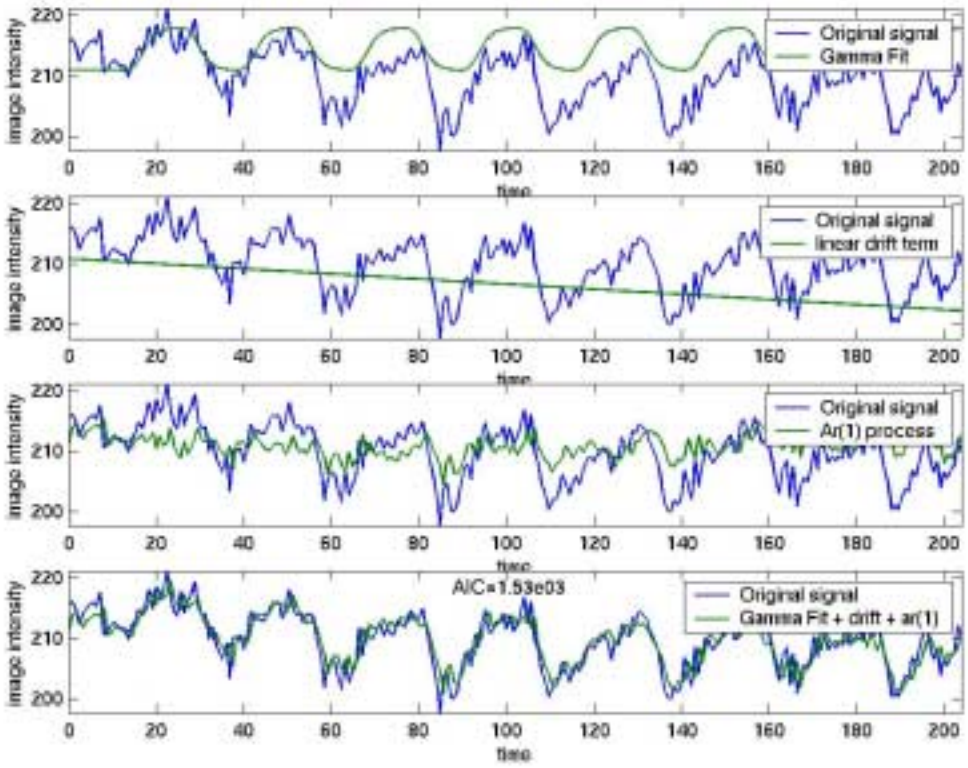


Figure 9 This figure illustrates the effect of each component in the convolution model within the context of regression. The top panel shows the reconstructed model alone, superimposed on the observation. The subsequent figure shows the linear trend estimated in the method, while the next reconstructs the estimated AR(1) process only. Finally the bottom panel shows the reconstruction when all the regression parameters have been taken into consideration. For this pixel, the AIC was calculated to be 1.53e03 to facilitate a simple comparison with the harmonic regression, which for this pixel was larger, hence inferior.

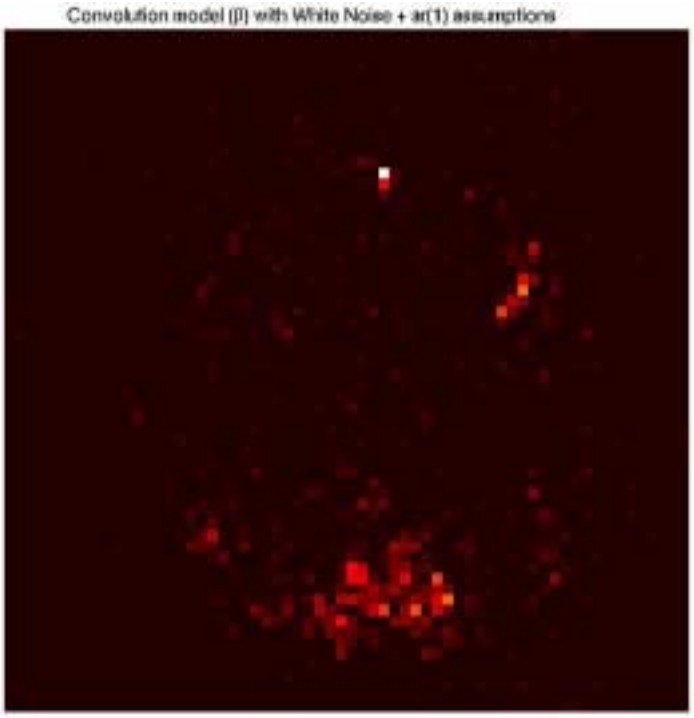


Figure 10 Raw activation map of the visual paradigm using the simple convolution model and assuming AR(1) noise in the regression.

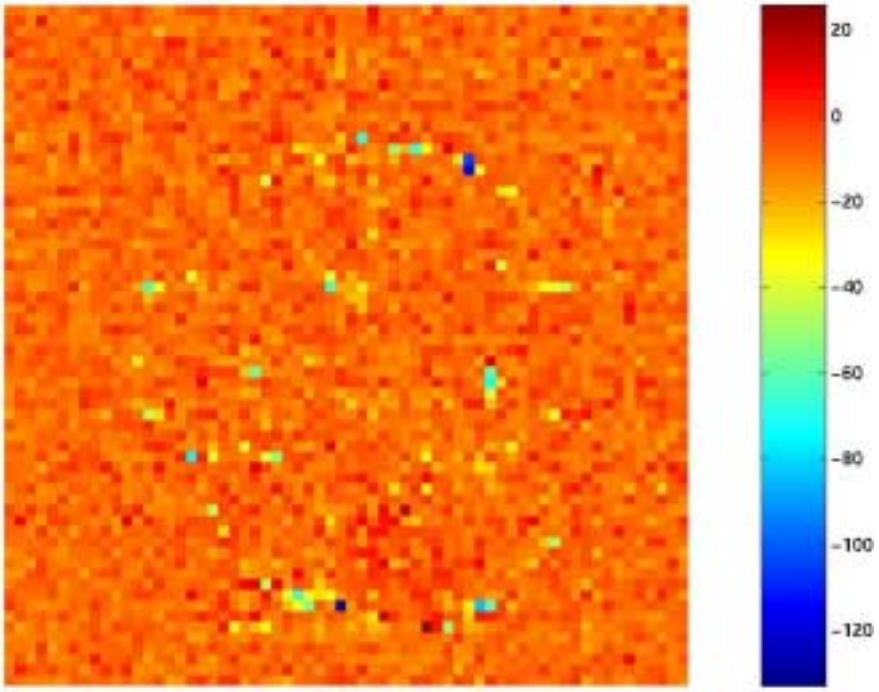


Figure 11 Difference between AIC map of the convolution model and the harmonic model, respectively. This map indicates the convolution model has the smaller AIC of the two methods.

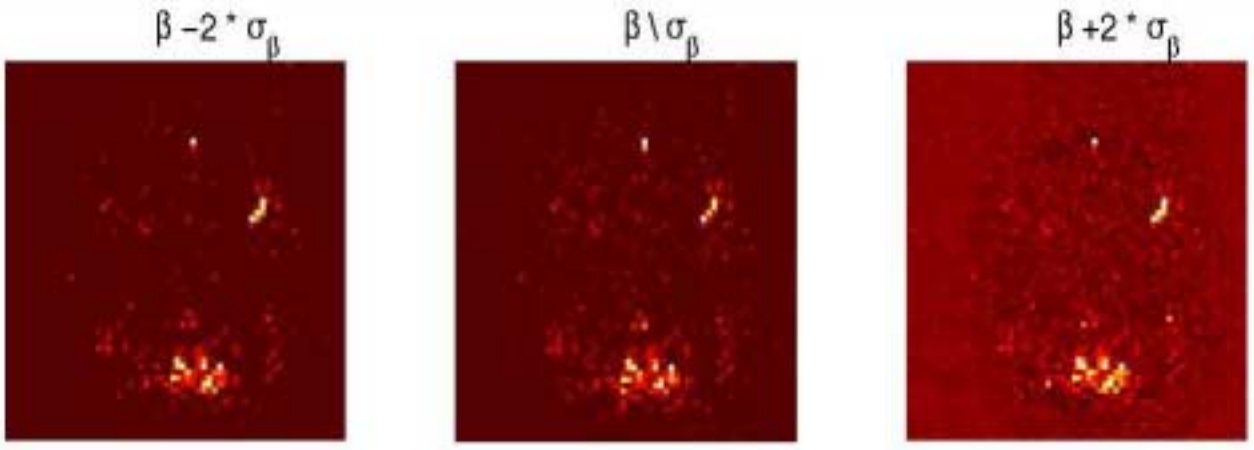


Figure 12 Activation maps for the convolution model at the upper (left), and lower (right) extremities of the fisher information confidence limits. The central map lies in the middle of this bound and has been normalised, in this instance, by the variance of β.

References

Casella G, Berger RL. *Statistical Inference*. In Kimmel J. (Ed). Belmont, CA, Duxbury Press 1990.

Purdon P, Solo V, Brown EN, Weisskoff P. Locally regularized spatio-temporal modeling and model comparison for functional MRI. *NeuroImage* 2001;**14**:912-23.

Ryan TP. *Modern Regression Methods*. Wiley:New York 1997.

Solo V, Purdon P, Brown E, Weisskoff R. A signal estimation approach to functional MRI. *IEEE Trans. Medical Imaging* 2001;**20**:26-35.