An fMRI Activation Method Using Complex Data

Daniel B. Rowe[1] and Brent R. Logan[2,*]
Department of Biophysics[1] and Division of Biostatistics[2]

Division of Biostatistics
Medical College of Wisconsin

Technical Report 45

February 2004

Division of Biostatistics
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI 53226
Phone:(414) 456-8280

An fMRI activation method using complex data

Daniel B. Rowe[1*] and Brent R. Logan[2]

Department of Biophysics[1] and Division of Biostatistics[2]

Medical College of Wisconsin

Milwaukee, WI USA

Abstract

In functional magnetic resonance imaging, voxel time courses after Fourier "image reconstruction" are complex valued as a result of phase imperfections due to magnetic field inhomogeneities. Nearly all fMRI studies derive functional "activation" based on magnitude time courses [1, 2]. Here we propose to directly model the entire complex or bivariate data rather than just the magnitude data. A nonlinear multiple regression model is used to model activation on the complex signal, and a likelihood ratio test is derived to determine activation in each voxel. We investigate the performance of the model on a real dataset, then compare the magnitude and complex models under varying signal to noise ratios in a simulation study with varying activation effects.

1 Introduction

In magnetic resonance imaging, we aim to image the density of "spinning" protons in a real valued physical object. The equations of Physics work out that the Fourier transform (FT) of the proton spin density (PSD) is a spatial frequency spectrum. We will obtain the spatial frequency spectrum and perform an inverse Fourier transform (IFT) to obtain the proton

*Corresponding Author: Daniel B. Rowe, Department of Biophysics, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226, dbrowe@mcw.edu.

spin density. This is done by taking successive measurements in time of a real valued signal, a voltage in a wire. The time axis is transformed to the spatial frequency or k-space axis. This physical signal or voltage is real valued, but it is "complex demodulated." In measuring the signal, there can be either one or two A to D converters. If there is a single A to D converter, successive signal measurements are alternately multiplied by either a cosine or a sine to obtain real (inphase) and imaginary (quadrature) parts. These two measurements are then shifted either half a step forward or backward to temporally align them. If there are two A to D converters, two measurements are then taken at the same time with one multiplied by a cosine and the other by a sine. This discretely measured complex valued signal is the discrete FT of the PSD. A discrete IFT is applied to the discretely measured signal. The original object or PSD is real valued but due to phase imperfections, a complex image of PSD's is produced [3].

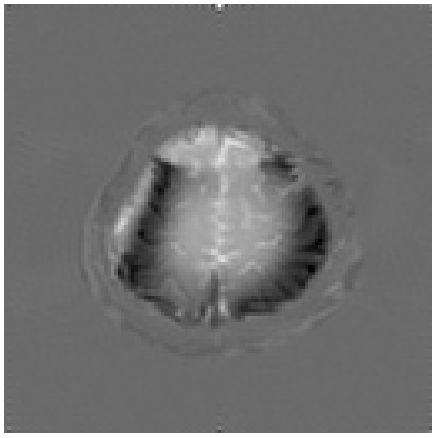
After Fourier image reconstruction, each voxel contains a time course of real and imaginary components of the measured PSD. Magnitude images are produced by taking the square root of the sum of squares of the real and imaginary parts of the measured PSD in each voxel at each time point. Nearly all fMRI studies obtain a statistical measure of functional activation based on magnitude image time courses. When this is done, phase information in the data is discarded. This is illustrated in Figure 1, where the real, imaginary, magnitude, and phase images are shown at a single point in time, for the example dataset discussed later.

Magnitude models typically assume normally distributed errors; alternatively, one can assume that the original real and imaginary components of the PSD have normally distributed errors. Independent normally distributed errors on the measured complex signal or equivalently complex PSD translates to a Rician distributed magnitude image that is approximately normal for large signal to noise ratios.

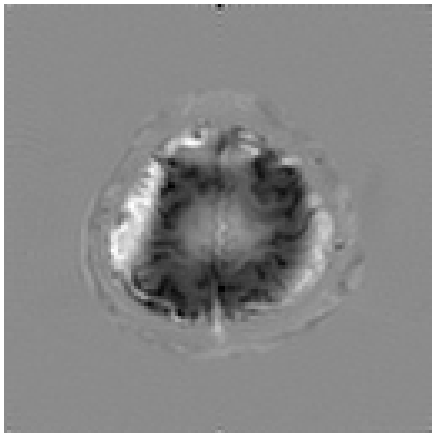
When computing magnitude image time courses and activations, the signal to noise ratio (SNR) may not be large enough for this approximate normality to hold. This is increasingly true with higher voxel resolutions. In addition, phase information or half of the numbers is discarded. A more accurate model should properly model the noise and use all the information contained in the real and imaginary components of the data.

Previous models for complex activation have been proposed [4, 5, 6]. Previous simple

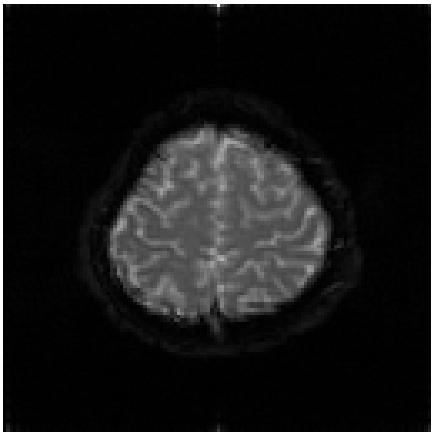
Figure 1: Real, imaginary, magnitude, and phase images at a fixed point in time.



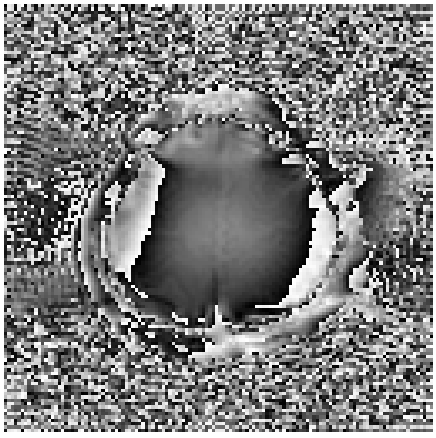
(a) Real image



(b) Imaginary image



(c) Magnitude image



(d) Phase image

linear regression models by Scharf and Friedlander (1994) and Lai and Glover (1997) did not accurately model the phase, while we correctly account for it through a nonlinear multiple regression model. A subsequent model by Nan and Nowak (1999) correctly assumed that the phase imperfections for the baseline and signal were the same, but was limited to a single baseline and signal model because of their model parameterization. In addition, their model did not directly estimate the regression coefficients or phase angle. We reparameterize and extend the model proposed by Nan and Nowak (1999) to a multiparameter baseline and signal model. We formulate the hypothesis test in terms of contrasts, which allows for more

elaborate hypothesis testing such as deconvolution and comparisons between multiple task conditions. Finally, our parameterization allows us to estimate the phase angle directly instead of the sine and cosine of the phase angle. We compare the results of the proposed model to a strict magnitude model in terms of thresholded activation maps on a real dataset. Finally, simulations are performed comparing our model to a magnitude model for various signal to noise ratios and task related effects.

2 Model

In MRI/fMRI, we aim to image a real valued physical object $\rho(x, y)$ and obtain a measured object $\rho_m(x, y)$ by measuring a 2D complex valued signal $s_m(k_x, k_y)$ at spatial frequencies (k_x, k_y). This signal consists of a true complex valued signal $s(k_x, k_y)$ plus a random complex noise term $\delta(k_x, k_y)$ with real and imaginary components that are assumed to be independent and identically normally distributed. Even if there were no phase imperfections, it is necessary to observe the imaginary parts of this signal because we phase encode for proper image formation. After image reconstruction, we obtain a complex valued measured object plus complex valued noise.

Neglecting the voxel location and focusing on a particular voxel, the complex valued image measured over time in a given voxel is

$$\rho_{mt} \quad = \quad [\rho_{Rt} + \eta_{Rt}] + i[\rho_{It} + \eta_{It}]$$

where $(\eta_{Rt}, \eta_{It})' \sim \mathcal{N}(0, \mathbf{\Sigma})$ and $\mathbf{\Sigma} = \sigma^2 I_2$. The distributional specification is on the real and imaginary parts of the image and not on the magnitude.

A nonlinear multiple regression model is introduced individually for each voxel that includes a phase error θ in which at time t, the measured proton spin density is given by

$$\rho_{mt} \quad = \quad [x_t'\beta \cos\theta + \eta_{Rt}] + i[x_t'\beta \sin\theta + \eta_{It}] \tag{2.1}$$

where $\rho_t = x_t'\beta = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_q x_{qt}$. The phase imperfection in Equation 2.1 is assumed to be fixed but unknown, not changing over time, but measured with error. Just as in Nan and Nowak (1999), we have also found this phase specification to be reasonable.

In fMRI, we take repeated measurements over time while a subject is performing a task. In each voxel, we compute a measure of association between the observed time course and a preassigned reference function that characterizes the experimental paradigm.

2.1 Magnitude Activation

The typical method to compute activations [1, 2] is to use the magnitude $|\rho_{mt}|$ which is denoted by y_t and written as

$$y_t = \left[(x_t'\beta\cos\theta + \eta_{Rt})^2 + (x_t'\beta\sin\theta + \eta_{It})^2\right]^{\frac{1}{2}}. \tag{2.2}$$

The magnitude model in Equation 2.2 discards any information contained in the phase, given by

$$\phi_t \;=\; \tan^{-1}\left[\frac{\rho_{It}+\eta_{It}}{\rho_{Rt}+\eta_{Rt}}\right].$$

The magnitude is not normally distributed but is Rician distributed. Both the magnitude and the phase are approximately normal for large SNR's [7, 8] as outlined in the appendix. The special case of the Rician where there is no signal is known as the Rayleigh distribution. It is known [3] that a histogram of noise outside the brain without any signal is Rayleigh distributed.

The Rician distribution is approximately normal for large signal to noise ratios (small relative error variance). This can be shown by completing the square in Equation 2.2 and proceeding as follows

$$\begin{aligned} y_t &= \left\{[x_t'\beta]^2 + [\eta_{Rt}^2+\eta_{It}^2] + 2[x_t'\beta][\eta_{Rt}\cos\theta + \eta_{It}\sin\theta]\right\}^{\frac{1}{2}} \\ &= [x_t'\beta]\left\{1 + \frac{2[\eta_{Rt}\cos\theta+\eta_{It}\sin\theta]}{[x_t'\beta]} + \frac{[\eta_{Rt}^2+\eta_{It}^2]}{[x_t'\beta]^2}\right\}^{\frac{1}{2}} \\ &\approx x_t'\beta + \epsilon_t \end{aligned} \tag{2.3}$$

where $\epsilon_t = \eta_{Rt}\cos\theta + \eta_{It}\sin\theta \sim N(0,\sigma^2)$. Again, the $\cos\theta$ and $\sin\theta$ arise from phase imperfections. If there were no phase imperfection, then $\theta = 0$. In this derivation, the approximation $\sqrt{1+u} \approx 1 + u/2$ was used for $|u| \ll 1$. This model can also be written as

$$\begin{array}{ccccccc} y & = & X & & \beta & + & \epsilon \\ n\times 1 & & n\times(q+1) & & (q+1)\times 1 & & n\times 1 \end{array} \tag{2.4}$$

where $\epsilon \sim \mathcal{N}(0, \sigma^2\mathbf{\Phi})$ and $\mathbf{\Phi}$ is the temporal correlation matrix, often taken to be $\mathbf{\Phi} = I_n$ after suitable pre-processing of the data.

The unconstrained maximum likelihood estimates of the vector of regression coefficients $\hat{\beta}$ and the error variance $\hat{\sigma}^2$ are derived in the appendix and given by

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y, \\ \hat{\sigma}^2 &= (y - X\hat{\beta})'(y - X\hat{\beta})/n \ . \end{aligned} \tag{2.5}$$

To construct a generalized likelihood ratio test of the hypothesis $H_0 : C\beta = 0$ vs. $H_a : C\beta \neq 0$, we maximize the likelihood under the constrained null hypothesis as in the appendix. This leads to constrained MLE's

$$\begin{aligned} \tilde{\beta} &= \mathbf{\Psi}\hat{\beta}, \\ \tilde{\sigma}^2 &= (y - X\tilde{\beta})'(y - X\tilde{\beta})/n, \end{aligned} \tag{2.6}$$

where

$$\mathbf{\Psi} = I_{q+1} - (X'X)^{-1}C'[C(X'X)^{-1}C']^{-1}C. \tag{2.7}$$

Then the likelihood ratio statistic is given by

$$-2\log\lambda_M = n\log\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right). \tag{2.8}$$

This has an asymptotic χ^2_r distribution, where r is the rank of C, and is asymptotically equivalent to the usual $t-$ or F-tests associated with statistical parametric maps. For example, with a model with β_0 representing an intercept, β_1 representing a linear drift over time, and β_2 representing an effect of a stimulus. Then to test whether the coefficient for the reference function or stimulus is 0, set $C = (0, 0, 1)$, so that the hypothesis is $H_0 : \beta_2 = 0$. The LR test has an asymptotic χ^2_1 distribution and is asymptotically equivalent to the usual t tests for activation given by

$$t_2 = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)}.$$

We use the χ^2 representation for ease of comparability with the complex activation model.

2.2 Complex Activation

Alternatively, we can represent the observed data at time point t as a 2×1 vector instead of as a complex number

$$\begin{pmatrix} y_{Rt} \\ y_{It} \end{pmatrix} = \begin{pmatrix} x_t'\beta\cos\theta \\ x_t'\beta\sin\theta \end{pmatrix} + \begin{pmatrix} \eta_{Rt} \\ \eta_{It} \end{pmatrix}, \quad t = 1, \ldots, n \ .$$

This model can also be written as

$$\begin{matrix} y \\ 2n \times 1 \end{matrix} = \begin{matrix} \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix} \\ 2n \times 2(q+1) \end{matrix} \begin{matrix} \begin{pmatrix} \beta\cos\theta \\ \beta\sin\theta \end{pmatrix} \\ 2(q+1) \times 1 \end{matrix} + \begin{matrix} \eta \\ 2n \times 1 \end{matrix} \tag{2.9}$$

where it is specified that the observed vector of data $y = (y_R', y_I')'$ is the vector of observed real values stacked on the vector of observed complex values and the vector of errors $\eta = (\eta_{Rt}', \eta_{It}')' \sim \mathcal{N}(0, \mathbf{\Sigma} \otimes \mathbf{\Phi})$ is similarly defined. Here we assume that $\mathbf{\Sigma} = \sigma^2 I_2$ and $\mathbf{\Phi} = I_n$.

Due to the multiparameter baseline and signal model in Equation 2.9, this is a generalization of the simple linear regression model by Nan and Nowak (1999) where there is only a mean and signal reference function. Previous simple linear regression models by Scharf and Friedlander (1994) and Lai and Glover (1997) did not accurately model the phase, while we correctly account for it through a nonlinear multiple regression model. Our generalization allows for more elaborate hypothesis testing frameworks, such as deconvolution and comparisons between task conditions.

As with the magnitude model, we can obtain unrestricted maximum likelihood estimates of the parameters as derived in the appendix to be

$$\begin{aligned} \hat{\theta} &= \frac{1}{2}\tan^{-1}\left[\frac{2\hat{\beta}_R'(X'X)\hat{\beta}_I}{\hat{\beta}_R'(X'X)\hat{\beta}_R - \hat{\beta}_I'(X'X)\hat{\beta}_I}\right] \\ \hat{\beta} &= \hat{\beta}_R\cos\hat{\theta} + \hat{\beta}_I\sin\hat{\theta}, \\ \hat{\sigma}^2 &= \frac{1}{2n}\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \hat{\beta}\cos\hat{\theta} \\ \hat{\beta}\sin\hat{\theta} \end{pmatrix}\right]'\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \hat{\beta}\cos\hat{\theta} \\ \hat{\beta}\sin\hat{\theta} \end{pmatrix}\right], \end{aligned} \tag{2.10}$$

where

$$\begin{aligned} \hat{\beta}_R &= (X'X)^{-1}X'y_R, \\ \hat{\beta}_I &= (X'X)^{-1}X'y_I. \end{aligned}$$

Note that the estimate of the regression coefficients is a weighted average of estimates from the real and imaginary parts.

The maximum likelihood estimates under the constrained null hypothesis $H_0 : C\beta = 0$ are derived in the appendix and given by

$$\begin{aligned}
\tilde{\theta} &= \frac{1}{2}\tan^{-1}\left[\frac{2\hat{\beta}_R'\mathbf{\Psi}(X'X)\hat{\beta}_I}{\hat{\beta}_R'\mathbf{\Psi}(X'X)\hat{\beta}_R - \hat{\beta}_I'\mathbf{\Psi}(X'X)\hat{\beta}_I}\right] \\
\tilde{\beta} &= \mathbf{\Psi}[\hat{\beta}_R\cos\tilde{\theta} + \hat{\beta}_I\sin\tilde{\theta}], \\
\tilde{\sigma}^2 &= \frac{1}{2n}\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \tilde{\beta}\cos\tilde{\theta} \\ \tilde{\beta}\sin\tilde{\theta} \end{pmatrix}\right]'\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \tilde{\beta}\cos\tilde{\theta} \\ \tilde{\beta}\sin\tilde{\theta} \end{pmatrix}\right] \quad , (2.11)
\end{aligned}$$

where $\mathbf{\Psi}$ is as defined in Equation 2.7 for the magnitude model.

This formulation of the model requires us to correctly deal with the phase angle. An alternative formulation is to let $\alpha_1 = \cos\theta$ and $\alpha_2 = \sin\theta$. Then the model is

$$y = \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix} \quad \begin{pmatrix} \alpha_1\beta \\ \alpha_2\beta \end{pmatrix} \quad + \quad \eta, \qquad \alpha_1^2 + \alpha_2^2 = 1 \ . \tag{2.12}$$

With the model formulation in Equation 2.12 we can identify it as a reduced rank regression model [9] with a sum of squares equal to 1 constraint on the α coefficients. In the same way as before, the parameters can be estimated under the unconstrained model as derived in the appendix to be

$$\begin{aligned}
\hat{\alpha}_1 &= \hat{\beta}'(X'X)\hat{\beta}_R/[(\hat{\beta}'(X'X)\hat{\beta}_R)^2 + (\hat{\beta}'(X'X)\hat{\beta}_I)^2]^{1/2} \\
\hat{\alpha}_2 &= \hat{\beta}'(X'X)\hat{\beta}_I/[(\hat{\beta}'(X'X)\hat{\beta}_R)^2 + (\hat{\beta}'(X'X)\hat{\beta}_I)^2]^{1/2} \\
\hat{\beta} &= \hat{\alpha}_1\hat{\beta}_R + \hat{\alpha}_2\hat{\beta}_I, \\
\hat{\sigma}^2 &= \frac{1}{2n}\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \hat{\alpha}_1\hat{\beta} \\ \hat{\alpha}_2\hat{\beta} \end{pmatrix}\right]'\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \hat{\alpha}_1\hat{\beta} \\ \hat{\alpha}_2\hat{\beta} \end{pmatrix}\right].
\end{aligned} \tag{2.13}$$

Again note that the estimate of the regression coefficients is a weighted average of estimates from the real and imaginary parts.

Similarly, the maximum likelihood estimates under the constrained null hypothesis H_0 :

$C\beta = 0$ are derived in the appendix and given by

$$\begin{aligned}
\tilde{\alpha}_1 &= \tilde{\beta}'(X'X)\hat{\beta}_R/[(\tilde{\beta}'(X'X)\hat{\beta}_R)^2 + (\tilde{\beta}'(X'X)\hat{\beta}_I)^2]^{1/2} \\
\tilde{\alpha}_2 &= \tilde{\beta}'(X'X)\hat{\beta}_I/[(\tilde{\beta}'(X'X)\hat{\beta}_R)^2 + (\tilde{\beta}'(X'X)\hat{\beta}_I)^2]^{1/2} \\
\tilde{\beta} &= \mathbf{\Psi}(\tilde{\alpha}_1\hat{\beta}_R + \tilde{\alpha}_2\hat{\beta}_I), \\
\tilde{\sigma}^2 &= \frac{1}{2n}\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \tilde{\alpha}_1\tilde{\beta} \\ \tilde{\alpha}_2\tilde{\beta} \end{pmatrix}\right]'\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \tilde{\alpha}_1\tilde{\beta} \\ \tilde{\alpha}_2\tilde{\beta} \end{pmatrix}\right]
\end{aligned} \tag{2.14}$$

In computing maximum likelihood estimates, an iterative maximization known as the Iterative Conditional Modes (ICM) algorithm [10, 11] is used.

Then for either formulation (Equation 2.9 or 2.12) the generalized likelihood ratio statistic for the complex activation model is

$$-2\log\lambda_C = 2n\log\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right) \ . \tag{2.15}$$

This statistic has an asymptotic χ^2_r distribution similar to the magnitude model statistic in Equation 2.8.

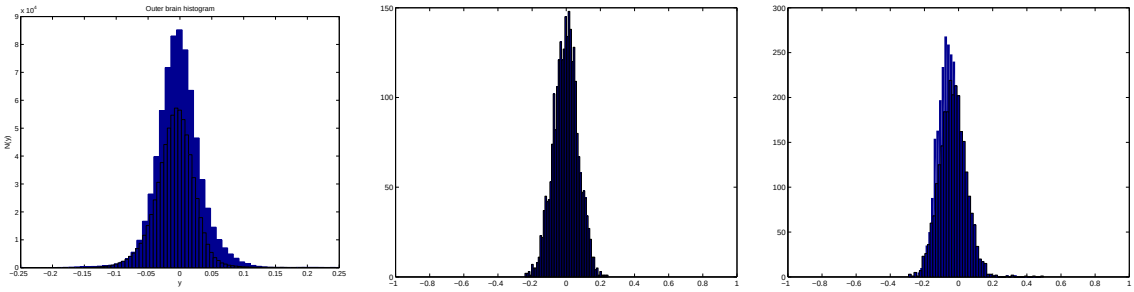
3 Application to fMRI dataset

A bilateral finger tapping experiment was performed in a block design with 16s off followed by eight epochs of 16s on and 16s off. Scanning was performed using a 1.5T GE Signa in which 5 axial slices of size 96×96 were acquired. In image reconstruction, the acquired data was zero filled to 128×128. After reconstruction, each voxel has dimensions in mm of $1.5625 \times 1.5625 \times 5$, with TE= 47ms. Observations were taken every TR= 1000ms so that there are 272 in each voxel. Data from a single axial slice through the motor cortex was selected for analysis. Pre-processing using an ideal filter was performed to remove respiration and low frequency physiological noise in addition to the removal of the first three points to omit machine warm-up effects.

First we checked the validity of the complex model assumptions by examining observations outside the brain where there is no task-related activation. Proper modeling of the noise is essential prior to modeling the signal. The phase angle was plotted against time to

investigate stability of the phase over time; this was relatively constant over time, confirming the observation of Nan and Nowak (1999), and is omitted for brevity. Next histograms of the real (solid) and imaginary (striped) components are constructed separately and superimposed on one another in Figure 2(a). These appear to both be approximately normally distributed with similar variances. Figure 2(b) contains histograms of the correlations between the real and imaginary components, again computed from the time series outside the brain. These are distributed closely around 0, indicating that the assumption of independence of the real and imaginary components is reasonable. Finally, an autoregressive order 1 (AR(1)) model was fit to the outside the brain time series and the correlation parameter ρ for the complex (solid) and magnitude (striped) was estimated from the data and histograms superimposed on one another. These are presented in Figure 2(c) where a 5% Bonferroni adjusted threshold is $\rho_{DW} = .26$ for the magnitude model and $\rho_{DW} = .19$ for the complex model. These thresholds were determined via Monte Carlo simulation by generating data sets of the same length and model. Most of the correlations are between -0.2 and 0.2, indicating little temporal autocorrelation of the data without signal.

Figure 2: No signal assessment of complex model.



(a) Histogram of real and imaginary data

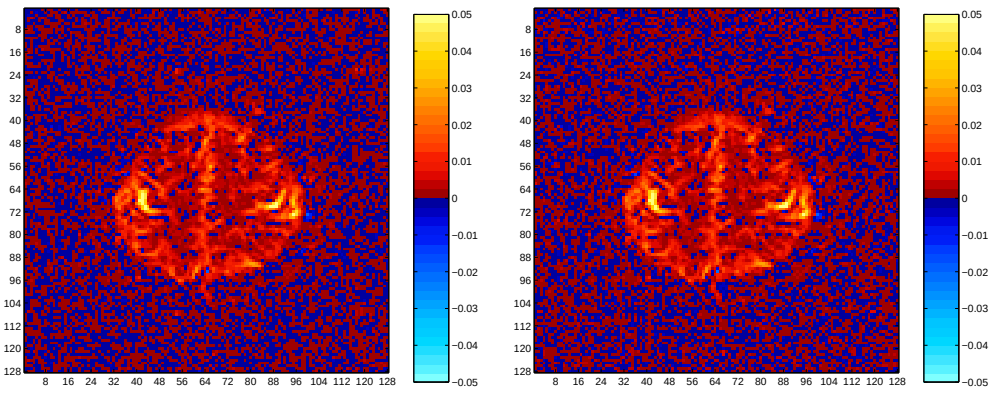
(b) Histogram of Correlations between real and imaginary measurements

(c) Histogram of magnitude and complex autocorrelation estimates

After verifying the model assumptions, we next model the signal and compare the results of fitting the complex and magnitude models. The linear magnitude and nonlinear complex multiple regression models were fit to the data with an intercept, a zero mean time trend, and a ± 1 square wave reference function. Parameter estimates of the task-related activation

β_2 are given in Figure 3 for (a) the magnitude model and (b) the complex model. These coefficient estimates are very similar between the two models.

Figure 3: Estimates of the reference function coefficients, β_2's.



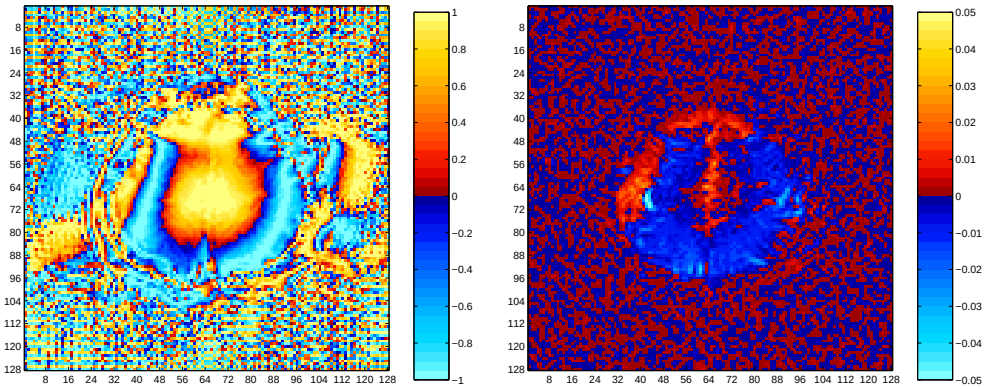
(a) Magnitude model

(b) Complex model

As previously noted, the estimated β_2 coefficients for the complex model in Figure 3(b) under the alternative hypothesis are weighted averages between the estimated value from the real and imaginary parts. This weighting is displayed in Figure 4 where the α_1 weights are in Figure 4(a), the estimated coefficient values from the real part β_{R2} are in Figure 4(b), the α_2 weights are in Figure 4(c), and the estimated coefficient values from the real part β_{I2} are in Figure 4(d).

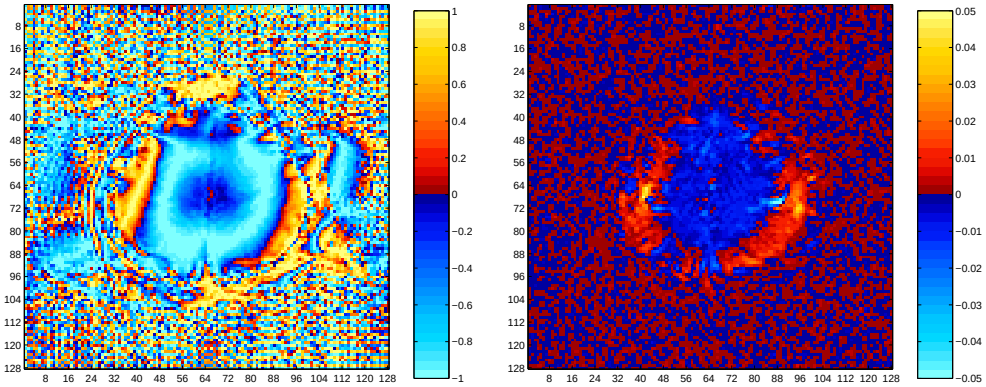
After fitting the model, residuals were used to re-evaluate the assumptions for the complex model. Histograms of the real (solid) and imaginary (striped) components are constructed from the residual time courses for all voxels separately and superimposed on one another in Figure 5(a). These appear to both be approximately normally distributed with similar variances. Figure 5(b) contains histograms of the correlations between the real and imaginary components, computed from the residual time courses for all voxels. These are distributed closely around 0, indicating that the assumption of independence of the real and imaginary components is reasonable. An autoregressive order 1 (AR(1)) model was fit to the residual time series in every voxel and the correlation parameter ρ for the complex (solid) and magnitude (striped) was estimated from the data and histograms superimposed on one another. These are presented in Figure 5(c) where again 5% Bonferroni adjusted

Figure 4: Complex real and imaginary estimated β_2's, and weights, α's.



(a) real part weights, α_1's

(b) real part estimates, β_{R2}'s



(c) imaginary part weights, α_2's

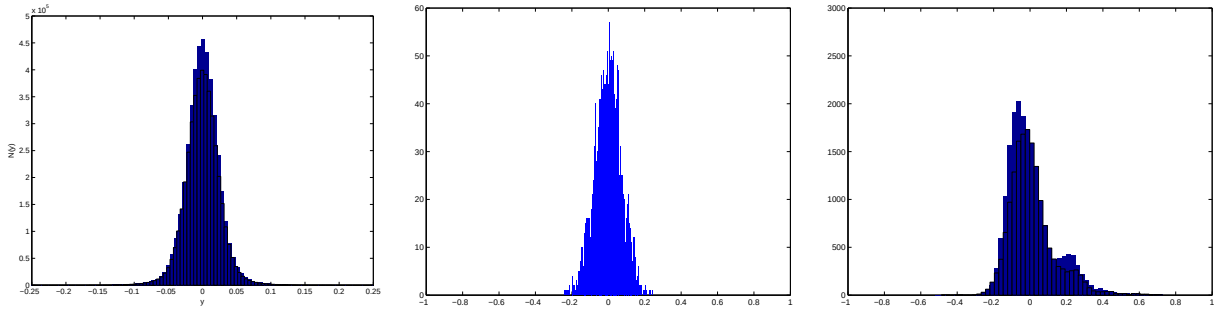
(d) imaginary part estimates, β_{I2}'s

magnitude and complex thresholds can be applied as previously described. The temporal autocorrelation present in the residuals is similar for the magnitude and complex models. Most of the correlations are between $-.20$ and 0.20, indicating that most of the voxels have little temporal autocorrelation.

Next we looked for significant task-related activation using a 5% false discovery rate threshold. This was done by applying the Benjamini-Hochberg procedure [12, 13, 14] to the voxel p-values obtained from the χ^2_1 approximation from the likelihood ratio statistic. Images of significant activation are given in Figure 6 for the magnitude and complex models.

While the activation images are similar due to the large SNR, note that the complex model appears to have sharper or more well-defined activation regions which align better

Figure 5: Residual histogram assessment of complex model.



(a) Real and imaginary data

(b) Real and imaginary correlations

(c) Magnitude and complex autocorrelations

with the gray matter at which the activation is to be.

4 fMRI Simulation

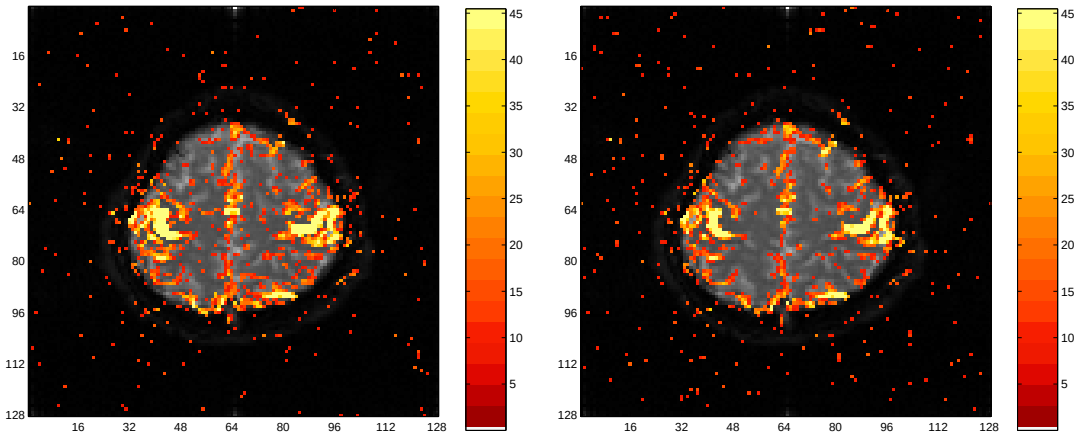
Data is generated to simulate the same bilateral finger tapping fMRI block design experiment with $n = 269$ points where the true activation structure is known so that the two activation methods can be evaluated. A 128×128 slice is selected for analysis within which four 7×7 ROI's as lightened in Figure 7 are designated to have activation.

For this slice, simulated fMRI data is constructed according to a multiple regression model which consists of an intercept, a time trend for all voxels but also a reference function x_{2t} for voxels in each ROI which is related to a block experimental design. This model dictates that for voxel i at time t,

$$y_{it} = [(\beta_0 + \beta_1 t + \beta_2 x_{2t})\alpha_{1i} + \eta_{Rit}] + i[(\beta_0 + \beta_1 t + \beta_2 x_{2t})\alpha_{2i} + \eta_{Iit}],$$

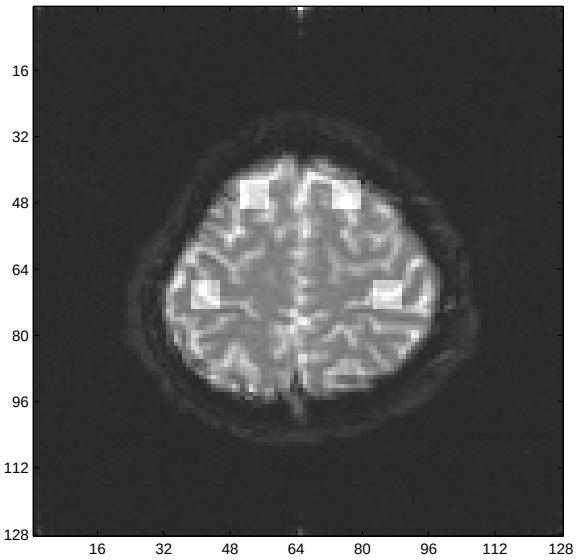
where η_{Rit}, η_{Iit} are i.i.d. N(0,σ^2). In this simulation study, α_1 and α_2 are voxel dependent and taken from the estimated values for the example dataset shown in Figure 4 (a) and (c), while $\beta_1 = 0.00001$ and $\sigma = 0.04909$ are assumed constant across voxels with values taken from a "highly active" voxel in the activation region of the sample dataset. The coefficient for the reference function β_2 is zero outside the ROI. Inside each ROI β_2 has constant value determined by an effect to noise ratio (ENR=β_2/σ) of $1, 0.5, 0.25, 0.125$, going from left to

Figure 6: Activation images using the LR test, thresholded at a 5% false discovery rate.



(a) Magnitude model (b) Complex model

Figure 7: Anatomical with ROI's.

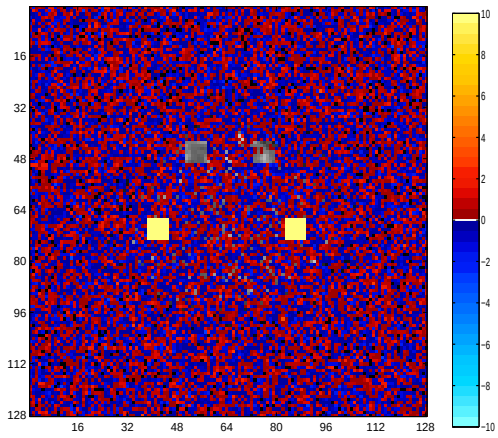


right and top to bottom. To investigate the effect of the signal to noise ratio (SNR) typically defined to be the mean divided by the standard deviation of a voxel time course. Note that the magnitude of β_0 observed in the real dataset is generally much larger than β_1 or β_2, indicating that it is the dominant feature in the SNR in addition to being the time course

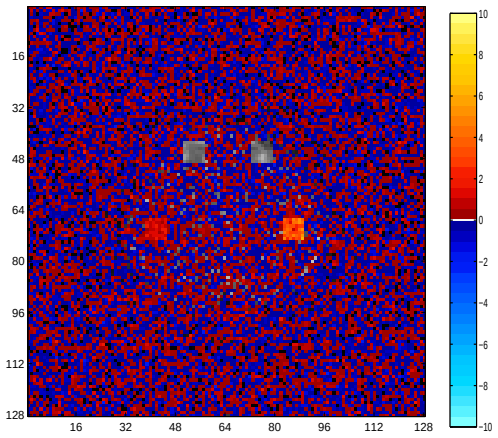
mean. Therefore since the variance is held fixed, we parameterize the SNR by varying β_0 so that the ratio SNR$= \beta_0/\sigma$ takes on values between 1 and 30, where 30 is approximately the value of SNR found in "highly active" voxels for the example dataset, and other values represent decreasing SNR.

In each voxel for a given model and SNR, 1000 simulated images were generated and thresholded using an unadjusted threshold with a 5% type I or per comparison error (PCE) rate, the Benjamini-Hochberg procedure with a 5% false discovery rate (FDR), and the Bonferroni procedure with a 5% familywise error (FWE) rate. For each thresholding method, the power, or relative frequency over the 1000 simulated images with which each voxel was detected as active, was recorded. Absolute differences in power between the complex and magnitude models were calculated for each voxel, mapped to a color scale, and shown in Figures 8 through 10 for the three thresholding procedures. Voxels with zero difference in power were assigned the voxel anatomical grey scale.

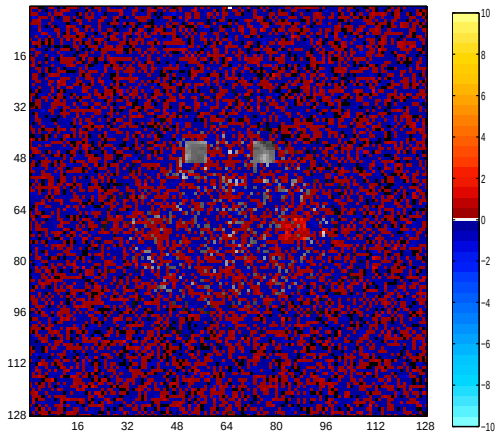
Figure 8: Differences in power between the models varying SNR, 5% PCE threshold.



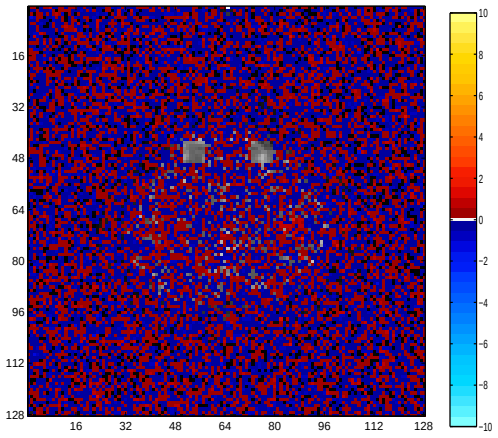
(a) SNR = 1



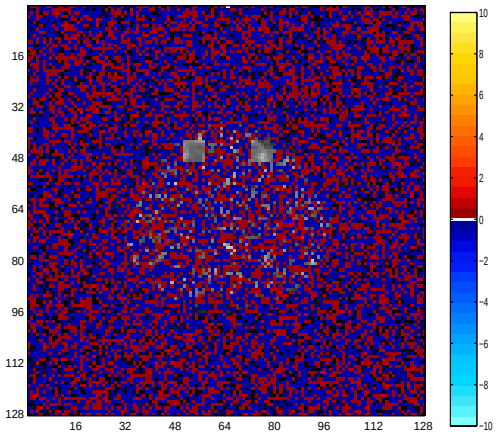
(b) SNR = 2.5



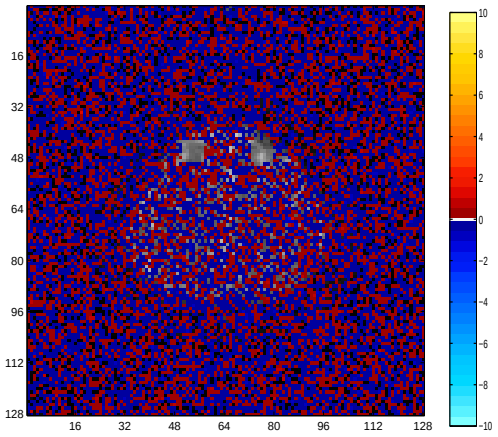
(c) SNR = 5



(d) SNR = 7.5

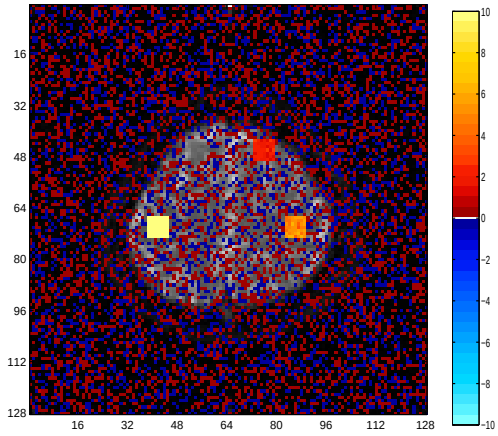


(e) SNR = 10

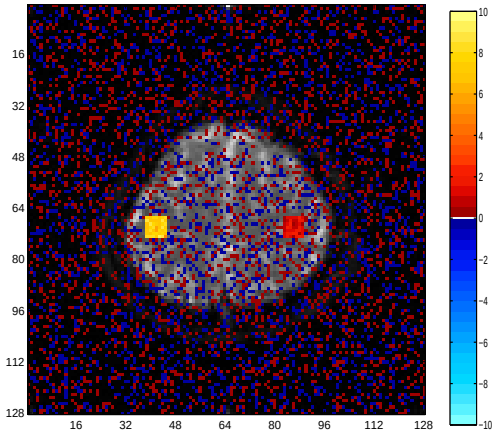


(f) SNR = 30

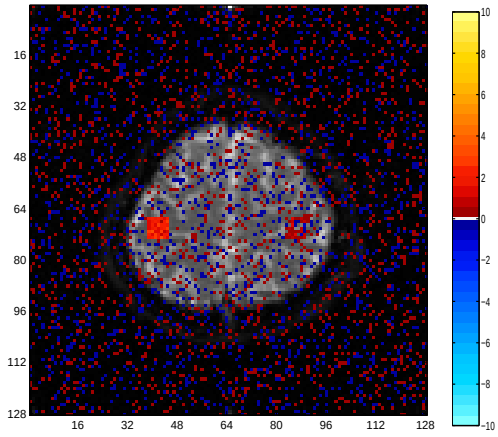
Figure 9: Differences in power between the models varying SNR, 5% FDR threshold.



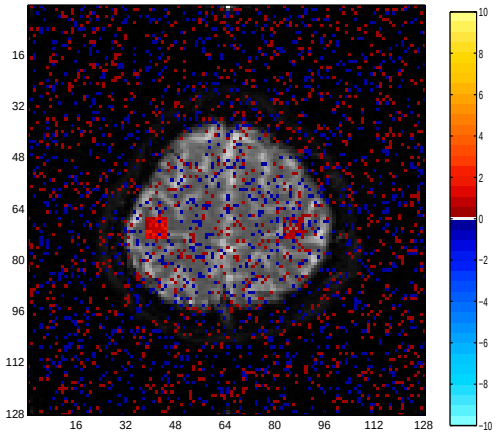
(a) SNR = 1



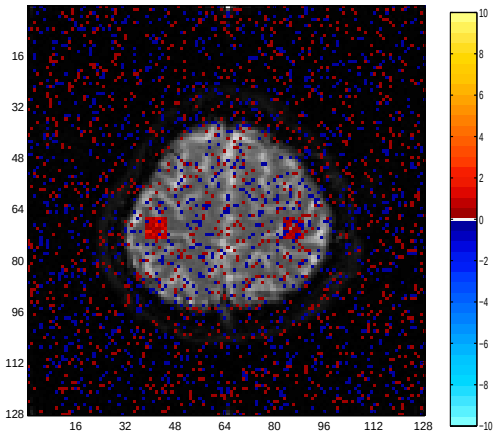
(b) SNR = 2.5



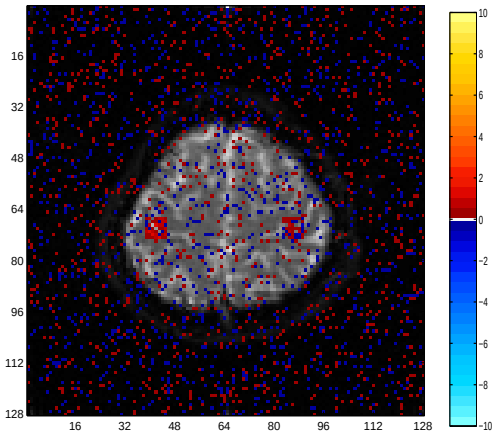
(c) SNR = 5



(d) SNR = 7.5

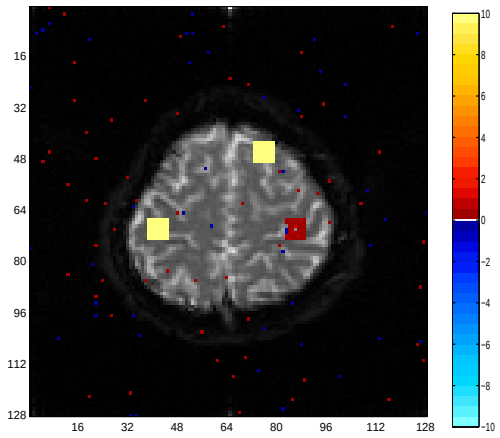


(e) SNR = 10

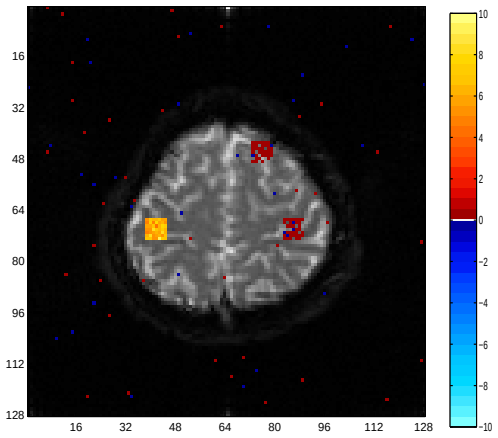


(f) SNR = 30

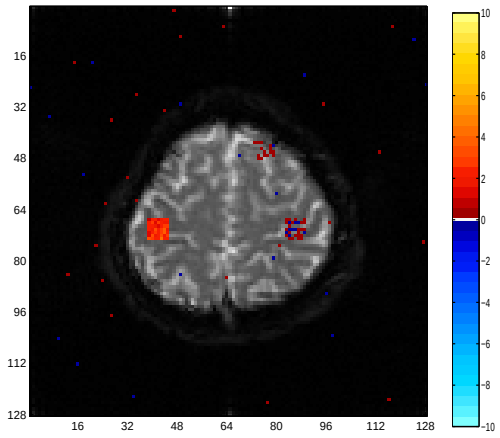
Figure 10: Differences in power between the models varying SNR, 5% FWE threshold.



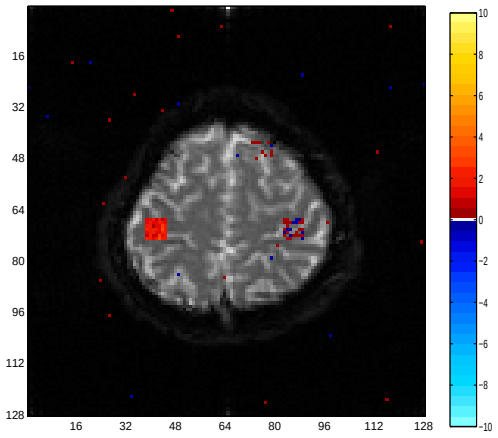
(a) SNR = 1



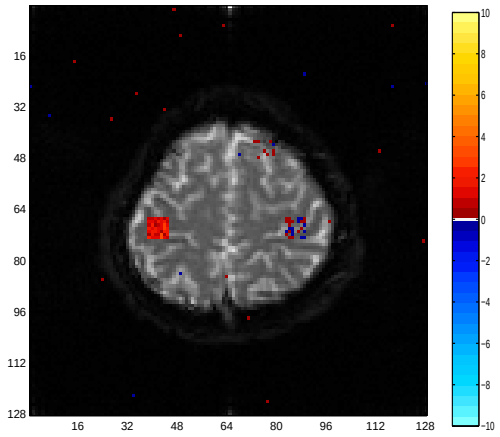
(b) SNR = 2.5



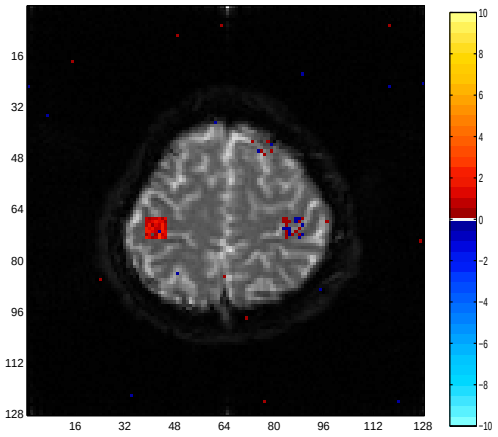
(c) SNR = 5



(d) SNR = 7.5



(e) SNR = 10

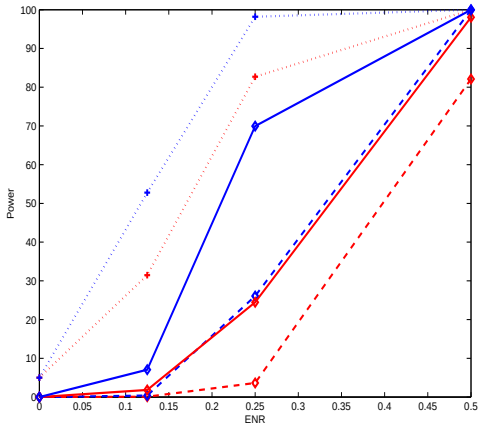


(f) SNR = 30

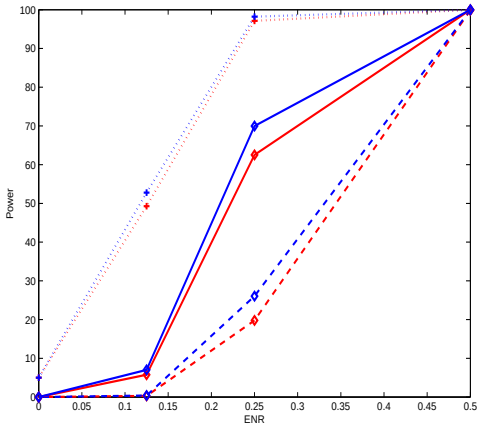
Note that there are little differences between the complex and magnitude models for the ENR= 0.5 to 1 range; however, this is because the power is approximately 1. For less strong effects, the differences are sensitive to the SNR and the complex model is generally useful for low SNR.

To further illustrate the power improvement of the complex model over the magnitude model for low SNR's, we plotted the power curves as a function of the effect to noise ratios. This was done for the three thresholding procedures (5% PCE, 5% FDR, 5% FWE) and the complex (blue) and magnitude (red) models. These power curves are given in Figure 11 where for all ENR's, the curves are from top to bottom for the 5% PCE (dotted), 5% FDR (solid), and 5% FWE (dashed) thresholds. These power curves illustrate similar results as before, that the complex model power curve is higher than the magnitude model power curve for low SNR, but the lines are quite close for higher SNR.

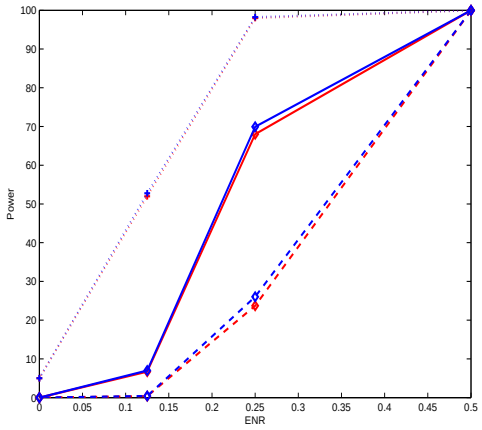
Figure 11: Power versus ENR for complex (blue) and magnitude (red) models.



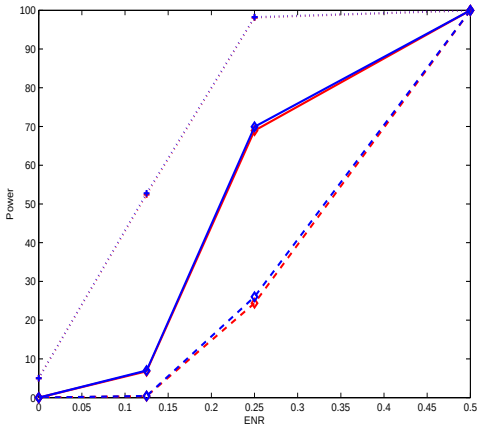
(a) SNR = 1



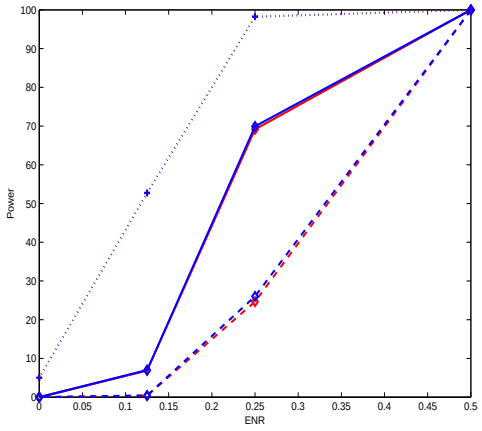
(b) SNR = 2.5



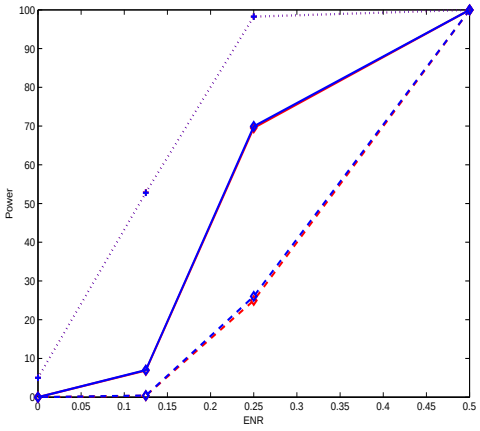
(c) SNR = 5



(d) SNR = 7.5

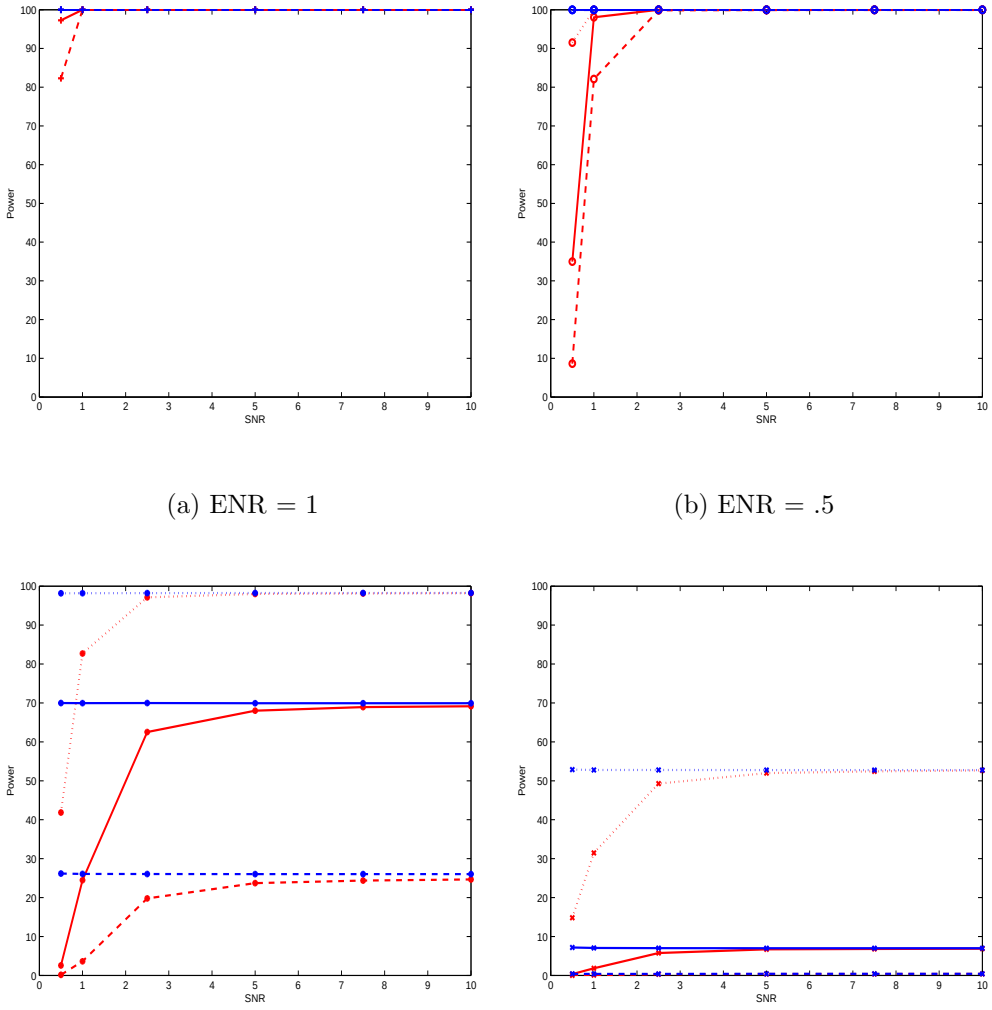


(e) SNR = 10



(f) SNR = 30

Figure 12: ENR power versus SNR for complex (blue) and magnitude (red) models.



(a) ENR = 1

(b) ENR = .5

(c) ENR = .25

(d) ENR = .125

To reiterate the advantage of the complex model over the magnitude model for low SNR's, in Figure 12 we plotted power versus SNR (.5, 1, 2.5, 5, 7.5, 10) for the four ENR's. The dotted curves represent the 5% PCE threshold, the solid curves the 5% FDR threshold, and the dashed curves the 5% FWE threshold. For example in Figure 12(c) the solid blue line with an asterisk is the complex model FDR power curve for an ENR of .25 while the solid red line with an asterisk is the corresponding magnitude model FDR power curve for an ENR of .25. It is evident for a given ENR that the complex model power curve is constant irrespective of SNR while the magnitude model power curve falls rapidly as the SNR decreases.

5 Conclusions

A complex data fMRI activation model was presented as an alternative to the typical magnitude data model. Activation statistics were derived from generalized likelihood ratio tests for both models. Activation from both models were presented for real fMRI data, then simulations were performed to compare the power to detect activation regions between the two models for several signal to noise ratios with varying effects.

It was found that for large signal to noise ratios, both models were comparable. However, for smaller signal to noise ratios, the complex activation model demonstrated superior power of detection over the magnitude activation model. This strongly indicates that modeling the complex data may become more useful as voxel sizes get smaller, since this decreases the SNR.

A Magnitude and Phase Distributions

The distribution of the magnitude and phase can be derived as follows. Let $y_R = \rho\cos\theta + \eta_R$ and $y_I = \rho\cos\theta + \eta_I$ where η_R and η_I are normally distributed with mean zero and variance σ^2. Then, make a change of variable from (y_R, y_I) to polar coordinates $r^2 = y_R^2 + y_I^2$ and $\phi = \tan^{-1}(y_I/y_R)$ or $y_R = r\cos\phi$ and $y_I = r\sin\phi$. The Jacobian of this transformation is $J(y_R, y_I \rightarrow r, \phi) = r$. The joint distribution of r and ϕ using trigonometric identities becomes

$$p(r,\phi|\rho,\theta,\sigma^2) \;=\; \frac{r}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}\left[r^2+\rho^2-2\rho r\cos(\phi-\theta)\right]} \;.$$

A.1 Magnitude Distribution

The marginal distribution of the magnitude r is found by integrating out the phase ϕ

$$p(r|\rho,\theta,\sigma^2) \;=\; \frac{r}{\sigma^2} e^{-\frac{1}{2\sigma^2}\left[r^2+\rho^2\right]} \int_{\phi=-\pi}^{\pi} \frac{1}{2\pi} e^{\frac{1}{\sigma^2}\rho r\cos(\phi-\theta)} d\phi$$

where the integral factor often denoted $I_o(r\rho/\sigma^2)$ is the zeroth order Bessel function of the first kind. The normal limiting distribution for large SNR or $\rho \rightarrow \infty$, is found by using the asymptotic form $I_o(r\rho/\sigma^2) \approx \exp(r\rho/\sigma^2)/\sqrt{(2\pi r\rho/\sigma^2)}$ of the Bessel function [15]. Additionally, in this limit, it is assumed that the exponential form of the normal distribution drops off more rapidly compared to the variation in the ratio $\sqrt{r/\rho}$ left as a factor. The distribution of the magnitude becomes the normal distribution with mean ρ and variance σ^2.

The Rayleigh limiting distribution for zero SNR or $\rho = 0$, is found by noting that $I_o(0) = 1$. The distribution of the magnitude becomes

$$p(r|\rho,\sigma^2) \;=\; \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \;.$$

A.2 Phase Distribution

The marginal distribution of the phase ϕ is found by integrating out the magnitude r

$$p(\phi|\rho,\theta,\sigma^2) \;=\; \frac{e^{-\frac{\rho^2}{2\sigma^2}}}{2\pi}\left[1 + \frac{\rho}{\sigma}\sqrt{2\pi}\cos(\phi-\theta) e^{\frac{\rho^2\cos^2(\phi-\theta)}{2\sigma^2}} \int_{-\infty}^{\frac{\rho\cos(\phi-\theta)}{\sigma}} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \;.\right]$$

The normal limiting distribution for large SNR or $\rho \to \infty$, is found by multiplying through, noting that the first term is approximately zero, that the difference between ϕ and θ is small so that the cosine of their difference is approximately one, and the sine of their difference is approximately their difference. The distribution of the phase becomes the normal distribution with mean θ and variance $(\sigma/\rho)^2$.

The uniform limiting distribution for zero SNR or $\rho = 0$, is found by noting that the integral factor goes to unity. The distribution of the phase becomes

$$p(\phi) \;=\; \frac{1}{2\pi}\,.$$

The distribution of the magnitude and phase for intermediate values of SNR can be found by numerical integration or Monte Carlo simulation. The complex model presented in this paper does not make these large SNR approximations.

B Generalized Likelihood Ratio Tests

In applications using multiple regression including fMRI, we often wish to test linear contrast hypothesis (for each voxel) such as

$$\begin{array}{rcccrcl} H_0: & C\beta & = & \gamma & vs \quad H_1: & C\beta & \neq & \gamma \\ & \sigma^2 & > & 0 & & \sigma^2 & > & 0\,, \end{array}$$

where C is an $r \times (q+1)$ matrix of full row rank and γ is an $r \times 1$ vector.

Magnitude Model

The likelihood ratio statistic is computed by maximizing the likelihood $p(y|\beta, \sigma^2, X)$ with respect to β and σ^2 under the null and alternative hypotheses. Denote the maximized values under the null hypothesis by $(\tilde{\beta}, \tilde{\sigma}^2)$ and those under the alternative hypothesis as $(\hat{\beta}, \hat{\sigma}^2)$. These maximized values are then substituted into the likelihoods and the ratio taken.

With the aforementioned distributional specifications, the likelihood of the model is

$$p(y|\beta, \sigma^2, X) = (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}} e^{-\frac{h}{2\sigma^2}}, \tag{B.1}$$

where

$$\begin{aligned} h &= (y - X\beta)'(y - X\beta) \\ &= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta \ . \end{aligned}$$

Unrestricted MLE's

Maximizing this likelihood with respect to the parameters is the same as maximizing the logarithm of the likelihood with respect to the parameters. In the case of β it is the same as minimizing the h term in the exponent. These derivatives are

$$\begin{aligned} \left.\frac{\partial}{\partial\beta}h\right|_{\beta=\hat{\beta},\sigma^2=\hat{\sigma}^2} &= 2(X'X)\hat{\beta} - 2X'y \\ \left.\frac{\partial}{\partial\sigma^2}\log[p(y|X,\beta,\theta,\sigma^2)]\right|_{\beta=\hat{\beta},\sigma^2=\hat{\sigma}^2} &= -\frac{n}{2}\frac{1}{\hat{\sigma}^2} + \frac{\hat{h}}{2}\frac{1}{(\hat{\sigma}^2)^2} \ . \end{aligned}$$

where $\hat{h}$ is h with MLE's substituted in. By setting these derivatives equal to zero and solving, we get the MLE's under the unrestricted model given in Equation 2.5.

Restricted MLE's

Maximizing this likelihood with respect to the parameters is the same as maximizing the logarithm of the likelihood with respect to the parameters. In the case of β it is the same as minimizing the h term in the exponent with the restriction in the form of a Lagrange multiplier as

$$\begin{aligned} h &= (y - X\beta)'(y - X\beta) - 2\Psi'(C\beta - \gamma) \\ &= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta - 2\Psi'C\beta + 2\Psi'\gamma \ . \end{aligned}$$

These derivatives are

$$\begin{aligned} \left.\frac{\partial}{\partial\beta}h\right|_{\beta=\tilde{\beta},\psi=\tilde{\psi},\sigma^2=\tilde{\sigma}^2} &= 2(X'X)\tilde{\beta} - 2X'y \\ \left.\frac{\partial}{\partial\psi}h\right|_{\beta=\tilde{\beta},\psi=\tilde{\psi},\sigma^2=\tilde{\sigma}^2} &= 2(C\tilde{\beta} - \gamma) \\ \left.\frac{\partial}{\partial\sigma^2}\log[p(y|X,\beta,\theta,\sigma^2)]\right|_{\beta=\tilde{\beta},\psi=\tilde{\psi},\sigma^2=\tilde{\sigma}^2} &= -\frac{n}{2}\frac{1}{\tilde{\sigma}^2} + \frac{\tilde{h}}{2}\frac{1}{(\tilde{\sigma}^2)^2} \end{aligned}$$

where $\tilde{h}$ is h with MLE's substituted in. By setting these derivatives equal to zero and solving, we get the MLE's under the restricted model given in Equation 2.6. Note that $\hat{\sigma}^2 = \hat{h}/n$ and $\tilde{\sigma}^2 = \tilde{h}/n$. Then the generalized likelihood ratio is

$$\lambda = \frac{p(y|\tilde{\beta}, \tilde{\sigma}^2, X)}{p(y|\hat{\beta}, \hat{\sigma}^2, X)} = \frac{(\tilde{\sigma}^2)^{-n/2} e^{-\tilde{h}n/(2\tilde{h})}}{(\hat{\sigma}^2)^{-n/2} e^{-\hat{h}n/(2\hat{h})}}, \tag{B.2}$$

and Equation 2.8 for the GLRT follows.

B.1 Complex Model with θ

Just as in the magnitude model, the likelihood ratio statistic is computed by maximizing the likelihood $p(y|\beta, \sigma^2, X)$ with respect to β and σ^2 under the null and alternative hypotheses. Denote the maximized values under the null hypothesis by $(\tilde{\beta}, \tilde{\sigma}^2)$ and those under the alternative hypothesis as $(\hat{\beta}, \hat{\sigma}^2)$. These maximized values are then substituted into the likelihoods and the ratio taken. With the aforementioned distributional specifications, the likelihood of the model is

$$p(y|X, \beta, \theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{2n}{2}} e^{-\frac{h}{2\sigma^2}} \tag{B.3}$$

where

$$\begin{aligned} h &= \frac{1}{2n}\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \beta\cos\theta \\ \beta\sin\theta \end{pmatrix}\right]'\left[y - \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix}\begin{pmatrix} \beta\cos\theta \\ \beta\sin\theta \end{pmatrix}\right] \\ &= \beta'(X'X)\beta - 2\beta'(X'X)[\hat{\beta}_R\cos\theta + \hat{\beta}_I\sin\theta] + \hat{\beta}_R'(X'X)\hat{\beta}_R + \hat{\beta}_I'(X'X)\hat{\beta}_I \\ &\quad + y_R'[I_n - X(X'X)^{-1}X']y_R + y_I'[I_n - X(X'X)^{-1}X']y_I \end{aligned}$$

Unrestricted MLE's

Maximizing this likelihood with respect to the parameters is the same as maximizing the logarithm of the likelihood with respect to the parameters. In the case of β and θ it is the

same as minimizing the h term in the exponent.

$$\begin{aligned}
\frac{\partial}{\partial\beta}h\bigg|_{\beta=\hat{\beta},\theta=\hat{\theta},\sigma^2=\hat{\sigma}^2} &= 2(X'X)\hat{\beta} - 2(X'X)\left[\hat{\beta}_R\cos\hat{\theta} + \hat{\beta}_I\sin\hat{\theta}\right] \\
\frac{\partial}{\partial\theta}h\bigg|_{\beta=\hat{\beta},\theta=\hat{\theta},\sigma^2=\hat{\sigma}^2} &= -2\hat{\beta}'(X'X)\left[(-\sin\hat{\theta})\hat{\beta}_R + (\cos\hat{\theta})\hat{\beta}_I\right] \\
\frac{\partial}{\partial\sigma^2}\log[p(y|X,\beta,\theta,\sigma^2)]\bigg|_{\beta=\hat{\beta},\theta=\hat{\theta},\sigma^2=\hat{\sigma}^2} &= -\frac{2n}{2}\frac{1}{\hat{\sigma}^2} + \frac{\hat{h}}{2}\frac{1}{(\hat{\sigma}^2)^2}
\end{aligned}$$

where $\hat{h}$ is h with MLE's substituted in. By setting these derivatives equal to zero and solving, we get the MLE's under the unrestricted model given in Equation 2.10.

Restricted MLE's

Maximizing this likelihood with respect to the parameters is the same as maximizing the logarithm of the likelihood with respect to the parameters. In the case of β and θ, it is the same as minimizing the h term in the exponent with the restriction in the form of a Lagrange multiplier as

$$\begin{aligned}
h &= \beta'(X'X)\beta - 2\beta'(X'X)[\hat{\beta}_R\cos\theta + \hat{\beta}_I\sin\theta] + \hat{\beta}_R'(X'X)\hat{\beta}_R + \hat{\beta}_I'(X'X)\hat{\beta}_I \\
&\quad + y_R'[I_n - X(X'X)^{-1}X']y_R + y_I'[I_n - X(X'X)^{-1}X']y_I + 2\psi'(C\beta - \gamma)\ .
\end{aligned}$$

Note that the maximization is performed by Lagrange multipliers and the appropriate term has been added to h

$$\begin{aligned}
\frac{\partial}{\partial\beta}h\bigg|_{\beta=\tilde{\beta},\theta=\tilde{\theta},\psi=\tilde{\psi},\sigma^2=\tilde{\sigma}^2} &= 2(X'X)\tilde{\beta} - 2(X'X)\left[\tilde{\beta}_R\cos\tilde{\theta} + \tilde{\beta}_I\sin\tilde{\theta}\right] \\
&\quad + 2C'\tilde{\psi} \\
\frac{\partial}{\partial\theta}h\bigg|_{\beta=\tilde{\beta},\theta=\tilde{\theta},\psi=\tilde{\psi},\sigma^2=\tilde{\sigma}^2} &= -2\tilde{\beta}'(X'X)\left[(-\sin\tilde{\theta})\tilde{\beta}_R + (\cos\tilde{\theta})\tilde{\beta}_I\right] \\
\frac{\partial}{\partial\psi}h\bigg|_{\beta=\tilde{\beta},\psi=\tilde{\psi},\psi=\tilde{\psi},\sigma^2=\tilde{\sigma}^2} &= 2(C\tilde{\beta} - \gamma) \\
\frac{\partial}{\partial\sigma^2}\log[p(y|X,\beta,\theta,\sigma^2)]\bigg|_{\beta=\tilde{\beta},\theta=\tilde{\theta},\psi=\tilde{\psi},\sigma^2=\tilde{\sigma}^2} &= -\frac{2n}{2}\frac{1}{\tilde{\sigma}^2} + \frac{\tilde{h}}{2}\frac{1}{(\tilde{\sigma}^2)^2}
\end{aligned}$$

where $\tilde{h}$ is h with MLE's substituted in. By setting these derivatives equal to zero and solving, we get the MLE's under the restricted model given in Equation 2.11.

Note that $\hat{\sigma}^2 = \hat{h}/(2n)$ and $\tilde{\sigma}^2 = \tilde{h}/(2n)$. Then the generalized likelihood ratio is

$$\lambda = \frac{p(y|\tilde{\beta}, \tilde{\sigma}^2, \tilde{\theta}, X)}{p(y|\hat{\beta}, \hat{\sigma}^2, \hat{\theta}, X)} = \frac{(\tilde{\sigma}^2)^{-2n/2} e^{-2\tilde{h}n/(2\tilde{h})}}{(\hat{\sigma}^2)^{-2n/2} e^{-2\hat{h}n/(2\hat{h})}} , \tag{B.4}$$

and Equation 2.15 for the GLRT follows.

B.2 Complex Model with α_1 and α_2

Alternatively, the model can be written with $\alpha_1 = \cos\theta$ and $\alpha_2 = \sin\theta$

Unrestricted MLE's

The term in the exponent is

$$\begin{aligned} h &= \beta'(X'X)\beta - 2\beta'(X'X)[\hat{\beta}_R\alpha_1 + \hat{\beta}_I\alpha_2] + \hat{\beta}_R'(X'X)\hat{\beta}_R + \hat{\beta}_I'(X'X)\hat{\beta}_I \\ &\quad + y_R'[I_n - X(X'X)^{-1}X']y_R + y_I'[I_n - X(X'X)^{-1}X']y_I - 2\delta(\alpha_1^2 + \alpha_2^2 - 1) . \end{aligned}$$

Note the Lagrange multiplier constraint that $\alpha_1^2 + \alpha_2^2 = 1$.

Maximizing this likelihood with respect to the parameters is the same as maximizing the logarithm of the likelihood with respect to the parameters. In the case of β, α_1, and α_2 it is the same as minimizing the h term in the exponent.

$$\begin{aligned} \frac{\partial}{\partial\beta} h\Big|_{\beta=\hat{\beta},\alpha_1=\hat{\alpha}_1,\alpha_2=\hat{\alpha}_2,\delta=\hat{\delta},\sigma^2=\hat{\sigma}^2} &= 2(X'X)\hat{\beta} - 2(X'X)\left[\hat{\beta}_R\hat{\alpha}_1 + \hat{\beta}_I\hat{\alpha}_2\right] \\ \frac{\partial}{\partial\alpha_1} h\Big|_{\beta=\hat{\beta},\alpha_1=\hat{\alpha}_1,\alpha_2=\hat{\alpha}_2,\delta=\hat{\delta},\sigma^2=\hat{\sigma}^2} &= -2\hat{\beta}'(X'X)\hat{\beta}_R - 2\hat{\delta}(2\hat{\alpha}_1) \\ \frac{\partial}{\partial\alpha_2} h\Big|_{\beta=\hat{\beta},\alpha_1=\hat{\alpha}_1,\alpha_2=\hat{\alpha}_2,\delta=\hat{\delta},\sigma^2=\hat{\sigma}^2} &= -2\hat{\beta}'(X'X)\hat{\beta}_I - 2\hat{\delta}(2\hat{\alpha}_2) \\ \frac{\partial}{\partial\delta} h\Big|_{\beta=\hat{\beta},\alpha_1=\hat{\alpha}_1,\alpha_2=\hat{\alpha}_2,\delta=\hat{\delta},\sigma^2=\hat{\sigma}^2} &= -2(\hat{\alpha}_1^2 + \hat{\alpha}_2^2 - 1) \\ \frac{\partial}{\partial\sigma^2}\log[p(y|X,\beta,\alpha_1,\alpha_2,\sigma^2)]\Big|_{\beta=\hat{\beta},\alpha_1=\hat{\alpha}_1,\alpha_2=\hat{\alpha}_2,\delta=\hat{\delta},\sigma^2=\hat{\sigma}^2} &= -\frac{2n}{2}\frac{1}{\hat{\sigma}^2} + \frac{\hat{h}}{2}\frac{1}{(\hat{\sigma}^2)^2} \end{aligned}$$

where $\hat{h}$ is h with MLE's substituted in. By setting these derivatives equal to zero and solving, we get the MLE's under the unrestricted model given in Equation 2.13.

Restricted MLE's

The term in the exponent is

$$\begin{aligned} h &= \beta'(X'X)\beta - 2\beta'(X'X)[\hat{\beta}_R\alpha_1 + \hat{\beta}_I\alpha_2] + \hat{\beta}_R'(X'X)\hat{\beta}_R + \hat{\beta}_I'(X'X)\hat{\beta}_I \\ &\quad + y_R'[I_n - X(X'X)^{-1}X']y_R + y_I'[I_n - X(X'X)^{-1}X']y_I \\ &\quad - 2\delta(\alpha_1^2 + \alpha_2^2 - 1) + 2\psi'(C\beta - \gamma) \ . \end{aligned}$$

Maximizing this likelihood with respect to the parameters is the same as maximizing the logarithm of the likelihood with respect to the parameters. In the case of β, α_1, and α_2 it is the same as minimizing the h term in the exponent. Note that the maximization is performed by Lagrange multipliers and the appropriate term has been added to h

$$\begin{aligned} \frac{\partial}{\partial\beta}h\Big|_{\beta=\tilde{\beta},\alpha_1=\tilde{\alpha}_1,\alpha_2=\tilde{\alpha}_2,\delta=\tilde{\delta},\psi=\hat{\psi},\sigma^2=\tilde{\sigma}^2} &= 2(X'X)\tilde{\beta} \\ &\quad -2(X'X)\left[\tilde{\beta}_R\tilde{\alpha}_1 + \tilde{\beta}_I\tilde{\alpha}_2\right] \\ &\quad +2C'\tilde{\psi} \\ \frac{\partial}{\partial\psi}h\Big|_{\beta=\tilde{\beta},\alpha_1=\tilde{\alpha}_1,\alpha_2=\tilde{\alpha}_2,\delta=\tilde{\delta},\psi=\hat{\psi},\sigma^2=\tilde{\sigma}^2} &= 2(C\tilde{\beta} - \gamma) \\ \frac{\partial}{\partial\alpha_1}h\Big|_{\beta=\tilde{\beta},\alpha_1=\tilde{\alpha}_1,\alpha_2=\tilde{\alpha}_2,\delta=\tilde{\delta},\psi=\hat{\psi},\sigma^2=\tilde{\sigma}^2} &= -2\tilde{\beta}'(X'X)\tilde{\beta}_R - 2\tilde{\delta}(2\tilde{\alpha}_1) \\ \frac{\partial}{\partial\alpha_2}h\Big|_{\beta=\tilde{\beta},\alpha_1=\tilde{\alpha}_1,\alpha_2=\tilde{\alpha}_2,\delta=\tilde{\delta},\psi=\hat{\psi},\sigma^2=\tilde{\sigma}^2} &= -2\tilde{\beta}'(X'X)\tilde{\beta}_I - 2\tilde{\delta}(2\tilde{\alpha}_2) \\ \frac{\partial}{\partial\delta}h\Big|_{\beta=\tilde{\beta},\alpha_1=\tilde{\alpha}_1,\alpha_2=\tilde{\alpha}_2,\delta=\tilde{\delta},\psi=\hat{\psi},\sigma^2=\tilde{\sigma}^2} &= -2(\tilde{\alpha}_1^2 + \tilde{\alpha}_2^2 - 1) \\ \frac{\partial}{\partial\sigma^2}\log[p(y|X,\beta,\alpha_1,\alpha_2,\sigma^2)]\Big|_{\beta=\tilde{\beta},\alpha_1=\tilde{\alpha}_1,\alpha_2=\tilde{\alpha}_2,\delta=\tilde{\delta},\psi=\hat{\psi},\sigma^2=\tilde{\sigma}^2} &= -\frac{2n}{2}\frac{1}{\tilde{\sigma}^2} + \frac{\tilde{h}}{2}\frac{1}{(\tilde{\sigma}^2)^2} \end{aligned}$$

where $\tilde{h}$ is h with MLE's substituted in. By setting these derivatives equal to zero and solving, we get the MLE's under the restricted model given in Equation 2.14.

C Prewhitening

In many applications of regression, the errors may be temporally autocorrelated resulting in correct estimation of the regression coefficients but inflated estimation of the residual error variance.

C.1 Magnitude Model

In the multiple regression magnitude model, the observation error covariance matrix may not be the identity matrix. A common practice is to estimate $\mathbf{\Phi}$ with $\hat{\mathbf{\Phi}}$, prewhiten, then repeat the analysis. For example, an AR(1) temporal autocorrelation (Markov) matrix with autocorrelation parameter ρ is estimated by $\hat{\rho}$, and $\hat{\mathbf{\Phi}}$ formed. Then by obtaining the factorization $\hat{\mathbf{\Phi}} = PP'$, and premultiplying

$$\begin{aligned} Py &= PX \quad \beta \quad + \quad P\epsilon \\ y_* &= X_* \quad \beta \quad + \quad \epsilon_* \, . \end{aligned} \tag{C.1}$$

Now, $\epsilon_* \sim \mathcal{N}(0, \sigma^2 I_n)$ and the data is analyzed according to the magnitude multiple regression model.

C.2 Complex Model

As in the magnitude regression model, in the multiple regression complex model, the observation error covariance matrix may not be the identity matrix. Again we can estimate $\mathbf{\Phi}$ with $\hat{\mathbf{\Phi}}$, prewhiten, then repeat the analysis. For example, an AR(1) temporal autocorrelation (Markov) matrix with autocorrelation parameter ρ_R for the real part and ρ_I for the imaginary part are estimated by $\hat{\rho}_R$ and $\hat{\rho}_R$, their average taken to obtain $\hat{\rho}$ and $\hat{\mathbf{\Phi}}$ formed. Then again by obtaining the factorization $\hat{\mathbf{\Phi}} = PP'$, and premultiplying

$$\begin{aligned} \begin{pmatrix} Py_R \\ Py_I \end{pmatrix} &= \begin{pmatrix} PX & 0 \\ 0 & PX \end{pmatrix} \begin{pmatrix} \beta\cos\theta \\ \beta\sin\theta \end{pmatrix} + \begin{pmatrix} P\eta_R \\ P\eta_I \end{pmatrix} \\ \begin{pmatrix} y_{R*} \\ y_{I*} \end{pmatrix} &= \begin{pmatrix} X_* & 0 \\ 0 & X_* \end{pmatrix} \begin{pmatrix} \beta\cos\theta \\ \beta\sin\theta \end{pmatrix} + \begin{pmatrix} \eta_{R*} \\ \eta_{I*} \end{pmatrix} \end{aligned} \tag{C.2}$$

Now, $\eta_* = (\eta_{R*}, \eta_{I*})' \sim \mathcal{N}(0, \Sigma \otimes I_n)$ and the data is analyzed according to the complex multiple regression model.

References

[1] P. Bandettini, A. Jesmanowicz, E. Wong, and J.S. Hyde, "Processing strategies for time-course data sets in functional MRI of the human brain," *Magnetic Resonance in Medicine*, vol. 30, no. 2, pp. 161–173, 1993.

[2] R.W. Cox, A. Jesmanowicz, and J.S. Hyde, "Real-time functional magnetic resonance imaging," *Magnetic Resonance in Medicine*, vol. 33, no. 2, pp. 230–236, 1995.

[3] E.M. Haacke, R. Brown, M. Thompson, and R. Venkatesan, *Magnetic Resonance Imaging: Principles and Sequence Design*, John Wiley and Sons, New York, 1999.

[4] S. Lai and G.H. Glover, "Detection of BOLD fMRI signals using complex data," *Proceedings of the ISMRM*, p. 1671, 1997.

[5] L.L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 2146–2157, 1994.

[6] F.Y. Nan and R.D. Nowak, "Generalized likelihood ratio detection for fMRI using complex data," *IEEE Transactions on Medical Imaging*, vol. 18, no. 4, pp. 320–329, 1999.

[7] S.O. Rice, "Mathematical analysis of random noise," *Bell system Tech. J.*, vol. 23, pp. 282, 1944, Reprinted by N. Wax, Selected papers on Noise and Stochastic Process, Dover Publication, 1954. QA273W3.

[8] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy data," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, 1995.

[9] G.C. Reinsel and R.P. Velu, *Multivariate Reduced-Rank Regression: Theory and Applications (Lecture Notes in Statistics, Vol 136)*, Springer Verlag, New York, 1998.

[10] Dennis V. Lindley and Adrian F. M. Smith, "Bayes estimates for the linear model," *Journal of the Royal Statistical Society B*, vol. 34, no. 1, 1972.

[11] D.B. Rowe, "Bayesian source separation for reference function determination in fMRI," *Magnetic Resonance in Medicine*, vol. 45, no. 5, pp. 374–378, 2001.

[12] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society B*, vol. 57, pp. 289–300, 1995.

[13] C.R. Genovese, N.A. Lazar, and T. Nichols, "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *Neuroimage*, vol. 15, pp. 772–786, 2002.

[14] B.R. Logan and D.B. Rowe, "An evaluation of thresholding techniques in fMRI analysis," *Neuroimage*, vol. 22, no. 1, pp. 95–108, 2004.

[15] A.H. Anderson and J.E. Kirsch, "Analysis of noise in phase contrast MR imaging," *Medical Physics*, vol. 23, no. 6, pp. 857–869, 1996.