Improved Activation Detection via Rice-Distributed fMRI Time Series Modeling

Daniel W. Adrian* Ranjan Maitra† Daniel B. Rowe‡

Abstract

Two developments in fMRI magnitude time series modeling, namely, the incorporation of temporal dependence and the Ricean distribution, have been separated by a distributional "mismatch": such time series modeling is largely based upon Gaussian-distributional-based extensions to the general linear model, which precludes its use under Ricean modeling. We bridge this gap by extending independent AR(p) errors to the latent, Gaussian-distributed real and imaginary components from which the Ricean-distributed magnitudes are computed by augmenting the observed magnitude data by missing phase data in an EM algorithm framework. We use the EM algorithm for parameter estimation and extend it to compute approximate standard errors and test statistics for activation and AR order detection. When compared to the standard Gaussian AR(p) model, this "AR(p) Ricean model" produces less-biased parameter estimates and similar performance on an experimental fMRI dataset.

Key Words: EM algorithm, empirical information matrix, hemodynamic response function, Monte Carlo integration, Rice distribution, von-Mises distribution

1. Introduction

Functional magnetic resonance imaging (fMRI) is a popular method for studying brain function because it is noninvasive, requires no exposure to radiation, and is widely available. The imaging modality is built on the fact that when neurons fire in response to a stimulus or a task, the blood oxygen levels in neighboring vessels changes, effecting the magnetic resonance (MR) signal on the order of 2-3% (Lazar, 2008), due to the differing magnetic susceptibilities of oxygenated and deoxygenated hemoglobin. This difference is behind the so-called Blood Oxygen Level Dependent (BOLD) contrast (Ogawa et al., 1990; Belliveau et al., 1991; Kwong et al., 1992; Bandettini et al., 1993) which is used as a surrogate for neural activity and is used to acquire time-course sequences of images, with the time-course in accordance with the stimulus and resting periods.

The general strategy to detect regions of neural activation is to fit, at each voxel, a model — commonly a general linear model (Friston et al., 1995) — to the time series observations against a transformation of the input stimulus: this transformation is the expected BOLD response and is effectively modeled in terms of a convolution of the stimulus time course with the hemodynamic response function (HRF), which measures the delay and dispersion of the BOLD response to an instantaneous neuronal activation (Friston et al., 1994; Glover, 1999). This provides the setting for the application of techniques such as Statistical Parametric Mapping (SPM) (Friston et al., 1990), where the time series at each voxel is reduced to a test statistic which summarizes the association between each voxel time course and the expected BOLD response (Bandettini et al., 1993). The resulting map is then thresholded to identify voxels that are significantly activated (Worsley et al., 1996; Genovese et al., 2002; Logan and Rowe, 2004).

*Department of Statistics, Grand Valley State University, Allendale, MI 49401

†Department of Statistics and Statistical Laboratory, Iowa State University, Ames, IA 50011

‡Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, WI 53233

Most statistical analyses focus on the magnitude data computed from the complex-valued measurements resulting from Fourier reconstruction (Jezzard and Clare, 2001) and discard the phase information. Because the real and imaginary measurements are well-modeled as two independent normal random variables with the same variance (Wang and Lei, 1994), these magnitude measurements follow the Rice distribution (Rice, 1944; Gudbjartsson and Patz, 1995). However, standard analyses assume that magnitude data are Gaussian-distributed, even though the Gaussian approximation of the Rice distribution is only valid at high signal-to-noise ratios (SNRs). This factor is increasingly important because the SNR is proportional to voxel volume (Lazar, 2008); thus an increase in the fMRI spatial resolution will correspond to a lowering of the SNR, making the Gaussian distributional approximation for the magnitude data less tenable.

In its simplest form, analysis of magnitude fMRI time series assumes no autocorrelation: however the naïveté of this assumption is widely recognized. There are many reasons for this: one is that the hemodynamic response disperses (or "smears," in fMRI jargon) neural activation. The hemodynamic (or BOLD) response to a single neural activation takes 15 to 20 seconds (Lazar, 2008), which is much longer than the sampling intervals of many fMRI techniques – 100 ms-5 s for echo-planar imaging (EPI) techniques (Friston et al., 1994). Further, the neuronal response, which can be modeled as a point response or a delta function (Friston et al., 1994), is itself very fast when compared to the BOLD response. Since fMRI experiments measure the BOLD response over time, the above discussion means that the observed time series within each voxel are correlated. Friston et al. (1994) also contend that the neuronal process is composed of "intrinsic" neuronal activities in addition to the stimulus-related response. Consequently, the authors say, autocorrelations in the observed time series arise from two neural components, both measured through the hemodynamic response: one that is experimentally induced owing to the stimulus and another that is due to intrinsic neuronal activity. The first component is modeled by convolution of the stimulus time course with the HRF, as discussed previously, while the second is modeled with autocorrelation. Additional sources of autocorrelation are also provided by the subject's cardiac and respiratory cycles (Friston et al., 2000).

Precise modeling of this temporal correlation is essential to maintaining assumed significance levels in tests for activation (Purdon and Weisskoff, 1998). Many analyses extend the linear model by introducing autocorrelated errors (Lazar, 2008). Prewhitening these errors is a common procedure, based on estimated autoregressive (AR) (Bullmore et al., 1996; Marchini and Ripley, 2000) or autoregressive moving average (ARMA) (Locascio et al., 1997) models, which produces the most efficient estimators. However, this approach can bias significance levels (Friston et al., 2000; Woolrich et al., 2001), so temporal (Worsley and Friston, 1995) and spatial (Worsley et al., 2002) smoothing have been recommended for more robustness. Likelihood-based activation statistics, based on incorporating an AR temporal correlation structure into the likelihood function, have also been proposed as a less-biased alternative to prewhitening approaches (den Dekker et al., 2009).

The above approaches all make Gaussian distributional assumptions for the observed magnitude time series, which as discussed before, is not appropriate, even approximately, at low SNR. This has led to the development of Rice-distributed magnitude-data models (den Dekker and Sijbers, 2005; Rowe, 2005; Solo and Noh, 2007; Zhu et al., 2009) which have, understandably, shown improved power of detection over their Gaussian counterparts at low SNR. These models, however, assume independence in the time series; incorporating autocorrelation in Rice-based models is impeded by the fact that the above approaches, such as ARMA modeling and prewhitening, are based on the Gaussian distribution.

In this paper, we develop a Ricean model for fMRI magnitude time series which incorporates AR(p) dependence. Due to the previously discussed "mismatch" between Gaussian-

based time series techniques and Ricean-distributed magnitude data, we do not model the magnitudes directly and instead utilize the fMRI data acquisition process as follows. Because the Rice distributed magnitude observations are computed from Gaussian distributed real and imaginary components, we apply the AR(p) dependence to this latent complex-valued data. In Section 2, because this complex-valued data is composed of observed magnitudes and "missing" phase data, we present the model through an EM algorithm (Dempster et al., 1977) framework and illustrate its use in parameter estimation; we also illustrate computation of approximate standard errors of these parameter estimates and tests for activation and AR order detection through extensions of the EM algorithm. We compare these AR(p)-Ricean-model parameter estimates to those based upon a Gaussian AR(p) model for simulated fMRI data in Section 3 and discuss our results in Section 4.

2. Methodology

We focus on the magnitude time series at a voxel, which we denote as $\boldsymbol{r} = (r_1, \ldots, r_n)$, where n is the number of scans. As discussed in Section 1, we incorporate autocorrelation into a Ricean-distributed model for $\boldsymbol{r}$ by applying AR(p) errors to the real and imaginary (Gaussian) time series, denoted $\boldsymbol{y}_R = (y_{R1}, \ldots, y_{Rn})$ and $\boldsymbol{y}_I = (y_{I1}, \ldots, y_{In})$, respectively. After transforming the distribution of $(\boldsymbol{y}_R, \boldsymbol{y}_I)$ to the magnitude-phase variables $(\boldsymbol{r}, \boldsymbol{\phi})$, where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)$, we apply the EM algorithm to obtain maximum likelihood estimates (MLEs) of model parameters, which we denote by $\boldsymbol{\tau}$. Because the phase data $\boldsymbol{\phi}$ is discarded in "magnitude-only" data analysis, the observed, missing, and complete data of EM algorithm terminology are represented by $\boldsymbol{r}$, $\boldsymbol{\phi}$, and $(\boldsymbol{r}, \boldsymbol{\phi})$, respectively.

2.1 Parameter estimation via the EM algorithm

An iteration of the EM algorithm consists of the Expectation and Maximization steps (or the E- and M-steps). At the $(k+1)$th iteration, $k \geq 0$, the E-step calculates the objective function $Q(\boldsymbol{\tau}; \boldsymbol{\tau}^{(k)}) = \mathrm{E}_{\boldsymbol{\phi}|\boldsymbol{r},\boldsymbol{\tau}^{(k)}}[\log f(\boldsymbol{r}, \boldsymbol{\phi}; \boldsymbol{\tau})]$, the expectation of the complete-data log-likelihood with respect to the conditional distribution $\boldsymbol{\phi}|\boldsymbol{r}$ at the current parameter estimates $\boldsymbol{\tau}^{(k)}$. The M-step calculates the updated parameter values $\boldsymbol{\tau}^{(k+1)}$ by maximizing $Q(\boldsymbol{\tau}; \boldsymbol{\tau}^{(k)})$ with respect to $\boldsymbol{\tau}$; that is, $\boldsymbol{\tau}^{(k+1)} = \mathrm{argmax}_{\boldsymbol{\tau}}\, Q(\boldsymbol{\tau}; \boldsymbol{\tau}^{(k)})$. In the following paragraphs, we illustrate the E- and M-steps involved in computing MLEs for AR(p)-Ricean-model parameters.

We begin our E-step description with the complete-data log-likelihood function, which results from applying AR(p) errors to the complex-valued data model of Rowe and Logan (2004),

$$\begin{pmatrix} \boldsymbol{y}_R \\ \boldsymbol{y}_I \end{pmatrix} = \begin{pmatrix} \boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}\cos\theta \\ \boldsymbol{\beta}\sin\theta \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_R \\ \boldsymbol{\eta}_I \end{pmatrix}, \tag{1}$$

where the $n \times q$ design matrix $\boldsymbol{X}$ models effects such as the baseline signal level, signal drift, and the expected BOLD response, and θ represents the constant mean of the phase time series $\boldsymbol{\phi}$. (Note that since magnitude time series contain no phase information, θ is neither known nor estimated.) The error terms $\boldsymbol{\eta}_R = (\eta_{R1}, \ldots, \eta_{Rn})$ and $\boldsymbol{\eta}_I = (\eta_{I1}, \ldots, \eta_{In})$ are independent AR(p) time series parameterized by AR coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)$ and white noise variance σ^2; that is, for $t = 1, \ldots, n$, $\eta_{Rt} = \sum_{i=1}^{p} \alpha_i \eta_{R,t-i} + \epsilon_{Rt}$ and $\eta_{It} = \sum_{i=1}^{p} \alpha_i \eta_{I,t-i} + \epsilon_{It}$, with $\epsilon_{Rt}, \epsilon_{It} \sim$ iid $N(0, \sigma^2)$. We also denote γ_j as the lag-j autocovariance, $j = 0, \ldots, p$, and use $(\gamma_0, \ldots, \gamma_p)$ as an alternative parameterization of $(\boldsymbol{\alpha}, \sigma^2)$, obtained via the Yule-Walker equations (Shumway and Stoffer, 2006). Additionally, we define $\boldsymbol{R}_n$ such that $\mathrm{Cov}(\boldsymbol{\eta}_R) = \mathrm{Cov}(\boldsymbol{\eta}_I) = \sigma^2 \boldsymbol{R}_n$. The AR order p, which is

assumed known in this model formulation, can be chosen using the order selection procedure described in Section 2.2. Thus, the complete-data log-likelihood is

$$\log f(\boldsymbol{y}_R, \boldsymbol{y}_I; \boldsymbol{\tau}) = -n \log \sigma^2 - \log |\boldsymbol{R}_n| - h/2\sigma^2, \tag{2}$$

where $h = \tilde{\boldsymbol{\alpha}}' \boldsymbol{D} \tilde{\boldsymbol{\alpha}}$, $\tilde{\boldsymbol{\alpha}}$ being the $(p+1)$-vector $(1, -\alpha_1, \ldots, -\alpha_p)$ and $\boldsymbol{D}$ the $(p+1)\times(p+1)$ symmetric matrix with (i,j)th entry $d_{ij} = \sum_{t=1}^{n-i-j} [\eta_{R,t+i}\eta_{R,t+j} + \eta_{I,t+i}\eta_{I,t+j}]$, $0 \leq i, j \leq p$ (Pourahmadi, 2001). To transform the complete log-likelihood function from real-imaginary to magnitude-phase variables, we apply the relations $y_{Rt} = r_t \cos\phi_t$ and $y_{It} = r_t \sin\phi_t$, $t = 1, \ldots, n$, to d_{ij}, which produces $d_{ij} = \sum_{t=1}^{n-i-j} r_{t+i} r_{t+j} \cos(\phi_{t+i} - \phi_{t+j}) - \mu_{t+i} r_{t+j} \cos(\phi_{t+j} - \theta) - \mu_{t+j} r_{t+i} \cos(\phi_{t+i} - \theta) + \mu_{t+i}\mu_{t+j}$, where $\mu_t = \boldsymbol{x}_t' \boldsymbol{\beta}$, $\boldsymbol{x}_t'$ being the tth row of $\boldsymbol{X}$. In view of (2) and this expression for d_{ij}, we note that the E-step involves two categories of expectations: the univariate expectations $\mathrm{E}[\cos(\phi_t - \theta)|r_t, \boldsymbol{\tau}^{(k)}]$, $t = 1, \ldots, n$, and the bivariate expectations $\mathrm{E}[\cos(\phi_t - \phi_{t+j})|r_t, r_{t+j}, \boldsymbol{\tau}^{(k)}]$, $j = 1, \ldots, p$, $t = 1, \ldots, n-j$.

The former expectations are with respect to the von-Mises distribution $VM(\cdot,\cdot)$, which is defined in Appendix A. It can be shown that $\phi_t | r_t, \boldsymbol{\tau}^{(k)}$ is von-Mises by transforming (y_{Rt}, y_{It}), which are independent and follow $N(\mu_t^{(k)} \cos\theta, \gamma_0^{(k)})$ and $N(\mu_t^{(k)} \sin\theta, \gamma_0^{(k)})$ distributions, to the variables (r_t, ϕ_t). This gives $f(\phi_t | r_t, \boldsymbol{\tau}^{(k)}) \propto \exp\{\frac{\mu_t^{(k)} r_t}{\gamma_0^{(k)}} \cos(\phi_t - \theta)\}$, where $\mu_t^{(k)} = \boldsymbol{x}_t' \boldsymbol{\beta}^{(k)}$, which, as seen in Appendix A, is the probability density function (PDF) of the $VM(\theta, \mu_t^{(k)} r_t / \gamma_0^{(k)})$ distribution. Thus, again following from Appendix A, the univariate expectations $\mathrm{E}[\cos(\phi_t - \theta)|r_t, \boldsymbol{\tau}^{(k)}] = A(\mu_t^{(k)} r_t / \gamma_0^{(k)})$, $t = 1, \ldots, n$, where $A(\cdot) = I_1(\cdot)/I_0(\cdot)$, $I_j(\cdot)$ being the jth order modified Bessel function of the first kind (Abramowitz and Stegun, 1965).

The bivariate expectations can be approximated via univariate Monte Carlo integration as follows. First, we generate $\psi^{(1)}, \ldots, \psi^{(m)} \sim$ iid $VM(0, \mu_t^{(k)} r_t / \gamma_0^{(k)})$, which can be efficiently accomplished through the rejection sampling algorithm of Best and Fisher (1979). The expectation $\mathrm{E}[\cos(\phi_t - \phi_{t+j})|r_t, r_{t+j}, \boldsymbol{\tau}^{(k)}]$ is then approximated as

$$\frac{1}{m} \sum_{i=1}^{m} \frac{A(K^{(i)})}{K^{(i)}} [\kappa^* \cos\psi^{(i)} + \delta], \tag{3}$$

where $K^{(i)} = \sqrt{\kappa^{*2} + \delta^2 + 2\kappa^* \delta \cos\psi^{(i)}}$, $\kappa^* = r_{t+j}(\gamma_0^{(k)} \mu_{t+j}^{(k)} - \gamma_j^{(k)} \mu_t^{(k)})/(\gamma_0^{2(k)} - \gamma_j^{2(k)})$, and $\delta = \gamma_j^{(k)} r_t r_{t+j} / (\gamma_0^{2(k)} - \gamma_j^{2(k)})$. See Appendix B for a derivation of (3).

To summarize, the E-step replaces the h-term in (2) by its expectation, which we denote by $h^{(k)} = \tilde{\boldsymbol{\alpha}}' \boldsymbol{D}^{(k)} \tilde{\boldsymbol{\alpha}}$, where $\boldsymbol{D}^{(k)}$ is a $(p+1)\times(p+1)$ symmetric matrix with (i,j)th entry

$$d_{ij}^{(k)} = \sum_{t=1}^{n-i-j} \Big\{ r_{t+i} r_{t+j} \mathrm{E}[\cos(\phi_{t+i} - \phi_{t+j})|r_{t+i}, r_{t+j}, \boldsymbol{\tau}^{(k)}] - \mu_{t+i} r_{t+j} A(r_{t+j} \mu_{t+j}^{(k)} / \gamma_0^{(k)}) - \mu_{t+j} r_{t+i} A(r_{t+i} \mu_{t+i}^{(k)} / \gamma_0^{(k)}) + \mu_{t+i}\mu_{t+j} \Big\}. \tag{4}$$

Also, to prepare for the M-step, we represent the $\boldsymbol{\beta}$-dependent portion of $h^{(k)}$ as $h^{(k)}(\boldsymbol{\beta}) = \boldsymbol{\beta}' \boldsymbol{X}' \boldsymbol{R}_n^{-1} \boldsymbol{X} \boldsymbol{\beta} - 2\boldsymbol{\beta}' \boldsymbol{X}' \boldsymbol{R}_n^{-1} \boldsymbol{X} \boldsymbol{u}^{(k)}$, where $\boldsymbol{u}^{(k)}$ is a vector of length n with tth entry $u_t^{(k)} = r_t A(\mu_t^{(k)} r_t / \gamma_0^{(k)})$, $t = 1, \ldots, n$.

Because the M-step does not have a closed form, we obtain $\boldsymbol{\tau}^{(k+1)}$ through three conditional maximization steps as described in the ECM algorithm of Meng and Rubin (1993).

First, given $\boldsymbol{\beta}^{(k)}$, we calculate $\boldsymbol{\alpha}^{(k+1)}$ from the equations (which modify Miller, 1995)

$$\sum_{j=1}^{p} \left(d_{ij}^{(kk)} + 2j\gamma_{|j-i|}^{(kk)} \right) \alpha_j = d_{i0}^{(kk)}, \;\; i = 1, \ldots, p, \tag{5}$$

where $d_{ij}^{(kk)}$ substitutes $\mu_t^{(k)}$ for μ_t, $t = 1, \ldots, n$, in (4) and $\gamma_j^{(kk)} = d_{0j}^{(kk)}/(2n)$ estimates the lag-j autocovariance, $j = 0, \ldots, p$. Second, conditioned on $\boldsymbol{\alpha}^{(k+1)}$, we calculate $\boldsymbol{\beta}^{(k+1)}$ as $\boldsymbol{\beta}^{(k+1)} = (\boldsymbol{X}'\boldsymbol{R}_n^{-1(k+1)}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}_n^{-1(k+1)}\boldsymbol{u}^{(k)}$, which follows from minimizing the expression for $h^{(k)}(\boldsymbol{\beta})$ in the previous paragraph, where $\boldsymbol{R}_n^{-1(k+1)}$ is calculated from $\boldsymbol{\alpha}^{(k+1)}$ (Pourahmadi, 2001). Finally, we calculate $\sigma^{2(k+1)} = h^{(k,k+1)}/(2n)$, where $h^{(k,k+1)}$ substitutes $(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}^{(k+1)})$ for $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in $h^{(k)}$. To complete the EM algorithm, we calculate $\gamma_j^{(k+1)}$, $j = 0, \ldots, p$, from $(\boldsymbol{\alpha}^{(k+1)}, \sigma^{2(k+1)})$ and the Yule-Walker equations for use in the next E-step.

To compute starting values, we used the independent Ricean model, which itself employs an EM algorithm (Solo and Noh, 2007), to compute $(\boldsymbol{\beta}^{(0)}, \sigma^{2(0)})$ and set $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$. To save computation time, we use fewer Monte Carlo samples m for initial iterations and increase m as the algorithm moves closer to convergence, as advocated by Wei and Tanner (1990). Also, due to the randomness associated with each EM step, we assumed convergence if the convergence criterion was satisfied for three (instead of two) consecutive iterations (Booth and Hobert, 1999), which produced the (approximate) MLE $\hat{\boldsymbol{\tau}}$.

2.2 Calculation of standard errors and test statistics

We utilize the Fisher information matrix for calculation of approximate standard errors for the MLEs. However, because the observed-data likelihood function is intractable in this case, we use the empirical information matrix (Meilijson, 1989), an estimate of the Fisher information matrix which can be calculated from the complete-data likelihood and its expectation under the E-step. More specifically, the empirical information matrix is the sum over independent observations of outer products of the score statistics, where each (observed-data) score statistic can be calculated as the expectation of the corresponding complete-data score statistic with respect to the distribution of the missing data conditioned on the observed data. In our context, we do not have independence, but we can exploit the fact that AR(p) processes have a conditional independence structure which provides similar factoring of the likelihood: when conditioned on the first p observations, the complete data consists of $n - p$ conditionally independent (complex-valued) observations; that is, $(r_t, \phi_t)|(r_{t-1}, \phi_{t-1}), \ldots, (r_{t-p}, \phi_{t-p})$, which we abbreviate as $[\boldsymbol{r}, \boldsymbol{\phi}]_{t|t-p}$, are independent for $t = p + 1, \ldots, n$. Similarly denoting $[\boldsymbol{r}]_{t|t-p}$ as $r_t|r_{t-1}, \ldots, r_{t-p}$, the empirical information matrix is given by $\boldsymbol{I}_e(\hat{\boldsymbol{\tau}}; \boldsymbol{r}) = \sum_{t=p+1}^{n} \boldsymbol{s}([\boldsymbol{r}]_{t|t-p}; \hat{\boldsymbol{\tau}})\boldsymbol{s}'([\boldsymbol{r}]_{t|t-p}; \hat{\boldsymbol{\tau}})$, where the score statistic $\boldsymbol{s}([\boldsymbol{r}]_{t|t-p}; \hat{\boldsymbol{\tau}})$ can be calculated as $\mathrm{E}_{\boldsymbol{\phi}|\boldsymbol{r}, \hat{\tau}}[\boldsymbol{s}([\boldsymbol{r}, \boldsymbol{\phi}]_{t|t-p}; \boldsymbol{\tau})]_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$, the expectation of $\boldsymbol{s}([\boldsymbol{r}, \boldsymbol{\phi}]_{t|t-p}; \boldsymbol{\tau}) = \partial/\partial\boldsymbol{\tau} \log L([\boldsymbol{r}, \boldsymbol{\phi}]_{t|t-p}; \boldsymbol{\tau})$. More details on this calculation are given in Appendix C. For verification, we also estimated the Fisher information matrix using Louis's method (Louis, 1982). Following standard practice, approximate standard errors for $\hat{\tau}_i$, $i = 1, \ldots, q + p + 1$, are given by $\{\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\tau}}; \boldsymbol{r})_{\tau_i\tau_i}\}^{1/2}$, the square-root of the diagonal entry of $\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\tau}}; \boldsymbol{r})$ corresponding to $\boldsymbol{\tau}_i$.

Wald statistics for activation and order detection follow from the information matrix. We generally pose the test for activation as $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0}$ vs. $H_a : \boldsymbol{C}\boldsymbol{\beta} \neq \mathbf{0}$, which has corresponding Wald statistic $(\boldsymbol{C}\hat{\boldsymbol{\beta}})'[\boldsymbol{C}\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\tau}}; \boldsymbol{r})_{\boldsymbol{\beta}\boldsymbol{\beta}}\boldsymbol{C}']^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}})$, where $\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\tau}}; \boldsymbol{r})_{\boldsymbol{\beta}\boldsymbol{\beta}}$ refers to the $q \times q$ block of $\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\tau}}; \boldsymbol{r})$ corresponding to $\boldsymbol{\beta}$. This statistic has an asymptotic χ^2 null distribution with degrees of freedom equal to the rank of $\boldsymbol{C}$. We utilize a sequential hypothesis testing procedure (similar to forward model selection in regression modeling)

for order detection, in which the detected order $\hat{p} = i' - 1$, where i' is the smallest i, $i \geq 1$, such that a test of $H_0 : \alpha_i = 0$ vs. $H_a : \alpha_i \neq 0$ does not reject H_0. A Wald statistic for this test is given by $\hat{\alpha}_i^2 / \boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\tau}}; \boldsymbol{r})_{\alpha_i \alpha_i}$, where $\hat{\alpha}_i$ and $\boldsymbol{I}_e^{-1}(\hat{\boldsymbol{\tau}}; \boldsymbol{r})_{\alpha_i \alpha_i}$ are computed under the AR(i) Ricean model, which follows an asymptotic χ_1^2 null distribution.

2.3 Gaussian Autoregressive model

We compare the parameter estimates and test statistics derived under the AR(p) Ricean model to those based on a standard Gaussian AR(p) model. In this Gaussian model, $\boldsymbol{r} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{X}$ is the same as before and $\boldsymbol{\epsilon}$ follows an AR(p) dependence structure parameterized by $\boldsymbol{\alpha}$ and σ^2. The MLEs of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and σ^2 can be obtained according to Cochrane and Orcutt (1949). For comparison with the test statistics derived in Section 2.2, the Gaussian AR(p) model test statistics are also calculated as Wald statistics which utilize the empirical information matrix.

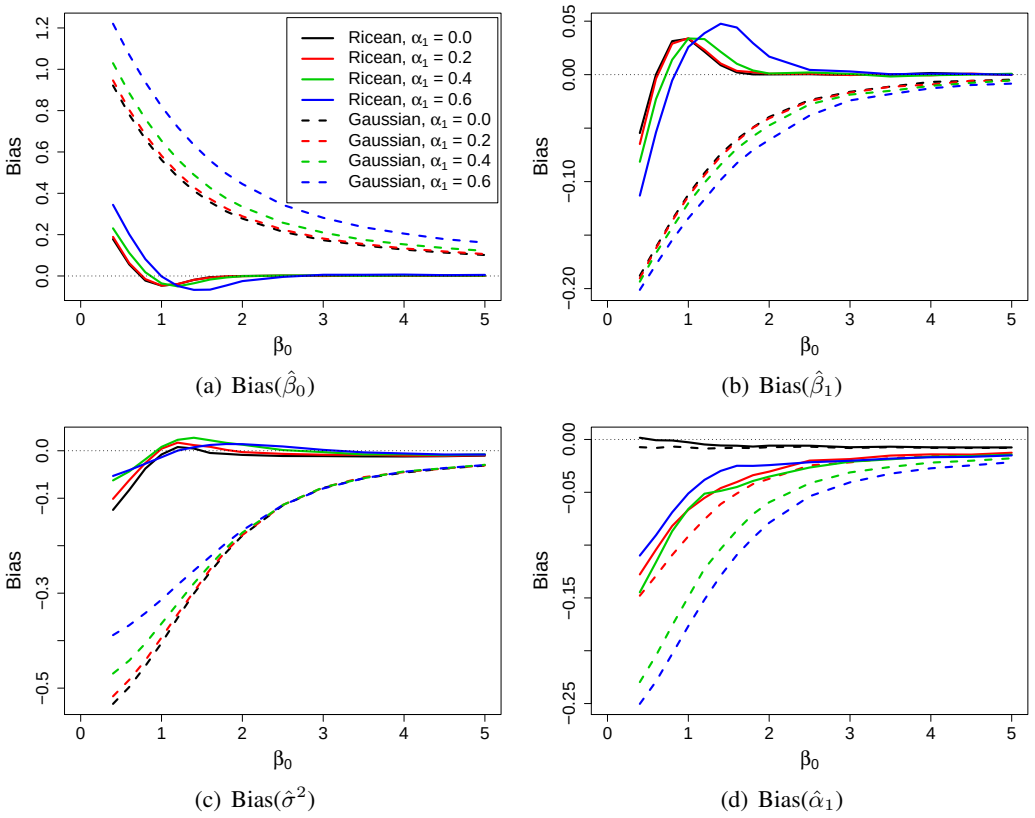
3. Experimental Evaluations

We generated Ricean-distributed magnitude time series of length $n = 256$ by simulating from (1) and computing $r_t = \sqrt{y_{Rt}^2 + y_{It}^2}$ for $t = 1, \ldots, 256$. The design matrix $\boldsymbol{X}$ contained $q = 2$ columns, which included an intercept term to model the baseline signal and a ± 1 square wave alternating every 16 time points to model the expected BOLD response. Thus, in $\boldsymbol{\beta} = (\beta_0, \beta_1)$, only β_1 was activation-related, so the test for activation was posed as $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ and the corresponding activation test statistic follows an asymptotic χ_1^2 null distribution. We maintained $\sigma^2 = 1.0$ over all simulations for simple interpretation of the signal-to-noise ratio SNR $= \beta_0/\sigma$. We varied β_0 from 0.4 to 10 to examine low SNR values (most fMRI data has SNR above 10) and used $\beta_1 = 0.2$ and 0.0 to represent activated and non-activated voxel time series, respectively. We applied AR(1) dependence with AR coefficient values $\alpha_1 = 0.2, 0.4$, and 0.6 and fit AR(1) models to all time series. We simulated 100,000 magnitude time series at each collection of parameter values and computed MLEs, standard error estimates, and (activation and AR order detection) test statistics under Ricean and Gaussian models as described in Section 2. Implementation of the AR(1) Ricean-model EM algorithm included ten preliminary iterations generating $m = 2$ Monte Carlo samples per expectation, followed by iterations utilizing $m = 10$ samples until convergence was reached.

First, we examine the biases of the Ricean and Gaussian-model MLEs, which are plotted against β_0 (or alternatively, SNR) in Figure 1. Of the two models, the Ricean-model estimates show less bias; the Gaussian-model estimates become increasingly biased as the SNR decreases due to the worsening Gaussian approximation of the Rice distribution discussed in Section 1. Further, we note that these (Gaussian-model) biases increase with α_1 and attribute this to decreasing SNR as well: as α_1 increases, the SNR decreases due to the increasing variance of AR(1) processes, which is given by $\gamma_0 = \sigma^2/(1 - \alpha_1^2)$. Though the opposite is true for the Gaussian-model estimate of σ^2, we argue that this effect is "artificial" and due to the bias of $\hat{\alpha}_1$ increasing with α_1.

4. Discussion

In this paper, we developed the first Ricean model for fMRI magnitude time series that incorporates time dependence. We used an indirect (but natural) approach, applying AR(p) errors to the Gaussian-distributed real and imaginary components from which the magnitudes are computed. We modeled the latent complex-valued data by augmenting the



(a) Bias($\hat{\beta}_0$) (b) Bias($\hat{\beta}_1$)

(c) Bias($\hat{\sigma}^2$) (d) Bias($\hat{\alpha}_1$)

Figure 1: Biases of the AR(1) Ricean- and Gaussian-model parameter estimates are plotted against β_0 (or alternatively, SNR) for different values of α_1.

observed magnitude data by missing phase data according to an EM algorithm framework, which we used to calculate parameter estimates, standard errors, and test statistics for activation and AR order detection. We showed that this AR(p) Ricean model produces less-biased parameter estimates than its standard AR(p) Gaussian counterpart.

A. The von-Mises distribution

If φ follows the von-Mises distribution $VM(\theta_0, \kappa)$, where θ_0 is the mean direction and κ is the concentration parameter, the PDF of φ is given by

$$f(\varphi|\theta_0, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\varphi - \theta_0)], \tag{6}$$

for $\varphi \in (-\pi, \pi)$, $\theta_0 \in (-\pi, \pi)$, and $\kappa > 0$ (Mardia and Jupp, 2000), $I_j(\cdot)$ being the jth order modified Bessel function of the first kind (Abramowitz and Stegun, 1965). In the E-step, we use the von-Mises expectations $\mathrm{E}(\cos\varphi) = A(\kappa)\cos\theta_0$ and $\mathrm{E}(\sin\varphi) = A(\kappa)\sin\theta_0$, where $A(\cdot) = I_1(\cdot)/I_0(\cdot)$, as well as the location-family property that $\varphi \sim VM(\theta_0, \kappa) \implies (\varphi - \theta_0) \sim VM(0, \kappa)$.

B. Derivation of Monte Carlo approximation

The Monte Carlo approximation (3) for $\mathrm{E}[\cos(\phi_t - \phi_{t+j})|r_t, r_{t+j}, \boldsymbol{\tau}^{(k)}]$, $j = 1, \ldots, p$, $t = 1, \ldots, n-j$, follows from the expansion

$$\mathrm{E}_{\phi_t|r_t,\boldsymbol{\tau}^{(k)}}\{\cos(\phi_t - \theta)\mathrm{E}_*[\cos(\phi_{t+j} - \theta)] + \sin(\phi_t - \theta)\mathrm{E}_*[\sin(\phi_{t+j} - \theta)]\}, \tag{7}$$

where $\mathrm{E}_*[\cdot]$ denotes expectation with respect to $\phi_{t+j}|\phi_t, r_t, r_{t+j}, \boldsymbol{\tau}^{(k)}$. We show that the latter is von-Mises-distributed by transforming $(y_{Rt}, y_{R,t+j})$ and $(y_{It}, y_{I,t+j})$, which are independent and bivariate normal with means $(\mu_t^{(k)}, \mu_{t+j}^{(k)})\cos\theta$ and $(\mu_t^{(k)}, \mu_{t+j}^{(k)})\sin\theta$, respectively, and (the same) covariance matrix with diagonal and off-diagonal entries $\gamma_0^{(k)}$ and $\gamma_j^{(k)}$, to magnitude-phase variables. It can then be shown that

$$f(\phi_{t+j}|\phi_t, r_t, r_{t+j}, \boldsymbol{\tau}^{(k)}) \propto \exp\left[\kappa^* \cos(\phi_{t+j} - \theta) + \delta\cos(\phi_t - \phi_{t+j})\right], \tag{8}$$

where κ^* and δ are as in Section 2.1. After combining the bracketed portion of (8) into a single cosine term, it is evident that $\phi_{t+j}|\phi_t, r_t, r_{t+j}, \boldsymbol{\tau}^{(k)} \sim VM(\Psi(\phi_t), K(\phi_t))$, where $\Psi(\phi_t) = \arctan\{\delta\sin(\phi_t - \theta)/[\kappa^* + \delta\cos(\phi_t - \theta)]\}$ and $K(\phi_t) = \{\kappa^{*2} + \delta^2 + 2\kappa^*\delta\cos(\phi_t - \theta)\}^{1/2}$. Using this distribution to compute the expectations $\mathrm{E}_*[\cdot]$ in (7) and simplifying, we obtain

$$\mathrm{E}_{\phi_t|r_t,\boldsymbol{\tau}^{(s)}}\left\{\frac{A(K(\phi_t))}{K(\phi_t)}[\kappa^*\cos(\phi_t - \theta) + \delta]\right\}. \tag{9}$$

Because this expectation does not have a closed form, we approximate it by Monte Carlo integration, simulating from $(\phi_t - \theta)|r_t, \boldsymbol{\tau}^{(k)} \sim VM(0, \mu_t^{(s)} r_t/\gamma_0^{(k)})$ and averaging as is in (3).

C. Calculation of empirical information matrix

To illustrate the calculation of $\boldsymbol{s}([\boldsymbol{r}]_{t|t-p}; \hat{\boldsymbol{\tau}}) = \mathrm{E}_{\boldsymbol{\phi}|\boldsymbol{r},\hat{\tau}}[\partial/\partial\boldsymbol{\tau}\log f([\boldsymbol{r}, \boldsymbol{\phi}]_{t|t-p}; \boldsymbol{\tau})]_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}$, $t = p+1, \ldots, n$, we begin with deriving $\log f([\boldsymbol{r}, \boldsymbol{\phi}]_{t|t-p}; \boldsymbol{\tau})$. By transforming the distributions of $[\boldsymbol{y}_R]_{t|t-p}$ and $[\boldsymbol{y}_I]_{t|t-p}$ (following the notation in Section 2.2), which are independent and normal with respective means $\mu_t\cos\theta + \sum_{i=1}^p \alpha_i(y_{R,t-i} - \mu_{t-i}\cos\theta)$ and $\mu_t\sin\theta + \sum_{i=1}^p \alpha_i(y_{I,t-i} - \mu_{t-i}\sin\theta)$ and variances σ^2, to magnitude-phase variables, it can be shown that $\log f([\boldsymbol{r}, \boldsymbol{\phi}]_{t|t-p}; \boldsymbol{\tau}) = -\log\sigma^2 - h_t/(2\sigma^2)$, where $h_t = \tilde{\boldsymbol{\alpha}}'\boldsymbol{D}_t\tilde{\boldsymbol{\alpha}}$, $\boldsymbol{D}_t$ being a $(p+1)\times(p+1)$ matrix with (i,j)th-entry $d_t(i,j) = r_{t-i}r_{t-j}\cos(\phi_{t-i} - \phi_{t-j}) - \mu_{t-i}r_{t-j}\cos(\phi_{t-j} - \theta) - \mu_{t-j}r_{t-i}\cos(\phi_{t-i} - \theta) + \mu_{t-i}\mu_{t-j}$, $0 \le i, j \le p$, and $\tilde{\boldsymbol{\alpha}}$ as before. After computing derivatives with respect to the parameters, the expectation involves the two types of expectations described in Section 2.1.

References

Abramowitz, M. and Stegun, I. (1965), *Handbook of Mathematical Functions*, Dover Publications.

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., and Hyde, J. S. (1993), "Processing strategies for time-course data sets in functional MRI of the human brain," *Magnetic Resonance in Medicine*, 30, 161–173.

Belliveau, J. W., Kennedy, D. N., McKinstry, R. C., Buchbinder, B. R., Weisskoff, R. M., Cohen, M. S., Vevea, J. M., Brady, T. J., and Rosen, B. R. (1991), "Functional mapping of the human visual cortex by magnetic resonance imaging," *Science*, 254, 716–719.

Best, D. J. and Fisher, N. I. (1979), "Efficient simulation of the von Mises distribution," *Journal of the Royal Statistical Society. Series C (Applied Statistics).*, 28, 152–157.

Booth, J. and Hobert, J. (1999), "Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo algorithm," *Journal of the Royal Statistical Society B*, 61, 265–285.

Bullmore, E., Brammer, M., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., and Sham, P. (1996), "Statistical methods of estimation and inference for function MR image analysis," *Magnetic Resonance in Medicine*, 35, 261–277.

Cochrane, D. and Orcutt, G. (1949), "Applications of least squares regression to relationships containing autocorrelated errors," *Journal of the American Statistical Association*, 44, 32–61.

Dempster, A. P., Laird, N. M., and Rubin, D. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society Series B*, 23, 1–38.

den Dekker, A. J., Poot, D. H. J., Bos, R., and Sijbers, J. (2009), "Likelihood-based hypothesis tests for brain activation detection from MRI data disturbed by colored noise: a simulation study," *IEEE Transactions on Medical Imaging*, 28, 287–296.

den Dekker, A. J. and Sijbers, J. (2005), "Implications of the Rician distribution for fMRI generalized likelihood ratio tests," *Magnetic Resonance Imaging*, 23, 953–959.

Friston, K. J., Frith, C. D., Liddle, P. F., Dolan, R. J., Lammertsma, A. A., and Frackowiak, R. S. J. (1990), "The relationship between global and local changes in PET scans," *Journal of Cerebral Blood Flow and Metabolism*, 10, 458–466.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-B., Frith, C. D., and Frackowiak, R. S. J. (1995), "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, 2, 189–210.

Friston, K. J., Jezzard, P., and Turner, R. (1994), "Analysis of functional MRI time-series," *Human Brain Mapping*, 1, 153–171.

Friston, K. J., Josephs, O., Zarahn, E., Holmes, A. P., Rouqette, S., and Poline, J.-B. (2000), "To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis," *NeuroImage*, 12, 196–208.

Genovese, C. R., Lazar, N. A., and Nichols, T. E. (2002), "Thresholding of statistical maps in functional neuroimaging using the false discovery rate:," *NeuroImage*, 15, 870–878.

Glover, G. H. (1999), "Deconvolution of impulse response in event-related BOLD fMRI," *NeuroImage*, 9, 416–429.

Gudbjartsson, H. and Patz, S. (1995), "The Rician distribution of noisy data," *Magnetic Resonance in Medicine*, 34, 910–914.

Jezzard, P. and Clare, S. (2001), "Principles of nuclear magnetic resonance and MRI," in *Functional MRI: An Introduction to Methods*, eds. Jezzard, P., Matthews, P. M., and Smith, S. M., Oxford University Press, chap. 3, pp. 67–92.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H.-M., Brady, T. J., and Rosen, B. R. (1992), "Dynamic Magnetic Resonance Imaging of Human Brain Activity During Primary Sensory Stimulation," *Proceedings of the National Academy of Sciences of the United States of America*, 89, 5675–5679.

Lazar, N. A. (2008), *The Statistical Analysis of Functional MRI Data*, Springer.

Locascio, J. J., Jennings, P. J., Moore, C. I., and Corkin, S. (1997), "Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging," *Human Brain Mapping*, 5, 168–193.

Logan, B. R. and Rowe, D. B. (2004), "An evaluation of thresholding techniques in fMRI analysis," *NeuroImage*, 22, 95–108.

Louis, T. (1982), "Finding the observed information matrix when using the EM algorithm," *Journal of the Royal Statistical Society B*, 44, 226–233.

Marchini, J. L. and Ripley, B. D. (2000), "A new statistical approach to detecting significant activation in functional MRI," *NeuroImage*, 12, 366–380.

Mardia, K. V. and Jupp, P. E. (2000), *Directional Statistics*, Wiley.

Meilijson, I. (1989), "A fast improvement to the EM algorithm on its own terms," *Journal of the Royal Statistical Society. Series B (Methodological)*, 51, 127–138.

Meng, X.-L. and Rubin, D. B. (1993), "Maximum likelihood estimation via the ECM algorithm: a general framework," *Biometrika*, 80, 267–278.

Miller, J. W. (1995), "Exact Maximum Likelihood Estimation in Autoregressive Processes," *Journal of Time Series Analysis*, 16, 607–615.

Ogawa, S., Lee, T. M., Nayak, A. S., and Glynn, P. (1990), "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields," *Magnetic Resonance in Medicine*, 14, 68–78.

Pourahmadi, M. (2001), *Foundations of Time Series Analysis and Prediction Theory*, Wiley.

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," *Human Brain Mapping*, 6, 239–249.

Rice, S. O. (1944), "Mathematical analysis of random noise," *Bell Systems Technical Journal*, 23, 282.

Rowe, D. B. (2005), "Parameter estimation in the magnitude-only and complex-valued fMRI data models," *NeuroImage*, 25, 1124–1132.

Rowe, D. B. and Logan, B. R. (2004), "A complex way to compute fMRI activation," *NeuroImage*, 23, 1078–1092.

Shumway, R. H. and Stoffer, D. S. (2006), *Time Series Analysis and Its Applications*, Springer, second edition ed.

Solo, V. and Noh, J. (2007), "An EM algorithm for Rician fMRI activation detection," in *ISBI*, pp. 464–467.

Wang, T. and Lei, T. (1994), "Statistical analysis of MR imaging and its applications in image modeling," *Proceedings of the IEEE International Conference on Image Processing and Neural Networks*, 1, 866–870.

Wei, G. and Tanner, M. (1990), "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," *Journal of the American Statistical Association*, 85, 699–704.

Woolrich, M. W., Ripley, B. D., Brady, M., and Smith, S. M. (2001), "Temporal autocorrelation in univariate linear modeling of FMRI data," *NeuroImage*, 14, 1370–1386.

Worsley, K. J. and Friston, K. J. (1995), "Analysis of fMRI time-series revisited - again," *NeuroImage*, 2, 173–181.

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., and Evans, A. C. (2002), "A general statistical analysis for fMRI data," *NeuroImage*, 15, 1–15.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (1996), "A unified statistical approach for determining significant voxels in images of cerebral activation," *Human Brain Mapping*, 4, 58–73.

Zhu, H., Li, Y., Ibrahim, J. G., Shi, X., An, H., Chen, Y., Gao, W., Lin, W., Rowe, D. B., and Peterson, B. S. (2009), "Regression Models for Identifying Noise Sources in Magnetic Resonance Images," *Journal of the American Statistical Association*, 104, 623–637.