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 2 – Further description of the dataset. Figure S-1 shows plots of the real, imaginary, magnitude, and phase time series at a single voxel, at one of the voxels showing the most activation. It also displays the time course of the "on-off" fingertapping and the expected BOLD response obtained by convolving this 0/1 stimulus time course with the Glover (1999) hemodynamic response function (HRF). Although this BOLD response is present in the data, trends in the time series (also known as also scanner drift) are also present which must be accounted for.

S-1.1. *Detrending scanner drift.* Detrending scanner drift is a common preprocessing step in the statistical analysis of fMRI time series. Two sources of this 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 S-2 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. 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.[1] 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

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

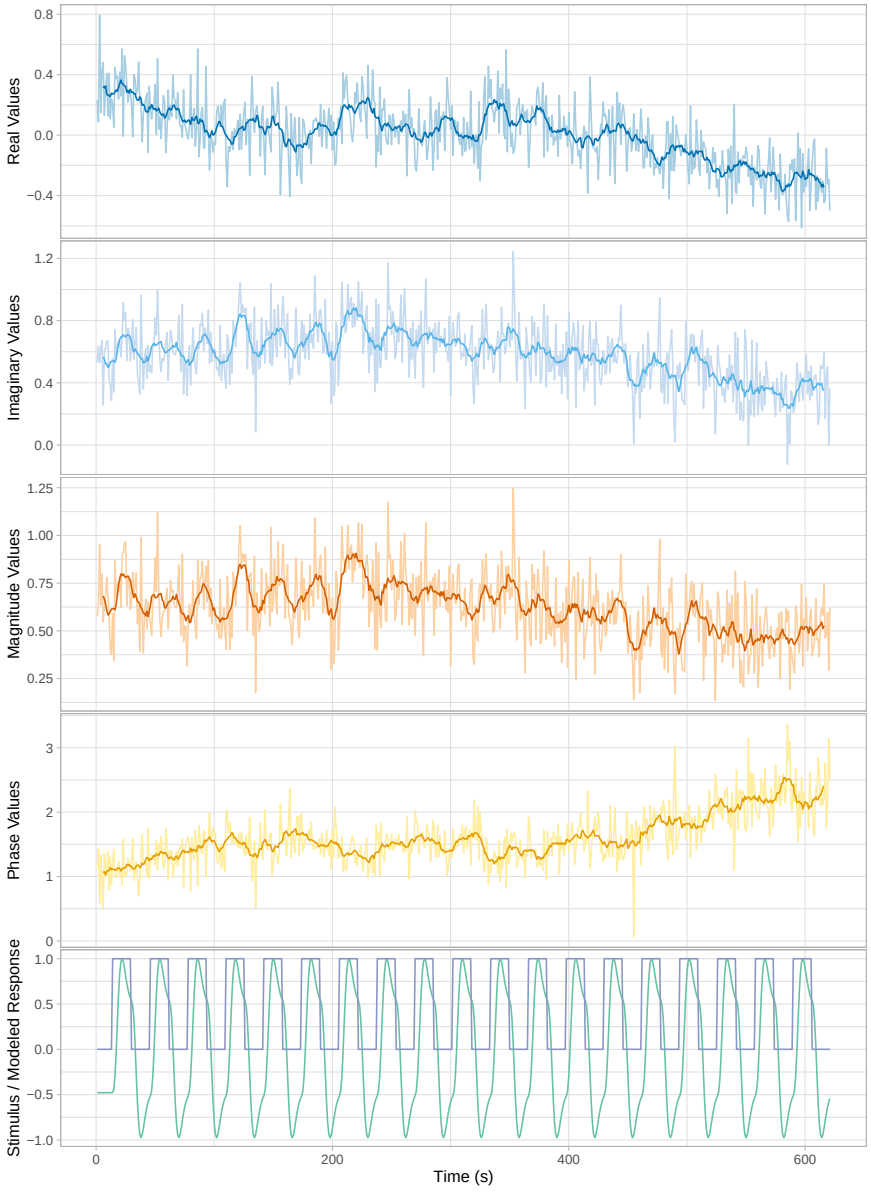
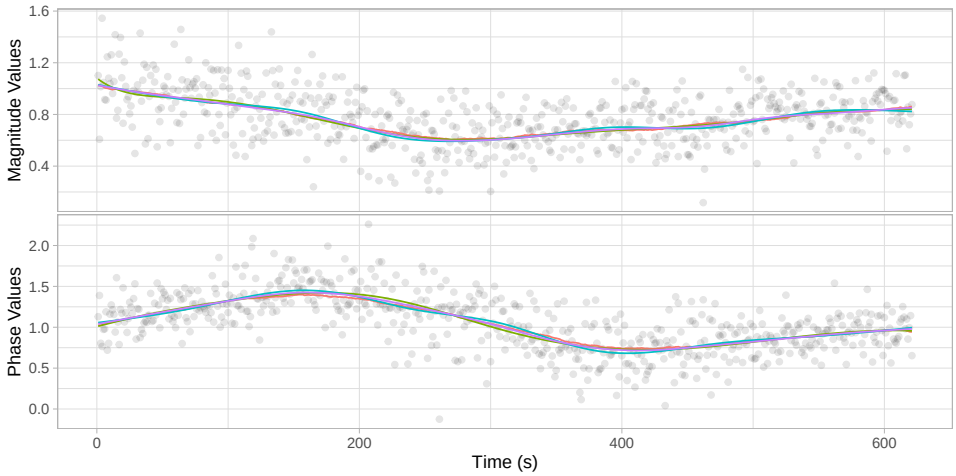


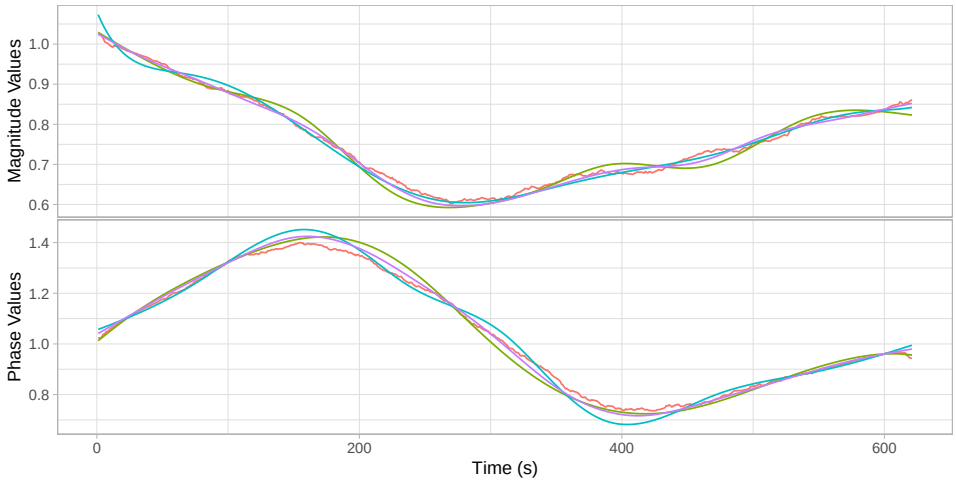
FIG S-1. *Time series of (from top) the real, imaginary, magnitude, and phase observations for one of the voxels showing the most activation. Lighter lines in each display represents the raw time series, while the darker lines show the result after applying a simple, central moving average filter with 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.*

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



(a) Raw data and fitted curves shown



(b) Only fitted curves shown

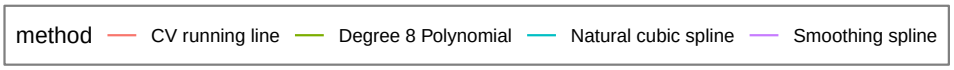


FIG S-2. *Comparison of four curve fitting methods on a selected voxel's time series for the purpose of detrending to correct scanner drift: CV running line (Adrian, Maitra and Rowe, 2018), degree 8 polynomial, natural cubic spline with 6 evenly spaced knots, smoothing spline with 8 effective degrees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

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.

For clarity, we 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 $\breve{y}_{\xi t} = y_{\xi t} - \widehat{y}_{\xi t} + \bar{\widehat{y}}_{\xi}$, for $\xi = R, I$, where $\bar{\widehat{y}}_{\xi}$ is the mean of the fitted values across the corresponding time series. (The same approach applied to magnitude-only data produces the magnitude-only detrended data.)

S-1.2. *Graphical summaries of SNR and CNR.* Figure S-3 shows spatial and frequency distributions of the signal-to-noise ratio (SNR) and the contrast-to-noise ratio (CNR) of the dataset. The SNRs and CNRs above were calculated as $\widehat{\beta}_0/\widehat{\sigma}$

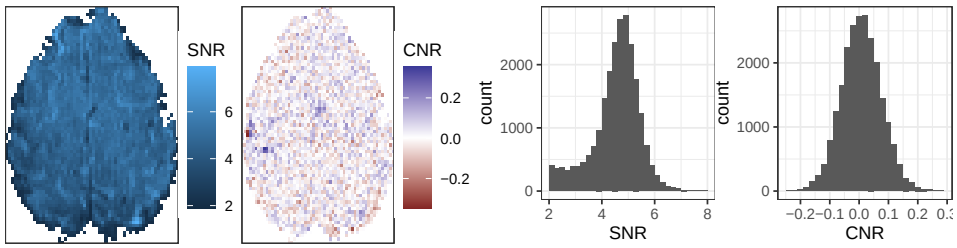


FIG S-3. *Graphical summaries of SNR and CNR: (left) images of slice 2; (right) frequency distributions*

and $\widehat{\beta}_1/\widehat{\sigma}$, respectively, where the previous parameter estimates were calculated from the magnitude voxel time series using ordinary least squares regression with an intercept term and the expected BOLD response in Figure S-1.

S-2. Supplement to Section 3 – Further Methodological Development.

S-2.1. *CV and MOG model methodology.*

S-2.1.1. *CV 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 $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0}$ vs. $H_a : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{0}$ 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 $\sigma^2\boldsymbol{R}_p = \mathrm{Cov}(\eta_{R1}, \ldots, \eta_{Rp}) = \mathrm{Cov}(\eta_{I1}, \ldots, \eta_{Ip})$, $\boldsymbol{R}_p^{-1}$ 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 $m = \mathrm{rank}(\boldsymbol{C})$.

S-2.1.2. *MOG model.* The log-likelihood function for the MOG model is given by $\log f(\boldsymbol{r};\boldsymbol{\tau}) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|\boldsymbol{R}_n| - \frac{1}{2\sigma^2}(\boldsymbol{r} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{R}_n^{-1}(\boldsymbol{r} - \boldsymbol{X}\boldsymbol{\beta})$, where $\boldsymbol{R}_n$ is such that $\sigma^2\boldsymbol{R}_n = \mathrm{Cov}(\boldsymbol{\epsilon})$. 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 $0 \le i, j \le p$, 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 $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{0}$ vs. $H_a : \boldsymbol{C}\boldsymbol{\beta} \neq \boldsymbol{0}$ 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 $\sigma^2\boldsymbol{R}_p = \mathrm{Cov}(\epsilon_1, \ldots, \epsilon_p)$.

S-2.2. *Relationships between CV, MOR, and MOG model densities.* We illustrate some relationships between the probability density functions (PDFs) of the CV, MOR, and MOG models for the observations at a single voxel and time-point. 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 CV model PDF.
2. **Fact 2**: For large SNR, the MOR model PDF approaches the MOG model PDF.

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

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

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-4 shows that the Ricean and Gaussian PDFs converge with increasing SNR.

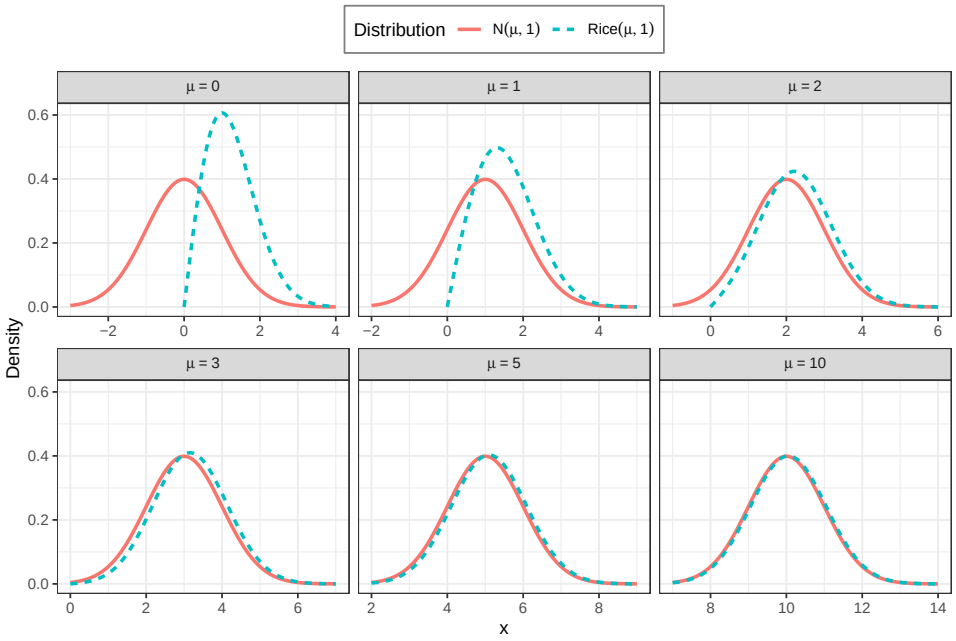


FIG S-4. *As the SNR (here, μ) increases, the Ricean and Gaussian PDFs converge.*

S-2.3. *Supplement to Section 3.2.1 – Methodological details of the EM algorithm corresponding to the AR(p) Ricean model.*

S-2.3.1. *Supplement to Section 3.2.1 – 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 $\mathbb{E}_{\phi_t|r_t;\boldsymbol{\tau}^{(k)}}[\cos(\phi_t - \theta)] = \mathbb{A}(r_t\mu_t^{(k)}/\gamma_0^{(k)})$, $t = 1, \ldots, n$.

S-2.3.2. *Supplement to Section 3.2.1 – 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 (5). 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 (5).

S-2.3.3. *Supplement to Section 3.2.1 – 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 2.2, 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-2.3.4. *Supplement to Section 3.2.2 – 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 (10). We denote the elements of $\boldsymbol{s}(r_t;\boldsymbol{\tau})$ that come from taking the partial derivatives in (9) 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 $d_t(i,j) = r_{t-i}r_{t-j}E_{t-i,t-j} - \mu_{t-i}u_{t-j} - \mu_{t-j}u_{t-i} + \mu_{t-i}\mu_{t-j}$, $0 \le i,j \le p$, 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-2.3.5. *Supplement to Section 3.2.2 – Computation time.* 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-5 shows the computation times required for parameter estimation of 1000 simulated time series under

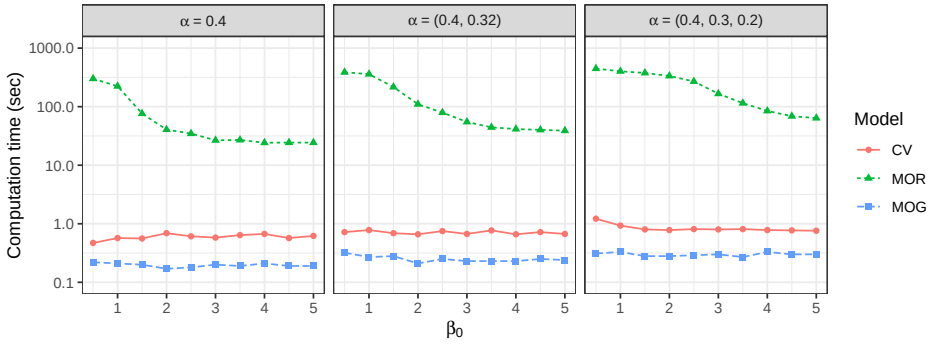


FIG S-5. *Computation times (sec) for parameter estimation per 1000 simulated time series under the three models. The EM algorithm used by the MOR model has much slower convergence than the other estimation schemes.*

the three models. We generated these time series under the CV model with $\boldsymbol{X}$ matrix described in Section 2.2, $\sigma^2 = 1$, $\boldsymbol{\beta} = (\beta_0, 0)'$ 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.

S-2.4. *Supplement to Section 3.3 – Further details regarding test statistics.*

S-2.4.1. *Supplement to Section 3.3 – 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 2. 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-3), 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,

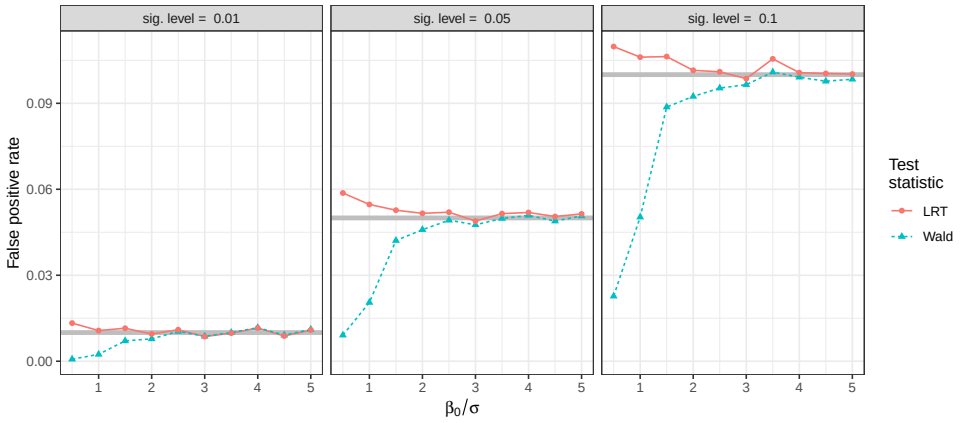


FIG S-6. *The false positive rates of the Wald test statistics fall sharply below the significance levels (shown by the thick grey lines) for β_0/σ values below 2, showing that LRT statistic is more reliable for low SNRs.*

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-2.4.2. *Supplement to Section 3.3 – Ricean AR(1) LRT statistic.* Here, we derive the expression for $f(r_t|r_{t-1};\boldsymbol{\tau})$ in (12). 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

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

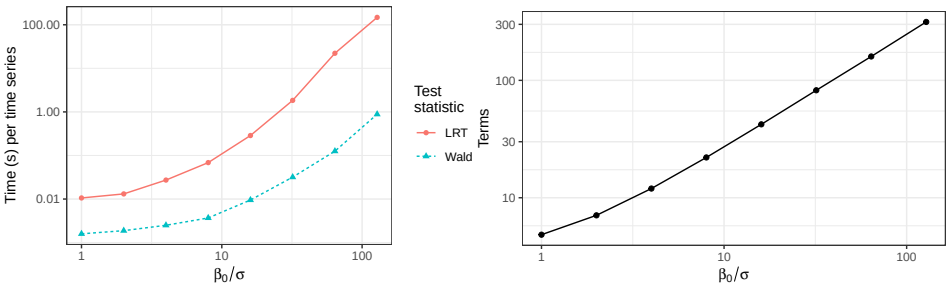
where $\omega_m = 1$ for $m = 0$ and $\omega_m = 2$ for $m \geq 1$. Also, using the result $\int_0^{2\pi} \cos[m(\phi_1 - \theta)] \exp[C_1 \cos(\phi_1 - \theta)] = \mathbb{I}_m(C_1)$, for $m \geq 0$ (Mardia and Jupp, 2000), we obtain $f(r_1, r_2)$. Dividing the result by (S-9), we obtain

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

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-2.4.3. *Supplement to Section 3.3 – 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 (12) 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 2.2, 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 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 (12) 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).



(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.)*

S-2.5. *Supplement to Section 3.4 – 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 3.4 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 CV model with $\boldsymbol{X}$ matrix described in Section 2.2, $\boldsymbol{\beta} = (5, 0)'$, $\sigma^2 = 1$, 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 $\delta = 0.01$ under the three models are shown in Figure S-8.[2]

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 $\log n = 6.43$ in this case 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 data-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 the CV model-based method. This difference can be attributed to the fact that CV model has twice the amount of data, which gives it more power in the sequential testing method.

S-3. Supplement to Section 4 – Further Simulation-based analyses.

S-3.1. *Supplement to Section 4.1 – Properties of parameter estimates.* Figure S-9 expands on the results in Figure 2, summarizing the biases, standard errors, and

[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.

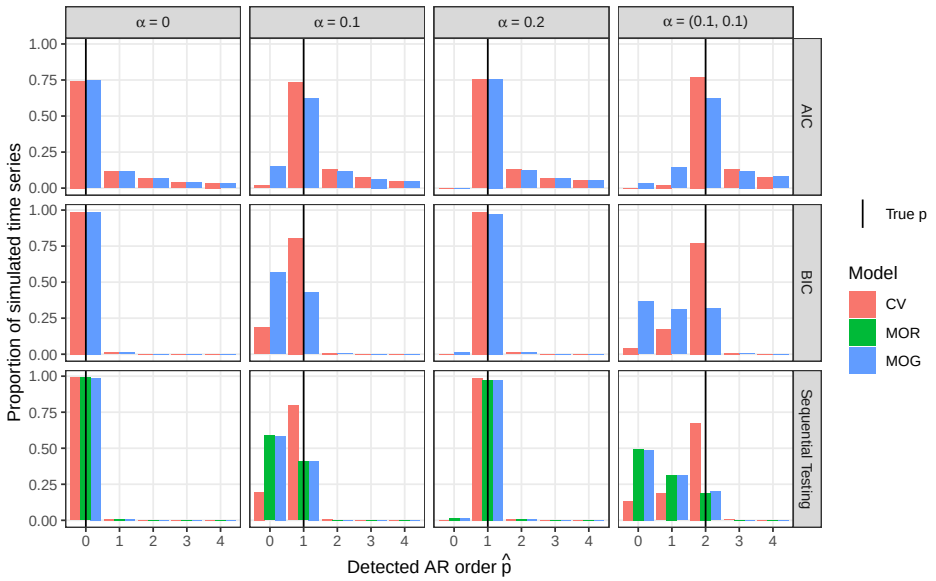


FIG S-8. *Proportion of simulated time series in which each AR order (0-4) is detected for four different sets of AR coefficients (columns), based on the AIC, BIC, and sequential testing methods (rows) and the three models (colored bars). Note that MOR model results are not shown for the AIC and BIC because the MOR model log-likelihood is not tractable for general p.*

root mean squared errors of $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\sigma}^2$, and $\widehat{\alpha}$ under the three models for simulated values of $\alpha = 0.2$, 0.4, 0.6, and 0.8. (Figure 2 only shows $\alpha = 0.4$.) Qualitative comparisons of the parameter estimate properties across models (which models are better/worse) are consistent across different values of α.

S-3.1.1. *Supplement to Section 4.2 – 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,

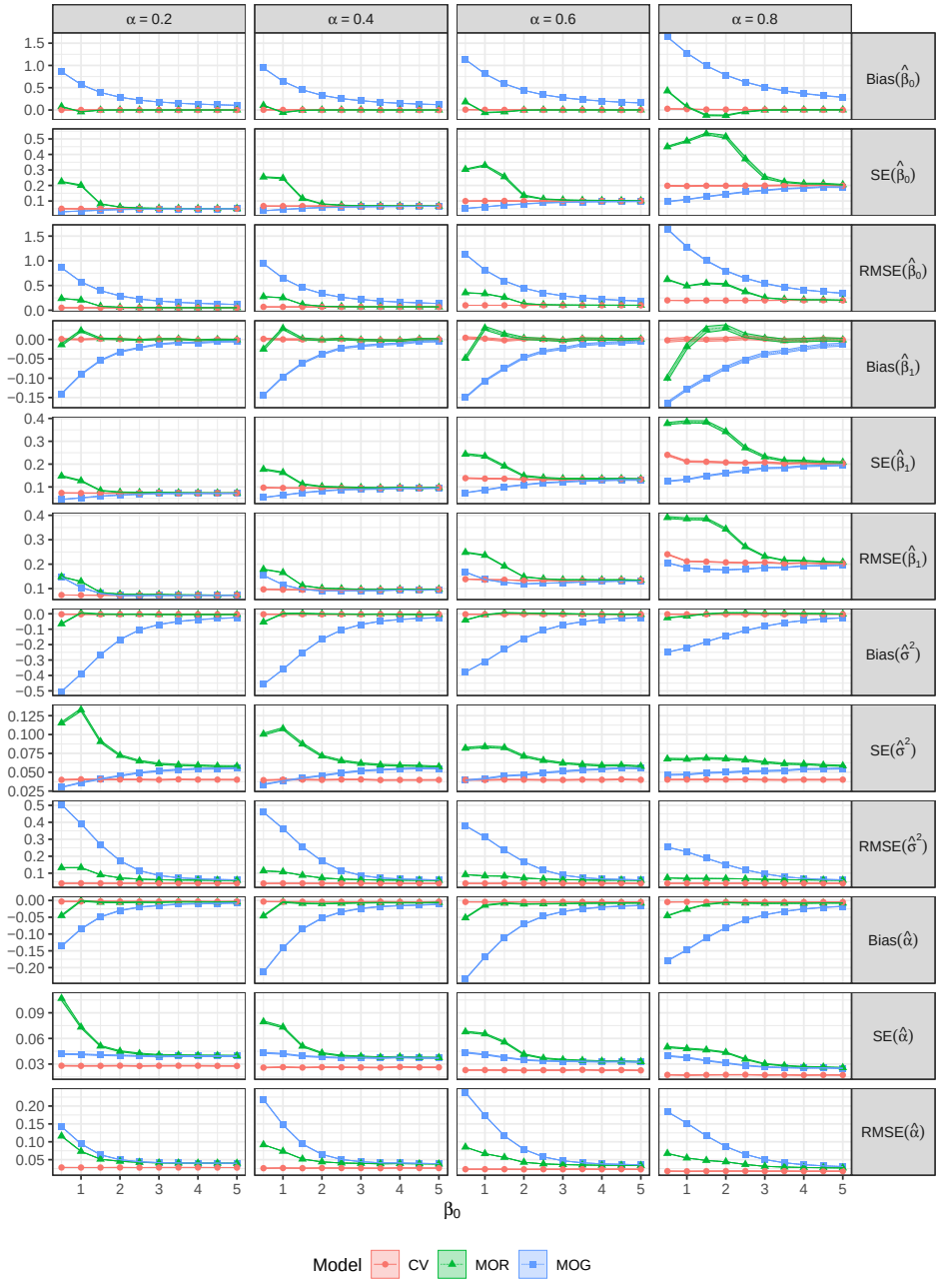


FIG S-9. *Biases, SEs (SE), and RMSEs (RMSE) of* $\widehat{\beta}_0$, $\widehat{\beta}_1$, $\widehat{\sigma}^2$, *and* $\widehat{\alpha}$ *under three models for simulated time series under* $\beta_1 = 0.2$, *and different values of* β_0 *and* α. *The shaded areas cover the estimate* $\pm$ *2 standard error intervals.*

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,

$$\log L(\boldsymbol{\tau}; r) = \log L_c(\boldsymbol{\tau}; r, \phi) - \log k(\phi|r; \boldsymbol{\tau}). \tag{S-19}$$

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:

$$\mathscr{I}(\boldsymbol{\tau}; r) = \mathscr{I}_c(\boldsymbol{\tau}; r) - \mathscr{I}_m(\boldsymbol{\tau}; r). \tag{S-20}$$

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

$$\log L(\boldsymbol{\tau}; r) = -\log \sigma^2 - \frac{r^2 + \mu^2}{2\sigma^2} + \log \mathbb{I}_0\left(\frac{\mu r}{\sigma^2}\right) \tag{S-21}$$

$$\log L_c(\boldsymbol{\tau}; r, \phi) = -\log \sigma^2 - \frac{r^2 + \mu^2}{2\sigma^2} + \frac{\mu r}{\sigma^2}\cos(\phi - \theta) \tag{S-22}$$

$$\log k(\phi|r; \boldsymbol{\tau}) = -\log I_0\left(\frac{\mu r}{\sigma^2}\right) + \frac{\mu r}{\sigma^2}\cos(\phi - \theta), \tag{S-23}$$

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

$$\mathscr{I}_m(\mu; r) = r^2 - r\mathbb{A}(\mu r)/\mu - r^2\mathbb{A}^2(\mu r), \tag{S-24}$$

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 5. Note that the fraction of the complete-data (total) information provided by the magnitude-only data increases as the SNR increases.

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 5 and the higher computation times of the MOR model at low SNRs in Figure S-5. As the SNR increases, the proportion of missing information decreases and so does the computation time.

S-4. Supplement to Section 5 – Further Analysis of low-SNR Dataset.

S-4.1. *Supplementary figures.* The following figures are presented in the order to which they are referred in the main article. Figure S-10 shows the frequency distributions of the voxel-wise detected AR orders under the three models. The

FIG S-10. *Frequency distributions of detected AR orders for the fingertapping dataset under the three models.*

orders were detected using the sequential testing procedure described in Section 3.4 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-11 expands on the results of Figure 6, showing activation maps for slices 1, 2, and 3 with inset maps showing the left central sulcus ROI. (Figure 6 only shows slice 2.) These additional slices are shown to demonstrate that the maps shown of slice 2 in Figure 6 show more activation than the other slices, but the patterns of activation are somewhat similar across all slices. In addition, Figure S-12 provides a graphical representation of the p-values in Table 2.

Figure S-13 refers to our study of adding extra noise to the original unsmoothed 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 6. The results verify that adding 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. Also, Figure S-14 shows a graphical representation of the detection proportions in Table 3.

References.

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.

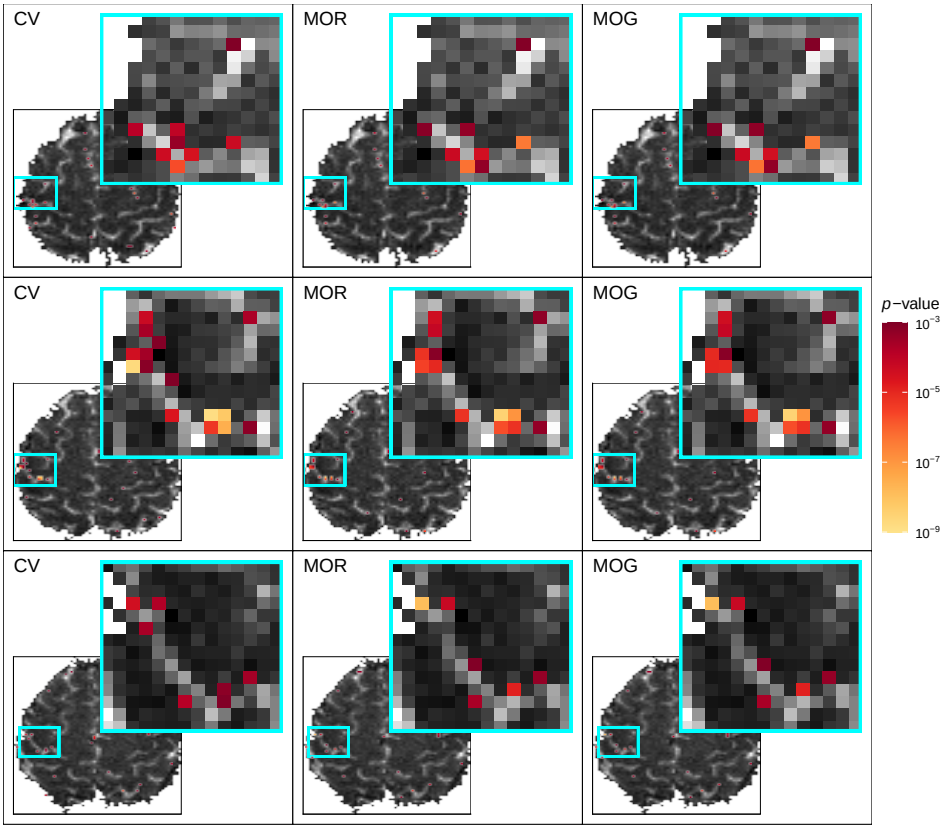


FIG S-11. *Activation maps for (by row, from top) slices 1, 2, and 3.*

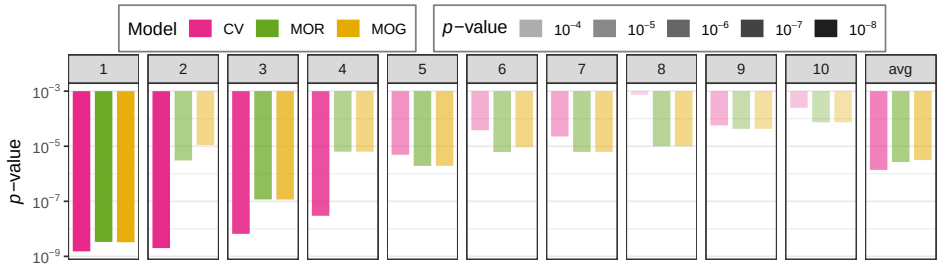


FIG S-12. *Graphical representation of the p-values in Table 2. (The shading represents the strength of the p-values.)*

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

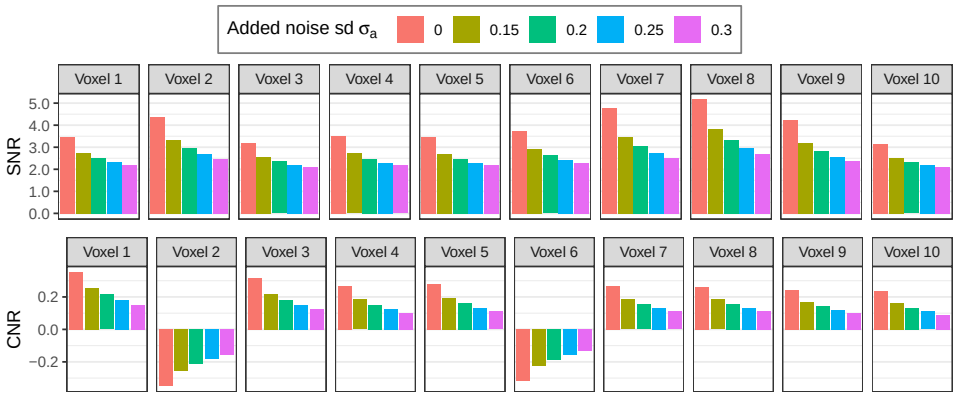


FIG S-13. *The SNRs and CNRs of the ten voxels identified in Figure 6 decrease when extra noise with standard deviation σ_a is added to the raw data ($\sigma_a = 0$ refers to raw data).*

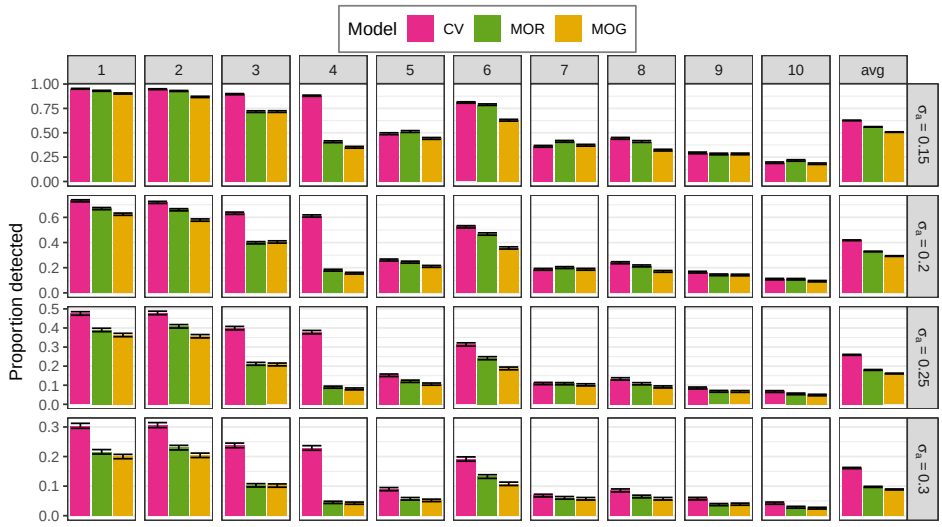


FIG S-14. *Proportion of replications the 10 voxels (and average proportion "avg") identified in Figure 6 (columns) detected as activated when noise with standard deviation σ_a (rows) was added to the raw data. The error bars show estimate $\pm$ 2 standard errors.*

data via the EM algorithm. *Journal of Royal Statistical Society Series B* **23** 1-38.

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

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

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.
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. Available at https://doi.org/10.1006/nimg.2002.1053
WATSON, G. N. (1948). *A Treatise on the Theory of Bessel Functions*. Cambridge University Press.

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