Submitted to the Annals of Applied Statistics
arXiv: math.PR/0000000

RICE-DISTRIBUTED AUTOREGRESSIVE TIME SERIES MODELING OF MAGNITUDE FUNCTIONAL MRI DATA

BY DANIEL W. ADRIAN[‡] AND RANJAN MAITRA[*,§] AND DANIEL B. ROWE[†,¶]

Grand Valley State University[‡] and Iowa State University[§] and Marquette University[¶]

Functional magnetic resonance imaging (fMRI) data generally consist of time series image volumes of the magnitude of complex-valued observations at each voxel. However, incorporating dependence and the Rice distribution – a more accurate model for the data – in the time series have been separated by a distributional "mismatch" because fMRI time series are currently modeled by Gaussian-distribution-based extensions to the general linear model, which precludes its use under Ricean modeling. We bridge this gap by applying pth-order autoregressive (AR) errors to the latent, Gaussian-distributed real and imaginary components from which the Ricean-distributed magnitudes are computed by augmenting the observed magnitude data with the missing phase data in an Expectation-Maximization (EM) algorithm framework. After parameter estimation via the EM algorithm, we compute AR order and test statistics for activation detection. Using simulated and experimental low-SNR fMRI data, we compare the performance of this Ricean time series model with those under the Gaussian AR(p) model, and also models on the entire complex-valued data. Our results show improved parameter estimation and activation detection, under the Ricean AR(p) model than its Gaussian counterpart. Further, models analyzing the complex-valued data detect activation better than magnitude-only models but only because they have more data and allow for incorporating nonspherical covariance structure. Thus, while our results here provide for the improved analysis of archived magnitude-only fMRI data, they also argue strongly against the currently routine practice of discarding the phase of the complex-valued fMRI time series, advocating instead for their inclusion in the analysis.

1. Introduction. Functional magnetic resonance imaging (fMRI) is a prominent non-invasive modality for studying human brain function. It is built upon the principle of the Blood Oxygen Level Dependent (BOLD) contrast (Bandettini *et al.*, 1993; Belliveau *et al.*, 1991; Kwong *et al.*, 1992; Ogawa *et al.*, 1990), where firing neurons lead to changes in the blood oxygen levels of neighboring vessels,

[*]Research supported in part by the the National Science Foundation CAREER Grant # DMS-0437555 and the National Institutes of Health (NIH) awards #R21EB016212 and #R21EB034184.

[†]Research supported in part by the National Institutes of Health (NIH) award #R21NS087450.

AMS 2000 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: bilateral finger-tapping motor experiment, EM algorithm, empirical information matrix, hemodynamic response function, Monte Carlo integration, Rice distribution, signal-to-noise ratio, von-Mises distribution

and the magnetic resonance (MR) signal fluctuates due to the differing magnetic susceptibilities of oxygenated and deoxygenated hemoglobin (Lazar, 2008). Scientists can gain insight on the functional structures of the brain by analyzing time courses of MR signals acquired while a subject performs a designed series of tasks.

The voxel-wise MR signal at each time point is originally complex-valued, containing real and imaginary (or equivalently, magnitude and phase) components. This complex-valued attribute is a consequence of how the data are acquired: the originally measured, complex-valued k-space data (Brown, Kincaid and Ugurbil, 1982; Ljunggren, 1983; Tweig, 1983) consist of the different frequency contributions to the signal from each voxel resulting from magnetic field gradients (Jezzard and Clare, 2001). Then, the application of the inverse Fourier transform (Jain, 1989), a complex-valued operation on the k-space data, separates these frequencies and localizes each voxel's measurements. However, despite the fact that the original signal is complex-valued, statistical analysis of fMRI data is almost always only on the magnitude data, with the acquired concomitant phase discarded. We refer to such analyses as "magnitude-only" (MO) statistical analyses, and note that this approach likely arises as a consequence of the default output of MR scanners that does not routinely include phase images, even though they can easily be collected by simply changing a preset variable in an input file (Yu *et al.*, 2018). Consequently, most fMRI data and analyses are MO, and, at least figuratively, do not use half of the originally available data.

One of the most common forms of MO analysis fits, at each voxel, a general linear model (Friston *et al.*, 1995) for the (preprocessed) time series observations in terms of a waveform representing the expected BOLD contrast. This waveform is the convolution of the stimulus time course with the hemodynamic response function (HRF), which gives the BOLD response to an instantaneous neuronal activation (Friston, Jezzard and Turner, 1994; Glover, 1999). These general linear models for magnitude fMRI time series also incorporate autoregressive (AR) (Bullmore *et al.*, 1996; den Dekker *et al.*, 2009; Marchini and Ripley, 2000) or autoregressive moving average (ARMA) (Locascio *et al.*, 1997) errors, due to several reasons. For one, the hemodynamic response to a single neural activation takes between 15 and 20 seconds (Lazar, 2008), which is much longer than the sampling intervals of many fMRI techniques — for instance, of between 100 milliseconds and five seconds for echo-planar imaging (EPI) techniques (Friston, Jezzard and Turner, 1994). Additional sources of autocorrelation are also provided by the subject's cardiac and respiratory cycles (Friston *et al.*, 2000), and by the common pre-processing step of temporal smoothing. From these model fits, the time series at each voxel is aggregated to a test statistic that measures the degree of activation, in the Statistical Parametric Mapping (SPM) framework of Friston *et al.* (1990). Thresholding methods are then applied to the SPM to identify activated voxels (Genovese, Lazar

and Nichols, 2002; Logan and Rowe, 2004; Worsley *et al.*, 1996).

The above MO approaches assume that the magnitude measurements follow a Gaussian distribution. However, it is well-known (Wang and Lei, 1994) that the complex k-space data are Gaussian distributed, and this distributional assumption is also preserved, by linearity, upon applying an inverse Fourier transform. Specifically, it is commonly assumed (Wang and Lei, 1994) that the real and imaginary measurements are independent normal random variables with the same variance and phase-coupled means, it follows that their magnitudes have a Rice distribution (Gudbjartsson and Patz, 1995; Rice, 1944). In general, MR images simultaneously acquired from multiple independent coils (Tristán-Vega, Aja-Fernández and Westin, 2012) can be shown to follow the non-central χ-distribution, with degrees of freedom equal to twice the number of coils (Wegmann, Eklund and Villani, 2017). The Rice distribution is the special case of a single coil and two degrees of freedom. The Gaussian MO model may be justified by the fact that the Rice distribution approaches the Gaussian distribution for large signal-to-noise ratios (SNRs). For magnitude fMRI time series, the SNR represents the ratio of the mean, that is, the non-activation-related, baseline signal to the standard deviation (SD) of the noise time series. There is also the contrast-to-noise ratio, or CNR, that is the ratio of the amplitude of the BOLD contrast to the noise SD. However, low-SNR fMRI data may occur in studies having small voxel sizes or voxels with a large degree of signal drop-out, such as those located near air/tissue boundaries. Out of this concern that the Gaussian assumption may not be adequate for such low-SNR data, Zhu *et al.* (2009) developed Rice-distributed models that ignored temporal dependence in the voxel-wise time series. Also, Solo and Noh (2007) demonstrated that Gaussian-model-based maximum likelihood (ML) estimates of parameters for simulated Ricean data are biased for SNRs under 5, with the bias increasing as the SNR decreases. However, Adrian, Maitra and Rowe (2013) showed that Rice- and Gaussian-model-based likelihood ratio tests for activation reported similar performance for SNRs as low as 0.6 in the absence of correlation in the time series.

A different approach utilizing the complex-valued (CV) data (that is, using both magnitude and phase data) has shown to have several advantages. Rowe and Logan (2004) introduced a general model for CV fMRI time series and showed through simulation that MO data-based activation tests have a substantial drop of power at SNRs below 3 while corresponding CV data-based tests have constant (and higher) power over all SNRs. Adrian, Maitra and Rowe (2018) demonstrated how to fit an extension of the Rowe and Logan (2004) model incorporating AR(p) time dependence and a general covariance structure for the real and imaginary errors. For experimental data, this CV model showed greater activation power than a corresponding Gaussian AR(p) MO model after spatial smoothing, but similar power

when the data were not smoothed. This differential behavior was attributed to the covariance structure of the two kinds of errors: while the unsmoothed data exhibited a spherical covariance structure under the twin assumptions of independence of real and imaginary errors and equal variances, the application of spatial smoothing induced a non-spherical covariance structure. The MO model does not use the necessary data to estimate the real/imaginary covariance structure; we will see in fact that it assumes sphericity. Because sphericity was an adequate assumption for the unsmoothed data, the MO and CV models performed similarly. However, the MO model was a poor fit for the nonspherical smoothed data and had worse performance than the CV model. Extensive simulations supported the better performance of CV model-based activation detection for nonspherical data. Beyond this empirical finding, the phase time series can be useful in the so-called "brain or vein" problem, that is, in determining whether a voxel showing a task-related magnitude change represents (desired) grey matter or (an undesired) draining vein (Hoogenrad *et al.*, 1998; Menon, 2002; Rowe, 2005a). It has been shown that task-related phase changes generally occur in the latter but not in the former.

In this paper, we explore some of the "missing ground" in these previous comparisons of fMRI time series models. Because the temporal correlation of fMRI time series has been well-established, we use AR errors of general order p, in contrast to previous comparisons between CV, MO Gaussian, and MO Ricean models that assume temporal independence (Adrian, Maitra and Rowe, 2013; Rowe, 2005a; Rowe and Logan, 2004; Solo and Noh, 2007). Also, the comparison in Adrian, Maitra and Rowe (2018) between the CV model with general (nonspherical) real/imaginary covariance and the MO Gaussian models, though both models are based on AR(p) errors, is missing two "intermediate models" that would help make it clear exactly which feature(s) of the CV model have the most impact. Because the model features of nonsphericity of the real/imaginary covariance, having twice the quantities in CV versus in MO data, and not utilizing the Gaussian distributional approximation of the (truly Rice-distributed) magnitudes are all present in the CV model but not the MO model in Adrian, Maitra and Rowe (2018), we were unable in that paper to separate which of these three features contribute to the CV model's improved performance. By adding two intermediate models in this paper, we are able to isolate each feature's contribution in a more scientific approach. One of these intermediate models models is a CV model that assumes a spherical real/imaginary covariance matrix (*i.e.*, independence and homogeneity of the real and imaginary errors at the same time-point). Comparing the nonspherical and spherical CV models isolates the nonsphericity model feature. The other intermediate model is a MO model with a Ricean distribution, but one that also models AR(p) temporal dependence in the errors. To our knowledge, this AR(p) Ricean model represents a novel methodological development, as all previously developed

Ricean models have assumed independent (or pre-whitened) time series. Comparing the CV spherical model to this MO Ricean model isolates the feature that CV data-based models have twice the quantities of MO models. Additionally, comparing the MO Ricean and Gaussian models isolates the influence of Gaussian approximation of the Rice-distributed magnitudes, and provides a fuller understanding of the strengths and shortcomings of the MO analyses relative to the CV analyses.

The remainder of the paper is structured as follows. Section 2 introduces a series of low-SNR images acquired from a finger-tapping experiment that is our motivating application. These images have a lower SNR than typical fMRI data due to their being acquired with the body coil instead of the head coil, and their analysis here is motivated by previous results that have indicated that CV model-based activation tests perform better than their MO model-based counterparts at low SNRs. Section 3 elaborates on the methodology behind the four models introduced above and features the novel development of the MO AR(p) Ricean model. Section 4 performs simulation experiments to study the validity of the methodological results and compares the results using different models, under a known ground truth. Section 5 presents the statistical analysis of the low-SNR dataset and its implications. We conclude with a discussion of the results and the paper. Our paper also has a supplement containing additional details regarding methodology, the simulation experiments, and the dataset analysis. Sections, figures, and equations in the supplement are referenced here with the prefix "S-".

2. A low-SNR fMRI finger-tapping experiment dataset.

2.1. *Data acquisition.* We develop our methods in the backdrop of a common sequential finger-tapping experiment. In our application, MR images were acquired with TR = 1s during a block design experiment with an initial 16s of rest followed by 19 epochs of 16s of right-hand finger tapping alternating with 16s of rest. Such finger-tapping experiments have application in noninvasive neurosurgical preparation (Lee *et al.*, 1999). Experiments that use finger tapping, sponge squeezing, or brushing of the palms as stimuli in block design are used to identify the location of hand function in candidate patients for resective surgical treatment for tumors and epilepsy (Lee, Jack and Riederer, 1998). While it is well-known that the central sulcus in the sensorimotor cortex is the location of hand function for normal healthy adults (Rumeau *et al.*, 1994), fMRI allows the "topographic relationship between a proposed surgical target and the location of specific

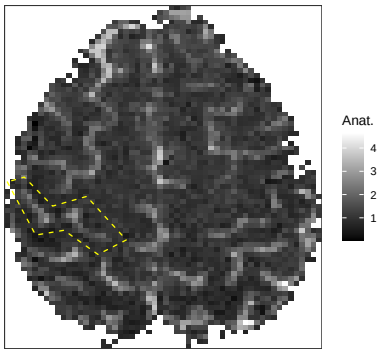


FIG 1. *Anatomical image of the second slice identifying the left central sulcus.*

functional areas" to be verified noninvasively (Lee *et al.*, 1999). For our dataset, due to the decussation of nerve fibers in the upper slices of the brain (Carpenter, 1991), it is expected that the right-hand finger tapping activates the left central sulcus of the brain, which is identified on the second slice[1] in Figure 1.

The data were acquired with the body coil. In general, the body coil has a large measurement field, and thus has lower SNR than specialized coils such as the head coil usually used in fMRI studies.[2] This sort of acquisition provides us with a low-SNR dataset of a well-studied experiment to serve as a marker for performance in low-SNR settings, which is where CV, MO Ricean and MO Gaussian models have been shown to diverge previously. Following standard practice, the first three images are excluded from our analysis due to machine "warmup" effects, leaving us with a temporal sequence of $n = 621$ images. Each volume image was composed of seven 2.5 mm thick 128×128 axial slices with a 24.0 cm FOV.

2.2. *Data processing pipeline.* For this dataset, the phase components of the time series images were not discarded but stored along with the magnitude images used in traditional fMRI analysis. The data processing flow included Nyquist ghost removal and correction for global zero-order off-resonance using three navigator echos (Jesmanowicz, Wong and Hyde, 1993; Nencka, Hahn and Rowe, 2008), image reconstruction from k-space by inverse Fourier transform (Kumar, Welti and Ernst, 1975; Rowe, 2016), and estimation and correction of the dynamic field using temporal off-resonance alignment of single-echo timeseries (TOAST) (Hahn, Nencka and Rowe, 2009, 2012). A binary mask of voxels above 12% of the maximum voxel signal magnitude was generated from the first magnitude image of the dataset (before discarding the first three images) to represent voxels within the brain. We also use the first magnitude image for the anatomical images on which we overlay our activation maps. Figure 2 shows images of the real, imaginary,

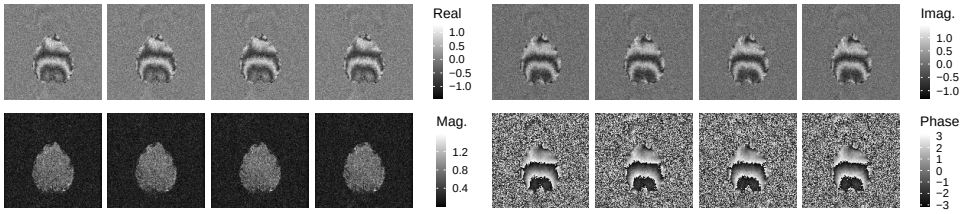


FIG 2. *Images (from top-left) of the real, imaginary, magnitude, and phase components of images in the second slice at the 14th, 22nd, 30th, and 38th time-points (from left-to-right, within each image data type).*

[1]We focus on the second slice because we will see later that it shows the strongest activation.

[2]When specialized receive-only coils are used, the body coil serves as the transmit coil. More recently, body coils have been used to acquire a separate reference image to obtain sensitivity maps.

magnitude, and phase components of the second slice at the $t = 14, 22, 30, 38$ time-points, representing the first cycle of tapping and rest periods. These images do not appear to change much in time because of the small amplitude of the BOLD response relative to the level of the baseline signal. Figure **??** shows plots of the real, imaginary, magnitude, and phase time series for one of the most strongly activated voxels.

Two further "preprocessing" steps included detrending to correct for scanner drift and spatial smoothing to increase the data's SNR and CNR. Two sources of drift are noise from the MR scanner and aliasing of cardiorespiratory cycles (Tanabe *et al.*, 2002), and the magnitude of these changes "often far exceeds" both the white noise and the amplitude of the task-related single change (Genovese, 2000). Our study of the dataset's time series suggests diverse, nonlinear shapes of drift profiles, not only for magnitude time series such as those reported in Genovese (2000), but for the real, imaginary, and phase time series as well. Figure **??** shows the plot of such a time series and compares four methods for fitting the trend: the CV running line (Adrian, Maitra and Rowe, 2018), a polynomial fit, a natural cubic spline, and a smoothing spline. (See Section **??** for more detail.) We determined that the smoothing spline was the preferred choice, but to study the robustness of our choice, we used both the CV running line and smoothing splines in practice.[3] An important last component of detrending CV time series is counter-rotating by the mean phase, producing a mean phase of zero for all voxel time series; this corrected for the global spatial differences seen in the real and imaginary components of Figure 2 and produced a more cohesive dataset for spatial smoothing.

After detrending, the resulting real, imaginary, and magnitude images at each time-point were spatially smoothed with 3-dimensional anisotropic Gaussian filters. Essentially, each measurement was recalculated as a weighted average of itself and neighboring measurements, with the weights determined by a 3-dimensional Gaussian distribution. We chose the filter to be anisotropic (*i.e.,* direction-dependent) because the intra-slice dimensions of the voxels is less than the interslice dimension; the standard deviation of the filter in each direction was chosen to be inversely proportional to these dimensions. To study the robustness of the analysis to different amounts of spatial smoothing, we used intraslice FWHMs of 1.5, 2, 3, and 4 voxels[4], while noting that such values are typically between 3 and 10 mm FWHM (Lazar, 2008). Thus, including the spatially unsmoothed datasets, we obtained a total of five CV and MO datasets. Figure **??** shows images of the SNRs and CNRs for the second slice using the different smoothing levels. We see that both the SNRs

[3]Indeed, the results were very similar for both detrending approaches, so we only present those based on smoothing splines.

[4]FWHM is the full width at half maximum, the width of the smoothing filter at half of its maximum density.

and CNRs increase with the amount of spatial smoothing, at the expense of spatial resolution and specificity.

After all the preprocessing, the CV and MO models were applied to each voxel time series. For each model, the design matrix $\boldsymbol{X}$ had $n = 621$ rows and $q = 2$ columns: one column was an intercept modeling the baseline MR signal and the other was a zero-centered waveform modeling the expected BOLD repsonse given by a convolution of the stimulus time course with the Glover (1999) hemodynamic response function. The bottom panel of Figure **??** shows a superposition of the block design stimulus time course with this expected BOLD response waveform. The next section presents these CV and MO time series models in detail.

3. Methodological development. As discussed in Section 1, we compare four models for fMRI time series in this paper with the goal of isolating the influence of three model features on activation detection performance. However, one of these models – the Rice-distributed autoregressive time series model – and its estimation, needs development, and we use this section to also do so.

3.1. *Statistical models for CV and MO time series.* We first introduce notation, focusing on a single voxel (and suppressing voxel-related subscripts). The CV measurement at time t can be denoted in real/imaginary form by $y_{Rt} + iy_{It}$ or in magnitude/phase form by $r_t \exp(i\phi_t) = r_t(\cos\phi_t + i\sin\phi_t)$. Trigonometric identities in the complex plane hold that $y_{Rt} = r_t\cos\phi_t$, $y_{It} = r_t\sin\phi_t$, $r_t = (y_{Rt}^2 + y_{It}^2)^{-\frac{1}{2}}$, and $\phi_t = \arctan_4(y_{It}, y_{Rt})$, the 4-quadrant arctangent (see Glisson, 2011, Page 348) corresponding to $\arctan(y_{It}/y_{Rt})$. We denote the real, imaginary, magnitude, and phase time series vectors by $\boldsymbol{y}_R = (y_{R1}, \ldots, y_{Rn})'$, $\boldsymbol{y}_I = (y_{I1}, \ldots, y_{In})'$, $\boldsymbol{r} = (r_1, \ldots, r_n)'$, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$, with n denoting the number of MR scans. The Rowe and Logan (2004) model states that

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

where the expected magnitude response $\boldsymbol{X\beta}$ is coupled with the constant phase location parameter θ. The columns of $\boldsymbol{X}$ represent various components of the magnitude signal including the baseline level and the expected BOLD contrast. The errors $\boldsymbol{\eta} = (\boldsymbol{\eta}_R', \boldsymbol{\eta}_I')' \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma} \otimes \boldsymbol{\Phi})$, where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Phi}$ are matrices of order 2 and n, specifying the real/imaginary and temporal covariances (the latter with an AR(p) structure), and the direct (Kronecker) product $\otimes$ implies separability of these covariances. An equivalent representation of (1) is by the magnitude and phase activation model (Rowe and Logan, 2004),

$$\begin{pmatrix} \boldsymbol{y}_R \\ \boldsymbol{y}_I \end{pmatrix} = \begin{pmatrix} \boldsymbol{CX} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{SX} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_R \\ \boldsymbol{\eta}_I \end{pmatrix}, \tag{2}$$

where $\boldsymbol{C}$ and $\boldsymbol{S}$ are diagonal matrices of order n, having tth entries $\cos\theta_t$ and $\sin\theta_t$, $t = 1, 2, \ldots, n$, where $\theta_t = \delta_0 + g(\boldsymbol{z}_t'\boldsymbol{\delta})$, with $g(x) = 2\arctan(x)$. So $\boldsymbol{\beta}$ models the magnitude-related activation and $\boldsymbol{\delta}$ models the phase-related activation. Based on this framework, we present four models:

1. **Complex-valued AR(p) nonspherical (CVNS) model**: Introduced in Adrian, Maitra and Rowe (2018), this model takes the form of (1) with an AR(p) structure for $\boldsymbol{\Phi}$ and a general (nonspherical) $\boldsymbol{\Sigma}$, allowing separate variances for the real and imaginary errors — denoted by σ_R^2 and σ_I^2, respectively — and correlation ρ between the real and imaginary errors. We denote the AR coefficients by $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$.
2. **Complex-valued AR(p) spherical (CVS) model**: This is a special case of the CVNS model, with the restriction that $\boldsymbol{\Sigma}$ is spherical — that is, $\rho = 0$ and $\sigma_R^2 = \sigma_I^2 \equiv \sigma^2$ (or $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_2$, where $\boldsymbol{I}_2$ is the identity matrix of order 2). Under this model, the lag-j autocovariances $\gamma_j = \mathrm{Cov}(\eta_{Rt}, \eta_{R,t-j}) = \mathrm{Cov}(\eta_{It}, \eta_{I,t-j})$, $j = 0, 1, \ldots, p$, are a useful reparameterization of $\boldsymbol{\alpha}$ and σ^2, with the Yule-Walker equations (Shumway and Stoffer, 2006) serving as an intermediary.
3. **Magnitude-only AR(p) Rician (MOR) model**: This model is the MO (marginal) counterpart of the CVS model; that is, under the MO Rician AR(p) model, the latent real and imaginary time series follow the CVS model. Under the MOR model, the probability density function (PDF) of magnitude measurements r_t is

$$f(r_t; \mu_t, \gamma_0) = \frac{r_t}{\gamma_0} \exp\left[\frac{-(r_t^2 + \mu_t^2)}{2\gamma_0}\right] \mathbb{I}_0\left(\frac{\mu_t r_t}{\gamma_0}\right), \tag{3}$$

   where $\mu_t = \boldsymbol{x}_t'\boldsymbol{\beta}$, $\boldsymbol{x}_t'$ is the tth row of $\boldsymbol{X}$, and $\mathbb{I}_0(\cdot)$ is the modified Bessel function of the first kind and the zeroth order.
4. **Magnitude-only AR(p) Gaussian (MOG) model**: This model assumes $\boldsymbol{r} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ follows an AR(p) structure.[5]

These four models are summarized in Table 1, which along with the three "model features" lists the key ways that differentiate them. These features are as follows:

1. *Nonsphericity*: The CVNS model has the ability to model *nonsphericity* in the covariance of the real and imaginary measurements, while the CVS model assumes sphericity, that is, $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_2$. Because the MO models only use the magnitudes, an aggregation of the real and imaginary measurements, they do not currently have the ability to model *nonsphericity*.

[5]For brevity, the distinction between the notations of corresponding parameters in different models is dropped.

		Model features		
Model	Model Abbrev.	Nonsphericity	Twice the Quantities	Ricean magnitudes
CV Nonspherical	CVNS	✓	✓	✓
CV Spherical	CVS		✓	✓
MO Ricean	MOR			✓
MO Gaussian	MOG			

TABLE 1

Summary of the four models compared in this paper for complex-valued (CV) and magnitude-only (MO) fMRI data and the features present in each model.

2. *Twice the quantities*: The CV data-based models use the real and imaginary measurements at each voxel and time-point, while the MO data-based models use only the magnitudes, so the former models use *twice the quantities* of the latter. (However, we note that this does not necessarily mean that twice the amount of useful information is present in the CV over the MO data, especially for our models of magnitude-related activation.[6])
3. *Ricean magnitudes*: The MOR model assumes the magnitude measurements are Ricean distributed, while the MOG model uses the Gaussian approximation. As previously discussed, the CVS model has marginally distributed *Ricean magnitudes*. The magnitude distribution in the CVNS model (Aalo, Efthymoglou and Chayawan, 2007) is more specialized, but because the CVNS model does not assume Gaussian-distributed magnitudes, we will consider this feature to be present.

Table 1 illustrates the sequential and hierarchical pattern of the three features present across the four models. For instance, the CVNS and CVS models both contain the *twice the quantities* and the *Ricean magnitudes* features, but differ in that the CVNS model features *nonsphericity* while the CVS model does not. Thus, we can isolate the influence of modeling nonsphericity (*i.e.*, on parameter estimates, activation detection, and so on) by comparing these two models. Similarly, comparing the (CVS, MOR) and (MOR, MOG) model pairs separately allows us to isolate the influence of the *twice the quantities* and *Ricean magnitudes* features.

3.2. *Parameter Estimation.* Of the four models introduced in Section 3.1, we spend the most time on the methodology of the MO Ricean AR(p) (or MOR) model due to its novelty; to our knowledge, this is the first development of a Ricean fMRI time series model allowing for dependence between successive errors. The other three models have seen more developed and so we refer our readers elsewhere for further explanation. The CV AR(p) nonspherical (or CVNS) model was illustrated

[6] Rowe (2005a) introduces fMRI models that allow for activation in both the magnitude and phase data.

in Adrian, Maitra and Rowe (2018), and we relegate discussions of the CV AR(p) spherical (CVS) and the MO Gaussian (MOG) models to Section **??**, as both are simplified cases of the CVNS model. In illustrating the methodology associated with each model, we first assume a known AR order p; because this order is not *a priori* known, we discuss methodology for order detection in Section 3.4.2. Also, due to our overriding interest in comparing these models, we derive relationships between the probability density functions (PDFs) of the CVS, MOR, and MOG models in Section **??**.

The methodology for the Ricean AR(p) model fits nicely into the framework of the EM algorithm (Dempster, Laird and Rubin, 1977; McLachlan and Krishnan, 2008) with $\boldsymbol{r}$, $\boldsymbol{\phi}$, and $(\boldsymbol{r}, \boldsymbol{\phi})$ playing the roles of the observed, missing, and complete data, respectively. And because the EM algorithm and its extensions serve as our "workhorse" methodology for the Ricean AR(p) model, we review it briefly here. An iteration of the EM algorithm consists of the Expectation (E-) and Maximization (M-) steps. At the $(k+1)$th iteration, the E-step calculates the objective function $Q(\boldsymbol{\tau}; \boldsymbol{\tau}^{(k)}) = \mathbb{E}_{\boldsymbol{\phi}|\boldsymbol{r},\boldsymbol{\tau}^{(k)}}[\log f(\boldsymbol{r}, \boldsymbol{\phi}; \boldsymbol{\tau})]$, 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 $\boldsymbol{\tau}^{(k+1)} = \operatorname{argmax}_{\boldsymbol{\tau}} Q(\boldsymbol{\tau}; \boldsymbol{\tau}^{(k)})$ by maximizing the objective function. We denote the vector of parameters by $\boldsymbol{\tau} = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \sigma^2)'$. The EM algorithm has well-known favorable properties such as monotone increase of the likelihood for each iteration and reliable global convergence (McLachlan and Krishnan, 2008). We next illustrate the specifics of the EM algorithm that we use for ML estimation.

3.2.1. *EM algorithm for ML estimation under the MO AR(p) Ricean model.* To compute starting values $\boldsymbol{\tau}^{(0)}$, we use the Gaussian AR(p) model as demonstrated in Section **??**. With the algorithm initialized, the E- and M-steps are as follow.

E-step. Borrowing from the representations of the AR(p) likelihood (Miller, 1995; Pourahmadi, 2001), the complete data log-likelihood can be shown to be

$$\log f(\boldsymbol{r}, \boldsymbol{\phi}; \boldsymbol{\tau}) = -n \log \sigma^2 - \log |\boldsymbol{R}_n| - h/2\sigma^2, \tag{4}$$

where $h = \tilde{\boldsymbol{\alpha}}' \boldsymbol{D} \tilde{\boldsymbol{\alpha}}$, with $\tilde{\boldsymbol{\alpha}}$ as 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 element

$$d_{ij} = \sum_{t=1}^{n-i-j} \begin{bmatrix} r_{t+i} r_{t+j} \cos(\phi_{t+i} - \phi_{t+j}) - \mu_{t+i} r_{t+j} \cos(\phi_{t+j} - \theta) \\ -\mu_{t+j} r_{t+i} \cos(\phi_{t+i} - \theta) + \mu_{t+i} \mu_{t+j} \end{bmatrix}, \tag{5}$$

where $\mu_t = \boldsymbol{x}_t' \boldsymbol{\beta}$, $\boldsymbol{x}_t'$ is the tth row of $\boldsymbol{X}$. In view of (4) and (5), the E-step involves two kinds of expectations: the univariate expectations $\mathbb{E}[\cos(\phi_t - \theta) | r_t; \boldsymbol{\tau}^{(k)}]$, $t = 1, \ldots, n$, and the bivariate expectations $\mathbb{E}[\cos(\phi_t - \phi_{t+j}) | r_t, r_{t+j}; \boldsymbol{\tau}^{(k)}]$, $j = 1, \ldots, p$,

$t = 1, \ldots, n-j$. The univariate expectations are with respect to the von-Mises distribution (Mardia and Jupp, 2000) and can be shown (Section **??**) to be

$$\mathbb{E}_{\phi_t|r_t;\boldsymbol{\tau}^{(k)}}[\cos(\phi_t - \theta)] = \mathbb{A}(\mu_t^{(k)} r_t / \gamma_0^{(k)}), \qquad t = 1, \ldots, n,$$

where $\mathbb{A}(\xi) = \mathbb{I}_1(\xi)/\mathbb{I}_0(\xi)$, with $\mathbb{I}_m(\xi)$ being the mth order modified Bessel function of the first kind (Abramowitz and Stegun, 1965) evaluated at ξ.

The bivariate expectations are more cumbersome to obtain. First, we reduce $\mathbb{E}_{\phi_t,\phi_{t+j}|r_t,r_{t+j};\boldsymbol{\tau}^{(k)}}[\cos(\phi_t - \phi_{t+j})]$ to the univariate expectation

$$\mathbb{E}_{\phi_t|r_t,\boldsymbol{\tau}^{(k)}}\left\{\frac{\mathbb{A}(K(\phi_t))}{K(\phi_t)}[\kappa\cos(\phi_t - \theta) + \delta]\right\}, \tag{6}$$

where $K(\phi_t) = [\kappa^2 + \delta^2 + 2\kappa\delta\cos(\phi_t - \theta)]^{1/2}$, $\kappa = r_{t+j}(\gamma_0^{(k)}\mu_{t+j}^{(k)} - \gamma_j^{(k)}\mu_t^{(k)})/b^{(k)}$, and $\delta = \gamma_j^{(k)} r_t r_{t+j}/b^{(k)}$, with $b^{(k)} = \gamma_0^{2(k)} - \gamma_j^{2(k)}$. (See Section **??** for more details.) Because (6) cannot be evaluated analytically, we approximate it via the Delta Method (Casella and Berger, 2002): $\mathbb{E}[f(X)] \approx f[\mathbb{E}(X)]$. When applied to (6), the Delta Method substitutes $\mathbb{A}(r_t\mu_t^{(k)}/\gamma_0^{(k)})$ for each instance of $\cos(\phi_t - \theta)$, including those in the $K(\phi_t)$ terms.

M-step. The global maxima of the objective function is not of closed form, so we obtain $\boldsymbol{\tau}^{(k+1)}$ through three conditional maximization steps as in the ECM algorithm of Meng and Rubin (1993). We modify the equations of Miller (1995) and Pourahmadi (2001) to

$$\sum_{j=1}^{p}\left(d_{ij}^{(k)} + 2j\gamma_{|j-i|}^{(k)}\right)\alpha_j = d_{i0}^{(k)}, \quad i = 1, \ldots, p, \tag{7}$$

where $d_{ij}^{(k)}$ is the E-step expectation of d_{ij} with μ_t evaluated at $\mu_t^{(k)}$ and $\gamma_j^{(k)} = d_{0j}^{(k)}/(2n)$. We first calculate $\boldsymbol{\alpha}^{(k+1)}$ from (7). Next, we calculate

$$\boldsymbol{\beta}^{(k+1)} = (\boldsymbol{X}'\boldsymbol{R}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}_n^{-1}\boldsymbol{u}^{(k)}, \tag{8}$$

where $\boldsymbol{R}_n^{-1}$ is obtained from $\boldsymbol{\alpha}^{(k+1)}$ (as in Pourahmadi, 2001) and $\boldsymbol{u}^{(k)}$ is a vector of n variables with tth element $u_t^{(k)} = r_t\mathbb{A}(r_t\mu_t^{(k)}/\gamma_0^{(k)})$. Note that it may be necessary to enforce the boundary conditions $\boldsymbol{X}\boldsymbol{\beta}^{(k+1)} \geq \boldsymbol{0}$, in which case (8) needs to be modified as discussed in Section **??**. Finally, we calculate $\sigma^{2(k+1)} = h^{(k+1)}/(2n)$, where $h^{(k+1)} = \tilde{\boldsymbol{\alpha}}'^{(k+1)}\boldsymbol{D}^{(k+1)}\tilde{\boldsymbol{\alpha}}^{(k+1)}$ and $\boldsymbol{D}^{(k+1)}$ is a matrix as before with terms $d_{ij}^{(k+1)}$ evaluated using $\mu_t^{(k+1)} = \boldsymbol{x}_t'\boldsymbol{\beta}^{(k+1)}$.

3.2.2. *Hybrid algorithm for ML estimation.* As the EM algorithm progresses through iterations, we monitor convergence using the maximum change in the parameter values across successive iterations. However, as is commonly known (McLachlan and Krishnan, 2008) the convergence of the EM algorithm is slow at low SNRs, especially when $\beta_0/\sigma < 2$). In order to speed up convergence, we employ the hybrid algorithm of Aitkin and Aitkin (1996) which alternates the EM iterations with those from a modified Newton-Raphson (NR) method where the Fisher information matrix is replaced by the empirical information matrix (Meilijson, 1989). The hybrid algorithm starts with five EM iterations before switching to the modified NR method until convergence of the parameters. In the latter case, we halve the NR step size up to five times.

Parameter updates from the modified Newton-Raphson method are given by

$$\boldsymbol{\tau}^{(k+1)} = \mathscr{I}_e^{-1}(\boldsymbol{\tau}^{(k)}; \boldsymbol{r})\boldsymbol{S}(\boldsymbol{r}; \boldsymbol{\tau}^{(k)}), \tag{9}$$

where $\mathscr{I}_e(\boldsymbol{\tau}; \boldsymbol{r})$ is the empirical information matrix and $\boldsymbol{S}(\boldsymbol{r}; \boldsymbol{\tau})$ is the score statistic. Both are constructed from the contributions to the score statistic at $t{=}p{+}1, p{+}2, \ldots, n$, denoted by $\boldsymbol{s}(r_t; \boldsymbol{\tau}) = \frac{\partial}{\partial \boldsymbol{\tau}} \log f(r_t | r_{t-1}, \ldots, r_{t-p}; \boldsymbol{\tau})$, which can be calculated from the complete data loglikelihood using the identity (adapted from McLachlan and Krishnan, 2008)

$$\boldsymbol{s}(r_t; \boldsymbol{\tau}) = \mathbb{E}_{\boldsymbol{\phi}|\boldsymbol{r};\boldsymbol{\tau}} \left[\frac{\partial}{\partial \boldsymbol{\tau}} \log f((r_t, \phi_t)|(r_{t-1}, \phi_{t-1}), \ldots, (r_{t-p}, \phi_{t-p}); \boldsymbol{\tau}) \right]. \tag{10}$$

These calculations, detailed in Section **??**, use quantities from the E-step. The empirical information matrix is calculated as

$$\mathscr{I}_e(\boldsymbol{\tau}; \boldsymbol{r}) = \sum_{t=p+1}^{n} \boldsymbol{s}(r_t; \boldsymbol{\tau})\boldsymbol{s}'(r_t; \boldsymbol{\tau}) - \frac{1}{n-p}\boldsymbol{S}(\boldsymbol{r}; \boldsymbol{\tau})\boldsymbol{S}'(\boldsymbol{r}; \boldsymbol{\tau}), \tag{11}$$

where $\boldsymbol{S}(\boldsymbol{r}; \boldsymbol{\tau}) = \sum_{t=p+1}^{n} \boldsymbol{s}(r_t; \boldsymbol{\tau})$.

3.3. *Calculation of test statistics under the MO AR(p) Ricean model.* We illustrate the calculations of Wald and likelihood ratio test statistics for a general test for activation, which posits $H_0 : \boldsymbol{C\beta} = \boldsymbol{0}$ against $H_a : \boldsymbol{C\beta} \neq \boldsymbol{0}$. Each test statistic is based on the MLEs $\widehat{\boldsymbol{\tau}}$ calculated by the above EM/NR hybrid algorithm.

3.3.1. *Wald test.* The Wald test statistic is given by

$$W = (\boldsymbol{C}\widehat{\boldsymbol{\beta}})' \left[\boldsymbol{C}\mathscr{I}_e^{-1}(\widehat{\boldsymbol{\tau}}; \boldsymbol{r})\boldsymbol{C}'\right]^{-1} (\boldsymbol{C}\widehat{\boldsymbol{\beta}}), \tag{12}$$

and asymptotically follows a null χ^2_m-distribution, where m is the rank of $\boldsymbol{C}$. It utilizes the empirical information matrix $\mathscr{I}_e$ of (11). However, our simulation studies

reported in Section **??** indicate that the Wald test statistic does not follow this null distribution for low-SNR time series, that is, when $\beta_0/\sigma < 2$. This shortcoming of the Wald test motivates the derivation of the likelihood ratio test (LRT) statistic described below.

3.3.2. *Likelihood ratio test for Ricean AR(1) model.* We derive an LRT for the Ricean AR(1) model, whose false positive rate better conforms with the significance level than the Wald test for low-SNR time series, as shown in Figure **??**. This LRT statistic is only for the Ricean AR(1) model, as the observed data loglikelihood is quite intractible for higher AR orders.[7] From standard results, the LRT statistic $\Lambda = 2[\ell(\widehat{\boldsymbol{\tau}}) - \ell(\tilde{\boldsymbol{\tau}})]$, where $\ell(\cdot)$ is the loglikelihood function $\log L(\cdot)$ and $\widehat{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{\tau}}$ represent the MLEs of $\boldsymbol{\tau}$ under H_a and H_0, respectively. Like the Wald statistic, the LRT statistic asymptotically follows a null χ^2_m-distribution. To derive the likelihood function $L(\boldsymbol{\tau}) = f(\boldsymbol{r};\boldsymbol{\tau})$ for the Ricean AR(1) model, we note that $f(\boldsymbol{r};\boldsymbol{\tau})$ can be factored as $f(r_1;\boldsymbol{\tau})\prod_{t=2}^{n} f(r_t|r_{t-1};\boldsymbol{\tau})$, where $f(r_1;\boldsymbol{\tau})$ is the Ricean PDF of (3). It can be shown (see Section **??**) that the conditional PDF $f(r_t|r_{t-1};\boldsymbol{\tau})$ is equal to

$$\frac{r_t}{\sigma^2} e^{C_0} \left[\mathbb{I}_0\left(\frac{r_{t-1}\mu_{t-1}}{\gamma_0}\right)\right]^{-1} \sum_{m=0}^{\infty} \omega_m \mathbb{I}_m(C_1)\mathbb{I}_m(C_2)\mathbb{I}_m(C_{12}), \tag{13}$$

where $C_0 = -[r_t^2 + \mu_t^2 + \alpha^2(r_{t-1}^2 + \mu_{t-1}^2) - 2\alpha\mu_{t-1}\mu_t]/(2\sigma^2)$, $C_1 = r_{t-1}(\mu_{t-1} - \alpha\mu_t)/\sigma^2$, $C_2 = r_t(\mu_t - \alpha\mu_{t-1})/\sigma^2$, and $C_{12} = \alpha r_{t-1}r_t/\sigma^2$. Also, in (13), $\omega_m = 1$ for $m = 0$ and $\omega_m = 2$ for $m \geq 1$. We caution that the number of terms needed for convergence of the sum in (13) increases with SNR, so calculation of this Ricean AR(1) loglikelihood becomes computationally prohibitive for high-SNR voxel time series. (See Section **??** for details.) In this case, we recommend use of the Wald statistic.

3.4. *Additional Methodological Issues.*

3.4.1. *Discussion of MO Ricean AR(p) model methodology.* The main purpose of developing the MOR model in this paper is to gain insight on the importance on the three model features vis-a-vis the comparison framework in Table 1. The methodology for the MOR model is less than straightforward due to the novel inclusion of AR(p) dependence to the Ricean model. For parameter estimation, the EM algorithm uses a Delta Method approximation in the E-step and a hybrid EM/NR algorithm at low SNRs. For testing, both a Wald and an LR test statistic were derived. Though the Wald test statistic may be derived for any AR order p, its

[7] Recall that the Wald test statistic can derived for any AR order.

false positive rate did not adhere to the significance level at low SNRs, which led to deriving the LR test for AR(1) time series.[8] We acknowledge that the strength of this methodology is not its simplicity nor its computation time. The latter is discussed in Section **??**, which shows that MOR model-based parameter estimation takes between 100 and 1000 times more computation time per voxel time series than the other models and is especially slow at low SNRs. It clearly is desirable to use the CVNS model; however, archival datasets are MO, and therefore our objective is to evaluate the more computationally expensive MOR model as an alternative to the currently used (simpler but inaccurate) MOG model.

3.4.2. *Choosing the order of the AR model.* We suggest a sequential testing approach for determining the AR order p. Starting with $k = 1$, and for increasing k, we posit $H_0 : p = k - 1$ vs. $H_a : p \geq k$ (or, in terms of the AR coefficients, $H_0 : \forall j \geq k, \alpha_j = 0$ vs. $H_a : \exists j \geq k : \alpha_j \neq 0$). The estimated AR order is then $\widehat{p} = k' - 1$, where k' is the first k in the sequence of tests for which H_0 is not rejected. An LRT statistic given by $2(\ell_k - \ell_{k-1})$, where ℓ_k is the optimized loglikelihood function for the AR(k) model, may be employed under the CVNS, CVS, and MOG models. For the MOR model, we use the Wald test statistic $\widehat{\alpha}_k^2 / \mathscr{I}_e^{-1}(\widehat{\boldsymbol{\tau}}, \boldsymbol{r})_{\alpha_k \alpha_k}$, where the denominator is the diagonal entry of the inverse empirical information matrix corresponding to α_k. From standard results, each test statistic (whether LRT or Wald) is asymptotically χ_1^2-distributed under $H_0 : p = k - 1$.

It can be shown that the significance level δ applied to each test controls "overdetecting" the order (*i.e.*, $\widehat{p} > p$) in the sense that $\delta = \mathbb{P}(\widehat{p} > p | \widehat{p} \geq p)$. See Adrian, Maitra and Rowe (2018, Section S-2.4) for a justification. Section **??** uses simulation to demonstrate that the sequential testing approach to detecting p gives similar results to approaches based on the AIC and BIC (Akaike, 1973; Schwarz, 1978). Note that we cannot use the latter approaches under the MOR model because the loglikelihood is not tractable for general p.

3.4.3. *Clarification of SNR and CNR.* Discussions of Ricean and complex-valued fMRI models often point to their advantages at low signal-to-noise ratios (SNRs). Thus, we seek to be clear about our definition of SNR and the definition of contrast-to-noise ratio (CNR) as well. While such definitions can be trivial in Gaussian-distributed magnitude-only models, they are less so for Ricean magnitude models or complex-valued models. Please see Section **??** for a discussion.

[8] It could be argued that the methodology was "overengineered" in that it was tested to work for any SNR down to zero, much lower than fMRI datasets in practice. As an example, even the "low-SNR" dataset introduced in Section 2 did not have low enough SNRs to require the hybrid algorithm or the LR test statistic.

3.4.4. *Phase-only activation.* Another benefit of utilizing the complex-valued data is that there exists activation information, also in the phase data. For instance, Rowe, Meller and Hoffman (2007) describe a phase-only data-based regression model for activation detection using the methods of Fisher and Lee (1992). We adapt this model to the context of this paper in Section **??**.

Our development in this section has laid the groundwork for our investigation of fMRI analyses using the four models. We now proceed with our evaluations.

4. Simulation experiments. As mentioned in Section 3.1, our purpose in comparing the four models is to evaluate the influence of three model features on model performance measures related to parameter estimation and activation detection. Specifically, the structure of Table 1 tells us that comparing the (1) (CVNS, CVS), (2) (CVS, MOR), and (3) (MOR, MOG) model pairs will provide isolated information about the impact of (1) *nonsphericity*, (2) *using twice the quantities*, and (3) the *Ricean magnitudes* features, respectively. We designed two sets of simulation experiments to make these comparisons in a setting of known truth. In the first set, we generated time series under the CVS model (therefore, also marginally under the MOR model), allowing us to make comparisons (2) and (3). The second set generated time series under the CVNS model and focused on comparison (1). In both cases, we used the $\boldsymbol{X}$ matrix of Section 2 so that our settings matched that of the finger-tapping experiment.

4.1. *Ricean modeling improves parameter estimates, activation detection.* We simulated low-SNR complex-valued time series under the CVS model — and therefore also magnitude time series under the MOR model — with white noise variance $\sigma^2 = 1$, AR(1) temporal dependence with AR parameters $\alpha = 0.2, 0.4, 0.6,$ and 0.8, baseline signal levels β_0 from 0.5 to 5.0, and activation levels $\beta_1 = 0.1, 0.2,$ and 0.3. For each combination of parameter values, we generated 10,000 voxel time series and fit each of the four models[9] under an assumed AR order of 1. Our model comparison analysis has two main components: properties of the parameter estimates and activation detection performance.

Figure 3 summarizes the properties of the parameter estimates, displaying the biases, standard errors (SEs), and Root Mean Squared Errors (RMSEs) of $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\sigma}^2$, and $\widehat{\alpha}$. Focusing first on the biases, it is immediately evident that the MOG model produces the most biased parameter estimates due to its Gaussian approximation of the truly *Ricean magnitudes*. Specifically, the biases of $\widehat{\beta}_0$ and $\widehat{\sigma}^2$ result from the mean and SD of the Rice distribution (which are the quantities that the

[9]The inclusion of the CVNS model only serves to assure that the "overfitting" involved in estimating ρ, σ_R^2, and σ_I^2 relative to the CVS model assumption that $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_2$ does not negatively affect model performance. (There seems to be very little, if any, effect, as the CVNS- and CVS-model-based results essentially coincide in all the figures presented here.)

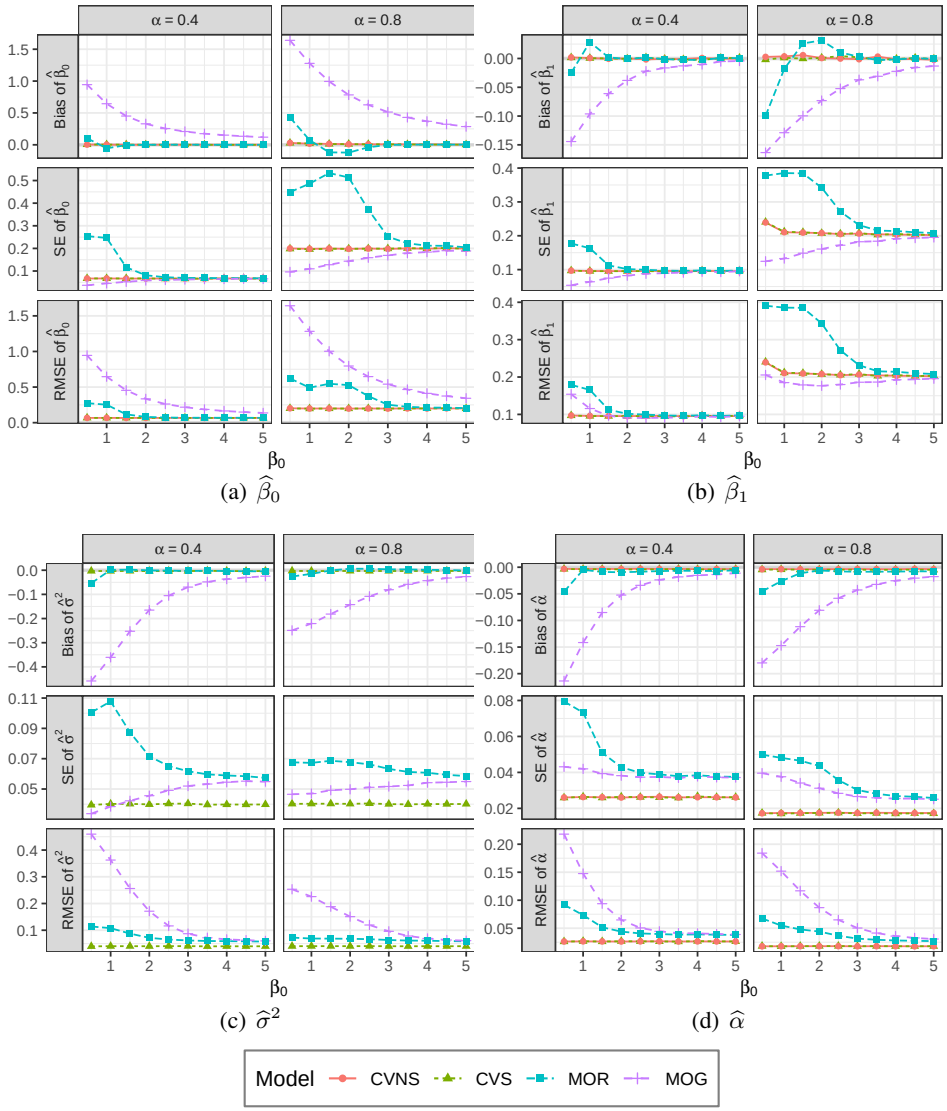


FIG 3. *Biases, SEs (SE), and RMSEs (RMSE) of (a)* $\widehat{\beta}_0$*, (b)* $\widehat{\beta}_1$*, (c)* $\widehat{\sigma}^2$*, and (d)* $\widehat{\alpha}$ *under the four models for the simulated time series under* $\beta_1 = 0.2$ *and different values of* β_0 *and* α*.*

MOG model estimates) being above and below the Ricean location and scale parameters, as shown in Figure **??**. (Bias of $\widehat{\beta}_1$ is also suggested by Figure **??** in the difference between the CNR and $\beta_1 = 0.2$ for $\alpha = 0$.) The MOG model biases decrease with increasing β_0 because the Gaussian approximation to the Rice distribution improves with SNR. These results match those already observed for the temporally independent case (Adrian, Maitra and Rowe, 2013; Solo and Noh,

2007), and, additionally, we see here that the MOG model-based estimate of the AR coefficient α is the most biased as well. We note that the MOR model-based parameter estimates are also biased for some cases where $\beta_0 \leq 2$, but much less so than their MOG model-based counterparts. The complex-valued model-based estimates, however, are unbiased in all cases. Switching our focus to the SEs, we note that the CV model-based SEs of $\widehat{\sigma}^2$ and $\widehat{\alpha}$ are lower than those for the MO models. In fact, the CV model-based SEs at $\beta_0 = 5$ are approximately a factor of $\sqrt{2}$ less than the MO SEs; this corresponds to a similar result observed in Rowe (2005b) about the sampling variances of $\widehat{\sigma}^2$ for the temporally independent case, suggesting that the *twice the quantities* feature is driving this difference. Overall, the RMSEs, which account for both bias and SE as $\text{RMSE}^2(\cdot) = \text{Bias}^2(\cdot) + \text{SE}^2(\cdot)$, are lowest for the complex-valued models (with the exception of $\widehat{\beta}_1$ at $\alpha = 0.8$) and are unrelated with β_0, suggesting that the CV models produce the most reliable parameter estimates (arguing against the current practice of discarding the phase data). However, given that archival datasets (of which there are massively many) do not have phase data stored, our results here also suggest the parameter estimation benefits of using the AR MOR model in place of the currently-used AR MOG model.

We focus next on comparing the activation detection performance of the models, calculating LRT statistics for the activation test of $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ for each simulated time series. To summarize the power of each LRT statistic, we calculated the partial area under the receiver operating characteristic curve or the pAUC (McClish, 1989; Zhou, Obuchowski and McClish, 2011). The pAUC is the area under the ROC curve — where the ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) — over a limited range of FPR values.[10] The rationale for using the pAUC rather than the (full) AUC, the area under the ROC curve for all FPR values from 0 to 1, is to exclude contributions to the curves from FPR values that are never used in practice, and to highlight differences between the methods in the part of the ROC that are most likely to be used in practice. For instance, using FPR values greater than 0.1 would allow for far too many false positives than practically ever used; indeed, our significance level threshold used on the real data in Section 5 is 0.001. Therefore, we calculated the pAUCs over a FPR range of 0 to 0.05. We calculated the pAUCs (separately for each combination of parameters β_0, β_1, and α) as the average of the TPRs for the significance levels $\delta = 0.0001, 0.0002, 0.0003, \ldots, 0.0500$; each TPR is the proportion of the 10,000 simulated test statistics greater than the $\chi^2_{1-\delta,1}$ quantile.

As shown in Figure 4, the pAUCs of simulated LRT statistics are consistently in the order (highest to lowest) of CVNS/CVS (which coincide), MOR, and MOG

[10]There are also pAUC versions that limit the TPR range (or both the FPR and TPR ranges), but we use a FPR-limited pAUC here.

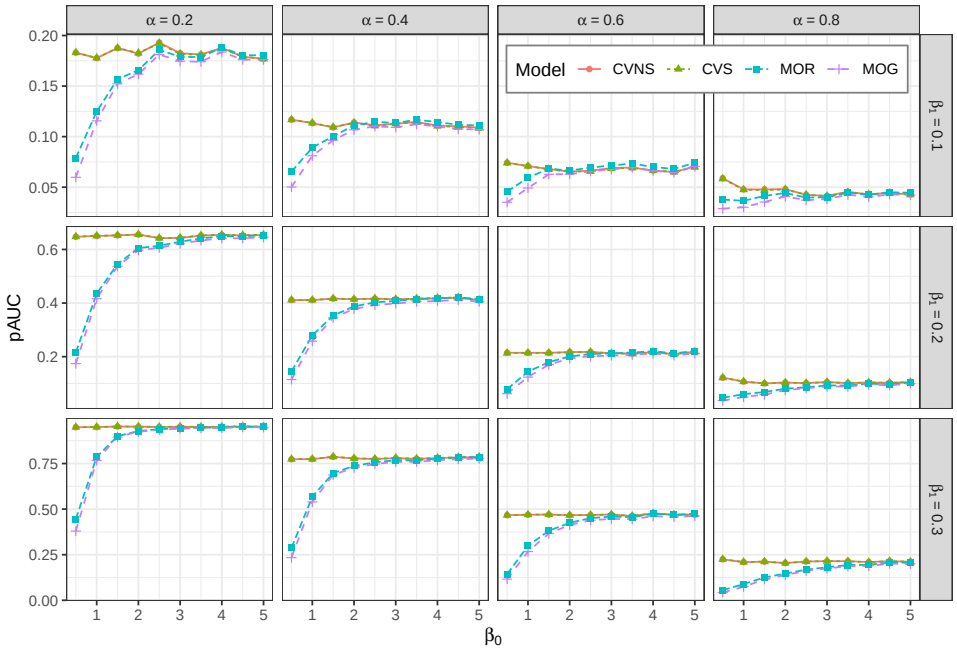


FIG 4. *The partial AUCs of the magnitude-only (MO) data-based LRT statistics decrease at low SNRs relative to their complex-valued (CV) data-based counterparts. Also, the Ricean (MOR) model-based pAUCs show improvements over those from the Gaussian (MOG) model.*

models. While the figure shows between-panel differences in the pAUCs due to the relationships between the CNR and the values of β_1 and α (positive and negative relationships, respectively), the within-panel patterns are quite similar. While the CV model-based pAUCs are constant as a function of β_0 (after accounting for simulation variability), the pAUCs of the MO model-based LRTSs decrease with β_0.

We attribute these differences to the model features in Table 1 as before. The *twice the quantities* feature of CV *vis-a-vis* MO seems to be driving the difference in detection performance: again, the MO data are missing the phases. The information in the phase data appears to be more valuable to the activation detection as β_0 (the SNR) decreases, as shown by the increasing discrepancy in CV/MO pAUCs. In fact, we may view this in terms of the "missing information principle" (Orchard and Woodbury, 1972). Section **??** describes how the missing information matrix comes from applying the EM algorithm framework of observed, missing, and complete data to the Fisher information matrix and may be computed under the CVS model. Interestingly, the curves of the "observed information" (in the MO data) and the "complete information" (in the CV data) in Figure **??** look similar to the pAUC curves in Figure 4. Both figures suggest that the phase data contain useful information about the activation, even though the activation itself occurs in

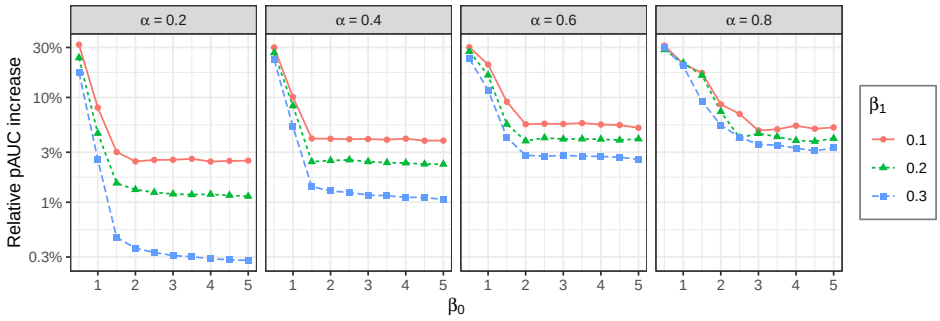


FIG 5. *Relative improvement in MOG model-based pAUCs by MOR model.*

the magnitude signal $\boldsymbol{X\beta}$.

In addition, the *Ricean magnitudes* lead to the increased MOR model-based pAUCs relative to the MOG model; after all, the MOR model uses the Rice distribution while the MOG model merely approximates it. Figure 5 focuses on the sizes of these improvements, displaying the per cent pAUC increases of the MOR model-based LRTSs relative to the MOG model, that is, (MOR - MOG) / MOG $\times$ 100%. We note that the sizes of MOR model improvements in pAUC increase as β_0 and β_1 decrease and α increases, which correspond to decreases in SNR and CNR. While it is not surprising that a lower SNR (with worse Gaussian approximation of the Ricean magnitudes) is associated with a larger MOR model improvement, it is worth noting that improvement is larger for lower activation levels (CNRs) as well.

To summarize, the results of our experiments suggest that modeling the *Ricean magnitudes* produces less-biased parameter estimates and better activation detection performance than the Gaussian approximation. Also, using *twice the quantities* in CV rather than MO data-based models leads to parameter estimates with less variation and additionally improved activation detection. Although using the CV data produces the best estimation and activation detection, utilizing the Ricean model when only magnitude (archival) data is available produces sizeable gains over the Gaussian approximation.

4.2. *Modeling real/imaginary covariance enhances activation detection.* Our other simulation experiment studied the influence of modeling the *nonsphericity* of $\boldsymbol{\Sigma}$, the covariance matrix of $(y_{Rt}, y_{It})'$. We generated time series from the CVNS model with various values of the real/imaginary correlation ρ and the real and imaginary variances σ_R^2 and σ_I^2, respectively; we set $\rho = 0, \pm 0.4$, and ± 0.8, and set the ratio $\sigma_R/\sigma_I = \{1.25^j : j \in \{-4, -3, \ldots, 4\}\}$, subject to the restriction that the average of the variances $(\sigma_R^2 + \sigma_I^2)/2$ was fixed at one. We also varied the central phase θ, with values 0 and $\pi/4$. The remaining parameters were held constant, with $\boldsymbol{\beta} = (5, 0.2)'$ and used a first-order AR process with coefficient $\alpha = 0.4$. As before, we generated 10,000 time series at each combination of parameter

values, calculated LRT statistics for activation, and calculated pAUCs to compare the activation detection performance of the four models.

Figure 6 shows these pAUCs: note that we exclude the CVS model here because

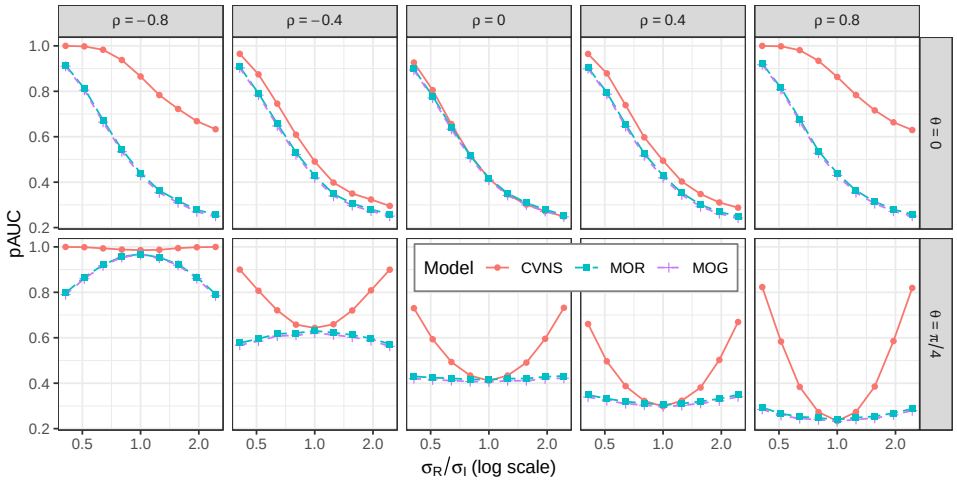


FIG 6. *The CVNS model-based pAUCs are higher than magnitude data-based pAUCs except when (1)* $\theta = 0$ *and* $\rho = 0$ *and (2)* $\theta = \pi/4$ *and* $\sigma_R/\sigma_I = 1$.

of its unreliable FPR (see Figure **??**). Thus, in CV data-based analyses, modeling the *nonpshericity* of $\mathbf{\Sigma}$ is essential for valid activation detection when $\mathbf{\Sigma}$ differs from $\sigma^2 \boldsymbol{I}_2$; MO data-based analyses do not face this issue (again see Figure **??**). However, the pAUCs (essentially, power) of the MO data-based analyses are substantially lower than their CVNS counterparts, as shown in Figure 6. In other words, the effect of modeling *nonsphericity* is improved activation detection. The size of improvement is related to the specification of θ and $\mathbf{\Sigma}$. When $\theta = 0$, there are large CVNS/MO differences when $|\rho| = 0.8$ but small differences when $\rho = 0$. When $\theta = \pi/4$, these differences increase as σ_R/σ_I gets farther from 1.[11]

We interpret these results by comparing θ, the direction of activation in the complex plane, to the constant density elliptical contours of the bivariate Gaussian distribution of $(y_{Rt}, y_{It})'$ provided by $\mathbf{\Sigma}$. The $(\theta, \mathbf{\Sigma})$ cases producing small differences are those where θ is "aligned" with the ellipses of $\mathbf{\Sigma}$; in other words, the ellipses are symmetric around θ. (A more precise interpretation of this "alignment" is that $(\cos\theta, \sin\theta)'$ is one of the eigenvectors of $\mathbf{\Sigma}$.) For $\theta = 0$, this occurs when $\rho = 0$ and the ellipses are aligned with the real and imaginary axes; for $\theta = \pi/4$, this is when $\sigma_R = \sigma_I$ and the ellipses are symmetric to the "45 degree" line.

To augment this geometric interpretation, a simulation study in Section **??** suggests that the superior CVNS-model-based pAUCs may be linked to correlations

[11]The pAUC values across different panels are influenced by varying CNRs. For instance, for $\theta = \pi/4$, the pAUCs are higher for negative ρs than for positive ρs due to the dependence of CNR on ρ displayed in Figure **??**.

between the magnitude and phase time series. That is, the cases where the CVNS and MO model-based pAUCs are the closest, that is, $\theta = 0$ with $\rho = 0$ and $\theta = \pi/4$ with $\sigma_R = \sigma_I$, are also those with zero correlation between r_t and ϕ_t (compare Figures 6 and **??**). Furthermore, the size of the CVNS model-based pAUC improvement seems to be connected to the size of the magnitude/phase correlation. Thus, we can postulate that when this correlation exists, the phase data provides important additional information about the magnitude-related activation.

In cases where the CV data are not available, such as, for archival data, using the MOR model instead of the MOG model on the MO data also produces higher pAUCs. Figure 7 shows these per cent pAUC gains. We note that the larger improvements in pAUC are for $\mathbf{\Sigma}$ values which produce lower-SNR time series – such as larger values of σ_R/σ_I when $\theta = 0$, for which the Gaussian approximation of the *Ricean magnitudes* is worse.

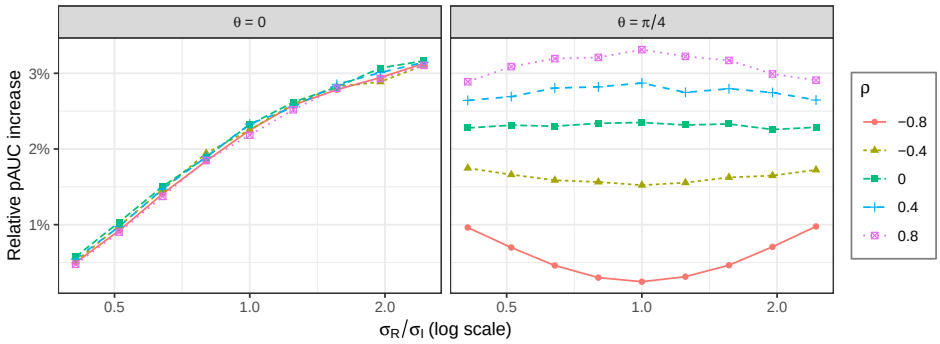


FIG 7. *Relative improvement in pAUCs upon using a MOR model instead of a MOG model.*

5. Results on low-signal fMRI dataset. In this section, we identify activation in the low-SNR fMRI dataset introduced in Section 2. Recall that there are five separate datasets to consider due to the different levels of spatial smoothing applied following the removal of the magnitude and phase temporal trends: the spatially unsmoothed dataset and datasets smoothed with FWHMs of 1.5, 2, 4, and 6 voxels. For each of these datasets, we compare the performance of the four models identified in Table 1. Working with each voxel time series of each dataset separately, we first detected the AR orders (see Figure **??**) and then tested for activation using $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$. We used LRT statistics for the CVNS, CVS, MOG, and AR(0) (independence in time series) MOR models and Wald test statistic for the AR(p) MOR model for $p \geq 1$, obtaining p-values from these test statistics under the null χ^2_1-distribution. (Note that the SNRs in the dataset were above the range of concern for the MOR model-based Wald test statistic that was discussed in Section 3.3.) To determine activation, we used a significance level threshold of 0.001 (Woo, Krishnan and Wager, 2014). For the four smoothed datasets, we also utilized

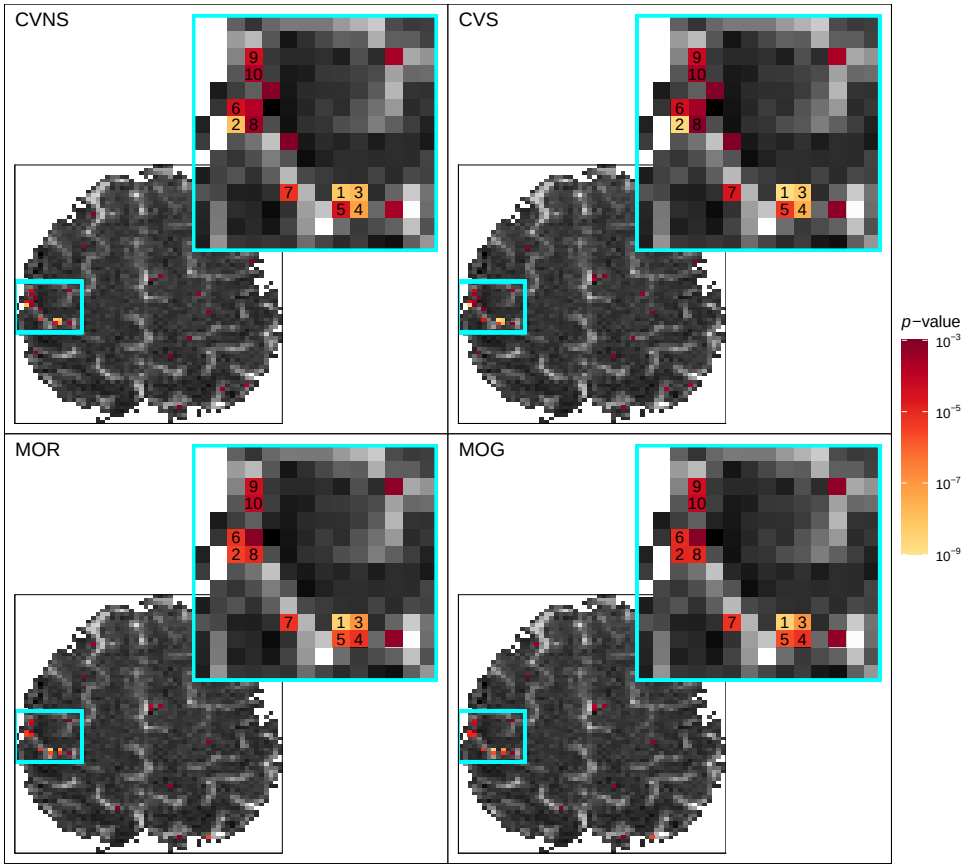
cluster thresholding with a third-order neighborhood structure and a size threshold of 11 voxels (Forman *et al.*, 1995). However, because cluster thresholding removed all activation for the unsmoothed dataset, we did not use it for determining activation (only for the unsmoothed dataset case). Thus, we present the results for the unsmoothed dataset and the smoothed datasets separately in what follows.

5.1. *Activation detection in unsmoothed data.* We display activation maps of the second slice[12] according to each of the four models in Figure 8(a). Each map shows a grayscale anatomical image (the magnitude image at the first time point) with the voxels having test statistics with p-values less than 0.001 colored according to the legend. The activation is rather sparse except for a region containing the left central sulcus (Figure 1), so we focus on this region of interest (ROI) in the inset maps. Recall that the central sulcus is the site guiding hand function for normal healthy adults. Specifically, we focus on the ten voxels with test statistics that provide the lowest minimum p-value over the four models, which are identified by numerals (from 1 = lowest to 10 = highest p-value) in the inset maps of Figure 8(a). This numbering carries over to Figure 8(b) where the size of the p-values can be compared more clearly across models. It is interesting to note that the CV models provide lower p-values than the MO models for three of the four voxels with the lowest p-valued test statistics (numbered 2, 3, and 4), though this phenomenon does not hold for all ten voxels. Because the SNRs are lowest for the unsmoothed data, we may attribute these lower CV model-based p-values to the low SNR and the simulation-based results seen in Figure 4.

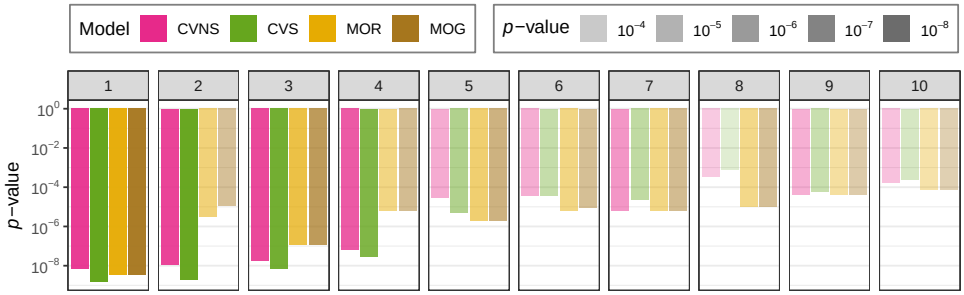
To further investigate the effect that low SNR has on activation detection, we added more noise to the acquired CV data. So, for the ten voxels identified in Figure 8(a) with acquired CV time series (y_{Rt}, y_{It}), $t = 1, \ldots, 621$, we obtained new synthetic data $y^*_{Rt} = y_{Rt} + w_{Rt}$ and $y^*_{It} = y_{It} + w_{It}$, with $w_{Rt}, w_{It} \sim$ iid $N(0, \sigma_a^2)$. Because a representative, data-based estimate of the noise SD is 0.15, we generated data using $\sigma_a = 0.15, 0.20, 0.25$, and 0.30. Adding noise to the original data in this manner reduces the SNR and CNR (see Figure **??**) as would occur with increasing spatial resolution of the MR scan. Figure 9 shows the proportions of the 10,000 generated datasets in which each of the ten voxels was detected under each model at the 0.001 significance level. Overall, the average detection rates (shown in the rightmost column) for the CV models are higher than the MO models across all the different added noise SDs. In addition, the MOR model shows greater detection rates than the MOG model.

5.2. *Activation detection in smoothed data.* Figure 10 has the activation maps of the second slice for the datasets that were spatially smoothed with FWHMs of

[12]We focus on the second slice because it shows the strongest activation. See Figure **??** for activation maps of the other slices.



(a) Activation maps with inset maps showing left central sulcus ROI; numbered voxels represent those having test statistics with the lowest p-values.



(b) The p-values of the test statistics for each model for the numbered voxels in (a).

FIG 8. *(a) Activation maps (slice 2) and (b) p-values from applying the four models to the ten voxels of the unsmoothed dataset.*

FIG 9. *Proportion of replications the 10 voxels identified in Figure 8(a) (columns) detected as activated when noise with SD σ_a (rows) was added to the raw data. (The rightmost column shows the average proportions over the 10 voxels.)*

1.5, 2, 4, and 6 voxels. Even though these smoothed data maps employ the additional cluster thresholding criteria not used on the unsmoothed data, it is immediately evident that these maps show much more activation. Moreover, the number of activated voxels increases as the amount of smoothing increases; this speaks to the increase in CNR with smoothing, as shown in Figure **??**. However, this greater CNR comes at the expense of spatial specificity. The left central sulcus (refer to Figure 1) is best identified by the data smoothed with FWHM = 1.5 or 2 voxels. Meanwhile, it could be argued that the activation regions in the maps based on FWHM = 4 and 6 voxels are too diffuse to be useful in practice.

The model-wise differences between the activation maps at each FWHM are much less pronounced. In fact, all four model-based maps are virtually indistinguishable at FWHMs of 1.5 and 2 voxels. For the maps based on smoothing at FWHMs of 4 and 6 voxels, the MO data-based maps show more activation than the CV data-based maps.[13] At first glance, these model comparisons may seem puzzling, given that the simulation experiments of Section 4 showed improved activation detection for the CV data-based methods over their MO data-based counterparts and better MOR model-based detection relative to the MOG model. In the following, we present several factors that we believe led to these results.

Perhaps most importantly, the SNR ranges of the smoothed datasets were higher

[13]These activation patterns by smoothing level and model are similar to those seen in the maps for other slices of the dataset, as shown in Figure **??**.

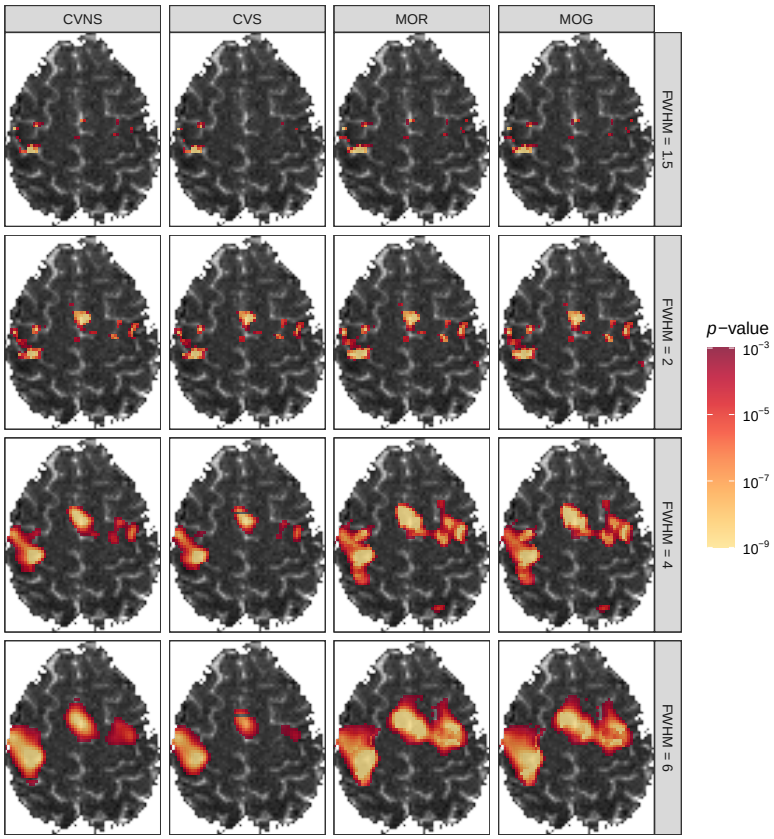


FIG 10. *Activation maps of the second slice calculated under the different models (columns) and smoothing levels (rows).*

than the SNRs showing the largest differences in the simulation experiments. Figure **??** shows that the SNRs are mostly above 8 for the data smoothed at FWHM = 1.5 voxels and the SNRs increase with wider smoothing kernels. In contrast, the CV vs. MO and MOR vs. MOG model-based pAUC gains decrease rapidly as β_0/σ values[14] increase in the region with $\beta_0/\sigma < 3$ (see Figures 4 and 5); for $\beta_0/\sigma \geq 3$, the pAUC gains are more modest, perhaps not large enough to make a visual difference on activation maps. This suggests that CV and MOR modeling will make more difference on images that are not spatially smoothed.

Additionally, we argue that the difference in the CV/MO maps for smoothing with FWHM = 4 and 6 voxels is related to differences in detected AR orders under the CV and MO data-based models. The detected orders, as shown in Figure **??**, strongly depended upon whether they were based on the CV or the MO data, with CV data-based models tending to find larger orders. A simulation study in

[14] And $\beta/\sigma <$ SNR in the presence of AR dependence.

Adrian, Maitra and Rowe (2018, Section 5.3) supports these results by demonstrating greater CV data-based order detection power and attributes it to *using twice the quantities* of MO data. This study also demonstrates that underdetecting the AR order (*i.e.*, finding $\widehat{p} < p$) can lead to more false positives in activation detection. To examine whether this phenomenon was behind the increased activation for the MO data-based maps relative to the CV maps, we set $\widehat{p} = 1$ for all voxels, leveling the order detection "playing field". The resulting maps shown in Figure **??** are notably more similar.

After accounting for the effect of order detection, we further interpret the activation maps – *i.e.*, now focusing on the AR(1) maps in Figure **??** – in terms of the CVNS model parameter MLEs (see Section **??**). For images smoothed at FWHMs of 1.5 and 2 voxels, the distributions of $\widehat{\rho}$ and $\widehat{\sigma}_R/\widehat{\sigma}_I$ are close to 0 and 1, respectively, suggesting that the CVS and MOR models adequately fit the data; this supports the model-wise similarities of these maps. However, greater levels of smoothing induce real/imaginary correlation and clustering of voxels where $\widehat{\sigma}_R \neq \widehat{\sigma}_I$. With spherical covariance becoming a less tenable model assumption, the unreliable CVS model false positive rate may account for the differences between the CVS model and the other three maps. Last, despite this nonsphericity, the CVNS model-based map does not clearly show greater activation than its MO counterparts (though it could be argued that some regions show higher CVNS p-values); with the distribution of $\widehat{\sigma}_R/\widehat{\sigma}_I$ centered around 1, the data, based on a central phase of $\theta_0 = \pi/4$, appear to be characterized by the lack of magnitude/phase correlation associated with the superior activation of the CVNS model (see Figure 6).

5.3. *Phase-only activation.* Section **??** displays the activation maps obtained when we applied the phase-only data-based methodology described in Section **??**. The resulting activation maps show a task-related phase change detected in the left central sulcus (in addition to the previously displayed magnitude changes), and it is the primary activation region identified.

6. Discussion. In this paper, we developed a Ricean model for fMRI magnitude time series that incorporates autoregressive time dependence. Our approach applies AR(p) errors to the Gaussian-distributed real and imaginary components from which the magnitudes are computed. We estimated model parameters from the MO data using the EM algorithm with the phase portion of the latent complex-valued data playing the role of missing data. We then extended the EM algorithm to derive Wald and LRTs for activation and AR order detection.

We compared this AR(p) Ricean model to three other models within the framework established in Section 3.1. Roughly, we expected the performance of the Ricean model to fall below that of the two CV data-based models (due to the CV data having *twice the quantities*) and above that of the MOG model (due to model-

ing the *Ricean magnitudes*). This was the case for simulated data: the MOG model produced the most biased parameter estimates, and the CV data-based parameter estimates showed the lowest RMSEs. Also, the CV data-based activation tests showed larger pAUCs than the two MO data-based tests, with the MOR model-based test producing the higher pAUCs of the two MO tests. Whether considering parameter estimates or activation detection, differences between the models increased with decreasing SNR.

The two CV data-based models differed in whether they estimated the *nonsphericity* of the real/imaginary covariance matrix. Simulations indicated that activation detection based on the nonspherical model outperformed its spherical counterpart (and the MO data-based models) when the central phase direction differed from both eigenvectors of the real/imaginary covariance matrix, which is the case when magnitude and phase time series are correlated. In addition, the test based on sphericity was not even reliable in terms of false positive rate when nonsphericity was present.

Further, we tested our methodology on a finger-tapping experiment, acquiring the data using the body coil to study a low-SNR dataset. Differing levels of spatial smoothing led to datasets with differing SNRs. For the unsmoothed dataset with the lowest SNR, the CV models produced lower p-values for activation than the MO models for three of four voxels in the right central sulcus region displaying the strongest activation. Further, we were able to consistently demonstrate improved activation detection by the CV models when we added extra simulated white noise to the experimental voxel time series; and in this case, the Ricean model showed better performance than the MOG model as well. For the smoothed datasets with higher SNR, the activation maps produced by the different models are similar to a large extent, with dissimilarities that may be attributed to the different AR orders selected by different models.

Future directions for this research can explore fMRI time series models that allow for activation in magnitude and phase (Rowe, 2005a). The models we have focused upon in this article (even the CV data-based ones) have assumed task-related magnitude changes. We touched upon a phase-only data-based model for task-related phase changes in Section 3.4.4; it may be worth exploring a single model that allows for both task-related magnitude and phase changes. In addition, it may also be worth expanding the modeling of Ricean time series presented here to χ-distributed time series, as discussed in Wegmann, Eklund and Villani (2017).

REFERENCES

AALO, V. A., EFTHYMOGLOU, G. P. and CHAYAWAN, C. (2007). On the Envelope and Phase Distributions for Correlated Gaussian Quadratures. *IEEE Communications Letters* **11** 985-987.

ABRAMOWITZ, M. and STEGUN, I. (1965). *Handbook of Mathematical Functions*. Dover Publications.

ADRIAN, D. W., MAITRA, R. and ROWE, D. B. (2013). Ricean over Gaussian modeling in magnitude fMRI analysis – Added complexity with neglible practical benefits. *Stat* **2** 303-316.

ADRIAN, D. W., MAITRA, R. and ROWE, D. B. (2018). Complex-valued time series modeling for improved activation detection in fMRI studies. *Annals of Applied Statistics* **12** 1451-1478.

AITKIN, M. and AITKIN, I. (1996). A hybrid EM/Gauss-Newton algorithm for maximum likelihood in mixture distributions. *Statistics and Computing* **6** 127-130.

AKAIKE, H. (1973). *2nd International Symposium on Information Theory* Information theory and an extension of the maximum likelihood principle 267-281. Akademiai Kiado.

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.

BROWN, T. R., KINCAID, B. M. and UGURBIL, K. (1982). NMR chemical shift imaging in three dimensions. In *Proceedings of the National Academy of Sciences, USA* **79** 3523-3526.

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.

CARPENTER, M. B. (1991). *Core Text of Neuroanatomy*. Williams and Wilkins.

CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*, Second edition ed. Thomson Learning.

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.

FISHER, N. I. and LEE, A. J. (1992). Regression Models for an Angular Response. *Biometrics* **48** 665-677.

FORMAN, S. D., COHEN, J. D., FITZGERALD, M., EDDY, W. F., MINTUN, M. A. and NOLL, D. C. (1995). Improved Assessment of Significant Activation in Functional Magnetic Resonance Imaging (fMRI): Use of a Cluster-Size Threshold. *Magnetic Resonance in Medicine* **33** 636-647.

FRISTON, K. J., JEZZARD, P. and TURNER, R. (1994). Analysis of functional MRI time-series. *Human Brain Mapping* **1** 153-171.

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., JOSEPHS, O., ZARAHN, E., HOLMES, A. P., ROUQUETTE, 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. (2000). A Bayesian Time-Course Model for Functional Magnetic Resonance Imaging Data. *Journal of the American Statistical Association* **95** 691–703. Available at http://www.jstor.org/stable/2669445

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.

GLISSON, T. H. (2011). *Introduction to Circuit Analysis and Design*. Springer, The Netherlands.

GLOVER, G. H. (1999). Deconvolution of impulse response in event-related BOLD fMRI. *NeuroIm-*

age **9** 416-429.

GUDBJARTSSON, H. and PATZ, S. (1995). The Rician distribution of noisy data. *Magnetic Resonance in Medicine* **34** 910-914.

HAHN, A. D., NENCKA, A. S. and ROWE, D. B. (2009). Improving robustness and reliability of phase-sensitive fMRI analysis using temporal off-resonance alignment of single-echo timeseries (TOAST). *NeuroImage* **44** 742-752.

HAHN, A. D., NENCKA, A. S. and ROWE, D. B. (2012). Enhancing the Utility of Complex-Valued Functional Magnetic Resonance Imaging Detection of Neurobiological Processes Through Postacquisition Estimation and Correction of Dynamic B_0 Errors and Motion. *Human Brain Mapping* **33** 288-306.

HOOGENRAD, F. G., REICHENBACH, J. R., HAACKE, E. M., LAI, S., KUPPUSAMY, K. and SPRENGER, M. (1998). In vivo measurement of changes in venous blood-oxygenation with high resolution functional MRI at .95 Tesla by measuring changes in susceptibilty and velocity. *Magnetic Resonace in Medicine* **39** 97-107.

JAIN, A. K. (1989). *Fundaments of Digital Image Processing*. Prentice Hall.

JESMANOWICZ, A., WONG, E. C. and HYDE, J. S. (1993). Phase Correction for EPI Using Internal Reference Lines. In *Proceedings from the International Society of Magnetic Resonance in Medicine* **12** 1239.

JEZZARD, P. and CLARE, S. (2001). Principles of nuclear magnetic resonance and MRI. In *Functional MRI: An Introduction to Methods* (P. Jezzard, P. M. Matthews and S. M. Smith, eds.) 3 67-92. Oxford University Press.

KUMAR, A., WELTI, D. and ERNST, R. R. (1975). NMR Fourier Zeugmatography. *Journal of Magnetic Resonance* **18** 69-83.

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.

LEE, C. C., JACK, C. R. and RIEDERER, S. J. (1998). Mapping of the Central Sulcus with Functional MR: Active versus Passive Activation Tasks. *American Journal of Neuroradiology* **19** 847-852.

LEE, C. C., WARD, H. A., SHARBROUGH, F. W., MEYER, F. B., MARSH, W. R., RAFFEL, C., SO, E. L., CASCINO, G. D., SHIN, C., XU, Y., RIEDERER, S. J. and JACK, C. R. (1999). Assessment of Functional MR Imaging in Neurosurgical Planning. *American Journal of Neuroradiology* **20** 1511-1519.

LJUNGGREN, S. (1983). A Simple Graphical Representation of Fourier-Based Imaging Methods. *Journal of Magnetic Resonance* **54** 338-343.

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.

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.

MCCLISH, D. K. (1989). Analyzing a Portion of the ROC Curve. *Medical Decision Making* **9** 190-195. Available at https://pubmed.ncbi.nlm.nih.gov/2668680/

MCLACHLAN, G. J. and KRISHNAN, T. (2008). *The EM Algorithm and Extensions*. 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.

MENON, R. S. (2002). Postacquisition suppression of large-vessel BOLD signals in high-resolution fMRI. *Magnetic Resonace in Medicine* **47** 1-9.

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

NENCKA, A. S., HAHN, A. D. and ROWE, D. B. (2008). The Use of Three Navigator Echoes in Cartesian EPI Reconstruction Reduces Nyquist Ghosting. In *Proceedings from the International Society of Magnetic Resonance in Medicine* **16** 3032.

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.

ORCHARD, T. and WOODBURY, M. A. (1972). A missing information principle: theory and applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **1** 697-715. University of California Press, Berkeley, California.

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

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

ROWE, D. B. (2005a). Modeling both the magnitude and phase of complex-valued fMRI data. *NeuroImage* **25** 1310-1324.

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

ROWE, D. B. (2016). *Handbook of Neuroimaging Data Analysis* Image Reconstruction in Functional MRI 205-232. Chapman & Hall/CRC.

ROWE, D. B. and LOGAN, B. R. (2004). A complex way to compute fMRI activation. *NeuroImage* **23** 1078-1092.

ROWE, D. B., MELLER, C. P. and HOFFMAN, R. G. (2007). Characterizing phase-only fMRI data with an angular regression model. *Journal of Neuroscience Methods* **161** 331-341.

RUMEAU, C., TZOURIO, N., MURAYAMA, N., PERETTI-VITON, P., LEVRIER, O., JOLIOT, M., MAZOYER, B. and SALAMON, G. (1994). Location of Hand Function in the Sensorimotor Cortex: MR and Functional Correlation. *American Journal of Neuroradiology* **15** 567-572.

SCHWARZ, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics* **6** 461-464.

SHUMWAY, R. H. and STOFFER, D. S. (2006). *Time Series Analysis and Its Applications*, Second Edition ed. Springer.

SOLO, V. and NOH, J. (2007). An EM algorithm for Rician fMRI activation detection. In *ISBI* 464-467.

TANABE, J., MILLER, D., TREGELLAS, J., FREEDMAN, R. and MEYER, F. G. (2002). Comparison of Detrending Methods for Optimal fMRI Preprocessing. *Neuroimage* **15** 902–907. https://doi.org/10.1006/nimg.2002.1053

TRISTÁN-VEGA, A., AJA-FERNÁNDEZ, S. and WESTIN, C.-F. (2012). Least squares for diffusion tensor estimation revisited: Propagation of uncertainty with Rician and non-Rician signals. *NeuroImage* **59** 4032-4043. Copyright - Copyright Elsevier Limited Feb 15, 2012; Last updated - 2016-11-01. Available at http://search.proquest.com.ezproxy.gvsu.edu/docview/1834299434?accountid=39473

TWEIG, D. B. (1983). The k-trajectory formulation of the NMR imaging process with applications in analysis and synthesis of imaging methods. *Medical Physics* **10** 610-21.

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.

WEGMANN, B., EKLUND, A. and VILLANI, M. (2017). Bayesian Rician Regression for Neu-

roimaging. *Frontiers in Neuroscience* **11**.

WOO, C.-W., KRISHNAN, A. and WAGER, T. D. (2014). Cluster-extent based thresholding in fMRI analyses. *NeuroImage* **91** 412-419.

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.

YU, C.-H., PRADO, R., OMBAO, H. and ROWE, D. (2018). A Bayesian Variable Selection Approach Yields Improved Detection of Brain Activation from Complex-Valued fMRI. *Journal of the American Statistical Association* **113** 1395-1410.

ZHOU, X. H., OBUCHOWSKI, N. A. and MCCLISH, D. K. (2011). *Statistical methods in diagnostic medicine*, Second edition ed. Wiley.

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.

DEPARTMENT OF STATISTICS
GRAND VALLEY STATE UNIVERSITY
ALLENDALE, MI 49401-9403
USA
E-MAIL: adriad1@gvsu.edu

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IA 50011-1090
USA
E-MAIL: maitra@iastate.edu

DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES
MARQUETTE UNIVERSITY
MILWAUKEE, WI 53233
USA
E-MAIL: daniel.rowe@marquette.edu

Submitted to the Annals of Applied Statistics

SUPPLEMENT TO "RICE-DISTRIBUTED AUTOREGRESSIVE TIME SERIES MODELING OF MAGNITUDE FUNCTIONAL MRI DATA"

BY DANIEL W. ADRIAN[‡] AND RANJAN MAITRA[*,§] AND DANIEL B. ROWE[†,¶]

Grand Valley State University[‡] and Iowa State University[§] and Marquette University[¶]

S-1. Supplement to Section ?? – Further description of the dataset. In addition to the spatial structure emphasized in the images of Figure **??**, Figure S-1 shows plots of the real, imaginary, magnitude, and phase time series at a single voxel, which happens to be one of the voxels showing the most activation.

As discussed in Section **??**, using detrending to correct for scanner drift is an important preprocessing step. Figure S-2 shows a plot of the magnitude and phase time series for a selected voxel and compares the fit from four curve fitting methods. One is the "CV running line" method introduced in Adrian, Maitra and Rowe (2018) that fits linear models to both the magnitude and phase time series (simultaneously), using a moving window that only considers time points within 64 seconds of the fitted time. The other three are more established curve fitting methods: a polynomial of degree 8, a natural cubic spline with 6 evenly spaced knots, and a smoothing spline with 8 effective degrees of freedom (Hastie, Tibshirani and Friedman, 2009). These last three methods are all based on fitting separate curves to the real and the imaginary time series. As shown in Figure S-2(a), the four methods all capture changes in the global nonlinear trend well. However, a closer look in Figure S-2(b) shows some differences. For one, the CV running line does not produce a smooth curve, which may introduce additional variance to the detrended time series. Overall, it appears that the smoothing spline is the "Goldilocks' choice" as it seems to do the best in terms of not taking the largest or smallest fitted value across all time points for both time series. (This property seems to hold across a variety of other voxel time series as well.)

For clarity, we formally describe the detrending process using notation. First, for the real and imaginary data, y_{Rt} and y_{It}, at a single voxel and time t, we use one of the four methods to obtain the fitted trend values $\widehat{y}_{Rt}$ and $\widehat{y}_{It}$. The detrended values are then calculated as $\check{y}_{\xi t} = y_{\xi t} - \widehat{y}_{\xi t} - \bar{\widehat{y}}_{\xi}$, for $\xi = R, I$, where $\bar{\widehat{y}}_{\xi}$ is the mean of the fitted values across the corresponding time series.[1] To produce a zero

[*]Research supported in part by the the National Science Foundation CAREER Grant # DMS-0437555 and the National Institutes of Health (NIH) awards #R21EB016212 and #R21EB034184.

[†]Research supported in part by the National Institutes of Health (NIH) award #R21NS087450.

[1]The same approach applied to magnitude-only data produces the magnitude-only detrended data.

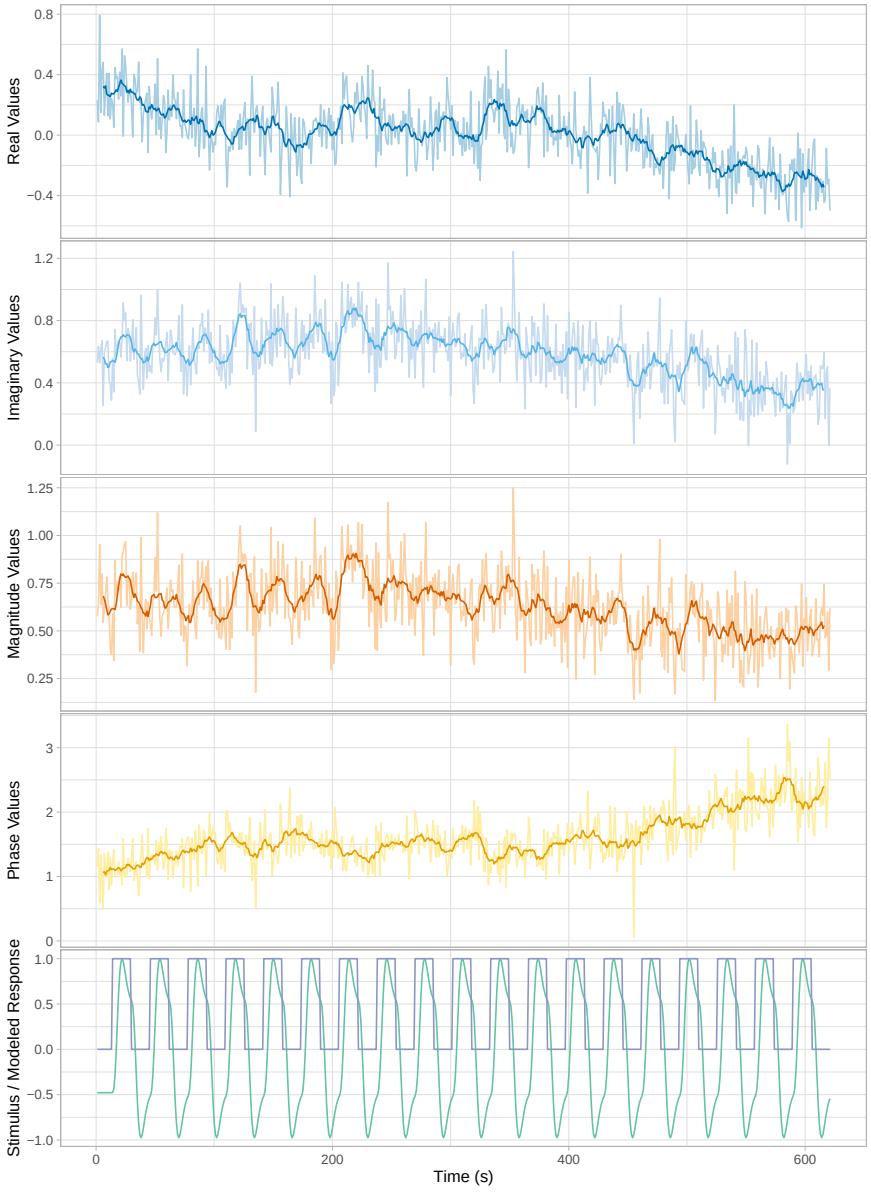
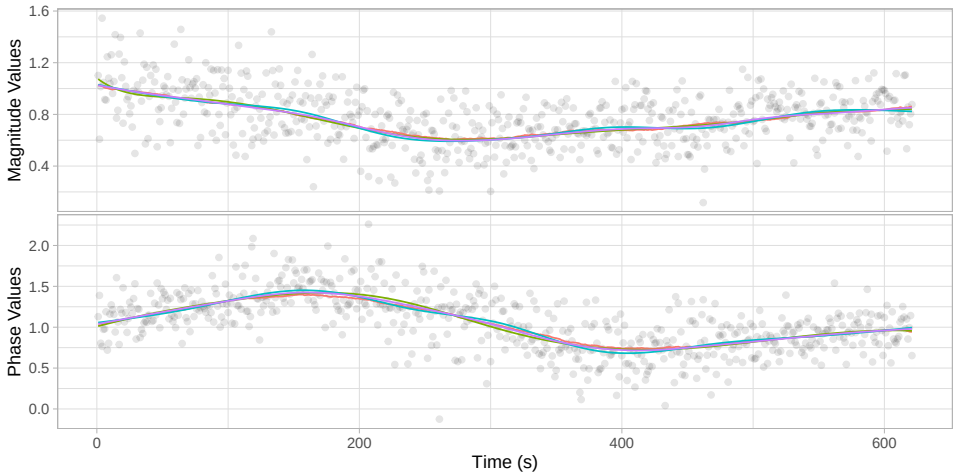
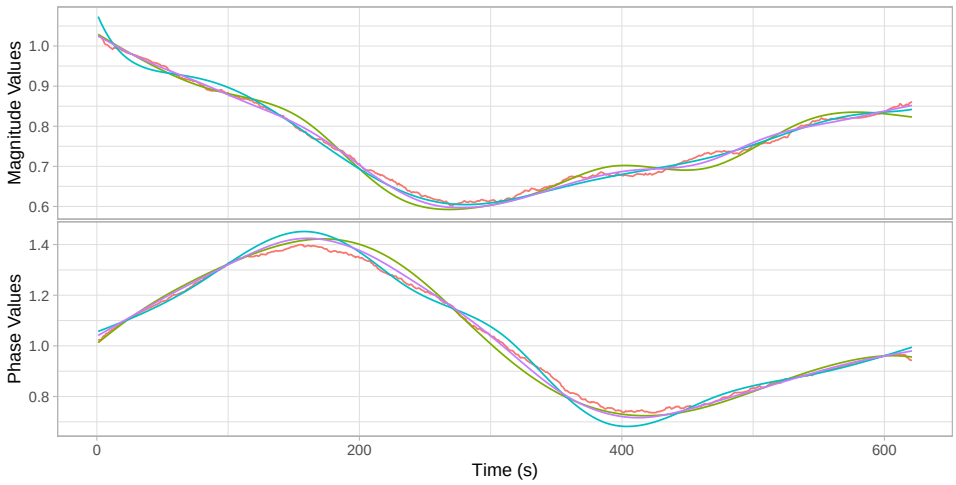


FIG S-1. *Time series of (from top) the real, imaginary, magnitude, and phase observations for one of the voxels showing the most activation. Lighter lines in each display represents the raw time series, while the darker lines show the result after applying a simple, central moving average filter with 5 nearest neighbors. The bottom display is of the 0/1 block design of the stimulus superimposed with the stimulus/HRF convolution, after zero-centering and unit scaling.*

mean phase for the time series, we calculate the detrended magnitude $\check{r}_t = (\check{y}_{Rt}^2 + \check{y}_{It}^2)^{1/2}$ and phase $\check{\phi}_t = \arctan_4(\check{y}_{It}, \check{y}_{Rt})$. Then mean phase is then calculated as



(a) Raw data and fitted curves shown



(b) Only fitted curves shown

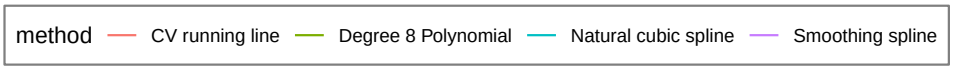


FIG S-2. *Comparison of four curve fitting methods on a selected voxel's time series for the purpose of detrending to correct scanner drift: CV running line (Adrian, Maitra and Rowe, 2018), degree 8 polynomial, natural cubic spline with 6 evenly spaced knots, smoothing spline with 8 effective degrees of freedom.*

$\bar{\check{\phi}} = \arctan_4(\bar{S}, \bar{C})$, where $\bar{S} = (1/n)\sum_{t=1}^{n} \sin \check{\phi}_t$ and $\bar{C} = (1/n)\sum_{t=1}^{n} \cos \check{\phi}_t$ (Mardia and Jupp, 2000). The final, zero mean phase, detrended time series are then calculated as $\tilde{y}_{Rt} = \check{r}_t \cos(\check{\phi}_t - \bar{\check{\phi}})$ and $\tilde{y}_{It} = \check{r}_t \sin(\check{\phi}_t - \bar{\check{\phi}})$.

Figure S-3 shows images of the SNRs and CNRs at the different levels of spatial

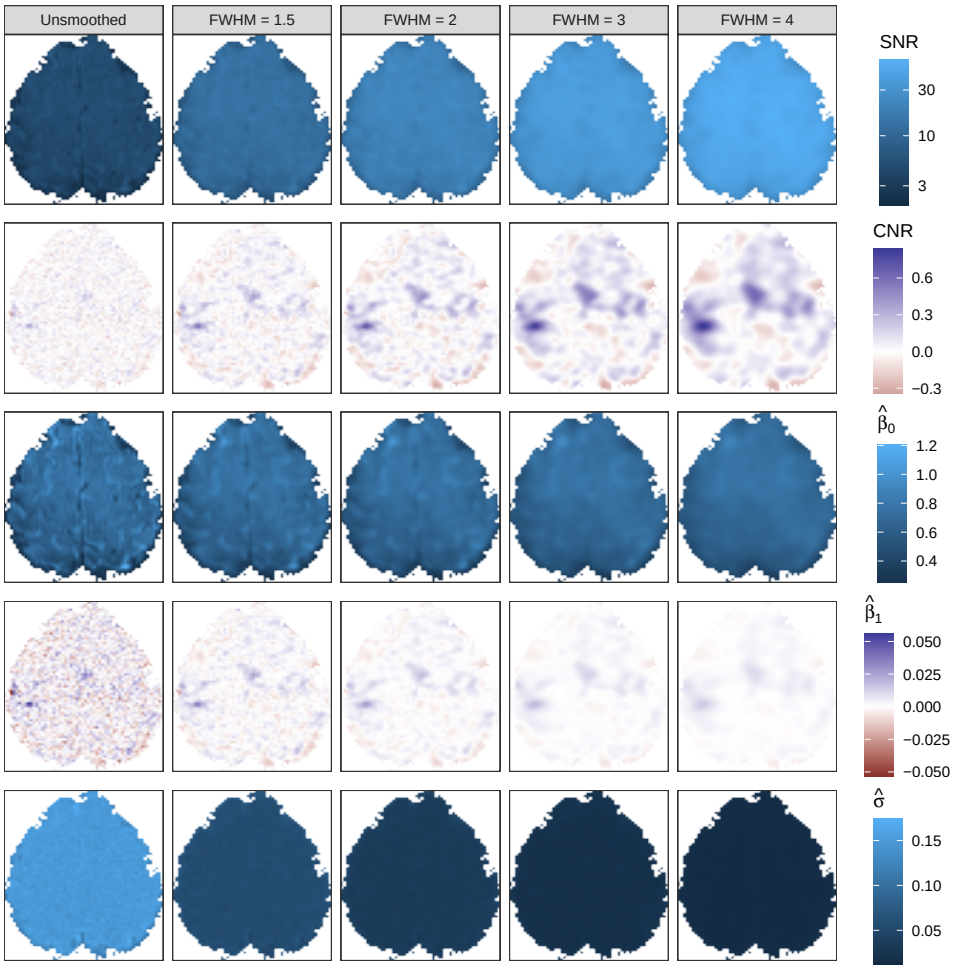


FIG S-3. *Images of the SNRs, CNRs, and the estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$, and $\widehat{\sigma}$ at each level of spatial smoothing (columns) for the second slice.*

smoothing. The SNRs and CNRs above were calculated as $\widehat{\beta}_0/\widehat{\sigma}$ and $\widehat{\beta}_1/\widehat{\sigma}$, respectively, where the previous parameter estimates were calculated from the magnitude voxel time series using ordinary least squares. In addition, Figure S-4 shows the frequency distribution of the voxel-wise SNRs by smoothing level.

S-2. Supplement to Section ?? – Further Methodological Development.

S-2.1. *CV AR(p) spherical (CVS) and MO AR(p) Gaussian (MOG) model methodology.*

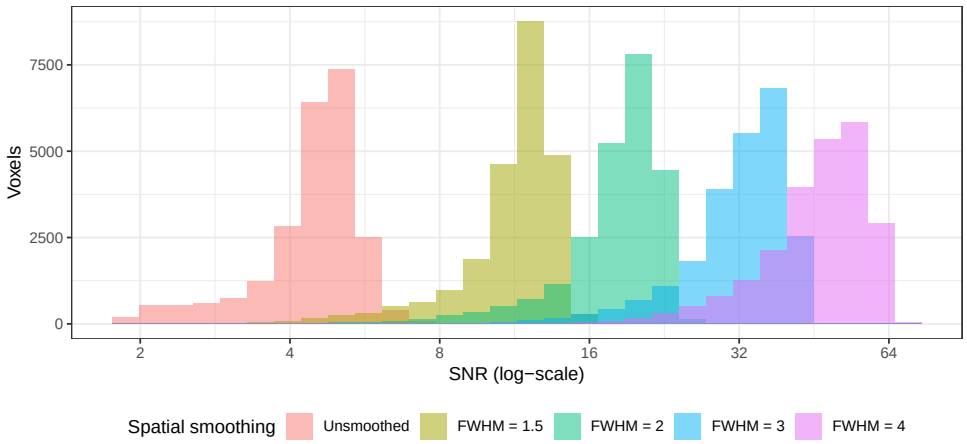


FIG S-4. *The voxel-wise SNRs by smoothing level.*

S-2.1.1. *CVS model.* The log-likelihood function is given by

$$\log f(\boldsymbol{y}_R, \boldsymbol{y}_I; \boldsymbol{\tau}) = -n \log \sigma^2 - \log |\boldsymbol{R}_n| - h/(2\sigma^2), \tag{S-1}$$

where

$$h = \begin{pmatrix} \boldsymbol{y}_R - \boldsymbol{X}\boldsymbol{\beta}\cos\theta \\ \boldsymbol{y}_I - \boldsymbol{X}\boldsymbol{\beta}\sin\theta \end{pmatrix}' \begin{pmatrix} \boldsymbol{R}_n^{-1} & 0 \\ 0 & \boldsymbol{R}_n^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_R - \boldsymbol{X}\boldsymbol{\beta}\cos\theta \\ \boldsymbol{y}_I - \boldsymbol{X}\boldsymbol{\beta}\sin\theta \end{pmatrix}. \tag{S-2}$$

The maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_R \cos\widehat{\theta} + \widehat{\boldsymbol{\beta}}_I \sin\widehat{\theta}$, where $\widehat{\boldsymbol{\beta}}_R = (\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{y}_R$, $\widehat{\boldsymbol{\beta}}_I = (\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{y}_I$, and $\widehat{\boldsymbol{R}}_n^{-1}$ is a function of $\widehat{\boldsymbol{\alpha}}$, the MLE of $\boldsymbol{\alpha}$, according to the $(2p+1)$-diagonal matrix given in Pourahmadi (2001). Further, the MLEs of θ and σ^2 are given by

$$\widehat{\theta} = \frac{1}{2}\arctan\left[\frac{2\widehat{\boldsymbol{\beta}}_R'\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_I}{\widehat{\boldsymbol{\beta}}_R'\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_R - \widehat{\boldsymbol{\beta}}_I'\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_I}\right] \tag{S-3}$$

and $\widehat{\sigma}^2 = \widehat{h}/(2n)$, where $\widehat{h}$ evaluates the parameters in (S-2) at their MLEs. We obtain $\widehat{\boldsymbol{\alpha}}$ by solving the system of equations (Miller, 1995)

$$\widehat{d}_{0k} = \sum_{j=1}^{p}(\widehat{d}_{jk} + 2j\widehat{\gamma}_{|j-k|})\widehat{\alpha}_j, \tag{S-4}$$

for $k = 1, \ldots, p$, with $\widehat{d}_{ij} = \sum_{t=1}^{n-i-j}\widehat{\eta}_{R,t+i}\widehat{\eta}_{R,t+j} + \widehat{\eta}_{I,t+i}\widehat{\eta}_{I,t+j}$, $0 \le i, j \le p$, and $\widehat{\gamma}_k = \widehat{d}_{0k}/(2n)$, wherein $\widehat{\eta}_{Rt} = y_{Rt} - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}\cos\widehat{\theta}$ and $\widehat{\eta}_{It} = y_{It} - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}\sin\widehat{\theta}$,

$t = 1, \ldots, n$. In practice, ML estimation consists of alternately updating $(\widehat{\theta}, \widehat{\boldsymbol{\beta}})$ and $(\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{R}}_n^{-1})$ in a Cochrane and Orcutt (1949)-type procedure until convergence. The LRT statistic for the test of $H_0 : \boldsymbol{C\beta} = \boldsymbol{0}$ vs. $H_a : \boldsymbol{C\beta} \neq \boldsymbol{0}$ is given by

$$\Lambda_{CVS,p} = 2n \log\left(\frac{\tilde{\sigma}^2}{\widehat{\sigma}^2}\right) - 2\log\left(\left|\tilde{\boldsymbol{R}}_p^{-1}\right| / \left|\widehat{\boldsymbol{R}}_p^{-1}\right|\right), \tag{S-5}$$

where $\boldsymbol{R}_p$ is such that $\sigma^2 \boldsymbol{R}_p = \text{Cov}(\eta_{R1}, \ldots, \eta_{Rp}) = \text{Cov}(\eta_{I1}, \ldots, \eta_{Ip})$, $\boldsymbol{R}_p^{-1}$ is a function of $\boldsymbol{\alpha}$ as in Pourahmadi (2001), and the "hats" and "tildes" denote quantities maximized with respect to H_a and H_0, respectively. It can be shown that $\Lambda_{CVS,p}$ follows an asymptotic χ^2_m null distribution, where $m = \text{rank}(\boldsymbol{C})$.

S-2.1.2. *MOG model.* The log-likelihood function for the MOG model is given by $\log f(\boldsymbol{r}; \boldsymbol{\tau}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|\boldsymbol{R}_n| - \frac{1}{2\sigma^2}(\boldsymbol{r} - \boldsymbol{X\beta})'\boldsymbol{R}_n^{-1}(\boldsymbol{r} - \boldsymbol{X\beta})$, where $\boldsymbol{R}_n$ is such that $\sigma^2\boldsymbol{R}_n = \text{Cov}(\boldsymbol{\epsilon})$. The MLEs of $\boldsymbol{\beta}$ and σ^2 are given by $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{R}}_n^{-1}\boldsymbol{r}$ and $\widehat{\sigma}^2 = (\boldsymbol{r} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'\widehat{\boldsymbol{R}}_n^{-1}(\boldsymbol{r} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})/n$, respectively. We obtain $\widehat{\boldsymbol{\alpha}}$ by solving the system of equations $\sum_{j=1}^{p}\{\widehat{d}_{ij} + (j/n)\widehat{d}_{0,|i-j|}\}\widehat{\alpha}_j = \widehat{d}_{0i}$, $i = 1, \ldots, p$, where $\widehat{d}_{ij} = \sum_{t=1}^{n-i-j}\widehat{\epsilon}_{t+i}\widehat{\epsilon}_{t+j}$, for $0 \leq i, j \leq p$, and $\widehat{\epsilon}_t = r_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}$, $t = 1, \ldots, n$. The estimation procedure begins with $\widehat{\boldsymbol{R}}_n = \boldsymbol{I}_n$ and then iteratively updates $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\alpha}}$, and $\widehat{\boldsymbol{R}}_n^{-1}$ until convergence. The LRT statistic for the test of $H_0 : \boldsymbol{C\beta} = \boldsymbol{0}$ vs. $H_a : \boldsymbol{C\beta} \neq \boldsymbol{0}$ is given by

$$\Lambda_{MOG,p} = n\log(\tilde{\sigma}^2/\widehat{\sigma}^2) - \log\left(\left|\tilde{\boldsymbol{R}}_p^{-1}\right| / \left|\widehat{\boldsymbol{R}}_p^{-1}\right|\right), \tag{S-6}$$

where $\boldsymbol{R}_p$ is such that $\sigma^2\boldsymbol{R}_p = \text{Cov}(\epsilon_1, \ldots, \epsilon_p)$.

S-2.2. *Relationships between CVS, MOR, and MOG model densities.* We illustrate some relationships between the probability density functions (PDFs) of the CVS, MOR, and MOG models for the observations at a single voxel and timepoint. Before we get into the derivations, let us state these relationships:

1. **Fact 1**: The MOR model PDF is the marginal PDF for the magnitude in the CVS model PDF.
2. **Fact 2**: For large SNR, the MOR model PDF approaches the MOG model PDF.

Derivation of Fact 1. Under the CVS model (and suppressing subscripts for time), the PDF is

$$f(y_R, y_I; \mu, \gamma_0, \theta) = (2\pi\gamma_0)^{-1}\exp\left[-\frac{(y_R - \mu\cos\theta)^2 + (y_I - \mu\sin\theta)^2}{2\gamma_0}\right]. \tag{S-7}$$

Transforming this PDF for the real/imaginary data to the magnitude/phase data yields

$$f(r,\phi;\mu,\gamma_0,\theta) = \frac{r}{2\pi\gamma_0}\exp\left[\frac{-(r^2+\mu^2)}{2\gamma_0}\right]\exp\left[\frac{\mu r}{\gamma_0}\cos(\phi-\theta)\right]. \tag{S-8}$$

The MO Ricean PDF then arises from integrating out ϕ in (S-8). That is, because $\int_{-\pi}^{\pi}\exp[\mu r\gamma_0\cos(\phi-\theta)]d\phi = 2\pi\mathbb{I}_0(\mu r/\gamma_0)$, the Ricean PDF (Rice, 1944) is

$$f(r;\mu,\gamma_0) = \frac{r}{\gamma_0}\exp\left[\frac{-(r^2+\mu^2)}{2\gamma_0}\right]\mathbb{I}_0\left(\frac{\mu r}{\gamma_0}\right). \tag{S-9}$$

Derivation of Fact 2. It can be shown the Ricean PDF approaches the Gaussian PDF at large SNRs – that is, for large values of $\mu/\sqrt{\gamma_0}$. We use the approximation (Abramowitz and Stegun, 1965) that for large values of x,

$$\mathbb{I}_0(x) = (2\pi x)^{-1/2}e^x\{1+1/(8x)+\mathcal{O}(x^{-2})\}. \tag{S-10}$$

Thus, for large SNR, which also implies large values of $\mu r/\gamma_0$, substituting (S-10) into the Ricean PDF (S-9) yields the Gaussian PDF

$$f(r;\mu,\gamma_0) = (2\pi\gamma_0)^{-1/2}\exp[-(r-\mu)^2/(2\gamma_0)] \tag{S-11}$$

times two additional terms, $(r/\mu)^{1/2}$ and $[1+(1/8)(\mu r/\gamma_0)^{-1}]$ that approach unity for large SNRs. (The former approaches unity because $|r-\mu|$ is on the order of $\sqrt{\gamma_0} \ll \mu_t$.) Indeed, Figure S-5 shows that the Ricean and Gaussian PDFs converge with increasing SNR.

S-2.3. *Supplement to Section* **??** *– Methodological details of the EM algorithm corresponding to the AR(p) Ricean model.*

S-2.3.1. *Supplement to Section* **??** *– Univariate Expectations.* To find the univariate expectations $\mathbb{E}_{\phi_t|r_t;\boldsymbol{\tau}^{(k)}}[\cos(\phi_t-\theta)]$, we show that the distribution of ϕ_t conditional on r_t is von Mises. From standard results, the conditional PDF $f(\phi_t|r_t;\boldsymbol{\tau})$ is the joint PDF $f(\phi_t,r_t;\boldsymbol{\tau})$ in (S-8) divided by the marginal PDF $f(r_t;\boldsymbol{\tau})$ in (S-9). Thus, the conditional PDF is

$$f(\phi_t|r_t;\boldsymbol{\tau}) = \left[2\pi\mathbb{I}_0\left(\frac{\mu_t r_t}{\gamma_0}\right)\right]^{-1}\exp\left[\frac{\mu_t r_t}{\gamma_0}\cos(\phi_t-\theta)\right], \tag{S-12}$$

which is the von Mises PDF with location parameter θ and concentration parameter $\mu_t r_t/\gamma_0$ (Mardia and Jupp, 2000). It then follows from properties of the von Mises distribution that the univariate expectations $\mathbb{E}_{\phi_t|r_t;\boldsymbol{\tau}^{(k)}}[\cos(\phi_t-\theta)] = \mathbb{A}(r_t\mu_t^{(k)}/\gamma_0^{(k)})$, $t=1,\ldots,n$.

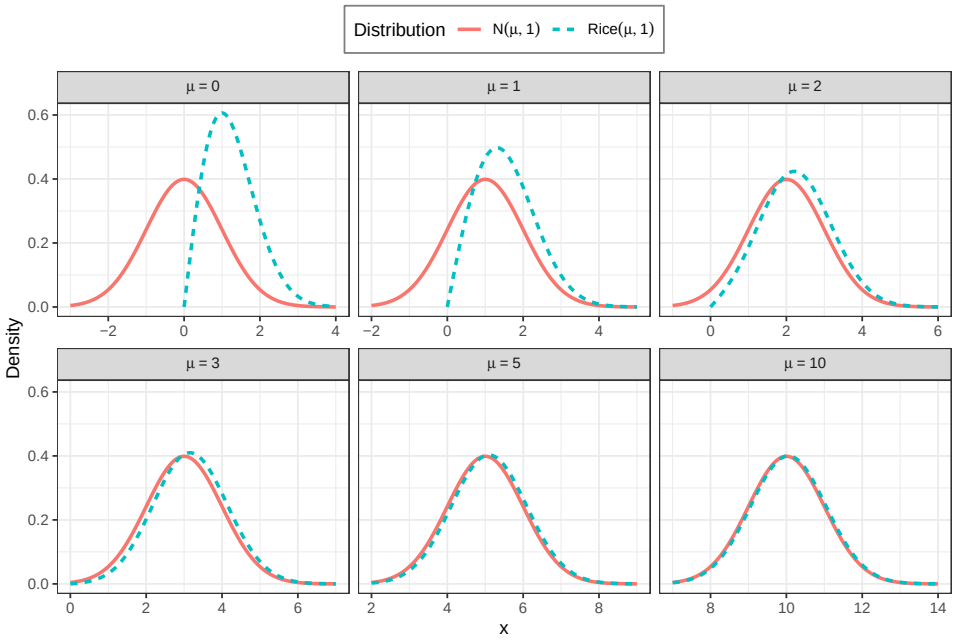


FIG S-5. *As the SNR (here, μ) increases, the Ricean and Gaussian PDFs converge.*

S-2.3.2. *Supplement to Section* **??** *– Bivariate Expectations.* Here, we show that the bivariate expectations $\mathbb{E}[\cos(\phi_t - \phi_{t+j})|r_t, r_{t+j}, \boldsymbol{\tau}^{(k)}]$ can be reduced to the univariate expectations in (**??**). Our strategy is to take the bivariate expectation as the "iterated expectations" $\mathbb{E}_{\phi_t|r_t}[\mathbb{E}_{\phi_{t+j}|\phi_t,r_t,r_{t+j}}\{\cos(\phi_{t+j} - \phi_t)\}]$. First, expanding the cosine term yields

$$\mathbb{E}_{\phi_t|r_t}\Big\{\cos(\phi_t - \theta)\mathbb{E}_{\phi_{t+j}|\phi_t,r_t,r_{t+j}}[\cos(\phi_{t+j} - \theta)] + \sin(\phi_t - \theta)\mathbb{E}_{\phi_{t+j}|\phi_t,r_t,r_{t+j}}[\sin(\phi_{t+j} - \theta)]\Big\}. \tag{S-13}$$

We now derive the conditional PDF of ϕ_{t+j} given ϕ_t, r_t, r_{t+j}. Starting with the distributions of $(y_{Rt}, y_{R,t+j})$ and $(y_{It}, y_{I,t+j})$, which are independent and bivariate normal, and using its magnitude and phase transformations, it can be shown that

$$f(\phi_{t+j}|\phi_t, r_t, r_{t+j}) \propto \exp[\kappa\cos(\phi_{t+j} - \theta) + \delta\cos(\phi_{t+j} - \phi_t)], \tag{S-14}$$

where $\kappa = r_{t+j}(\gamma_0\mu_{t+j} - \gamma_j\mu_t)/b$ and $\delta = \gamma_j r_t r_{t+j}/b$, with $b = \gamma_0^2 - \gamma_j^2$. It can then be shown that $\phi_{t+j}|\phi_t, r_t, r_{t+j}$ follows the von Mises distribution by writing the bracketed portion of (S-14) as $K\cos(\phi_{t+j} - \Psi)$ where $K = [\kappa^2 + \delta^2 + 2\kappa\delta\cos(\phi_t - \theta)]^{1/2}$ and Ψ is such that $\sin(\Psi - \theta) = \delta\sin(\phi_t - \theta)/K$ and $\cos(\Psi - \theta) = [\kappa + \delta\cos(\phi_t - \theta)]/K$. Thus, the conditional distribution of $(\phi_{t+j} - \theta)$ given ϕ_t, r_t, r_{t+j} is von Mises with location parameter $\Psi - \theta$ and concentration parameter K. It follows that $E_{\phi_{t+j}|\phi_t,r_t,r_{t+j}}[\cos(\phi_{t+j} - \theta)] = A(K)\cos(\Psi - \theta)$ and

$E_{\phi_{t+j}|\phi_t,r_t,r_{t+j}}[\sin(\phi_{t+j}-\theta)] = A(K)\sin(\Psi-\theta)$ (Mardia and Jupp, 2000). Substituting these expectations into (S-13) and using the earlier expressions for the sine and cosine of $(\Psi-\theta)$, we obtain (**??**).

S-2.3.3. *Supplement to Section* **??** *– Maximizing with respect to constraints.* To find $\boldsymbol{\beta}^{(k+1)} = \text{argmax}_{\boldsymbol{\beta}}\, Q(\boldsymbol{\alpha}^{(k+1)}, \boldsymbol{\beta}, \sigma^{2(k)}; \boldsymbol{\tau}^{(k)})$ as part of the (conditional) M-step, we must maximize with respect to the constraint $\boldsymbol{X}\boldsymbol{\beta}^{(k+1)} \geq \mathbf{0}$. In the following, we illustrate this constrained maximization for the $\boldsymbol{X}$ matrix defined in Section **??**, which has two columns: the first is an intercept containing all ones and the second is the expected BOLD response, which we denote by $\boldsymbol{b}$. It can be shown that $\boldsymbol{X}\boldsymbol{\beta} \geq \mathbf{0}$ if and only if $\boldsymbol{A}\boldsymbol{\beta} \geq \mathbf{0}$, where $\boldsymbol{A}$ contains only two rows of $\boldsymbol{X}$: the rows $\boldsymbol{a}_1' = (1, \min(\boldsymbol{b}))$ and $\boldsymbol{a}_2' = (1, \max(\boldsymbol{b}))$. To maximize with respect to $\boldsymbol{A}\boldsymbol{\beta} \geq \mathbf{0}$, we first calculate the unrestricted maximizer

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{R}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}_n^{-1}\boldsymbol{u}^{(k)}. \tag{S-15}$$

If $\boldsymbol{A}\widehat{\boldsymbol{\beta}} \geq \mathbf{0}$, then $\boldsymbol{\beta}^{(k+1)} = \widehat{\boldsymbol{\beta}}$. Otherwise, let $J = \{j : \boldsymbol{a}_j'\widehat{\boldsymbol{\beta}} < 0\}$. Calculate $\tilde{\boldsymbol{\beta}}_j = \boldsymbol{\Omega}_j\widehat{\boldsymbol{\beta}}$ for each $j \in J$, where $\boldsymbol{\Omega}_j = \boldsymbol{I}_2 - [\boldsymbol{a}_j'(\boldsymbol{X}'\boldsymbol{R}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{a}_j]^{-1}(\boldsymbol{X}'\boldsymbol{R}_n^{-1}\boldsymbol{X})^{-1}\boldsymbol{a}_j\boldsymbol{a}_j'$. Then $\boldsymbol{\beta}^{(k+1)}$ is the $\tilde{\boldsymbol{\beta}}_j$ which maximizes $Q(\boldsymbol{\alpha}^{(k+1)}, \tilde{\boldsymbol{\beta}}_j, \sigma^{2(k)}; \boldsymbol{\tau}^{(k)})$.

S-2.4. *Supplement to Section* **??** *– Empirical Information Matrix.* We illustrate the calculation of $\boldsymbol{s}(r_t;\boldsymbol{\tau})$, $t = p+1, \ldots, n$, the contributions to the score statistic from time t from which the empirical information matrix $\mathscr{I}_e(\boldsymbol{\tau};\boldsymbol{r})$ is constructed as in (**??**). We denote the elements of $\boldsymbol{s}(r_t;\boldsymbol{\tau})$ that come from taking the partial derivatives in (**??**) with respect to σ^2, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ by $s_{\sigma^2}(t)$, $\boldsymbol{s}_{\boldsymbol{\alpha}}(t)$, and $\boldsymbol{s}_{\boldsymbol{\beta}}(t)$, respectively. It can be shown that $s_{\sigma^2}(t) = [\tilde{\boldsymbol{\alpha}}'\boldsymbol{D}_t\tilde{\boldsymbol{\alpha}} - 2\sigma^2]/(2\sigma^4)$, where $\boldsymbol{D}_t$ is a matrix of order $(p+1)$ having (i,j)th entry $d_t(i,j) = r_{t-i}r_{t-j}E_{t-i,t-j} - \mu_{t-i}u_{t-j} - \mu_{t-j}u_{t-i} + \mu_{t-i}\mu_{t-j}$, $0 \leq i,j \leq p$, with $E_{rs} = \mathbb{E}_{\boldsymbol{\phi}|\boldsymbol{r};\boldsymbol{\tau}}[\cos(\phi_r - \phi_s)]$, $u_t = r_t A(r_t\mu_t/\gamma_0)$, and $\mu_t = \boldsymbol{x}_t'\boldsymbol{\beta}$. Further, $\boldsymbol{s}_{\boldsymbol{\alpha}}(t) = \boldsymbol{D}_{t(0)}'\tilde{\boldsymbol{\alpha}}/\sigma^2$, where $\boldsymbol{D}_{t(0)}$ is the matrix $\boldsymbol{D}_t$, but without the first row. Finally, $\boldsymbol{s}_{\boldsymbol{\beta}}(t) = \tilde{\boldsymbol{\alpha}}'\boldsymbol{D}_{t,\boldsymbol{\beta}}\tilde{\boldsymbol{\alpha}}/(2\sigma^2)$, where $\boldsymbol{D}_{t,\boldsymbol{\beta}}$ is the partial derivative of $\boldsymbol{D}_t$ with respect to $\boldsymbol{\beta}$, with (i,j)th entry $d_{t,\boldsymbol{\beta}}(i,j) = (u_{t-i} - \mu_{t-i})\boldsymbol{x}_{t-j} + (u_{t-j} - \mu_{t-j})\boldsymbol{x}_{t-i}$.

S-2.5. *Supplement to Section* **??** *– Further details regarding test statistics.*

S-2.5.1. *Supplement to Section* **??** *– Comparing false positive rates.* The following simulation experiment examines the basic utility of the Wald and likelihood ratio test (LRT) statistics in terms of whether they follow their theoretical null distributions. To mimic the finger-tapping experiment, we generated magnitude time series from the AR(1) Ricean model with the two-column $\boldsymbol{X}$ matrix described in Section **??**. Of the parameter $\boldsymbol{\beta} = (\beta_0, \beta_1)$ corresponding to $\boldsymbol{X}$, only β_1 is

activation-related; thus, the activation test is $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$, and we set $\beta_1 = 0$ to examine the null distributions of the test statistics, which theoretically should be χ^2_1. To examine an SNR range similar to that in the dataset (see Figure S-4), we set $\sigma = 1$ and varied β_0 from 0.5 to 5.0. We set $\alpha_1 = 0.3$, generated 10,000 time series for each β_0 value, and calculated the Wald and LRT statistics. Figure S-6 shows the proportions of test statistics in which H_0 was rejected (in effect,

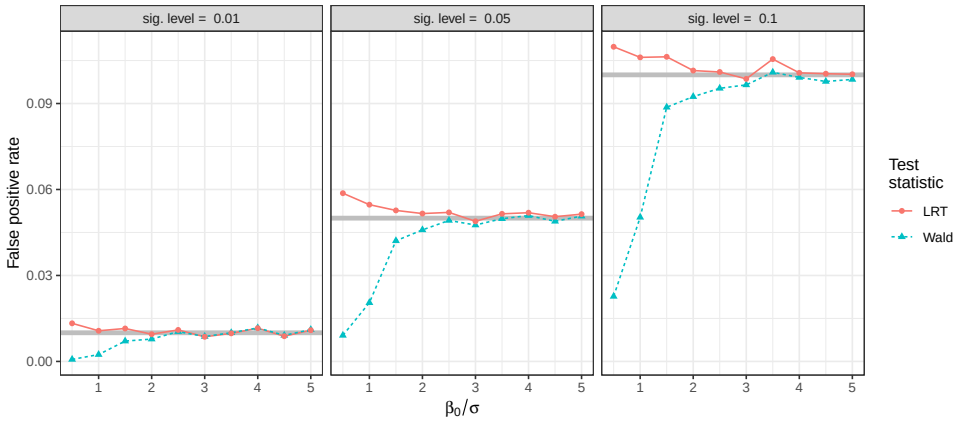


FIG S-6. *The false positive rates of the Wald test statistics fall sharply below the significance levels (shown by the thick grey lines) for β_0/σ values below 2, showing that LRT statistic is more reliable for low SNRs.*

the false positive rates) based on the theoretical χ^2_1 null distribution at significance levels of 0.01, 0.05, and 0.10. If the test statistic truly followed the theoretical null distribution, each false positive rate should be close to the significance level (with small discrepancies explained by simulation variability). However, it is evident that the Wald test is unusable at β_0 values below 2 due to its false positive rates falling sharply below the significance level. Overall, the LRT statistic seems more reliable due to its false positive rates better conforming with the significance level.

S-2.5.2. *Supplement to Section* **??** *– Ricean AR(1) LRT statistic.* Here, we derive the expression for $f(r_t|r_{t-1};\boldsymbol{\tau})$ in (**??**). For notational simplicity, we focus on $f(r_2|r_1)$. Starting with $y_{R2}|y_{R1} \sim N(\mu_2 \cos\theta + \alpha(y_{R1} - \mu_1 \cos\theta), \sigma^2)$ and $y_{R1} \sim N(\mu_1 \cos\theta, \sigma^2/(1-\alpha^2))$, and similarly for the imaginary component, and transforming to magnitude and phase, it can be shown that

(S-16)

$$f(r_1, \phi_1, r_2, \phi_2) \propto \exp[C_1 \cos(\phi_1 - \theta) + C_2 \cos(\phi_2 - \theta) + C_{12} \cos(\phi_1 - \phi_2)],$$

where $C_1 = r_1(\mu_1 - \alpha\mu_2)/\sigma^2$, $C_2 = r_2(\mu_2 - \alpha\mu_1)/\sigma^2$, and $C_{12} = \alpha r_1 r_2/\sigma^2$. First, we integrate with respect to ϕ_2. Writing $C_2 \cos(\phi_2 - \theta) + C_{12} \cos(\phi_1 - \phi_2) = K \cos(\phi_2 - \psi)$, where $K = [C_2^2 + C_{12}^2 + 2C_2 C_{12} \cos(\phi_1 - \theta)]^{1/2}$, and using

$\int_0^{2\pi} \exp[K\cos(\phi_2 - \psi)]d\phi_2 = 2\pi\mathbb{I}_0(K)$, we apply the Neumann Addition Formula (Watson, 1948) to write

$$\mathbb{I}_0(K) = \sum_{m=0}^{\infty} \omega_m \mathbb{I}_m(C_2)\mathbb{I}_m(C_{12})\cos[m(\phi_1 - \theta)], \tag{S-17}$$

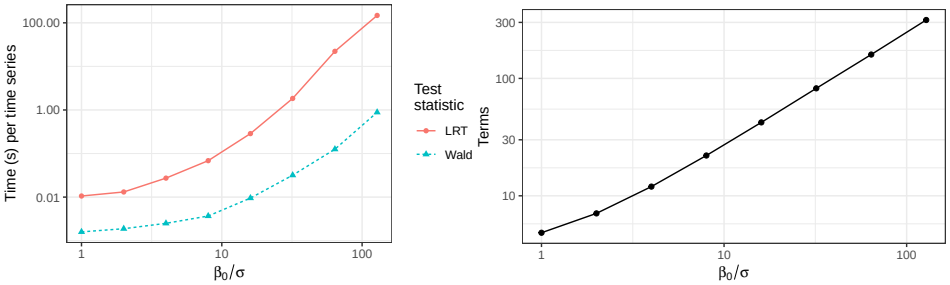
where $\omega_m = 1$ for $m = 0$ and $\omega_m = 2$ for $m \geq 1$. Also, using the result $\int_0^{2\pi} \cos[m(\phi_1 - \theta)]\exp[C_1\cos(\phi_1 - \theta)] = \mathbb{I}_m(C_1)$, for $m \geq 0$ (Mardia and Jupp, 2000), we obtain $f(r_1, r_2)$. Dividing the result by (S-9), we obtain

$$f(r_2|r_1) = \frac{r_2}{\sigma^2} e^{C_0} \left[\mathbb{I}_0\left(\frac{r_1\mu_1}{\gamma_0}\right)\right]^{-1} \sum_{m=0}^{\infty} \omega_m \mathbb{I}_m(C_1)\mathbb{I}_m(C_2)\mathbb{I}_m(C_{12}), \tag{S-18}$$

where $C_0 = -[r_2^2 + \mu_2^2 + \alpha^2(r_1^2 + \mu_1^2) - 2\alpha\mu_1\mu_2]/(2\sigma^2)$.

S-2.5.3. *Supplement to Section* **??** *– Calculating the AR(1) log-likelihood function at high SNRs is computationally prohibitive.* The following simulation study demonstrates that calculation of the Ricean AR(1) log-likelihood function (**??**) becomes increasingly computationally prohibitive as the SNR increases. To vary the SNR over the range of values seen in the dataset over different amounts of spatial smoothing, we generated magnitude time series from the Ricean AR(1) model with the $\boldsymbol{X}$ matrix described in Section **??**, and β_0 varied over $\{2^j : j \in \mathbb{Z}, 0 \leq j \leq 7\}$. Other parameter values of $\sigma^2 = 1$, $\alpha = 0.3$, and $\beta_1 = 0$ were held constant.

After the parameter estimates for each simulated time series were calculated using the hybrid EM/NR algorithm, we timed the calculation of the Wald and LRT statistics. Timing results are from a Intel Core i5-6300M CPU 64-bit processor running `C` within `R` (R Core Team, 2020). Figure S-7(a) shows that the calculation time



(a) Calculation time (b) Number of terms

FIG S-7. *(a) The calculation time for the LRT statistic is greater than the Wald test statistic, becoming prohibitive for high SNRs. (b) The average number of terms of the sum* $\sum_{m=0}^{\infty} \omega_m I_m(C_1)I_m(C_2)I_m(C_{12})$ *needed for convergence increases with SNR as well. (Note: Both plots use log scales.)*

of the LRT statistic is uniformly higher than the Wald test statistic and increases with the SNR. The computational time starts to become prohibitive from SNRs at around 30, when it is aproximately 1 second per time series, and increases further from there. As suggested by Figure S-7(b), this increase in computation time is due to the increase in terms of the sum $\sum_{m=0}^{\infty} \omega_m \mathbb{I}_m(C_1)\mathbb{I}_m(C_2)\mathbb{I}_m(C_{12})$ in (**??**) necessary for convergence (defined as a change of less than 10^{-10}). As a result, we will restrict our use of the LRT statistic to SNRs below 10 and use the Wald test statistic otherwise (where the false positive rate problem illustrated in Figure S-6 does not appear to be an issue).

S-2.6. *Supplement to Section* **??** *– Discussion of MO Ricean AR(p) model methodology.* The computation time under the MOR model is much greater than the other models because the convergence of the EM algorithm is slow, even with the acceleration provided by the hybrid scheme including Newton-Raphson steps. Figure S-8 shows the computation times required for parameter estimation of 1000 simulated time series under the four models. We generated these time series under the CVS model with $\boldsymbol{X}$ matrix described in Section **??**, $\sigma^2 = 1$, $\boldsymbol{\beta} = (\beta_0, 0)'$ for β_0 from 0.5 to 5.0, and AR coefficients of 0.4, $(0.4, 0.32)$, and $(0.4, 0.3, 0.2)$. In our calculation, we assumed the correct AR order was known. Per thousand time series, the computation times for MOR model range from 24 to 447 seconds, while the other models are most always under one second. It is interesting to note that the MOR model computation time decreases as the SNR increases, decreasing by a factor of 10 as β_0 increases from 0.5 to 5.0. Computation times also increase for all models as the AR order increases.

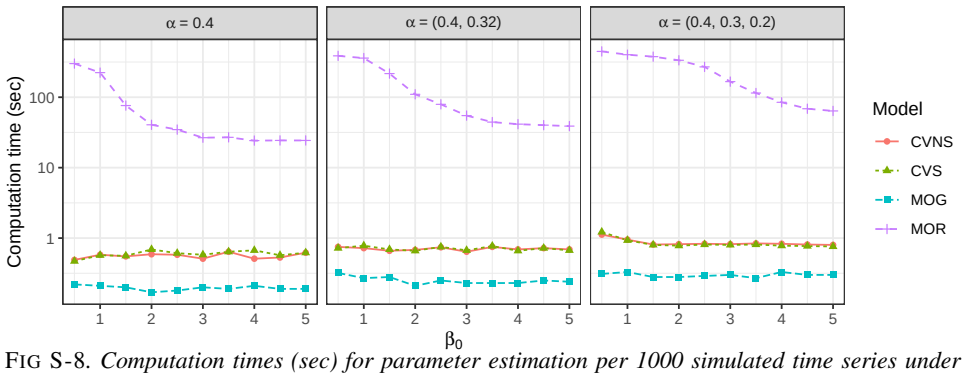


FIG S-8. *Computation times (sec) for parameter estimation per 1000 simulated time series under the four models. The EM algorithm used by the MOR model has much slower convergence than the other estimation schemes.*

S-2.7. *Supplement to Section* **??** *– Choosing the order of the AR model.* We performed a simulation study to compare the AR orders detected under the sequential testing method proposed in Section **??** with those detected under the AIC and

BIC (Akaike, 1973; Schwarz, 1978) – *i.e.,* more common model selection criteria. We generated 10,000 time series under the CVS model with $\boldsymbol{X}$ matrix described in Section **??**, $\boldsymbol{\beta} = (5, 0)'$, $\sigma^2 = 1$, and four sets of different AR coefficients: (1) $\alpha = 0$ for temporal independence, (2) $\alpha = 0.1$ and (3) $\alpha = 0.2$ for AR(1) dependence, and (4) $\boldsymbol{\alpha} = (0.1, 0.1)'$ for AR(2) dependence. The proportions of time series detecting each AR order $\widehat{p}$ based on the AIC, BIC, and sequential testing method using a significance level of $\delta = 0.1$ (and under the 4 models) are shown in Figure S-9.[2]

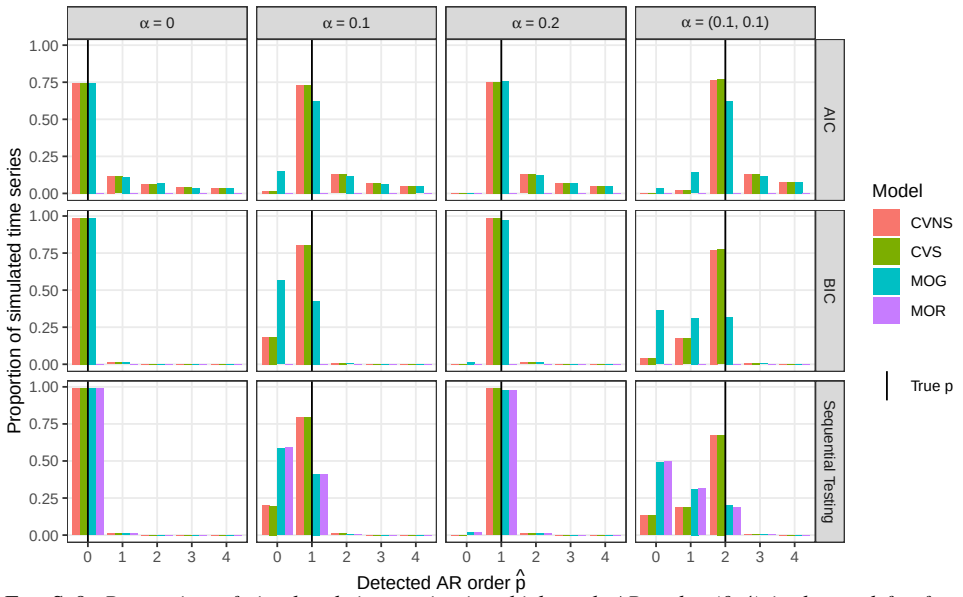


FIG S-9. *Proportion of simulated time series in which each AR order (0-4) is detected for four different sets of AR coefficients (columns), based on the AIC, BIC, and sequential testing methods (rows) and the four models (colored bars). Note that MOR model results are not shown for the AIC and BIC because the MOR model log-likelihood is not tractable for general p.*

Two main results are worthy of special attention: first, the sequential testing method in general detects a similar distribution of orders as the AIC and BIC. The BIC is more similar to the sequential method, which can be explained by the fact that the BIC penalty for each additional parameter of $\log n = 6.43$ (here) is closer to the sequential testing threshold of $\chi^2_{1,0.99} = 6.63$ than the AIC penalty of 2. The second result is that the MO-based order detection methods are more likely to have a negative bias, and under-detect orders (in this case, especially for $\alpha = 0.1$ and $\boldsymbol{\alpha} = (0.1, 0.1)'$) than CV-based methods. Applying the framework in Table **??**, this difference can be attributed to the fact that CV-based methods have twice the

[2]Note that MOR model results are not shown for the AIC and BIC because the MOR model log-likelihood is not tractable for general p.

amount of data, which gives them more power in the sequential testing method.

S-2.8. *Supplement to Section* **??** *– Clarification of SNR and CNR.* We will follow our previous definition of the SNR for a magnitude time series as the ratio of the mean baseline signal to the standard deviation of the noise. The CNR is the ratio of the size of the BOLD response (*i.e.*, the difference of the means under maximum activation and at baseline) to the standard deviation of the noise. An advantage of the MOG model under independence is that its parameters lend themselves easily to the specification of the SNR and CNR. That is, for the $\boldsymbol{X}$ matrix described in Section **??** with two columns – an intercept and a zero-centered waveform modeling the BOLD response, β_0 is the mean baseline signal level, β_1 is the size of the BOLD response, and σ is the noise standard deviation. Thus, the SNR is β_0/σ and the CNR is β_1/σ. However, this is not the case under more complicated models. For instance, if we introduce an AR(1) dependence to the noise with AR coefficient α and white noise variance σ^2 – hereafter abbreviated by AR$(1, \alpha, \sigma^2)$, the SNR and CNR are in terms of a noise variance of $\gamma_0 = \sigma^2/(1-\alpha^2)$.

Interpreting the SNR and CNR in our Ricean framework is complicated by the fact that the location and scale parameters of the Rice(ν, γ_0) distribution are *not* the its mean and variance. For $Y \sim$ Rice(ν, γ_0), $\mathbb{E}(Y) = \sqrt{\gamma_0 \pi/2}\mathbb{L}_{1/2}(\nu^2/(2\gamma_0))$ and $\text{Var}(Y) = (2\gamma_0 + \nu^2) - \pi\gamma_0/2\mathbb{L}^2_{1/2}(\nu^2/(2\gamma_0))$, where the Laguerre polynomial $\mathbb{L}_{1/2}(x) = e^{-x}[(1+x)\mathbb{I}_0(x/2) + x\mathbb{I}_1(x/2)]$ (Zhu *et al.*, 2009). (For an AR$(1, \alpha, \sigma^2)$ Ricean model, we would replace γ_0 by $\sigma^2/(1-\alpha^2)$ in the formulas for the mean and variance.) Note that we have so far only defined SNR and CNR for magnitude time series; for complex-valued time series we define them in terms of the SNR/CNR of the calculated magnitude time series.[3] For instance, the SNR for the AR$(1, \alpha, \sigma^2)$ CVS model would be based on the formulas for the mean and variance of the Rice distribution above.

Figure S-10 plots the mean, standard deviation, SNR, and CNR for the AR(1) Rice models with $\sigma^2 = 1$, $\beta_1 = 0.2$, β_0 varying from 0 to 5, and α varying from 0 to 0.8. Specifically, if $R_0 \sim$ Rice(β_0, γ_0) and $R_1 \sim$ Rice$(\beta_0 + \beta_1, \gamma_0)$, where $\gamma_0 = \sigma^2/(1-\alpha^2)$ as before, the mean, standard deviation, SNR, and CNR are calculated as $\mathbb{E}(R_0)$, $SD(R_0)$, $\mathbb{E}(R_0)/SD(R_0)$, and $[\mathbb{E}(R_1) - \mathbb{E}(R_0)]/SD(R_0)$, respectively. (Note that the mean is always greater than β_0 but approaches β_0, as β_0 increases.) The standard deviation is less than $\gamma_0^{1/2} = \sigma/(1-\alpha^2)^{1/2}$ but approaches it as β_0 increases as well. Focusing on the independent Rice model, we see that the SNR and CNR approach the values of β_0 and β_1, respectively as β_0 increases.

For the CVNS model, the SNR and CNR are impacted by the specification of (the non-spherical) $\boldsymbol{\Sigma}$. To illustrate this, consider the CVNS model with $\beta_0 = 5$,

[3] Indeed, because the complex-valued model captures the true underlying source of fMRI data, it could be claimed that the SNR should be defined in terms of it instead.

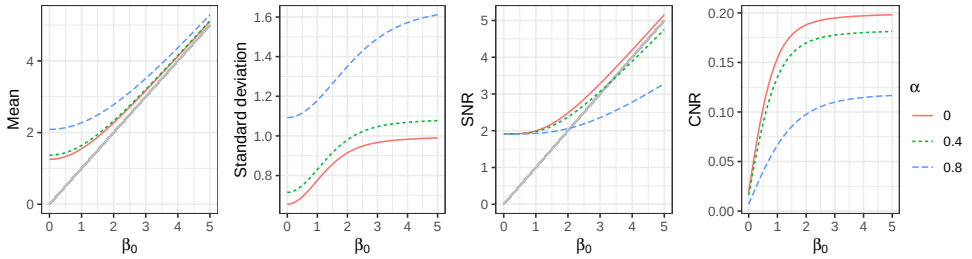
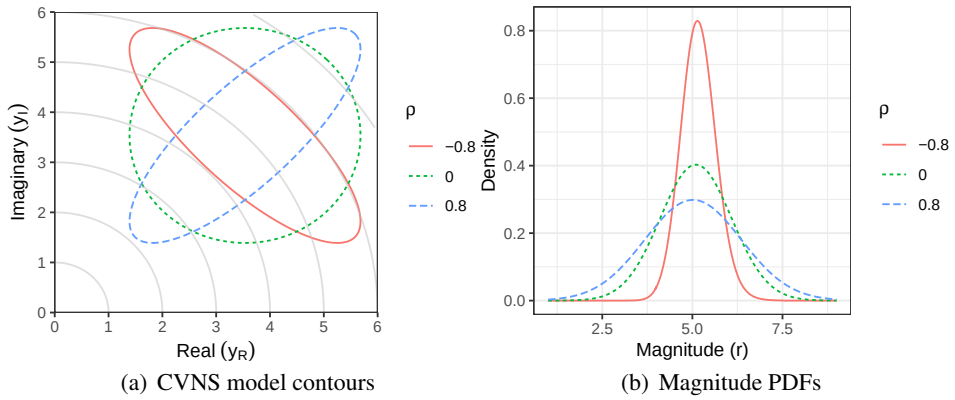
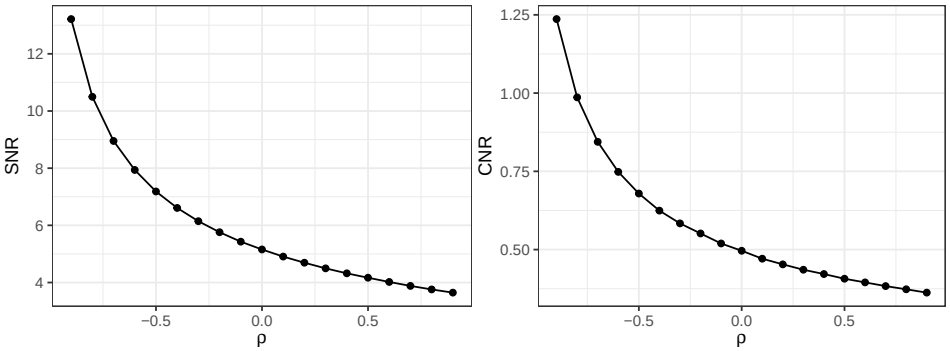


FIG S-10. *(From left) The mean, standard deviation, SNR, and CNR of the AR(1) Rice models with* $\sigma^2 = 1$, $\beta_1 = 0.2$, β_0 *from 0 to 5, and* α *from 0 to 0.8. (The diagonal solid grey lines represent a one-to-correspondence between the x- and y-axes.)*

$\theta = \pi/4$, $\sigma_R^2 = \sigma_I^2 = 1$ and varying real/imaginary correlation ρ. Figure S-11(a) displays 90% probability contours of this bivariate normal distribution for $\rho = 0$



(a) CVNS model contours (b) Magnitude PDFs



(c) SNR (d) CNR

FIG S-11. *(a) 90% probability contours for the CVNS model distribution with parameters* $\beta_0 = 5$, $\theta = \pi/4$, $\sigma_R^2 = \sigma_I^2 = 1$ *and varying real/imaginary correlation* ρ*, and the corresponding (b) magnitude probability density functions, (c) signal-to-noise ratio, and (d) contrast-to-noise ratio.*

and ± 0.8. These contours suggest that the distribution of the magnitudes based on

this CVNS model will have more spread for positive ρs than for negative ρs. This is verified by the magnitude probability density functions (Aalo, Efthymoglou and Chayawan, 2007) in Figure S-11(b). As a result, the SNRs and CNRs, as compared with the values for $\rho = 0$, increase and decrease as ρ becomes more negative and positive, respectively, as shown in Figure S-11(c)(d). The phase location θ also plays a role: effect of $\boldsymbol{\Sigma}$ on the SNRs and CNRs will vary depending on θ.

S-2.9. *Supplement to Section* **??** *– Phase-only activation.* The model assumes that each phase measurement ϕ_t is independent and follows the von Mises distribution with location parameter $\theta_t = \theta + g(\boldsymbol{w}_t'\boldsymbol{\delta})$ and concentration parameter κ. In the previous, $g(\cdot) = 2\arctan(\cdot)$ and $\boldsymbol{w}_t'$ is the tth row of a $n \times b$ matrix $\boldsymbol{W}$ that describes the phase change in time. Note that the matrix $\boldsymbol{W}$ does not contain the intercept column used in regression because the baseline phase is already modeled by θ. (In our case, $\boldsymbol{W}$ is a single column and is the same as the second column of $\boldsymbol{X}$ which comes from convolving the stimulus time course with the HRF.) Thus, the log-likelihood function is given by

$$\log L(\theta, \boldsymbol{\delta}, \kappa|\boldsymbol{\phi}) = -n\log(2\pi I_0(\kappa)) + \kappa \sum_{t=1}^{n} \cos\left(\phi_t - \theta - g(\boldsymbol{w}_t'\boldsymbol{\delta})\right). \tag{S-19}$$

An iterative algorithm for ML parameter estimation (Fisher and Lee, 1992) is as follows: after calculating the OLS-based starting value $\boldsymbol{\delta}^{(0)} = 0.5(\boldsymbol{W}'\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{\phi}$ (or utilizing another method), calculate the updated parameter estimates at the kth iteration as

$$\theta^{(k+1)} = \arctan_4(S^{(k)}, C^{(k)}) \tag{S-20}$$

$$\kappa^{(k+1)} = A^{-1}(R^{(k)}) \tag{S-21}$$

$$\boldsymbol{\delta}^{(k+1)} = \boldsymbol{\delta}^{(k)} + (\boldsymbol{W}'\boldsymbol{G}^2\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{G}^2\boldsymbol{z}. \tag{S-22}$$

In (S-20), $\arctan_4(y, x)$ is the 4-quadrant arctangent such that $\arctan_4(y, x) = \arctan(y/x)$, $S^{(k)} = (1/n)\sum_{t=1}^{n}\sin(\phi_t - g(\boldsymbol{w}_t'\boldsymbol{\delta}))$, and $C^{(k)} = (1/n)\sum_{t=1}^{n}\cos(\phi_t - g(\boldsymbol{w}_t'\boldsymbol{\delta}))$. In (S-21), $R^{(k)} = [S^{(k)2} + C^{(k)2}]^{1/2}$ and $A^{-1}(\cdot)$ is the inverse of the ratio of modified Bessel functions $\mathbb{A}(\cdot) = \mathbb{I}_1(\cdot)/\mathbb{I}_0(\cdot)$. The inverse can be well approximated (Mardia and Jupp, 2000) as

$$A^{-1}(R) = \begin{cases} 2R + R^3 + 5R^5/6 & R < 0.53 \\ -0.4 + 1.39R + 0.43/(1-R) & 0.53 \le R < 0.85 \\ \left[2(1-R) - (1-R)^2 - (1-R)^3\right]^{-1} & R \ge 0.85 \end{cases} \tag{S-23}$$

Last, in (S-22), $\boldsymbol{G}$ is a diagonal matrix with entries $g'(\boldsymbol{w}_t'\boldsymbol{\delta}) = 2/(1 + (\boldsymbol{w}_t'\boldsymbol{\delta})^2)$, $t = 1, \ldots, n$, and $\boldsymbol{z}$ is an n-vector with tth entry $z_t = \sin(\phi_t - \theta - g(\boldsymbol{w}_t'\boldsymbol{\delta}))/[A(\kappa)g'(\boldsymbol{w}_t'\boldsymbol{\delta})]$. The iterations proceed till convergence of (S-19).

Finally, a Wald test statistic for $H_0 : \boldsymbol{\delta} = \mathbf{0}$ vs. $H_a : \boldsymbol{\delta} = \mathbf{0}$ can be calculated from the MLE $\widehat{\boldsymbol{\delta}}$ and its asymptotic covariance matrix

(S-24)
$$\mathbb{V}\text{ar}(\widehat{\boldsymbol{\delta}}) = \frac{1}{\widehat{\kappa}\mathbb{A}(\widehat{\kappa})}\left\{(\boldsymbol{W}'\boldsymbol{G}^2\boldsymbol{W})^{-1} + \frac{(\boldsymbol{W}'\boldsymbol{G}^2\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{g}\boldsymbol{g}'\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{G}^2\boldsymbol{W})^{-1}}{n - \boldsymbol{g}'\boldsymbol{W}(\boldsymbol{W}'\boldsymbol{G}^2\boldsymbol{W})^{-1}\boldsymbol{W}'\boldsymbol{g}}\right\},$$

as per Fisher and Lee (1992), where $\boldsymbol{g}$ is a vector of the diagonal elements of $\boldsymbol{G}$.

S-3. Supplement to Section ?? – Further Simulation-based analyses.

S-3.1. *Supplement to Section* **??** *– Missing information matrix.* The Fisher information matrix is commonly used to estimate the standard errors of parameter estimates (Casella and Berger, 2002) and in-so-doing, quantify the amount of "information" given in the data about a parameter. In the framework of the EM algorithm, separate information matrices can be derived based on the complete, observed, and missing data (McLachlan and Krishnan, 2008). This extension allows us to quantify the amount of "missing information" in the missing data about a parameter (Orchard and Woodbury, 1972). In our context, recall that the magnitude, phase, and magnitude-phase constitute the observed, missing, and complete data, respectively. Thus, by deriving the missing information matrix, we can quantify the amount of "missing information" in the phase we miss out on when using magnitude-only data.

For simplicity, consider a single complex-valued measurement, with magnitude r and phase ϕ. We assume that the real and imaginary components y_R and y_I are independent and normally distributed, with means $\mu\cos\theta$ and $\mu\sin\theta$, respectively, and identical variances σ^2. The log-likelihood function of the observed data r is equal to that of the complete data (r, ϕ) minus the missing data ϕ; that is,

(S-25)
$$\log L(\boldsymbol{\tau}; r) = \log L_c(\boldsymbol{\tau}; r, \phi) - \log k(\phi|r; \boldsymbol{\tau}).$$

Differentiating with respect to $\boldsymbol{\tau}$ twice and obtaining its expectation with respect to ϕ, conditional on r, gives the following relationship between the information matrices:

(S-26)
$$\mathscr{I}(\boldsymbol{\tau}; r) = \mathscr{I}_c(\boldsymbol{\tau}; r) - \mathscr{I}_m(\boldsymbol{\tau}; r).$$

In words, the observed information is equal to the complete information minus the missing information.

It can be shown that observed-, complete-, and missing-data log-likelihoods are

$$\log L(\boldsymbol{\tau}; r) = -\log \sigma^2 - \frac{r^2+\mu^2}{2\sigma^2} + \log \mathbb{I}_0\left(\frac{\mu r}{\sigma^2}\right) \tag{S-27}$$

$$\log L_c(\boldsymbol{\tau}; r, \phi) = -\log \sigma^2 - \frac{r^2+\mu^2}{2\sigma^2} + \frac{\mu r}{\sigma^2}\cos(\phi-\theta) \tag{S-28}$$

$$\log k(\phi|r; \boldsymbol{\tau}) = -\log I_0\left(\frac{\mu r}{\sigma^2}\right) + \frac{\mu r}{\sigma^2}\cos(\phi-\theta), \tag{S-29}$$

respectively. For simplicity, suppose it is known that $\sigma^2 = 1$. Then, differentiating twice with respect to μ shows that the complete- and missing-data information matrices are $\mathscr{I}_c(\mu; r) = 1$ and $\mathscr{I}_m(\mu; r) = \frac{\partial^2}{\partial \mu^2} \log \mathbb{I}_0(\mu r)$, respectively. It can be shown that

$$\mathscr{I}_m(\mu; r) = r^2 - r\mathbb{A}(\mu r)/\mu - r^2\mathbb{A}^2(\mu r), \tag{S-30}$$

where $\mathbb{A}(\cdot) = \mathbb{I}_1(\cdot)/\mathbb{I}_0(\cdot)$ as before.

After averaging over the Rice(μ, 1) distribution of R using Monte-Carlo integration, these information matrices are displayed in Figure S-12. Note that the

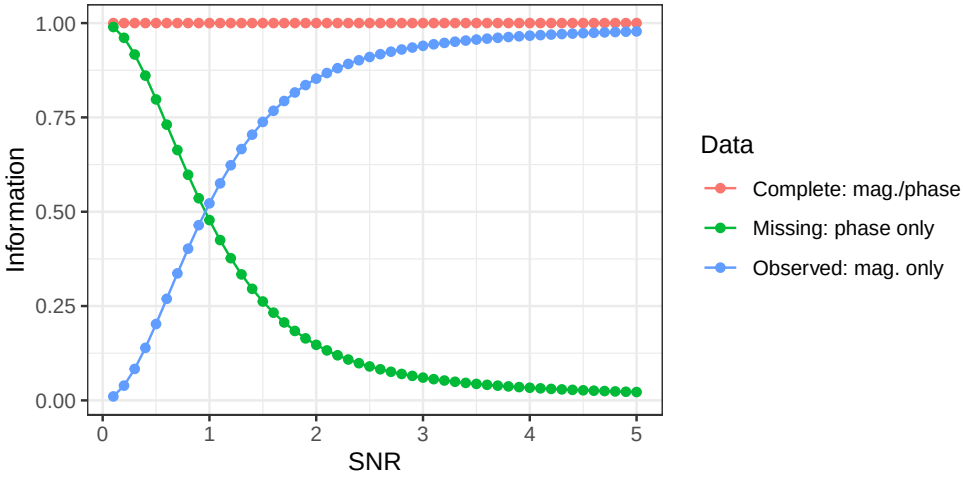


FIG S-12. *Plots of the information matrices as a function of the SNR. Note that the fraction of the complete-data (total) information provided by the magnitude-only data increases as the SNR increases.*

fraction of the complete-data (total) information provided by the magnitude-only data increases as the SNR increases. This is similar to the graphs in Figure **??** comparing the pAUCs of the complex-valued and the magnitude-only data-based test statistics.

Interestingly, the rate of convergence of the EM algorithm is a function of the missing- and complete-data information matrices (Dempster, Laird and Rubin,

1977). Specifically, defining the rate of convergence as $r_c = \lim_{k\to\infty} ||\boldsymbol{\tau}^{(k+1)} - \widehat{\boldsymbol{\tau}}||/||\boldsymbol{\tau}^{(k)} - \widehat{\boldsymbol{\tau}}||$, it can be shown that r_c is given by the largest eigenvalue of $\boldsymbol{\mathcal{I}}_c^{-1}(\widehat{\boldsymbol{\tau}}; \boldsymbol{r}, \boldsymbol{\phi})\boldsymbol{\mathcal{I}}_m(\widehat{\boldsymbol{\tau}}; \boldsymbol{r})$. This information ratio matrix measures the proportion of information about $\boldsymbol{\tau}$ that is missing by not also observing $\boldsymbol{\phi}$ in addition to $\boldsymbol{r}$ (McLachlan and Krishnan, 2008, Section 3.9.3). The greater the proportion of missing information, the slower the rate of convergence. We then see a connection between the large proportion of missing information at low SNRs in Figure S-12 and the higher computation times of the MOR model at low SNRs in Figure S-8. As the SNR increases, the proportion of missing information decreases and so does the computation time.

S-3.2. *Supplement to Section* **??** *– Further nonsphericity-related simulation studies.* We checked the false positive rates of the four model-based LRT statistics (LRTSs), assuming the same parameter estimates as Section **??** – except with the activation parameter β_1 set to zero. Figure S-13 displays the false positive rates

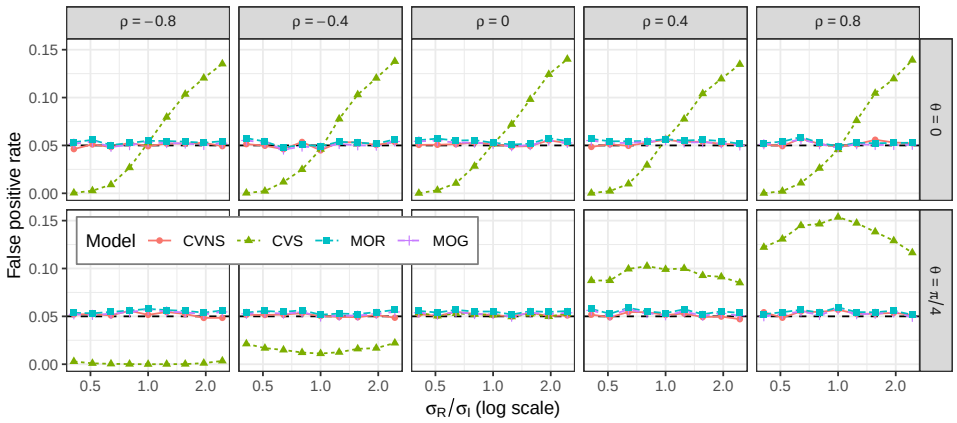
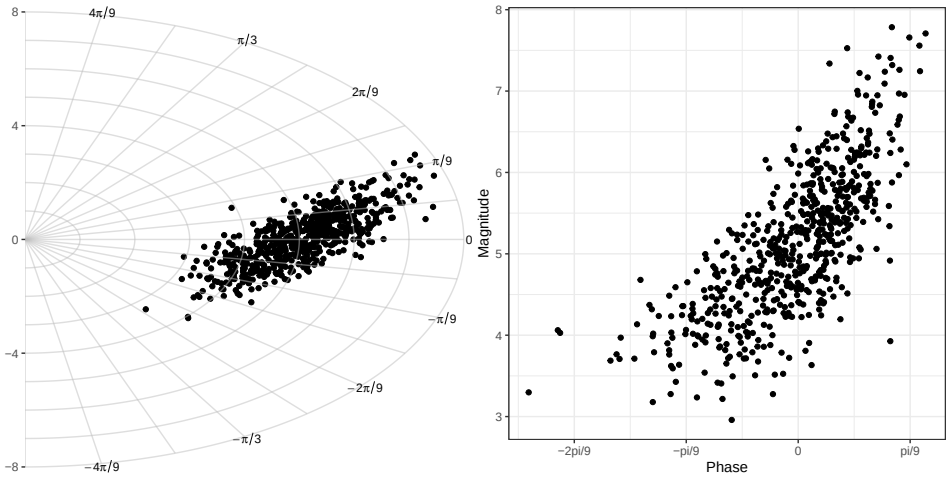


FIG S-13. *False positive rates of the model-based test statistics at the 0.05 significance level, based on simulated data.*

based on a significance level of 0.05, the proportion of the time series with LRTS greater than the 0.95 quantile of the χ_1^2 distribution. The false positive rate of the CVS model-based LRT statistic does not in general conform to the significance level of 0.05, except for certain special cases including the spherical case where $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_2$. This indicates that the CVS model-based LRTS is not of practical use when sphericity cannot be established.

We used simulation to augment our explanation of the relationship between the size of the CVNS model-based pAUC improvement over the MO model and the specification of θ and $\boldsymbol{\Sigma}$ (see Figure **??**), demonstrating that time series with non-spherical $\boldsymbol{\Sigma}$ possess correlated magnitude and phase time series. Further, we suggest that this correlation appears to be connected to superior activation detection

performance of the CVNS model. Figure S-14 shows plots of a single time series



(a) Polar (b) Cartesian

FIG S-14. *Magnitude/phase correlation is demonstrated by a simulated complex-valued time series (a) plotted in polar coordinates and (b) with magnitude-phase transformed to Cartesian coordinates.*

generated from the CVNS model with $\boldsymbol{\beta}_0 = (5, 0)'$, $\theta = 0$, $\sigma_R^2 = \sigma_I^2 = 1$, and $\rho = 0.8$. When the magnitude and phase in the polar-coordinate plot in Figure S-14(a)[4] are plotted on the Cartesian coordinates in Figure S-14(b), the substantial magnitude/phase correlation is clearly represented. We next showed how the magnitude/phase correlation depended on the specification of ρ, σ_R/σ_I, and θ by generating time series from the same parameter values as in Section **??**. Based on 10,000 time series, Figure S-15 shows the Pearson correlation[5] values, using the same layout as Figure **??** for comparison. Viewing Figures S-15 and **??** side-by-side yields interesting results: first, the cases where the CVNS and MO model-based pAUCs are the closest are exactly when the magnitude/phase correlation is zero. Second, when the CVNS model-based pAUCs are higher, the size of the difference seems to be connected to the size of the correlation. This correlation indicates that the phase data provides important additional information about the magnitude-related activation.

S-4. Supplement to Section ?? – Further Analysis of low-SNR Dataset.

[4]The concentric semi-circles represent the magnitudes values and the "spokes" represent the phase values.

[5]Using the linear-circular correlation as defined by Mardia and Jupp (2000, p. 245), which may be more appropriate in this case, provides similar results to the Pearson correlation.

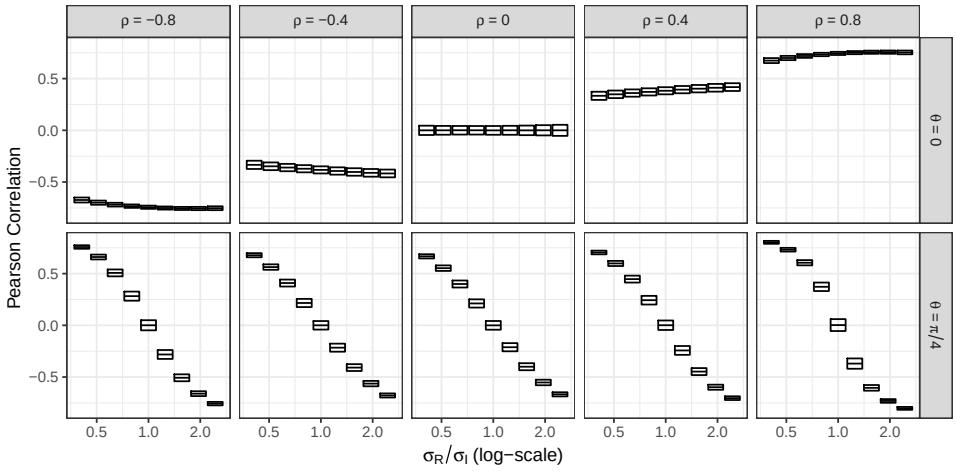


FIG S-15. *Pearson correlations between the magnitude and phase, summarized as mean ± one standard deviation, of simulated time series from the CVNS model.*

S-4.1. *Supplementary figures.* The figures presented are given in the order to which they are referred in the main article. Figure S-16 shows the frequency distributions of the voxel-wise detected AR orders for the five datasets and the four

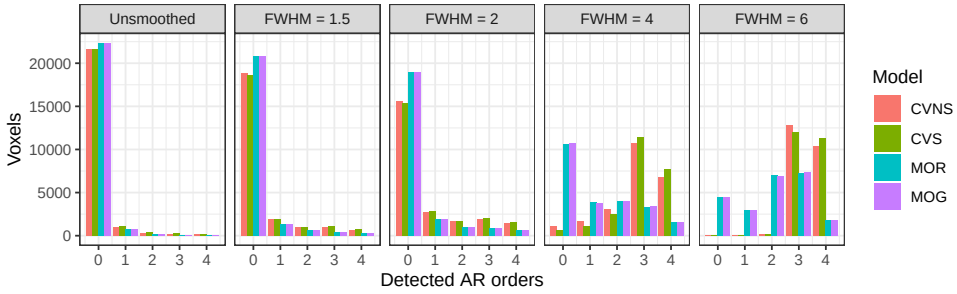


FIG S-16. *Frequency distributions of detected AR orders for the unsmoothed and smoothed datasets with FWHMs = 1.5, 2, 4, and 6 under the four models.*

models. The orders were detected using the sequential testing procedure described in Section **??** with a significance level of 0.01. The orders were limited to a maximum of 4 due to large computational times using the MOR model; if a larger cap on orders is desired in practice, we suggest using the MOG model as a surrogate, as it gives very similar results without such computational issues.

Figure S-17 shows the activation maps of all slices for the unsmoothed data. All slices are shown here to demonstrate that the maps shown of slice 2 in Figure **??** show more activation than the other slices (although, admittedly, the activation is sparse for all slices).

Figure S-18 refers to our study of adding extra noise to the original unsmoothed

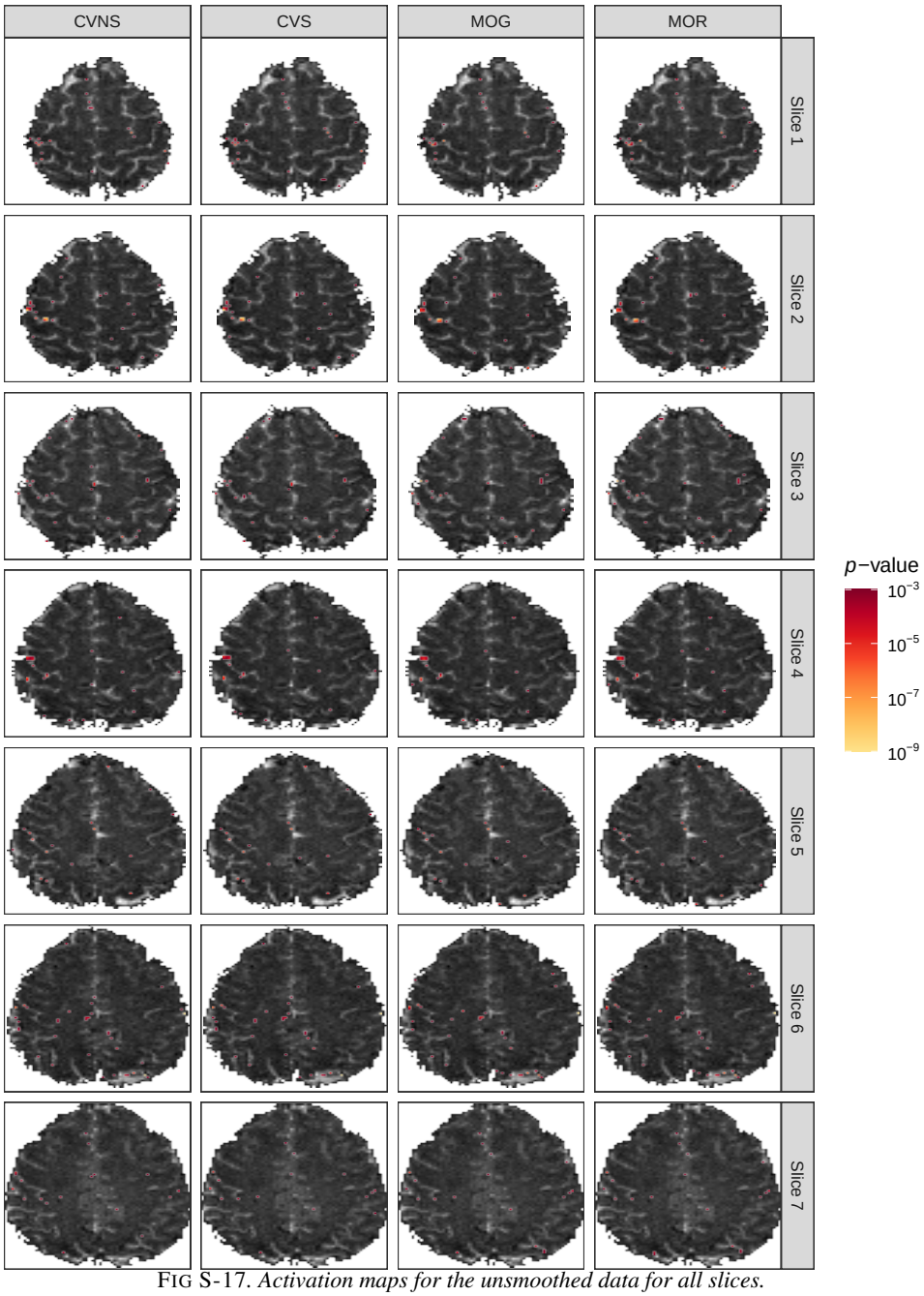


FIG S-17. *Activation maps for the unsmoothed data for all slices.*

data to further lower the SNR and CNR of the dataset. The plots show the average SNRs and CNRs for 10,000 simulation-based time series generated from each of the 10 voxel time series identified in Figure **????**. The results verify that adding

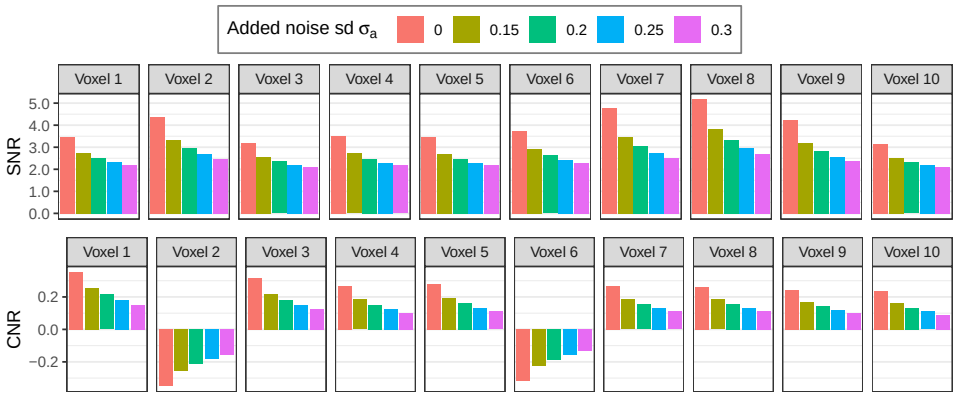


FIG S-18. *The SNRs and CNRs of the ten voxels identified in Figure* **????** *decrease when extra noise with standard deviation* σ_a *is added to the raw data (*$\sigma_a = 0$ *refers to raw data).*

extra noise to the raw data does indeed lower the SNRs and CNRs and that the effect is intensified as the noise standard deviation σ_a increases.

Figure S-19 shows the activation maps of slices 1 and 3 of the smoothed datasets with FWHMs = 1, 2, 4, and 6. These plots are provided as evidence that the acti-

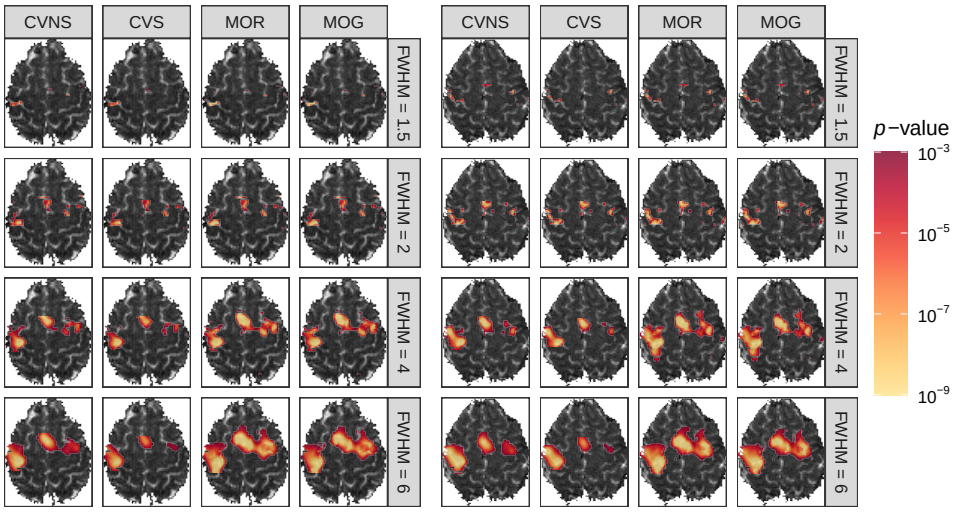


FIG S-19. *Activation maps of slices 1 (left) and 3 (right) calculated under the different models (columns) and smoothing levels (rows).*

vation maps for slice 2 in Figure **??** are representative of other slices. The comparisons made between the different model-based maps in regards to slice 2 may be extended to other slices as well.

Last, Figure S-20 shows activation maps of slice 2 calculated under an assumed AR order of 1 at all voxels. This is provided as evidence that differences between

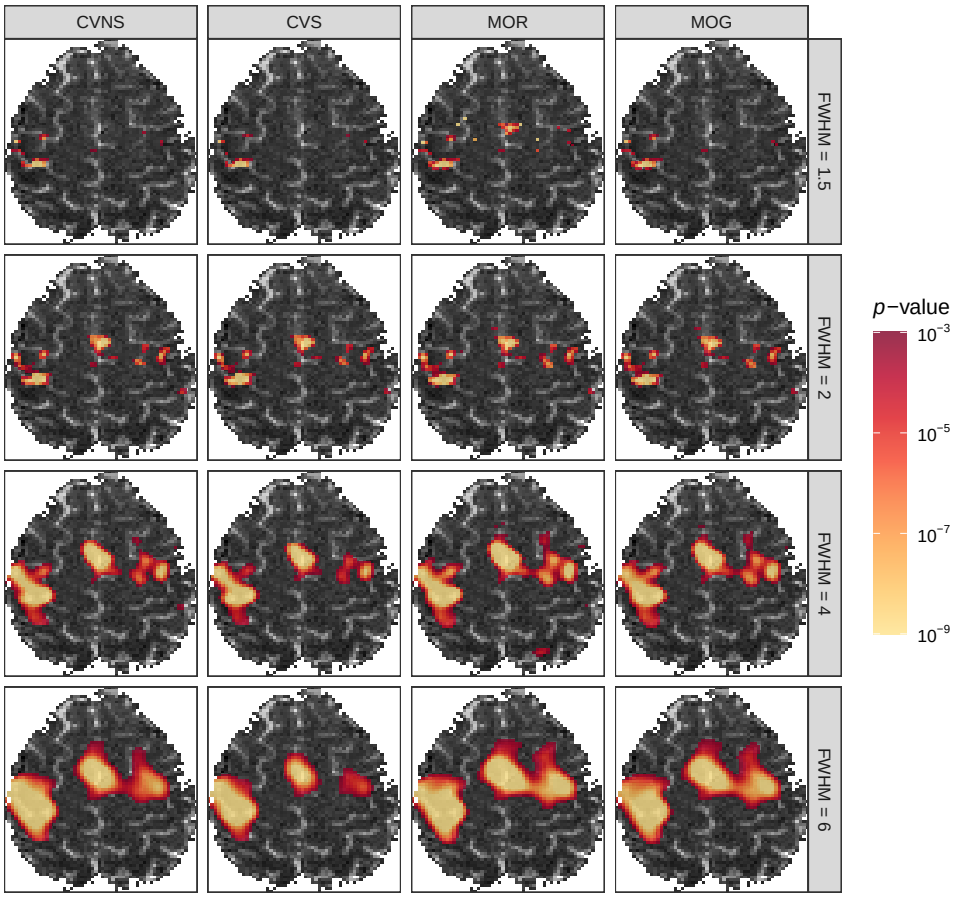


FIG S-20. *Activation maps of slice 2 calculated under an assumed AR order of 1 at all voxels.*

the complex-valued data-based activation maps and their magnitude-only data-based counterparts in Figure **??** may be attributed to differences in the detected AR orders. That is, the maps in Figure S-20 based on the same AR order for all models are notably more similar than the detected-order-based maps in Figure **??**.

S-4.2. *Supplement to Section* **??** *– parameter estimate summary.* Figures S-21 and S-22 show spatial and frequency distributions, respectively, of the CVNS model-based MLEs under an assumed AR order of 1 for all voxel time series. These summaries allow us to examine the effect of spatial smoothing on the parameter estimates. To summarize, for the unsmoothed data; it appears the AR(0) CVS model is adequate, as the distributions of $\widehat{\rho}$ and $\widehat{\alpha}$ are both centered closely around zero. When smoothing is applied, the spatial distributions of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are blurred and the noise standard deviations $\widehat{\sigma}_R$ and $\widehat{\sigma}_I$ decrease; the net effect is an increase in SNR and CNR, as shown in Figure S-3. More interestingly, smoothing induces

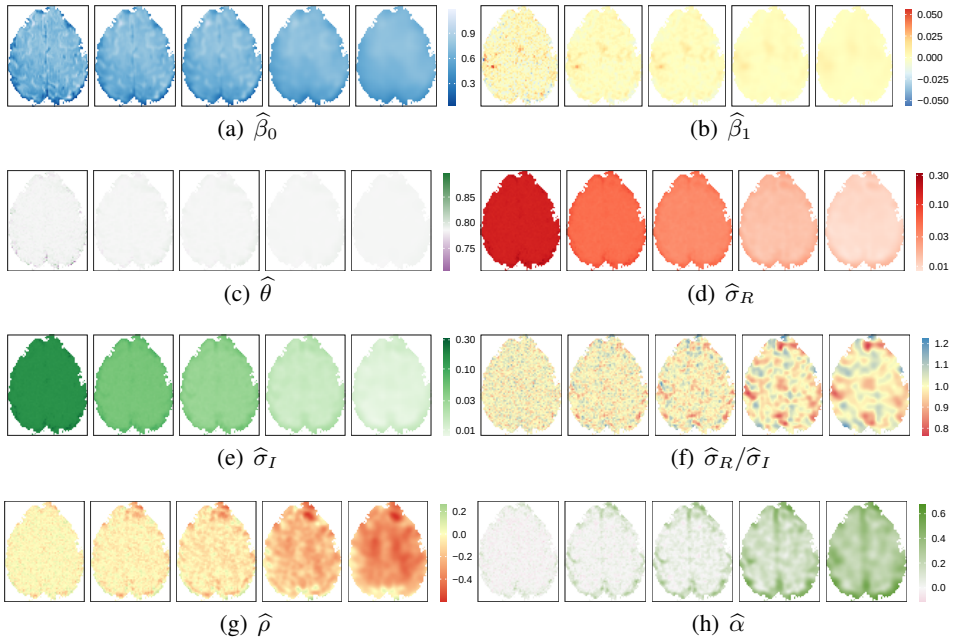


FIG S-21. *Images of the CVNS model parameter estimates for the second slice of (moving left to right within each subfigure) the unsmoothed data and the smoothed datasets with FWHMs of 1.5, 2, 4, and 6 voxels.*

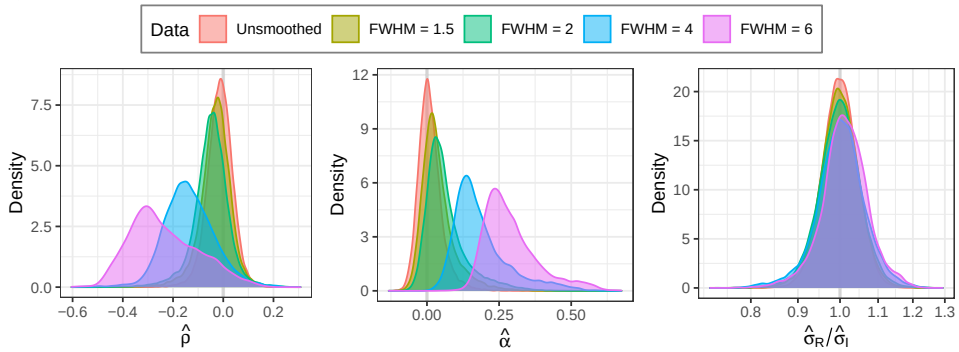


FIG S-22. *Density estimates of the distributions of the CVNS model MLEs of (left to right) $\widehat{\rho}$, $\widehat{\alpha}$, and $\widehat{\sigma}_R/\widehat{\sigma}_I$ for the five different datasets with different smoothing levels.*

correlation in the dataset, both real/imaginary correlation and temporal correlation, as the distributions of $\widehat{\rho}$ and $\widehat{\alpha}$ become more negative and positive, respectively, more so as the FWHM increases. However, the frequency distribution of $\widehat{\sigma}_R/\widehat{\sigma}_I$ is not much affected by spatial smoothing, still being closely centered around one, although the spatial distribution is more smooth.

S-4.3. *Supplement to Section* **??** *– phase-only activation.* Here, we present the results of applying the phase-only data-based model methodology described in S-2.9 to the finger-tapping dataset. Figure S-23 shows the task-related phase acti-

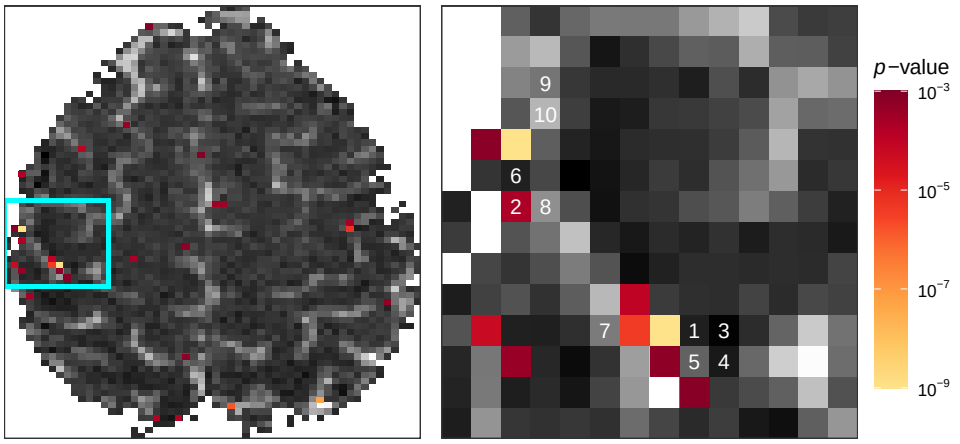


FIG S-23. *(Left) Phase-only data-based activation map of slice 2 of the unsmoothed dataset. (Right) Map zoomed in to same ROI around the left central sulcus as in Figure* **??***; the numbers 1-10 are the same as in Figure* **??***, showing the voxels with the smallest p-values for task-related magnitude activation (which interestingly do not often coincide with the voxels showing task-related phase activation).*

vation map for slice 2 of the unsmoothed dataset. Comparing to the four maps showing in the task-related magnitude activation in Figure **??**, the phase-only data-based map also identifies the left central sulcus as the region of primary activation. However, a closer look shows that very few of the magnitude-activated voxels in **??** (which are also labeled by the numbers 1-10 in Figure S-23) and the phase-activated voxels are in common. It is interesting to note that magnitude- and phase-activated voxels are often neighbors. This would fit the "brain or vein" paradigm previous discussed, in which phase-activation occurs in draining veins located close to the grey matter regions of interest.

Figure S-24 shows the phase-only data-based activation maps for slice 2 of the

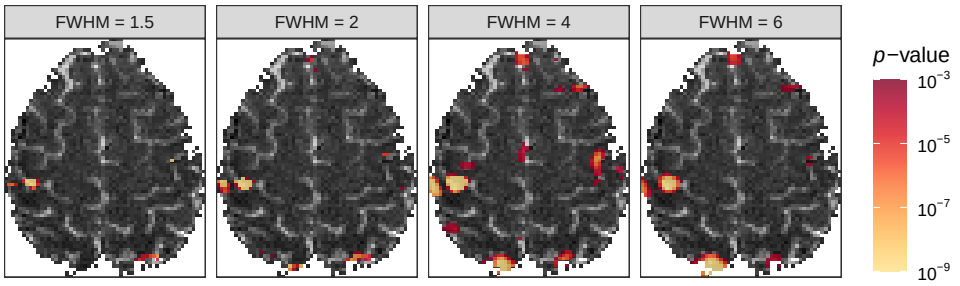


FIG S-24. *Phase-only data-based activation maps for slice 2 of the smoothed datasets.*

smoothed datasets. As in the magnitude-activation maps of Figure **??**, these phase-activation maps prominently identify the left central sulcus. It could even be argued that the latter due so more clearly.

References.

AALO, V. A., EFTHYMOGLOU, G. P. and CHAYAWAN, C. (2007). On the Envelope and Phase Distributions for Correlated Gaussian Quadratures. *IEEE Communications Letters* **11** 985-987.

ABRAMOWITZ, M. and STEGUN, I. (1965). *Handbook of Mathematical Functions*. Dover Publications.

ADRIAN, D. W., MAITRA, R. and ROWE, D. B. (2018). Complex-valued time series modeling for improved activation detection in fMRI studies. *Annals of Applied Statistics* **12** 1451-1478.

AKAIKE, H. (1973). *2nd International Symposium on Information Theory* Information theory and an extension of the maximum likelihood principle 267-281. Akademiai Kiado.

CASELLA, G. and BERGER, R. L. (2002). *Statistical Inference*, Second edition ed. Thomson Learning.

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.

FISHER, N. I. and LEE, A. J. (1992). Regression Models for an Angular Response. *Biometrics* **48** 665-677.

HASTIE, T., TIBSHIRANI, R. and FRIEDMAN, J. (2009). *The elements of statistical learning: data mining, inference and prediction*, 2 ed. Springer. Available at http://www-stat.stanford.edu/~tibs/ElemStatLearn/

MARDIA, K. V. and JUPP, P. E. (2000). *Directional Statistics*. Wiley.

MCLACHLAN, G. J. and KRISHNAN, T. (2008). *The EM Algorithm and Extensions*. Wiley.

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

ORCHARD, T. and WOODBURY, M. A. (1972). A missing information principle: theory and applications. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* **1** 697-715. University of California Press, Berkeley, California.

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

R CORE TEAM, (2020). R: A Language and Environment for Statistical Computing, Vienna, Austria.

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

SCHWARZ, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics* **6** 461-464.

WATSON, G. N. (1948). *A Treatise on the Theory of Bessel Functions*. Cambridge University Press.

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.

DEPARTMENT OF STATISTICS
GRAND VALLEY STATE UNIVERSITY
ALLENDALE, MI 49401-9403
USA
E-MAIL: adriand1@gvsu.edu

DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IA 50011-1090
USA
E-MAIL: maitra@iastate.edu

DEPARTMENT OF MATHEMATICAL AND STATISTICAL SCIENCES
MARQUETTE UNIVERSITY
MILWAUKEE, WI 53233
USA
E-MAIL: daniel.rowe@marquette.edu