Some contributions to Ricean and complex-valued modeling of functional MRI time series

Daniel Wright Adrian
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/etd
Part of the Statistics and Probability Commons

Recommended Citation
http://lib.dr.iastate.edu/etd/10318

This Dissertation is brought to you for free and open access by the Graduate College at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Some contributions to Ricean and complex-valued modeling of functional MRI time series

by

Daniel Wright Adrian

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Ranjan Maitra, Major Professor
Ulrike Genschel
William Meeker
Dan Nettleton
Daniel Rowe
Stephen Vardeman

Iowa State University
Ames, Iowa
2011

Copyright © Daniel Wright Adrian, 2011. All rights reserved.
DEDICATION

I would like to dedicate this dissertation to Julie, without whose love and support I would not have been able to complete this work. I would also like to thank my friends and family for their loving guidance during the writing of this work.
# TABLE OF CONTENTS

LIST OF TABLES ................................................................. vi

LIST OF FIGURES .............................................................. viii

ABSTRACT ................................................................. xi

CHAPTER 1. INTRODUCTION .............................................. 1

CHAPTER 2. IMPROVED ACTIVATION DETECTION VIA COMPLEX-VALUED AUTOREGRESSIVE MODELLING OF FMRI VOXEL TIME SERIES ................................................................. 3

Abstract ................................................................. 3

2.1 Introduction ................................................................. 3

2.2 Detecting Activation in a Finger-Tapping Experiment ........................................... 8

2.3 Methodological Development ................................................................. 9

2.3.1 Autoregressive modeling for complex-valued fMRI time series data ................ 10

2.3.2 Prewhitening-based approaches to calculating activation statistics ................... 12

2.3.3 Choosing the order of the autoregressive model ............................................ 13

2.3.4 Detecting voxels significantly activated by the stimulus ................................... 14

2.4 Experimental Evaluations ................................................................. 15

2.4.1 Complex-valued/magnitude-only activation detection at low SNR .................... 17

2.4.2 AR order detection errors and their consequences on activation detection .......... 18

2.4.3 Investigating bias in significance levels for prewhitening- and likelihood-based activation statistics ................................................................. 20

2.5 Application to fMRI dataset ................................................................. 23

2.6 Discussion ................................................................. 25
CHAPTER 3. SUPPLEMENT TO “IMPROVED ACTIVATION DETECTION VIA COMPLEX-VALUED AUTOREGRESSIVE MODELLING OF FMRI VOXEL TIME SERIES” ......................................................... 28
3.1 MLEs of parameters under Null Models .................................................. 28
  3.1.1 Restricted MLEs under magnitude-only autoregressive model .............. 28
  3.1.2 Restricted MLEs under complex-valued autoregressive model .............. 28
3.2 Independence of real and imaginary residuals .......................................... 29
3.3 Determining simulation parameters from fMRI dataset ............................ 30
3.4 Validation of AR order detection procedures from simulated voxel time series .. 30
3.5 Diagnostics for checking model assumptions ........................................... 33

CHAPTER 4. ON THE USE OF GAUSSIAN AND RICE DISTRIBUTIONS FOR FITTING MAGNITUDE FMRI TIME SERIES DATA ......................................................... 36
Abstract ........................................................................................................ 36
4.1 Introduction .............................................................................................. 36
4.2 Methodological Development ................................................................. 38
  4.2.1 Models for magnitude fMRI time series ............................................ 39
  4.2.2 Methods for evaluating activation statistics ....................................... 42
4.3 Experimental Evaluations ........................................................................ 44
  4.3.1 Parameter estimation and computation times ...................................... 45
  4.3.2 Evaluation of activation tests ............................................................. 46
4.4 Discussion .................................................................................................. 50

CHAPTER 5. ESTIMATING PARAMETERS FOR RICE-DISTRIBUTED TIME SERIES OBSERVATIONS WITH APPLICATIONS TO FMRI DATA ................................................................................. 52
Abstract ........................................................................................................ 52
5.1 Introduction .............................................................................................. 52
5.2 Methodology ............................................................................................ 55
  5.2.1 Parameter estimation via the EM algorithm ....................................... 56
5.2.2 Calculation of standard errors and test statistics .................. 59
5.2.3 Gaussian Autoregressive model ........................................ 60
5.3 Experimental Evaluations .................................................. 60
5.4 Detecting Activation in a Finger-Tapping Experiment .................... 65
5.5 Discussion ................................................................. 66
5.6 Appendix ................................................................. 67
  5.6.1 The von-Mises distribution .............................................. 67
  5.6.2 Derivation of Monte Carlo approximation ............................. 68
  5.6.3 Calculation of empirical information matrix .......................... 68

BIBLIOGRAPHY ............................................................... 70
## LIST OF TABLES

2.1 Summary of parameters used in the two contexts of simulation experiments. In the above, $\alpha^* = (0.17, 0.45, -0.11, -0.23)$ and the voxel group abbreviations “Back.”, “Nonact.”, and “Activ.” represent background, non-activated, and activated voxel groups, respectively.  

2.2 The proportions of simulated voxel time series detecting each AR order $\hat{p}$ under the complex-valued and magnitude-only model order detection procedures introduced in Section 2.3.3. The true order of 4 is shown in bold. Results are reported under both PCER and FDR thresholding and the PACF and LRT order detection test statistics.

3.1 The number of voxels detecting each AR order $\hat{p}$ for the finger-tapping dataset inside and outside the brain, using PACF/LR test statistics and PCER/FDR thresholding.

3.2 Proportion of simulated AR(0) and AR(1) complex-valued time series detecting orders $\hat{p} = 0, 1$, and greater than 1 for different values of $\alpha_1$ under the LRT/PACF order detection procedures with PCER level $\delta = 0.05$.

4.1 Summary of the models and LRT statistics presented in Section 4.2.1.

4.2 Computation times (in seconds) for 10,000 LRT statistics under each model.

4.3 True detection rates of the different LRT statistics, according to an $\alpha = 0.05$ significance level, at different $\beta_0$ (or, equivalently, SNR) values.
4.4 Comparisons of $\hat{\tau}_G$, the AUCs of $\Lambda_G$, to those of the other LRT statistics, as measured by the $z$-statistics described in Section 4.2.2. The notation follows that of Table 4.1 – e.g. $z_R$ refers to the $z$-statistic calculated from comparing the AUCs of $\Lambda_G$ and $\Lambda_R$. Bold statistics represent significant differences at the $\alpha = 0.05$ level, after a Bonferroni adjustment. . . . . 51

5.1 Standard error estimates of the AR(1) Ricean estimates (a) $\hat{\beta}_1$ and (b) $\hat{\alpha}_1$ calculated from magnitude time series generated with $\beta_1 = 0.0, \alpha_1 = 0.0$, and various values of $\beta_0$. The standard error $SE_{\text{boot}}$ is the standard deviation of the MLEs obtained from simulation (a bootstrap estimate), and $SE_{\text{emp},0.10}$ and $SE_{\text{emp},0.90}$ refer to the 10th and 90th percentiles of the distribution of the empirical-information-based standard errors. . . 63
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Images of the (a) real, (b) imaginary, (c) magnitude, and (d) phase data for time points 5, 9, 13, 17, 21, 25, 29, and 33 (moving left to right), which represent the first complete 32-s cycle of the finger-tapping experiment, containing 16-s periods of tapping and rest.</td>
</tr>
<tr>
<td>2.2</td>
<td>Plots of activation detection rates of complex-valued and Gaussian-distributed magnitude model LRT statistics against SNR for (a) the single voxel simulation context, using a $\delta = 0.0005$ PCER level, and (b) the brain slice context using an FDR level $q^* = 0.05$. The CNR is 0.35.</td>
</tr>
<tr>
<td>2.3</td>
<td>Images of activation detection rate in (SNR, CNR)-space for the (a),(d) complex-valued and (b), (e) magnitude-only model LRT statistics and (c), (f) images of the difference (“complex minus magnitude”) of these rates. Simulations are performed in (a)-(c) the single voxel context with $\delta = 0.0005$ and (d)-(f) the brain slice context with $q^* = 0.05$.</td>
</tr>
<tr>
<td>2.4</td>
<td>ROC curves for LRT activation statistics based on assigned orders $\hat{p} = 0,1,2,3,4$ and complex-valued and magnitude-only models under (a) PCER and (b) FDR thresholding.</td>
</tr>
<tr>
<td>2.5</td>
<td>ROC curves for complex-valued- and magnitude-only-model LRT activation statistics based on detected AR orders under (a) PCER and (b) FDR thresholding.</td>
</tr>
<tr>
<td>2.6</td>
<td>Plots of bias in (a) PCER (or Type I error rate) and (b) FDR versus nominal values for prewhitening and likelihood-based, magnitude-only and complex-valued model LRT statistics.</td>
</tr>
<tr>
<td>Section</td>
<td>Description</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>2.7</td>
<td>Images of the detected AR orders under (a) complex-valued and (b) magnitude-only data approaches for the finger-tapping dataset, using the LRT statistic with FDR thresholding at a $q^* = 0.05$ level.</td>
</tr>
<tr>
<td>2.8</td>
<td>(a) Anatomical image of the subject’s brain displaying the central sulci (in green), which contain the sensori-motor finger area cortices. Activation maps of the (b) complex-valued and (c) magnitude-only model LRT statistics (overlayed on top of the same anatomical image), thresholded at the 5% false discovery rate. (Note that activation maps are drawn after masking out voxels outside the brain, as determined by the anatomical image.)</td>
</tr>
<tr>
<td>3.1</td>
<td>Images of the detected AR orders $\hat{p}$ for the dataset, using the complex-valued model order detection methodology described in Section 2.3.3, under PACF/LR test statistics and PCER/FDR thresholding.</td>
</tr>
<tr>
<td>3.2</td>
<td>Images of complex-valued AR(4) model MLEs for the finger-tapping data set.</td>
</tr>
<tr>
<td>3.3</td>
<td>Images of the PACFs computed from the real and imaginary residuals of the finger-tapping dataset for lags 1 to 20.</td>
</tr>
<tr>
<td>3.4</td>
<td>Quantile-quantile plot of Box-Pierce $Q_{20}$-statistics for independent- and AR(4)-model-fit residuals of the simulated and finger-tapping data (“empirical”) versus a random sample from the null distribution of $Q_{20}$ under an AR(4) model, $\chi^2_{16}$.</td>
</tr>
<tr>
<td>4.1</td>
<td>Integrals of Taylor, Gaussian, Ricean, and truncated normal PDFs over positive support for different signal parameters $\mu$ and noise parameter $\sigma^2 = 1.0$.</td>
</tr>
<tr>
<td>4.2</td>
<td>(a)-(c) Biases, (d)-(f) standard errors (SE), and (g)-(i) root mean squared errors (RMSE) of the unrestricted MLEs under each model plotted against SNR($=\beta_0$). The models are labeled in (a) as in Table 4.1.</td>
</tr>
</tbody>
</table>
4.3 False detection rates of the different LRT statistics, according to an
$\alpha = 0.05$ significance level, are plotted against SNR. The legend follows
Table 4.1. ................................................................. 48

4.4 The (theoretical) variance of the Rice$(\mu, 1.0)$ distribution plotted against
$\mu$ (or alternatively, SNR), with estimates of the middle 95% of the dis-
tributions of $\hat{\sigma}_G^2$ (obtained from simulation) at $\mu = 0.0, 0.5, \ldots, 6.0$. A
horizontal line at $\sigma_G^2 = \sigma_R^2 = 1.0$ is given for comparison. ............... 49

5.1 Biases of the AR(1) Ricean- and Gaussian-model parameter estimates
are plotted against $\beta_0$ (or alternatively, SNR) for different values of $\alpha_1$. 62

5.2 False detection rates at a 0.05 significance level for the (a) activation
and (b) order detection statistics are plotted against $\beta_0$. ....................... 64

5.3 Areas under the ROC curve (AUCs) for the (a) activation and (b) order
detection statistics are plotted against $\beta_0$, where “*” denotes a statisti-
cally significant difference at the 0.05 significance level after a Bonferroni
adjustment. .............................................................. 64

5.4 Images of the magnitude data for time points 5, 9, 13, 17, 21, 25, 29, and
33 (moving left to right), which represent the first complete 32-s cycle
of the finger-tapping experiment, containing 16-s periods of tapping and
rest. .............................................................. 65

5.5 Images of the detected AR orders under (a) Ricean and (b) Gaussian
models overlayed on a contour plot of brain anatomy. ....................... 66

5.6 Images of the $q$-values associated with the voxelwise activation statistics
under (a) Ricean and (b) Gaussian models overlayed on a contour plot
of brain anatomy. .............................................................. 67
ABSTRACT

Although it is well-known that data from functional magnetic resonance imaging (fMRI) experiments are complex-valued as a result of Fourier reconstruction, the vast majority of statistical analyses focus only on the magnitudes of these complex-valued measurements and discard the phase information. Moreover, most “magnitude-only” analyses rely on a Gaussian-approximation to the Ricean-distributed magnitudes, which is not (even approximately) valid at low signal-to-noise ratios (SNRs). As a result, we advocate use of the entire complex-valued data in statistical modeling and extend the complex-valued-data model in Rowe and Logan (2004) by applying AR(p) dependence to the real and imaginary errors. Based on this complex-valued model, we develop a likelihood-ratio test (LRT) for detecting activated brain voxels (or volume elements) which outperforms an LRT based on a Gaussian-assumed AR(p) magnitude-only model for simulated and experimental data. For existing fMRI datasets with unrecoverable phase information, we advocate Ricean modeling of the magnitude data; to this end, we compare the performance of activation tests based on Ricean and Gaussian magnitude-only models. In addition, we develop tests based on an “AR(p) Ricean” model that augments the observed magnitude data with missing phase data in an EM algorithm framework. Somewhat surprisingly, the Ricean-based activation tests perform similarly to their Gaussian-based counterparts, even at low SNRs, which further supports the use of complex-valued data.
CHAPTER 1. INTRODUCTION

This dissertation contributes to the study of statistical modeling of functional magnetic resonance imaging (fMRI) data. Functional MRI is commonly used to study brain function because it is noninvasive, requires no exposure to radiation, and is widely available. The primary goal of statistical analysis of fMRI data is detecting the brain region(s) activated by a given stimulus or task. This is commonly done in two steps: first, the time series of measurements at each voxel (or volume element) is reduced to a test statistic which summarizes the association between each voxel time course and the expected response to the stimulus. Second, the resulting map of statistics is thresholded to identify voxels that are significantly activated. We focus on the first of these steps, developing activation statistics based complex-valued and Ricean modeling of fMRI voxel time series.

The vast majority of statistical analyses study the magnitude data computed from the complex-valued measurements resulting from Fourier reconstruction. This practice, which has carried over from structural MRI, discards the phase information. However, noting that both components of the data are acquired, Nan and Nowak (1999) and Rowe and Logan (2004) encourage use of both the magnitude and the phase (i.e., complex-valued) data in the analysis and demonstrate that such analyses have greater activation detection power than “magnitude-only” analyses at low signal-to-noise ratios (SNRs). In a paper in progress (Chapter 2), we extend the complex-valued data model in Rowe and Logan (2004) by applying AR($p$) dependence to the real and imaginary error vectors, and we apply this model to fMRI data from a finger-tapping experiment. Chapter 3 is a supplement to this paper.

While we hope this will help spur the adoption of complex-valued methodology, we note that since the practice to date has been to rely on magnitude-only fMRI datasets, there are a large number of available datasets for which the phase information has been discarded. As a result,
we also consider magnitude time series, which are commonly assumed to follow a Gaussian general linear model. However, because the real and imaginary measurements are well-modeled as two independent Gaussian random variables with the same variance, the magnitude data follow the Rice distribution, which is well-approximated by the Gaussian distribution at high SNRs, but not so when the SNR is low. Consequently, Rice-distributed magnitude-only models (den Dekker and Sijbers, 2005; Rowe, 2005b; Zhu et al., 2009) have shown improved power of detection over Gaussian models at low SNR. In Chapter 4, we expand upon these previous studies of Ricean and Gaussian models for low-SNR fMRI magnitude time series.

In Chapter 5, we delve deeper into Ricean models by incorporating temporal dependence. Previously, such efforts have been complicated by the fact that time series modeling is largely based upon the Gaussian distribution, a “mismatch” under Ricean models. However, we bridge this gap by applying AR($p$) errors to the Gaussian-distributed real and imaginary components – similar to the model of Chapter 2, except that the complex-valued data are unobserved in this magnitude-only context. This “AR($p$) Ricean” model depends on augmenting the observed magnitude data by missing phase data in an EM algorithm framework. We use the EM algorithm for parameter estimation and extend it to compute approximate standard errors and test statistics for activation and AR order detection. We compare the performance of this new model to the standard Gaussian AR($p$) model and apply both to the finger-tapping dataset.
CHAPTER 2. IMPROVED ACTIVATION DETECTION VIA
COMPLEX-VALUED AUTOREGRESSIVE MODELLING OF FMRI
VOXEL TIME SERIES

A paper in preparation
Daniel W. Adrian, Ranjan Maitra, and Daniel B. Rowe

Abstract

A complex-valued model with \( AR(p) \) errors is proposed as an alternative to the more common Gaussian-assumed magnitude-only \( AR(p) \) model for fMRI time series. Likelihood-ratio-test-based activation statistics are derived for both models and are compared in terms of activation detection and false discovery rates for simulated and experimental data. For simulated data, the complex-valued \( AR(p) \) model likelihood-ratio activation statistic shows superior power of activation detection at low signal-to-noise ratios and lower false discovery rates. Also, when applied to an experimental data set, the activation map produced by the complex-valued \( AR(p) \) model more clearly identifies the primary activation regions. Our results advocate the use of the complex-valued data and the Gaussian \( AR(p) \) model as a more efficient and reliable tool in fMRI experiments over the current practice of using only the magnitude dataset.

2.1 Introduction

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

Each MR image is obtained in a series of steps from the so-called $k$-space data which encodes different frequency contributions to each voxel. The different frequencies result from magnetic field gradients (Jezzard and Clare, 2001) and need to be inverted to localize measurements at each voxel. This is achieved by applying the inverse Fourier transform (Jain, 1989) on the $k$-space data, which results in a complex-valued observation at each voxel and each time-point. Thus, acquired fMRI (and MR) data at each voxel and time-point can, in reality, be written in terms of its real and imaginary (alternately, magnitude and phase) components. The real and imaginary components of the acquired voxel-wise MR signal are well-modeled as two independent normal random variables with the same variance (Wang and Lei, 1994). This implies that the magnitude measurements follow the Rice distribution (Rice, 1944; Gudbjartsson and Patz, 1995), which is well-approximated by the normal distribution at high signal-to-noise ratio (SNR), but not so when the SNR is low.

Acquired MR datasets have typically used the magnitude measurements at each voxel for display and analysis. This practice of using only the magnitude data while discarding the phase at each voxel has carried over to fMRI practice so much so that the vast majority of statistical analyses of such data completely ignore the phase data and base their inferences on only the magnitude time series at each voxel (Rowe and Logan, 2004, 2005). Thus, even though additional (phase) information is available, analysis in fMRI has almost exclusively focused on the time series of the magnitude MR data at each voxel. Indeed, as we discuss in our review of current fMRI practice, many of the methods used in such analyses assume that the magnitude time series are normally distributed, even though such observations may not all be obtained at high SNR.

Under the framework outlined above, the general strategy is to fit, at each voxel, a model
— commonly a general linear model (Friston et al., 1995) – to the time series observations against a transformation of the input stimulus: this transformation is the expected BOLD response and is effectively modeled in terms of a convolution of the stimulus time course with the hemodynamic response function (HRF), which measures the delay and dispersion of the BOLD response to an instantaneous neuronal activation (Friston et al., 1994; Glover, 1999). This provides the setting for the application of the Statistical Parametric Mapping (SPM) technique of Friston et al. (1990) which was originally developed to analyze Positron Emission Tomography (PET) time course data, but which has since been extended to become one of the most popular approaches to analyzing fMRI data. The time series at each voxel is thus reduced to a test statistic at each voxel, which summarizes the association between each voxel time course and the expected BOLD response (Bandettini et al., 1993). The resulting map is then thresholded to identify voxels that are significantly activated (Worsley et al., 1996; Genovese et al., 2002; Logan and Rowe, 2004).

In its simplest form, the above analysis assumes no autocorrelation within the time series: however it is widely realized that this assumption is not supported in reality. There are many reasons for this: one is that the hemodynamic response disperses (or “smears”, in fMRI jargon) neural activation. The hemodynamic (or BOLD) response to a single neural activation takes 15 to 20 seconds (Lazar, 2008), which is much longer than the sampling intervals of many fMRI techniques – 100 ms-5 s for echo-planar imaging (EPI) techniques (Friston et al., 1994). Further, the neuronal response, which can be modeled as a point response or a delta function (Friston et al., 1994), is itself very fast when compared to the BOLD response. Since fMRI experiments measure the BOLD response over time, the above discussion means that the observed time series within each voxel are correlated. Friston et al. (1994) also contend that the neuronal process is composed of “intrinsic” neuronal activities in addition to the stimulus-related response. Consequently, the authors say, autocorrelations in the observed time series arise from two neural components, both measured through the hemodynamic response: one that is experimentally induced owing to the stimulus and another that is due to intrinsic neuronal activity. The first component is modeled by the convolution of the stimulus time course with the HRF, as discussed previously, while the second is present even in the absence of stimuli. Note also
that some additional sources of autocorrelation are also provided by the subject’s cardiac and respiratory cycles (Friston et al., 2000).

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

The above approaches all make Gaussian distributional assumptions for the observed magnitude time series, which as discussed before, is not appropriate, even approximately, at low SNR. This has led to the development of Rice-distributed magnitude-only models (den Dekker and Sijbers, 2005; Rowe, 2005b; Zhu et al., 2009) which have, understandably, shown improved power of detection over their Gaussian counterparts at low SNR. Incorporating autocorrelation directly in the Rice-based models is however complicated, and the prewhitening approaches discussed above do not apply since they are based on Gaussian-distribution-based extensions of the linear model.

A different approach, advocated by Nan and Nowak (1999) and Rowe and Logan (2004), encourages use of both the magnitude and the phase (i.e., complex-valued) data in the analysis. Noting that both components of the data are all acquired, just not used, these authors have also demonstrated that complex-valued statistical analyses of voxel time series show a greater power of activation detection than Gaussian-distribution-assumed magnitude-only (henceforth referred to as magnitude-only in this paper, unless otherwise specified) analyses at low SNRs. In simulation studies that assume independent errors, complex-valued models have shown increased detection power over magnitude-only models at low SNR (of less than 5, and sometimes
even as high as 7.5), and the two have shown comparable detection power at high SNR (Rowe and Logan, 2004). In addition, magnitude-only models yield biased parameter estimates at low SNR: even for large SNR, the variance of the residual variance estimates is twice that obtained with the complex-valued model (Rowe, 2005b). These results are due to two shortcomings of magnitude-only data analysis; first, half the data is discarded, which causes the larger variance of residual variance estimates under the magnitude-only model. Secondly, as mentioned earlier, the approximate Gaussian distributional assumption for Rice-distributed magnitude data is poor at low SNR. This factor is increasingly important because the SNR is proportional to voxel volume (Lazar, 2008). Thus an increase in the fMRI spatial resolution will correspond to a lowering of the SNR, making the Gaussian distributional approximation for the magnitude data even less tenable.

Using the phase data in addition to the magnitude data in fMRI studies has proved valuable in other ways. For one, prewhitening can occur without distributional approximation for complex-valued data models (Rowe and Logan, 2004). Time course data on the phase have also proved useful in identifying voxels that show unwanted BOLD response due to the presence of large draining veins (Hoogenrad et al., 1998; Menon, 2002). These voxels show a task-related phase change, while voxels with a more random orientation of blood vessels do not (Rowe, 2005a). Using a complex-valued model with constant phase assumption (Rowe and Logan, 2004) has proved successful in biasing against such voxels with large draining veins (Nencka and Rowe, 2007). In addition, the complex-valued data can permit the analysis of the original, $k$-space data (Rowe et al., 2009), before preprocessing induces spatial and temporal correlation artifacts (Nencka et al., 2009), which may cloud conclusions in functional connectivity and fMRI studies.

In this paper, we further develop the complex-valued time series analysis of fMRI data. Our showcase application is a dataset from a finger-tapping experiment introduced in Section 2.2. We use this application as the context within which we introduce methodology that applies an AR($p$) dependence structure to the real and imaginary error vectors of the model in Rowe and Logan (2004). We derive likelihood-based activation statistics based on this model in Section 2.3 and compute them for both simulated and real fMRI datasets in Sections 2.4 and
respectively. We also compute similar activation statistics under a Gaussian-distributed magnitude-only model with AR(\(p\)) errors. After applying thresholding procedures, we compare the performance of the two statistics in terms of detection probability and control of false positive and false discovery rates. We discuss these results in Section 2.6. This paper also has an online supplement providing further details on methodology, experimental illustrations, performance evaluations and data analysis, which serves as Chapter 3 of this document.

### 2.2 Detecting Activation in a Finger-Tapping Experiment

Our showcase application for this paper comes from a commonly-performed bilateral sequential finger-tapping experiment, as studied in Rowe and Logan (2004). In this case, the MR images were acquired while the (normal healthy male) volunteer subject was instructed to either lie at rest or to rapidly tap fingers of both hands (hence bilateral) at the same time. The fingers were tapped sequentially in the order of index, middle, ring and little fingers. The experiment consisted of a block design with 16 s of rest followed by eight “epochs” of 16 s tapping alternating with 16 s of rest. MR scans were acquired once every second, resulting in 272 images. For this dataset, the complex and imaginary components of the time series images were not discarded, but stored along with the magnitude image commonly used in traditional fMR analysis. Figure 2.1 shows images of the real, imaginary, magnitude and phase data at time points \(t = 5, 9, \ldots, 33\), which constitute the first 32-s cycle containing 16-second time periods of tapping and rest on a single axial slice through the motor cortex consisting of \(128 \times 128\) voxels. (Note that traditionally, only the magnitude images are used in fMRI analysis, while the phase images are discarded.) For simplicity, we restrict attention in this paper to this two-dimensional slice of the dataset. These images appear not to change much in time because, as explained in Section 2.1, the BOLD stimulus response is very small compared to the overall MR signal in all fMRI experiments. A dataset on a well-studied paradigm such as this provides us with as close to a “known” detected activation area as is possible in fMRI: numerous studies have confirmed activation in the sensori-motor finger area cortex in the central sulcus. Thus, this dataset provides us with an ideal case study for both developing and evaluating new methodology.
2.3 Methodological Development

We focus on the complex-valued time series at a voxel, which comprises of real and imaginary time series observations, respectively denoted in this paper as $\mathbf{y}_R = (y_{R1}, \ldots, y_{Rn})'$ and $\mathbf{y}_I = (y_{I1}, \ldots, y_{In})'$, with $n$ being the number of scans. For notational simplicity here, we suppress voxel-related subscripts, and denote the voxel-wise magnitude time series data as $\mathbf{r} = (r_1, \ldots, r_n)'$, where $r_t = \sqrt{y_{Rt}^2 + y_{It}^2}$, $t = 1, \ldots, n$. We first briefly discuss the magnitude-only model. In doing so, we also introduce broadly the setup of our experiment.

As discussed in Section 2.1, magnitude-only fMRI time series observations at a voxel are often analyzed by extending the linear model $\mathbf{r} = \mathbf{X}\beta + \epsilon$ where $\epsilon$ is assumed to be multivariate normally distributed with an AR($p$) dependence structure (Marchini and Ripley, 2000; Bullmore et al., 1996; Worsley et al., 2002). The design matrix $\mathbf{X}$ is of order $n \times q$ with columns
representing the baseline signal, signal drift, and the expected BOLD response. The AR($p$) distribution of $\epsilon$ is parameterized by AR coefficients $\alpha = (\alpha_1, \ldots, \alpha_p)$ and white noise variance $\sigma^2$. Under this setting, the log-likelihood function is given by $\log L(\alpha, \beta, \sigma^2 | r) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2} \log |R_n| - \frac{1}{2\sigma^2} (r - X\beta)'R_n^{-1}(r - X\beta)$, where $R_n$ is the $n \times n$ matrix such that $\sigma^2 R_n = \text{Cov}(\epsilon)$.

Unrestricted maximum likelihood estimates (MLEs) of the parameters $\beta$ and $\sigma^2$ are then given by $\hat{\beta} = (X'R_n^{-1}X)^{-1}X'R_n^{-1}r$ and $\hat{\sigma}^2 = (r - X\hat{\beta})'R_n^{-1}(r - X\hat{\beta})/n$, respectively, with $R_n^{-1}$ given as a function of $\hat{\alpha}$, i.e. as the MLE of $\alpha$ (Pourahmadi, 2001). We obtain $\hat{\alpha}$ by solving the system of equations: $\sum_{j=1}^p (d_{jk} + j\hat{\gamma}_{j-k})\hat{\alpha}_j = \hat{d}_{0k}$, for $k = 1, \ldots, p$, as in Miller (1995), where $\hat{d}_{ij} = \sum_{t=1}^{t-i-j} \hat{\epsilon}_{t+i}\hat{\epsilon}_{t+j}$, for $0 \leq i, j \leq p$, and $\hat{\gamma}_{k} = \hat{d}_{0k}/n$, $k = 0, \ldots, p - 1$, is the lag $k$ sample autocovariance. In the preceding discussion, $\hat{\epsilon}_t = r_t - x_t'\hat{\beta}$, where $x_t'$ is the $t$th row of $X$, $t = 1, \ldots, n$. The estimation procedure, due to Cochrane and Orcutt (1949), begins with $\hat{R}_n = I_n$, the identity matrix of order $n \times n$, and then iteratively updates $\hat{\beta}$, $\hat{\alpha}$, and $\hat{R}_n^{-1}$ until convergence.

A general hypothesis test for activation can be framed as $H_0 : C\beta = 0$ vs. $H_a : C\beta \neq 0$. (Note that this formulation of the alternative allows for “negative activation” in response to the fMRI stimulus/task at the voxel.) The likelihood ratio test (LRT) statistic is given by

$$-2 \log \lambda_M = n \log \left(\frac{\tilde{\sigma}^2}{\sigma^2}\right) - \log \left(\frac{|\tilde{R}_p^{-1}|}{|\hat{R}_p^{-1}|}\right),$$

where $\tilde{\sigma}^2$ and $\tilde{\alpha}$ are restricted MLEs under $H_0$ (see the derivations in Section 3.1.1) and $\hat{R}_p^{-1}$ and $\tilde{R}_p^{-1}$ are functions of $\hat{\alpha}$ and $\tilde{\alpha}$, respectively, as in Pourahmadi (2001). The null distribution of the LRT statistic (2.1) is asymptotically $\chi^2_m$ with $m = \text{rank}(C)$.

### 2.3.1 Autoregressive modeling for complex-valued fMRI time series data

Following Rowe and Logan (2004), our model for complex-valued fMRI voxel time series is

$$\begin{pmatrix} y_R \\ y_I \end{pmatrix} = \begin{pmatrix} X & 0 \\ 0 & X \end{pmatrix} \begin{pmatrix} \beta \cos \theta \\ \beta \sin \theta \end{pmatrix} + \begin{pmatrix} \eta_R \\ \eta_I \end{pmatrix},$$

(2.2)

This formulation means that the real and imaginary time series have phase-coupled means according to a central phase $\theta$, fixed in the time series but allowed to vary between voxels.

As before, the $n \times q$ design matrix $X$ contains columns to model baseline level, signal drift,
and expected BOLD response. The real and imaginary error vectors, \( \eta_R \) and \( \eta_I \), are assumed to be independent and Gaussian-distributed with \( \text{Cov}(\eta_R) = \text{Cov}(\eta_I) = \Sigma \). Rowe and Logan (2004) specify that \( \Sigma = \sigma^2 I_n \), assuming that existing correlations in the time series have been removed by the prewhitening procedure outlined in their paper. However, we assign an AR\( (p) \) process to the real and imaginary errors, with AR coefficients \( \alpha \) and white noise variance \( \sigma^2 \). Define \( R_n \) such that \( \sigma^2 R_n = \text{Cov}(\eta_R) = \text{Cov}(\eta_I) \). Under this framework, the log-likelihood function is given by

\[
\log L(\alpha, \beta, \theta, \sigma^2 | y_R, y_I) = -n \log \sigma^2 - \log |R_n| - h/2\sigma^2, \tag{2.3}
\]

Similar to the derivations in Rowe and Logan (2004), the unrestricted MLE for \( \beta \) is \( \hat{\beta} = \hat{\beta}_R \cos \hat{\theta} + \hat{\beta}_I \sin \hat{\theta} \), where \( \hat{\beta}_R = (X'\hat{R}_n^{-1}X)^{-1}X'\hat{R}_n^{-1}y_R \) and \( \hat{\beta}_I = (X'\hat{R}_n^{-1}X)^{-1}X'\hat{R}_n^{-1}y_I \), and \( \hat{R}_n^{-1} \) is again a function of \( \hat{\alpha} \) as in Pourahmadi (2001). The MLE for \( \theta \) is given by

\[
\hat{\theta} = \frac{1}{2} \arctan \left[ \frac{2\hat{\beta}_R'X'\hat{R}_n^{-1}X\hat{\beta}_I}{\hat{\beta}_R'X'\hat{R}_n^{-1}X\hat{\beta}_R - \hat{\beta}_I'X'\hat{R}_n^{-1}X\hat{\beta}_I} \right], \tag{2.4}
\]

while that for \( \sigma^2 \) is \( \hat{\sigma}^2 = \hat{h}/2n \), where \( \hat{h} \) replaces parameters by their MLEs in (2.3). We obtain \( \hat{\alpha} \) by solving the system of equations:

\[
\hat{d}_{0k} = \sum_{j=1}^{p} (\hat{d}_{jk} + 2j\hat{\gamma}_{j-k})\hat{\alpha}_j, \tag{2.5}
\]

for \( k = 1, \ldots, p \), which is a slight modification of the system obtained in the Gaussian-assumed magnitude-only time series observations earlier. Further, \( \hat{d}_{ij} = \sum_{t=1}^{n-1} \hat{\eta}_{R,t+i}^j \hat{\eta}_{R,t+j}^i \), 0 \( \leq i, j \leq p \) in (2.5), and \( \hat{\gamma}_k = \hat{d}_{0k}/2n \) is the lag-\( k \) sample autocovariance, \( k = 0, \ldots, p - 1 \). Also, \( \hat{\eta}_{R,t} = y_{R,t} - x_t^i \hat{\beta} \cos \hat{\theta} \) and \( \hat{\eta}_{I,t} = y_{I,t} - x_t^i \hat{\beta} \sin \hat{\theta}, t = 1, \ldots, n \). The ML estimation procedure thus consists of iteratively updating \( (\hat{\theta}, \hat{\beta}), \hat{\alpha}, \) and \( \hat{R}_n^{-1} \) successively, proceeding until convergence.

Activation tests can be framed in the same way as before, i.e. by positing \( H_0 : C\beta = 0 \) against \( H_a : C\beta \neq 0 \). The LRT statistic for the complex-valued AR\( (p) \) model is given by

\[
-2\log \lambda_C = 2n \log \left( \frac{\mathbf{\hat{\sigma}}^2}{\sigma^2} \right) - 2 \log \left( |\mathbf{\hat{R}}_p^{-1}| / |\mathbf{\hat{R}}_p^{-1}| \right), \tag{2.6}
\]
where \( \tilde{\alpha} \) and \( \tilde{\sigma}^2 \) are restricted MLEs obtained under \( H_0 \) and derived in Section 3.1.2. Under \( H_0 \), the LRT statistic is again asymptotically \( \chi^2_m \). Note also that the matrices \( \tilde{R}_p^{-1} \) and \( \hat{R}_p^{-1} \) are functions of \( \hat{\alpha} \) and \( \tilde{\alpha} \), respectively, as in Pourahmadi (2001). Further, both (2.1) and (2.6) are modifications of the LRT statistics given in Rowe and Logan (2004) for AR\((p)\) rather than independent errors.

2.3.2 Prewhitening-based approaches to calculating activation statistics

A simpler alternative to the previous likelihood-based approaches for calculating LRT activation statistics under assumptions of known covariance matrices (for both magnitude-only and complex modeled cases) is prewhitening, but because of the restrictive assumptions, resulting inferences are at best “only approximately valid” (see den Dekker et al., 2009, who, as an alternative, proposed LRT-based activation statistics assuming magnitude-only data and Gaussian model.) Rowe and Logan (2004) outlined a prewhitening approach for complex-valued AR(1) datasets. Here, we formalize and extend the above to the case where the estimated covariance matrix is computed by fitting an AR\((p)\) model to the residuals. These estimates are used to “whiten” the data following which the MLEs and LRT-based activation statistics are calculated. The latter is illustrated in Rowe and Logan (2004), which assumes temporal independence within the time series at a voxel; thus, here we describe only how to obtain the “whitened” data. First, the AR order \( p \) is estimated using methodology described in Section 2.3.3. Estimates of AR coefficients can then be obtained from (2.5), following which \( R_n^{-1} \) can be estimated using \( R_n^{* -1} \) as in Section 2.3.1. Write \( R_n^{* -1} = QQ' \) in terms of its Cholesky factorization. Then \( (y_R^*, y_i^*)' = I_2 \otimes Q(y_R', y_i')' \), with \( \otimes \) representing the Kronecker product, is assumed to be multivariate normally distributed with mean vector \( (\cos \theta, \sin \theta)' \otimes (QX\beta) \) and covariance matrix \( \sigma^2 I_{2n} \), and forms the “whitened” data. In comparison with the likelihood-based procedure, which iteratively updates \( (\hat{\theta}, \hat{\beta}), \hat{\alpha}, \) and \( \hat{R}_n^{-1} \) until convergence of the ML estimates, prewhitening essentially “stops” after the first iteration. As a result, prewhitening requires less computation than the likelihood-based procedure, but, as mentioned above, makes “only approximately valid” inferences.
2.3.3 Choosing the order of the autoregressive model

The order of the AR\((p)\) models, whether for the magnitude-only or the complex-valued case, is not \textit{a priori} known and needs to be determined. We propose sequentially testing \(H_0 : \alpha_k = 0\) vs. \(H_a : \alpha_k \neq 0\), starting with \(k = 1\), for increasing \(k\). Let \(k'\) be the first \(k\) in the sequence of tests for which the null hypothesis can not be rejected. Then the estimated AR\((p)\) order is given by \(\hat{p} = k' - 1\). We propose two alternative test statistics for carrying out each test: the (sample) partial autocorrelation function (PACF) and, separately, another LRT statistic for order detection. Both of our test statistics are extensions to the complex-model case of the usual magnitude-only version. In the latter case, the PACF is calculated from the magnitude-only residuals \(\hat{e}\) assuming independence. \textit{Shumway and Stoffer (2006)} show that for an AR\((p)\) process of \(n\) observations, the lag-\(k\) sample PACF \(\hat{a}_{kk}\) has an asymptotic \(N(0, 1/n)\) distribution, for \(k > p\). The null distribution of the magnitude-only PACF statistic \(\hat{a}_{kk}^{(M)}\) is then approximately \(N(0, 1/n)\). Extension to the complex-valued fMRI time series case essentially involves combining the contributions from the real and imaginary residuals, resulting in our proposed PACF test statistic \(\hat{a}_{kk}^{(C)} = \hat{a}_{kk}^{(R)} + \hat{a}_{kk}^{(I)}\), the sum of the lag-\(k\) PACFs computed from the real and imaginary parts of the residuals. These residuals are computed as \(\hat{\eta}_R = y_R - X\hat{\beta} \cos \hat{\theta}\) and \(\hat{\eta}_I = y_I - X\hat{\beta} \sin \hat{\theta}\), respectively, where \(\hat{\beta}\) and \(\hat{\theta}\) are as in Section 2.3.1, with \(\hat{R}_n^{-1} = I_n\). Because these residuals are independent (see Section 3.2), the PACF statistic has a \(N(0, 2/n)\) distribution under \(H_0\).

The alternative LRT-based test statistic is given by the usual \(2(\hat{\ell}_k - \hat{\ell}_{k-1})\) where \(\hat{\ell}_k\) is the optimized log-likelihood for the (magnitude-only or complex-valued) AR\((k)\) model: from standard results, this test statistic is asymptotically \(\chi_1^2\)-distributed under \(H_0\).

The decision on whether to continue testing in the sequential procedure outlined above can be based on either standard per-comparison error rate (PCER) methodology or false discovery rate (FDR) thresholding (\textit{Benjamini and Hochberg, 1995}). The latter accounts for multiple significance assessments in order detection. For PCER thresholding, we base these single-test decisions by specifying the probability of Type I error, which is rejecting \(H_0 : \alpha_k = 0\) when \(k > p\). This probability, say \(\delta\), has the property that \(\delta = \text{Pr}(\hat{p} > p | \hat{p} \geq p)\), the probability that
the detected order is over-specified, given that it is not underspecified. To show this, recall that
the detected order $\hat{p} = k'-1$, where $k'$ is the first $k$ for which we are unable to reject $H_0 : \alpha_k = 0$.
Note two facts: first, rejecting $H_0 : \alpha_k = 0$ means that $k' > k \Rightarrow \hat{p} > k - 1$. Second, simply
testing $H_0 : \alpha_k = 0$ in the context of the procedure implies that $k' \geq k \Rightarrow \hat{p} \geq k - 1$. From the
definition of $\delta$, for $k > p$, $\delta = \Pr(\hat{p} > k - 1 | \hat{p} \geq k - 1)$. Substituting $k = p + 1$ yields the above
result.

We now describe simultaneous detection of the order in $M$ voxel time series using FDR
thresholding. For $m = 1, \ldots, M$, denote $\alpha_{mk}$ be the $k$th order AR coefficient. For increasing
$k$, starting at $k = 1$, we simultaneously test $H_0 : \alpha_{m(k)} = 0$ vs. $H_a : \alpha_{m(k)} \neq 0$, for $m^{(k)} = 1, \ldots, M_k$. For each of the $M_k$ voxel time series, $p$-values (i.e., the measure of evidence against
$H_0$ – not AR order) are computed from one of the discussed test statistics and decisions are
based on the “Bonferroni-type” FDR controlling procedure given in Benjamini and Hochberg
(1995). For $k = 1$, all $M$ voxel time series are tested; that is, $M_1 = M$. Let $k'_m$ be the first $k$
for which the time series at the $m$th voxel fails to reject $H_0$. Then the detected order for this
voxel time series is $\hat{p}_m = k'_m - 1$. For increasing $k$, the number of tested voxel time series $M_k$
decreases as voxels with $k'_m < k$ (whose AR order has already been determined) are excluded
from tests. That is, $M_{k+1} = M_k - G_k$, where $G_k$ is the number of voxels with “fail to reject
$H_0$” decisions for order $k$. Simultaneous tests continue for increasing $k$ until $M_k = 0$. The
FDR level $q^*$ is the rate at which the null hypothesis is rejected in error: thus, in the context
of simultaneous order detection, it is the rate at which the order is detected in error (actually
over-specified).

2.3.4 Detecting voxels significantly activated by the stimulus

Having detected the order of the fitted autoregressive models, our task now is to detect
activation, which is really the primary goal of our experiment. As mentioned previously, a
general test for activation for a single voxel time series is $H_0 : C\beta = 0$ vs. $H_a : C\beta \neq 0$. Each voxel is identified as activated if $H_0$ is rejected, whether for the magnitude-only or
complex-valued AR($p$) model: in either case, given the order $p$, the tests are asymptotically
$\chi^2_m$-distributed under $H_0$. Once again, activation decisions can be based on PCER or FDR.
thresholding, with the latter accounting for the multiple testing issues introduced when multiple voxels are considered.

### 2.4 Experimental Evaluations

We applied the methodology in Section 2.3, computing LRT activation statistics for simulated fMRI data under both magnitude-only and complex-valued models with AR(p) errors. We simulated complex-valued voxel time series in a manner so as to mimic the experiment of Section 2.2, obtaining magnitude time series versions of them in the same way as is done in fMRI. Our simulation setup used the model (2.2) with our $X$-matrix matching that of the experiment in Section 2.2. Specifically, our design matrix $X$ contained $q = 3$ columns: the first column denoted the intercept term, the second modeled linear drift in the signal, while the third modeled the expected BOLD response with a $\pm 1$ square wave. (Note that consequently, with $\beta = (\beta_0, \beta_1, \beta_2)'$, activation tests were equivalent to testing $H_0 : \beta_2 = 0$ against $H_a : \beta_2 \neq 0$: also, the LRT activation statistics are $\chi^2_1$-distributed under $H_0$.) The square wave was lagged five time points (i.e. 5 seconds) from the stimulus time course to model the lag induced by the BOLD response, as discussed in Section 2.1; compared with other lags, the lag of five produced the highest activation statistics in the experimental dataset. Due to concerns about the constant phase assumption, we removed the first 12 and the last four time points, leaving us with voxel time series of length $n = 256$. This $X$-matrix was used in simulating datasets and evaluating the performance of our methodology in this section as well as in the analysis of our dataset in Section 2.5.

We compared the performance of LRT activation statistics under magnitude-only and complex-valued AR(p) models in terms of maximizing and minimizing true and false detection rates, respectively. We emphasize three contexts of this comparison in Sections 2.4.1 - 2.4.3, also expanding upon previous work in each. First, we examine the case with low SNR, building upon simulation experiments that, under the assumption of temporally independent (or prewhitened) voxel time series, have shown superior detection rate of complex-valued model activation statistics over their Gaussian-distributed magnitude-only counterparts (Nan and Nowak, 1999; Rowe and Logan, 2004). We extend this comparison to AR(p) time series. Next, we focus on how the
activation detection performance of both LRT statistics is affected by errors in order detection. As discussed in Section 2.1, the effect of modeling temporal dependence on activation detection has a long history for magnitude-only fMRI time series (Lazar, 2008) which we also examine for complex-valued data. Last, we compare the bias in significance levels for likelihood-based and prewhitening-based activation statistics. Such bias, which has long been a criticism of prewhitening approaches to fMRI magnitude time series, is also computed for complex-valued data.

In each of the experiments described above, we simulated fMRI datasets and computed activation detection rates in two ways. First, we generated voxel time series from the same parameters repeatedly and used standard (PCER) thresholding to determine activation. We call this the “single voxel” simulation context. However, real fMRI datasets contain numerous voxel time series and activation detection must account for multiple testing. Thus, we also generated brain slices of $128 \times 128$ voxel time series and, in this “brain slice” context, used FDR thresholding to determine activation. These simulated brain slices were designed to represent the finger-tapping dataset of Section 2.2 and contained three groups of voxels: background (outside the brain) and inactivated and activated brain voxels. Each slice contained 275 activated voxels, a number estimated from the dataset. Further, we chose our parameter values to be the same for voxels in each group but different from those in other groups.

The parameter values used in each simulation context are given in Table 2.1, which are obtained from their estimates in the finger-tapping dataset as described in Section 3.3. In this section, values of $\beta_0$ and $\beta_2$ are parameterized through the signal-to-noise ratio, $\text{SNR} \equiv \beta_0/\sigma$, and the contrast-to-noise ratio, $\text{CNR} \equiv \beta_2/\sigma$, respectively. The SNR measures the size of baseline, non-BOLD signal relative to the noise level, and the CNR measures the relative size of the BOLD response. In Section 2.4.1, we simulate at SNR less than 10, but otherwise we use SNR = 50, the approximate estimate from the dataset in Figure 3.2(f). Since the low-level finger-tapping task has a higher CNR than high-level tasks of interest, such as cognition, we simulate at CNRs less than dataset estimates shown in Figure 3.2(g).
Simulation Thresholding Voxel Parameter values

<table>
<thead>
<tr>
<th>Simulation context</th>
<th>Thresholding procedure</th>
<th>Voxel group</th>
<th>Parameter values</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single voxel</td>
<td>PCER</td>
<td>——</td>
<td>$\beta_0$</td>
<td>-0.000026</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\beta_1$</td>
<td>$\sigma \times \text{SNR}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\beta_2$</td>
<td>0.0329</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$\sigma$</td>
<td>0.0329</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Back.</td>
<td>$\alpha^*$</td>
<td>0.000005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.02</td>
<td>0</td>
<td>0.0194</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nonact.</td>
<td>$\sigma \times \text{SNR}$</td>
<td>0.0329</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activ.</td>
<td>$\sigma \times \text{SNR}$</td>
<td>0.0329</td>
</tr>
<tr>
<td>Brain slice</td>
<td>FDR</td>
<td>Back.</td>
<td>0.02</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0.0194</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nonact.</td>
<td>$\sigma \times \text{SNR}$</td>
<td>0.0329</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Activ.</td>
<td>$\sigma \times \text{SNR}$</td>
<td>0.0329</td>
</tr>
</tbody>
</table>

Table 2.1 Summary of parameters used in the two contexts of simulation experiments. In the above, $\alpha^* = (0.17, 0.45, -0.11, -0.23)$ and the voxel group abbreviations “Back.,” “Nonact.,” and “Activ.” represent background, non-activated, and activated voxel groups, respectively.

2.4.1 Complex-valued/magnitude-only activation detection at low SNR

As noted in Section 2.1, SNR is proportional to voxel volume, so fMRI studies with increased spatial resolution will have lower SNR data. We simulate such fMRI datasets with SNR = 1, 2, …, 10 and CNR = 0.05, 0.10, …, 0.50, generating 100,000 single voxel time series and 100 brain slices at each (SNR, CNR) combination. Assuming correct order detection, LRT activation statistics are computed under complex-valued and magnitude-only models. Activation is detected at PCER and FDR thresholds of $\delta = 0.0005$ and $q^* = 0.05$, respectively, and detection rate is computed as the proportion of the activated simulated voxel time series (i.e. with positive CNR) detected as such. These activation detection rates are plotted against SNR for CNR = 0.35 in Figure 2.2, which shows striking similarity to those for simulated temporally independent voxel time series (compare with Rowe and Logan, 2004, Figure 12). The activation detection rate is constant in SNR for the complex-valued model LRT statistic, but decreases at low SNR under the Gaussian-distributed magnitude-only model. As discussed in Section 2.1, the latter is most likely owing to the poor Gaussian approximation of the Rice-distributed magnitude observations at low SNR. This distributional approximation appears more tenable for SNR $\geq 6$ because the magnitude-only model detection rate is constant over this range, though at a slightly lower level than the complex-valued model. We ascribe this slight difference to the disposal of the phase information under the magnitude-only model.

Figure 2.3 summarizes the relationship between detection rate and SNR for all the CNRs by displaying these detection rates in (SNR, CNR)-space. Images are presented for complex-
Figure 2.2 Plots of activation detection rates of complex-valued and Gaussian-distributed magnitude model LRT statistics against SNR for (a) the single voxel simulation context, using a $\delta = 0.0005$ PCER level, and (b) the brain slice context using an FDR level $q^* = 0.05$. The CNR is 0.35.

valued and magnitude-only LRT statistics and for the differences in their detection rates. The features in the detection rate by SNR relationship discussed in the previous paragraph are again present, most prominently for moderate CNRs; the differences in activation detection rates again vanish for low and high CNRs, as detection rates are then close to zero and one, respectively, regardless of model and SNR. Note that the negative differences in Figures 2.3(c) and (f) (which favor the magnitude-only model), though visually compelling, represent very small differences: the largest is less than 0.1%.

2.4.2 AR order detection errors and their consequences on activation detection

Before investigating the effects of AR order detection errors on the performance of the LRT activation statistics, we examine the rates of such errors under complex-valued and magnitude-only models. Using the parameters in Table 2.1 and a true AR order of $p = 4$, we simulate 100,000 single voxel time series and 100 brain slices with zero CNR and SNR = 50. We apply the order detection methods introduced in Section 2.3.3 (tested for other simulated datasets in Section 3.4), using PCER and FDR levels $\delta = q^* = 0.05$. The proportions of voxel time series detecting each order are shown in Table 2.2, which only includes in-brain voxels for the simulated brain slices. Magnitude-only model order detection procedures have error rates more than double those for complex-valued model procedures. This is not surprising, considering
that complex-valued model order detection methods use twice the amount of information than the magnitude-only model detection. (Note also that most order detection errors constitute underspecification.)

Based on assigned orders $\hat{p} = 0, 1, \ldots, 8$, we compute LRT activation statistics for simulated AR(4) voxel time series with SNR = 50 and CNR = 0.2. These activation statistics were thresholded at various PCER and FDR levels to obtain the receiver operating characteristic (ROC) curves in Figure 2.4, which plot true detection rate against false detection rate. In ROC plots, better performing statistics will be closer to the top and left, indicating higher true detection rates and lower false detection rates, respectively. Under this criterion, the complex-valued and magnitude-only model statistics based on the correct orders perform best while statistics based on under-detected orders show inferior performance, while those for over-detected orders are indistinguishable from the correct order curves (and therefore not shown); thus, it
Table 2.2  The proportions of simulated voxel time series detecting each AR order \( \hat{p} \) under the complex-valued and magnitude-only model order detection procedures introduced in Section 2.3.3. The true order of 4 is shown in bold. Results are reported under both PCER and FDR thresholding and the PACF and LRT order detection test statistics.

<table>
<thead>
<tr>
<th>( \hat{p} )</th>
<th>LRT</th>
<th>PACF</th>
<th>LRT</th>
<th>PACF</th>
<th>LRT</th>
<th>PACF</th>
<th>LRT</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.016</td>
<td>0.017</td>
<td>0.048</td>
<td>0.049</td>
<td>0.149</td>
<td>0.151</td>
<td>0.306</td>
<td>0.313</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.069</td>
<td>0.071</td>
<td>0.066</td>
<td>0.068</td>
<td>0.221</td>
<td>0.221</td>
<td>0.181</td>
<td>0.182</td>
</tr>
<tr>
<td>3</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
<td>0.024</td>
<td>0.025</td>
<td>0.020</td>
<td>0.021</td>
</tr>
<tr>
<td>4</td>
<td>0.865</td>
<td>0.866</td>
<td>0.886</td>
<td>0.882</td>
<td>0.575</td>
<td>0.572</td>
<td>0.493</td>
<td>0.484</td>
</tr>
<tr>
<td>5</td>
<td>0.046</td>
<td>0.043</td>
<td>0</td>
<td>0</td>
<td>0.030</td>
<td>0.029</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>( \geq 6 )</td>
<td>0.002</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
<td>0.002</td>
<td>0.002</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

appears that under-specifying the order has more severe consequences on activation detection than over-specifying it. Note also that for each assigned order \( \hat{p} \), the complex-valued model activation statistic shows (slightly) higher performance than its magnitude-only counterpart.

The previous results indicate that order detection errors effect the activation detection performance of the magnitude-only model more adversely than the complex-valued model. Order detection error rates were higher under the magnitude-only model and mostly constituted underspecification, which was shown to cause poorer activation detection. Further, the complex-valued model LRT showed higher performance when the order was controlled. Figure 2.5, which shows ROC curves for LRT statistics based on detected orders, demonstrates the consolidation of such effects.

2.4.3 Investigating bias in significance levels for prewhitening- and likelihood-based activation statistics

Historical concerns for prewhitening-based approaches to computing activation statistics have centered around bias in significance levels (Friston et al., 2000), with likelihood-based activation statistics suggested as a possible remedy (den Dekker et al., 2009). In this section, we compute both prewhitening- and likelihood-based LRT activation statistics, assuming perfect
order detection, and measure their bias in PCER and FDR levels. Our studies are over 100,000 simulated single voxel time series with zero CNR and 1000 simulated brain slices with CNR = 0.2, again at SNR = 50.

Figure 2.6 shows the significance level bias (“estimated minus nominal”) against nominal significance level for PCER and FDR thresholding. For PCER thresholding, the estimated significance level is the Type I error rate – i.e. the proportion of non-activated voxels detected as activated. For FDR thresholding, we calculate the proportion of detected voxels (as activated) which are non-activated. To summarize, all LRT activation statistics show bias, which may be partially attributed to the asymptotic specification of their null distributions. All bias is in the
positive direction, where more false detections occur than are nominally specified. However, the likelihood-based and complex-valued model statistics show less bias than the prewhitening-based and magnitude-only statistics, respectively. Note also that bias sizes are larger under FDR thresholding than PCER thresholding.

The results also present a third advantage of the complex-valued model LRT statistic over its magnitude-only counterpart, even when the effects of low SNR and order detection are controlled: the magnitude-only statistic has a higher false detection rate. False detection rates are also higher for prewhitening versus likelihood-based approaches, an example of the “only approximately valid” inference based on prewhitening (den Dekker et al., 2009).

The results of all our simulation experiments demonstrate three advantages of activation detection via the complex-valued model over the Gaussian-distributed magnitude-only model: higher (true) detection rate at low SNR, smaller decrease in detection performance due to order detection errors, and smaller false detection rate. The first, which is perhaps most striking, is due to the untenable Gaussian approximation to the Rice-distributed magnitudes at SNRs below 5. The SNR for finger-tapping dataset is well above this range, so we will not see such an effect for it, but, as mentioned in Section 2.1, the SNR will decrease for datasets incorporating more spatial resolution.
Figure 2.7 Images of the detected AR orders under (a) complex-valued and (b) magnitude-only data approaches for the finger-tapping dataset, using the LRT statistic with FDR thresholding at a $q^* = 0.05$ level.

2.5 Application to fMRI dataset

We detected activated voxels for the finger-tapping dataset under both the complex-valued and magnitude-only models. We used the model matrix $X$ described in Section 2.4 in this application. Our computation of functional activation had three steps: order detection, computation of LRT activation statistics, and thresholding. First, we detected the AR order for each voxel time series, applying the magnitude-only and complex-valued model procedures presented in Section 2.3.3. As shown in Figure 2.7, inside the brain, the complex-valued model primarily detected an order of four while the magnitude-only model mostly detected zero or two. Based on these detected orders, LRT activation statistics for the test of $H_0 : \beta_2 = 0$ vs. $H_a : \beta_2 \neq 0$ were calculated for both models. The voxel-wise $p$-values, computed from the $\chi^2_1$ null distribution, were thresholded at a $q^* = 0.05$ FDR level, determining whether each voxel was detected. The resulting activation maps for complex-valued and magnitude-only statistics are shown in Figures 2.8(b) and 2.8(c), respectively. On them, only detected activated voxels are colored – with intensities according to the size of the activation statistic – and are overlayed on top of the greyscale anatomical image. Thus, our displayed activation maps display both the location of voxels detected and their “degree” of activation, where larger activation statistics demonstrate stronger activation.

We also performed some diagnostic checks to investigate the complex-valued Gaussian
Figure 2.8 (a) Anatomical image of the subject’s brain displaying the central sulci (in green), which contain the sensori-motor finger area cortices. Activation maps of the (b) complex-valued and (c) magnitude-only model LRT statistics (overlayed on top of the same anatomical image), thresholded at the 5% false discovery rate. (Note that activation maps are drawn after masking out voxels outside the brain, as determined by the anatomical image.)

AR\((p)\) model assumptions for the dataset. Since the assumptions that the real and imaginary observations are normally distributed and uncorrelated, with the same variance have already been established in general for MR (see e.g. Wang and Lei, 1994) and specifically (see Figures 2 and 5 in Rowe and Logan, 2004) for this dataset, we focus on the assumption that the real and imaginary errors share a common AR\((p)\) dependence structure by noting the similarities of the images of the voxel-wise PACFs for the real and imaginary residuals shown in Section 3.5. We also investigated, in the same section, whether an AR\((p)\) model sufficiently removes this temporal dependence by computing Box-Pierce statistics (Box and Pierce, 1970) for independent and AR-model-fitted residuals from the finger-tapping dataset. Results reported there show that the AR model greatly reduces the autocorrelation present, but more sophisticated time series methods may be necessary to accurately model the dependence structure in the time series.

We now discuss our findings and the relative advantages of using the complex-valued model over magnitude-only analysis. As indicated earlier, the finger-tapping task has well-established fMRI-detected activation regions in the central sulci, which are identified on the anatomical image in Figure 2.8(a). We argue that the complex-valued model activation map of Figure 2.8(b) is visually preferable to its magnitude-only counterpart displayed in Figure 2.8(c). Although
both activation maps detect regions of voxels containing the central sulci, the one obtained using the complex-valued model identifies the central sulci more clearly. Also, voxels detected outside the central sulci in the complex map, better adhere to grey matter (shown lighter in Figure 2.8(a)), which is intrinsically where neural activation takes place. Our maps may also be compared to those in Figure 6 of Rowe and Logan (2004), which are computed (for the same dataset) under the assumption of temporal independence. Our maps, under both complex-valued and magnitude-only models, identify the central sulci more clearly, which we attribute to modeling the AR($p$) independence. Thus, we see improved detection and localization abilities in using the time series information, which is enhanced when we use the complex-valued observations over the magnitude-only datasets.

### 2.6 Discussion

In this paper, we have further developed the complex-valued time series analysis of fMRI data for use in fMRI data analysis. As explained here, fMRI datasets are really complex-valued when collected, but most analysis methods routinely discard the phase information, utilizing only the magnitude images in the data analysis. In doing so, current practice has been to assume a Gaussian distribution for the magnitude data, a supposition that is not even approximately correct for low SNR values. This last point is important to note because SNR (being proportional to voxel volume) decreases with increased spatial resolution. In this paper therefore, we have proposed an AR($p$) model for complex-valued time series, thus extending the independent model of Rowe and Logan (2004). Under this model framework, we derived an LRT statistic for detecting activated brain voxels. We compared its performance to a statistic similarly derived under a Gaussian-assumed magnitude-only linear model with AR($p$) errors. For low-SNR simulated data, the complex-valued statistic demonstrates notably higher activation detection rates than the Gaussian magnitude-only statistic, due to the inaccuracy of the normal approximation to the Rice-distributed magnitude data. This is potentially advantageous especially for the case of fMRI datasets collected at higher spatial resolutions and for datasets with higher-level cognitive tasks. In either cases, SNR and CNR values are lower and thus, there is a greater payoff for using the complex-valued approaches. Even for
high-SNR simulated data, the complex-valued approach yields lower AR order detection error rates (which negatively affect activation detection) and lower false activation detection rates, simply due to the availability of twice as many quantities in the complex-valued setting. For the finger-tapping dataset, the activation map for the complex-valued statistic more clearly identifies brain regions known to be associated with finger movement – evidence which also indicates a lower false detection rate. Aside from the major focus on complex-valued versus magnitude-only methods, we also demonstrated that prewhitening-based activation statistics produce higher false detection rates than likelihood-based statistics.

There are several aspects of our work that require further attention. For one, our AR($p$) modeling was seen to not be entirely adequate in modeling model the correlation structure with regard to the finger-tapping experimental dataset. This means that more sophisticated time series methods may be needed to completely model the temporal dependence, as revealed by the diagnostic checks. Possible extensions may be to add a moving average component to the autoregressive model and/or to introduce a seasonal component related to the periodicity in the application of the stimulus. More fundamentally, the errors may not be stationary and, if so, cannot be properly modeled using stationary methods. Secondly, we have evaluated and demonstrated performance on a dataset with high SNR in order to establish the validity of our methodology: it would be interesting to also evaluate performance on a low-SNR experimental dataset. There is some scope for optimism here, given the results of our simulation experiments and the fact that our modeling is more accurate than a Gaussian-approximated magnitude-only time series approach which is actually more suspect at lower SNR. Finally, while we hope that our methods and applications here will spur the adoption of complex-valued methodology for fMRI datasets, we note that since the practice to date has been to rely on magnitude-only fMRI datasets, there are a large number of available datasets for which the phase information has been discarded. For such datasets, temporal models that correctly model the time series in terms of the Rice distribution are needed. It is our view that complex-valued data analysis should become the norm in fMRI: however, for the these datasets, methods on Rice-distributed regression time series that accurately model the temporal correlation also need to be developed. Thus, we note that while we have presented a compelling case for incorporating complex-valued
analysis in fMRI, there are many issues that could benefit further with increased attention.
CHAPTER 3. SUPPLEMENT TO “IMPROVED ACTIVATION DETECTION VIA COMPLEX-VALUED AUTOREGRESSIVE MODELLING OF FMRI VOXEL TIME SERIES”

3.1 MLEs of parameters under Null Models

3.1.1 Restricted MLEs under magnitude-only autoregressive model

Under the magnitude-only AR($p$) model, the restricted MLEs under $H_0 : C\beta = 0$ follow the equations

\[
\hat{\beta} = \Psi(X'\bar{R}_n^{-1}X)^{-1}X'\bar{R}_n^{-1}r,
\]
\[
\hat{\sigma}^2 = (r - X\bar{\beta})'(\bar{R}_n^{-1}(r - X\bar{\beta}))/n,
\]
\[
\hat{d}_{0k} = \sum_{j=1}^p (\hat{d}_{jk} + j\hat{\gamma}_{j-k})\hat{a}_j, \quad k = 1,\ldots,p,
\]

where $\Psi$ and $\bar{R}_n^{-1}$ are as in Section 3.1.2 below. Further, $\hat{d}_{ij} = \sum_{t=1}^{n} i \hat{\varepsilon}_t j \hat{\varepsilon}_t + j$ for $0 \leq i, j \leq p$, where $\hat{\varepsilon}_t = r_t - x'_t\bar{\beta}$, $t = 1,\ldots,n$, and $\hat{\gamma}_k = \hat{d}_{0k}/n$, for $k = 0,\ldots,p - 1$.

3.1.2 Restricted MLEs under complex-valued autoregressive model

Under the complex-valued AR($p$) model, similar to the results in Rowe and Logan (2004), the restricted MLEs under $H_0 : C\beta = 0$ follow the equations

\[
\hat{\beta} = \Psi[\bar{\beta}_R \cos \bar{\theta} + \bar{\beta}_I \sin \bar{\theta}]
\]
\[
\hat{\theta} = \frac{1}{2} \arctan \left[ \frac{2\bar{\beta}_R'\Psi'X'\bar{R}_n^{-1}X\bar{\beta}_I}{\bar{\beta}_R'\Psi'X'\bar{R}_n^{-1}X\bar{\beta}_I - \bar{\beta}_I'\Psi'X'\bar{R}_n^{-1}X\bar{\beta}_R} \right],
\]
\[
\hat{d}_{0k} = \sum_{j=1}^p (\hat{d}_{jk} + 2j\hat{\gamma}_{j-k})\hat{a}_j, \quad k = 1,\ldots,p,
\]
where \( \Psi = I_q - (X'\tilde{R}_n^{-1}X)^{-1}C' \left[ C(X'\tilde{R}_n^{-1}X)^{-1}C' \right]^{-1} C \), \( \tilde{\beta}_R = (X'\tilde{R}_n^{-1}X)^{-1}X'\tilde{R}_n^{-1}y_R \), \( \tilde{\beta}_I = (X'\tilde{R}_n^{-1}X)^{-1}X'\tilde{R}_n^{-1}y_I \), and \( \tilde{R}_n^{-1} \) is a function of \( \tilde{\alpha} \) as in Pourahmadi (2001). Further, 
\[
\tilde{d}_{ij} = \sum_{t=1}^{n-i+j} \tilde{\eta}_{R,t+i+\tilde{\eta}_{R,t+j}} + \tilde{\eta}_{I,t+i+\tilde{\eta}_{I,t+j}}, 0 \leq i, j \leq p,
\]
and \( \tilde{\gamma}_k = \tilde{d}_{0k}/2n \), \( k = 0, \ldots, p - 1 \), where \( \tilde{\eta}_{R,t} = y_{R,t} - x'_t\tilde{\beta} \cos \tilde{\theta} \) and \( \tilde{\eta}_{I,t} = y_{I,t} - x'_t\tilde{\beta} \sin \tilde{\theta} \), \( t = 1, \ldots, n \).

### 3.2 Independence of real and imaginary residuals

We prove the independence of the real and imaginary parts of the residuals, which is necessary to derive the null distribution of the PACF statistic for AR order detection in Section 2.3.3. To prove independence, we show that the real residuals are a function of only the real observations and the imaginary residuals are a function of only the imaginary observations: independence of the residuals \( \tilde{\eta}_R \) and \( \tilde{\eta}_I \) then follows from the independence of \( y_R \) and \( y_I \) under model (2.2).

Specifically, \( \tilde{\eta}_R = (I_n - P_X)y_R \) and \( \tilde{\eta}_I = (I_n - P_X)y_I \), where \( P_X = X(X'X)^{-1}X' \) is the standard projection matrix. We show this for the real residuals; the proof for the imaginary residuals is similar. By definition, the real residual vector is \( \tilde{\eta}_R = y_R - X\tilde{\beta} \cos \tilde{\theta} \), where \( \tilde{\beta} \) and \( \tilde{\theta} \) are as in Section 2.3.1. We use \( \tilde{R}_n^{-1} = I_n \) here, as is also done in the calculation of the PACF statistic, but the proof also holds for general \( \tilde{R}_n^{-1} \). Now,

\[
\tilde{\beta} \cos \tilde{\theta} = (\tilde{\beta}_R \cos \tilde{\theta} + \tilde{\beta}_I \sin \tilde{\theta}) \cos \tilde{\theta} = \tilde{\beta}_R(\frac{1}{2} + \frac{1}{2} \cos 2\tilde{\theta}) + \tilde{\beta}_I(\frac{1}{2} \sin 2\tilde{\theta})
\]

(3.1)

\[
= \frac{1}{2H} \left[ \tilde{\beta}_R(H + \tilde{\beta}_R'X'X\tilde{\beta}_R - \tilde{\beta}_I'X'X\tilde{\beta}_I) + \tilde{\beta}_I(2\tilde{\beta}_R'X'X\tilde{\beta}_I) \right],
\]

(3.2)

where (3.1) substitutes the expression for \( \tilde{\beta} \) given in Section 2.3.1, (3.2) applies standard trigonometric identities, and (3.3) substitutes the following expressions for \( \sin 2\tilde{\theta} \) and \( \cos 2\tilde{\theta} \).

It can be shown from (2.4) that \( \sin 2\tilde{\theta} = (2\tilde{\beta}_R'X'X\tilde{\beta}_I)/H \) and \( \cos 2\tilde{\theta} = (\tilde{\beta}_R'X'X\tilde{\beta}_R - \tilde{\beta}_I'X'X\tilde{\beta}_I)/H \), where \( H = \tilde{\beta}_R'X'X\tilde{\beta}_R + \tilde{\beta}_I'X'X\tilde{\beta}_I \). The bracketed expression in (3.3) reduces to \( 2H\tilde{\beta}_R \), and thus \( \tilde{\beta} \cos \tilde{\theta} = \tilde{\beta}_R \). As a result, \( \tilde{\eta}_R = y_R - X\tilde{\beta}_R = (I_n - P_X)y_R \).
Table 3.1 The number of voxels detecting each AR order \( \hat{p} \) for the finger-tapping dataset inside and outside the brain, using PACF/LRT test statistics and PCER/FDR thresholding.

<table>
<thead>
<tr>
<th>( \hat{p} )</th>
<th>LRT</th>
<th>PACF</th>
<th>LRT</th>
<th>PACF</th>
<th>LRT</th>
<th>PACF</th>
<th>LRT</th>
<th>PACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>122</td>
<td>208</td>
<td>177</td>
<td>307</td>
<td>8187</td>
<td>8236</td>
<td>9400</td>
<td>9400</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td>10</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>598</td>
<td>784</td>
<td>710</td>
<td>960</td>
<td>4109</td>
<td>4126</td>
<td>3419</td>
<td>3409</td>
</tr>
<tr>
<td>3</td>
<td>27</td>
<td>9</td>
<td>20</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1731</td>
<td>1457</td>
<td>1835</td>
<td>1442</td>
<td>537</td>
<td>527</td>
<td>363</td>
<td>347</td>
</tr>
<tr>
<td>≥ 5</td>
<td>427</td>
<td>448</td>
<td>163</td>
<td>193</td>
<td>607</td>
<td>553</td>
<td>258</td>
<td>226</td>
</tr>
</tbody>
</table>

3.3 Determining simulation parameters from fMRI dataset

The AR orders and parameter values used in simulation experiments (shown in Table 2.1) are determined from their estimates for the finger-tapping dataset. We specify the AR order \( p \) by applying the complex-valued model order detection methodology in Section 2.3.3 (which is validated on simulated data in Section 3.4). The voxel-wise detected orders for the dataset are shown in Figure 3.1 and summarized in Table 3.1. All order detection procedures, including those with PACF/LRT statistics and PCER/FDR thresholding (with \( \delta = q^* = 0.05 \)), agree in detecting a majority of voxels with an order of zero in the background and an order of four inside the brain. As a result, we simulate background voxel time series with \( p = 0 \) and those inside the brain with \( p = 4 \). Then, to determine simulation parameters, we calculated the complex-valued model MLEs given in Section 2.3.1 under these orders, which are shown in Figure 3.2 for \( \hat{p} = 4 \). Parameters in Table 2.1 are simply medians of these MLEs.

3.4 Validation of AR order detection procedures from simulated voxel time series

We test the AR order detection methodology introduced in Section 2.3.3 on simulated fMRI voxel time series, validating that nominal PCER and FDR significance levels are achieved. Recall that, for PCER thresholding, the significance level \( \delta \) is also equal to \( \Pr(\hat{p} > p | \hat{p} \geq p) \).
Figure 3.1 Images of the detected AR orders $\hat{p}$ for the dataset, using the complex-valued model order detection methodology described in Section 2.3.3, under PACF/LR test statistics and PCER/FDR thresholding.

the probability of over-detection given that under-detection has not occurred. To validate this property, we detect orders with $\delta = 0.05$ for 100,000 complex-valued voxel time series simulated from model (2.2), with non-AR parameters as in Table 2.1 and AR(1) parameters $\alpha_1 = 0.00, 0.05, 0.10, 0.15, 0.20$. The previous property is shown to hold for both LRT and PACF statistics in the last columns of Table 3.2. Specifically, approximately 5 percent of the AR(0) time series have detected $\hat{p}$ greater than 0, and approximately 5 percent of the AR(1) time series not detecting $\hat{p} = 0$ (not under-detecting) have detected $\hat{p} > 1$. Further, as we should expect, $\hat{p} = 1$ is detected more frequently as $|\alpha_1|$ increases.

We also show that FDR levels are controlled in the order detection procedures by validating two properties of the FDR-controlling procedure in Benjamini and Hochberg (1995). The first such property is that when $H_0$ is true for all tests, the false discovery rate is equal to the family-wise error rate (FWER), the probability that at least one of the tests rejects $H_0$. In
Figure 3.2 Images of complex-valued AR(4) model MLEs for the finger-tapping data set.

In the order detection context, we illustrate this property by simulating AR(0) time series (i.e. $H_0 : \alpha_1 = 0$ is true) and forming blocks of them (of constant size, say, 5); then, we calculate the FWER as the proportion of blocks in which, under FDR thresholding, $\hat{p} \geq 1$ (i.e. $H_0 : \alpha_1 = 0$ is rejected) at least once. For an FDR level of $q^* = 0.05$, simulations showed an FWER of 0.055 for the LRT statistic and 0.053 for the PACF statistic. Second, the Benjamini and Hochberg (1995) FDR controlling procedure has expected FDR equal to the nominal level $q^*$ times the proportion of tests in which $H_0$ is true. We chose a proportion of one-half, simulating 100,000 AR(0) and AR(1) time series, and calculated FDR as the proportion of time series detecting $\hat{p} \geq 1$ (rejecting $H_0 : \alpha_1 = 0$) for which $p = 0$ ($H_0 : \alpha_1 = 0$ is true). Results for $q^* = 0.05$ showed observed FDRs at approximately the expected level of 0.025: they were 0.027 and 0.026 for the LRT and PACF statistics, respectively.
Table 3.2 Proportion of simulated AR(0) and AR(1) complex-valued time series detecting orders $\hat{p} = 0, 1$, and greater than 1 for different values of $\alpha_1$ under the LRT/PACF order detection procedures with PCER level $\delta = 0.05$.

### 3.5 Diagnostics for checking model assumptions

An assumption of the AR($p$) model for complex-valued fMRI time series in Section 2.3.1 is that the real and imaginary errors have the same autoregressive dependence structure. We check this assumption for the finger-tapping dataset with images of the PACFs computed from the real and imaginary residuals, which are shown in figure 3.3. These residuals $\hat{\eta}_R$ and $\hat{\eta}_I$ are computed under independence; that is, they are

$$
\hat{\eta}_R = y_R - X\hat{\beta}\cos\hat{\theta}, \quad \hat{\eta}_I = y_I - X\hat{\beta}\sin\hat{\theta},
$$

(3.4)

where $\hat{\beta}$ and $\hat{\theta}$ are as in Section 2.3.1 with $\hat{R}_n^{-1} = I_n$. Because these images of real and imaginary PACFs look similar at each lag, we are willing to accept this model assumption.

To check whether the AR($p$) model sufficiently removes the temporal dependence from the complex-valued voxel time series in the finger-tapping dataset, we compute Box-Pierce $Q$-statistics (Box and Pierce, 1970). The Box-Pierce statistic $Q_K$ for a (real-valued) time series of length $n$ is defined as $n \sum_{k=1}^{K} \hat{\rho}^2(k)$, where $\hat{\rho}^2(k)$ is the lag-$k$ sample autocorrelation of the model fit residuals. For truly AR($p$) time series, $Q_K$ is asymptotically $\chi^2_{K-p}$. The value of $K$ is chosen somewhat arbitrarily, but typically $K = 20$ (Shumway and Stoffer, 2006), which we use here. We compute $Q_{20}$ for residuals under independent and AR($p$) model fits, which gives us a measure of the reduction in autocorrelation due to the AR($p$) model fit. Under independence, the residuals are as in (3.4); the AR($p$)-model-fit residuals are the $(n-p)$-vectors $\hat{\eta}_R$ and $\hat{\eta}_I$, where $\hat{\epsilon}_{Rt} = \hat{\eta}_{R,t+p} - \sum_{k=1}^{p} \alpha_k \hat{\epsilon}_{R,t+p-k}$ and $\hat{\epsilon}_{It} = \hat{\eta}_{I,t+p} - \sum_{k=1}^{p} \alpha_k \hat{\epsilon}_{I,t+p-k}$, $t = 1, \ldots, n - p$. 

Figure 3.3 Images of the PACFs computed from the real and imaginary residuals of the finger-tapping dataset for lags 1 to 20.

where $\hat{\eta}_R$ and $\hat{\eta}_I$ are entries of residual vectors in (3.4). For each model fit residuals, two $Q$-statistics are computed: one for each of the real and imaginary residuals.

We compute $Q_{20}$ statistics for independent- and AR(4)-model residuals for voxels inside the brain (2916 in all), where $\hat{p} = 4$ is the order detected for the majority of in-brain voxels in Section 3.3. We also simulated truly AR(4) complex-valued voxel time series, with parameters as in table 2.1, and calculated $Q_{20}$ statistics for the same two kinds of residuals. The resulting four $Q$-statistics are compared to a random sample from the null distribution of $Q_{20}$ under an AR(4) model, $\chi^2_{16}$, in the quantile-quantile plot in figure 3.4. The AR(4)-model residual $Q$-statistics for the simulated data are close to the $\chi^2_{16}$ null distribution and are well-below the AR(4)-model residual $Q$-statistics for the dataset. This indicates the dataset still contains substantial autocorrelation after the AR model fit, and perhaps more complex methods are needed to remove the autocorrelation, such as incorporating moving average, integrated, or seasonal components in the time series model. However, the AR(4)-model residual $Q$-statistics
Figure 3.4 Quantile-quantile plot of Box-Pierce $Q_{20}$-statistics for independent- and AR(4)-model-fit residuals of the simulated and finger-tapping data ("empirical") versus a random sample from the null distribution of $Q_{20}$ under an AR(4) model, $\chi^2_{16}$.

for the dataset are much smaller than its independent-model counterparts, indicating that the AR model greatly reduces temporal autocorrelation.
CHAPTER 4. ON THE USE OF GAUSSIAN AND RICE DISTRIBUTIONS FOR FITTING MAGNITUDE FMRI TIME SERIES DATA

A paper in preparation
Daniel W. Adrian, Ranjan Maitra, and Daniel B. Rowe

Abstract

It is well-known that Gaussian modeling of fMRI magnitude time series, which are truly Rice-distributed, constitutes an approximation, especially at low signal-to-noise ratios (SNRs). Based on this fact, previous work has demonstrated that Ricean-based activation tests show increased performance over Gaussian-based tests at low SNRs (den Dekker and Sijbers, 2005; Rowe, 2005b). However, we identify limiting assumptions and approximations in this work and, removing them, provide an updated comparison of such Ricean and Gaussian modeling, incorporating recent advances in Ricean parameter estimation via the EM algorithm (Solo and Noh, 2007; Zhu et al., 2009). After evaluating such tests through ROC curve methodology, we find reasons to doubt the earlier findings (den Dekker and Sijbers, 2005; Rowe, 2005b) and instead conclude that the gains produced by Ricean-based activation tests, even at low SNRs, are marginal at best.

4.1 Introduction

Functional magnetic resonance imaging (fMRI) is a popular noninvasive method for studying the spatial characteristics of human brain function. The imaging modality depends on the fact that when neurons fire in response to a stimulus or task, the blood oxygen levels in
neighboring vessels change, effecting the magnetic resonance (MR) signal on the order of 2-3% (Lazar, 2008), due to the differing magnetic susceptibilities of oxygenated and deoxygenated hemoglobin. This difference causes the so-called Blood Oxygen Level Dependent (BOLD) contrast Ogawa et al. (1990); Belliveau et al. (1991); Kwong et al. (1992); Bandettini et al. (1993), which is used as a surrogate for neural activity and is used to acquire time-course sequences of images, in which the time-course is in accordance with the presentation of the stimulus. Such images are composed of MR measurements at each voxel, or volume element.

To detect regions of neural activation, the general strategy is to fit, at each voxel, a model — commonly a general linear model (Friston et al., 1995) — to the time series observations against the expected BOLD response. This provides the setting for the application of techniques such as Statistical Parametric Mapping (SPM) (Friston et al., 1990), where the time series at each voxel is reduced to a test statistic which summarizes the association between each voxel time course and the expected BOLD response (Bandettini et al., 1993). The resulting map is then thresholded to identify voxels that are significantly activated (Worsley et al., 1996; Genovese et al., 2002; Logan and Rowe, 2004).

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

Following this justification, previous work (Solo and Noh, 2007; den Dekker and Sijbers, 2005; Rowe, 2005b) has demonstrated disadvantages of Gaussian-based modeling for simulated low-SNR, Rice-distributed time courses. Specifically, it has been shown that Gaussian-model maximum likelihood estimates (MLEs) of Ricean parameters are increasingly biased with the
decrease in SNR (Solo and Noh, 2007). Also, den Dekker and Sijbers (2005) shows that a
Gaussian-based likelihood ratio test (LRT) for activation has lower detection rate than a Ricean-
based LRT, and the difference increases with decreasing SNR. Further, the paper argues that the
Gaussian-based activation test “should never be used for low-SNR fMRI time series” because
its false detection rate is non-constant as a function of SNR and fails to match the specified
significance level. In a similar result, Rowe (2005b) derives a Ricean-approximated-based LRT
statistic which takes higher mean values than its Gaussian counterpart.

However, we argue that den Dekker and Sijbers (2005) and Rowe (2005b), which pro-
provide important evidence in favor of Ricean modeling of fMRI data, make assumptions and
approximations which put their results into question. The former, perhaps because it uses
“nonstandard” and “exhaustive” numerical optimization techniques (Solo and Noh, 2007), as-
sumes that the noise variance is known and constant across all voxels when, typically, it is
estimated separately for each voxel time series (Friston et al., 1995). On the other hand, Rowe
(2005b) avoids such numerical difficulties through a Taylor-series-based approximation of the
Rice distribution (Rowe, 2005b), and we argue that use of the exact Rice distribution will yield
optimal Ricean-based results. The previous assumption and approximation are not necessary
when the Expectation Maximization (EM) algorithm (Dempster et al., 1977) is applied to ML
estimation of Ricean parameters (Solo and Noh, 2007; Zhu et al., 2009). However, a study of
Ricean-based LRTs based on this EM scheme is missing from the literature.

Therefore, we perform this study, focusing on Ricean and Gaussian modeling of low-SNR
magnitude fMRI time series and comparing the previous assumption- and approximation-based
LRTs (den Dekker and Sijbers, 2005; Rowe, 2005b) to assumption-free LRTs. In Section 4.2, we
review all of the competing models and LRTs and discuss methods to compare them. Section
4.3 computes and evaluates these LRTs for simulated data, and we discuss our findings in
Section 4.4.

4.2 Methodological Development

Due to our emphasis on Rice-distributed magnitude time series, we begin by deriving the
Ricean probability density function (PDF) of magnitude data. We focus on the time series of
magnitude measurements at a voxel, which we denote as \( r = (r_1, r_2, \ldots, r_n) \), with \( n \) being the number of scans. As discussed in Section 4.1, each magnitude measurement is literally computed as the magnitude \( r_t = \sqrt{y^2_{R,t} + y^2_{I,t}} \), \( t = 1, \ldots, n \), of the real and imaginary measurements \( y_{R,t} \) and \( y_{I,t} \), respectively. This complex-valued observation vector is well-modeled as \( y_{R,t} = x'_t \beta \cos \theta_t + \eta_{R,t} \) and \( y_{I,t} = x'_t \beta \sin \theta_t + \eta_{I,t} \), where \( x'_t \) is the \( t \)th row, \( t = 1, \ldots, n \), of an \( n \times q \) design matrix \( X \) which models the baseline signal, signal drift, and expected BOLD response, \( \theta_t \) is the phase imperfection, and \( \eta_{R,t} \) and \( \eta_{I,t} \) are independent \( \mathcal{N}(0, \sigma^2) \) random variables. Thus, the PDF of the real-imaginary variables \( (y_{R,t}, y_{I,t}) \) takes a bivariate Gaussian form. Starting from this PDF, the Ricean PDF of \( r_t \) results from a transformation to the magnitude-phase variables \( (r_t, \phi_t) \), where \( \phi_t = \arctan(y_{I,t}/y_{R,t}) \), which, after “integrating out” \( \phi_t \), takes the form

\[
 f(r_t|\beta_R, \sigma^2_R) = \frac{r_t}{\sigma^2_R} \exp\left\{ -\frac{r_t^2 + (x'_t \beta_R)^2}{2\sigma^2_R} \right\} \int_{-\pi}^{\pi} \frac{1}{2\pi} \exp\left[ r_t \left( \frac{x'_t \beta_R}{\sigma^2_R} \right) \cos(\phi_t - \theta_t) \right] d\phi_t, \tag{4.1}
\]

for \( r_t \geq 0 \), \( x'_t \beta_R \geq 0 \), and \( \sigma^2_R > 0 \), where the integral expression is equivalent to \( I_0(r_t x'_t \beta_R/\sigma^2_R) \), \( I_0(\cdot) \) being the zeroth order modified Bessel function of the first kind (Abramowitz and Stegun, 1965). We attach the subscript “R” to these Ricean parameters to differentiate them from parameters for other models that we introduce later. We use the notation \( r_t \sim \text{Rice}(x'_t \beta_R, \sigma^2_R) \) as shorthand for (4.1), where the first parameter defines the deterministic signal level and the second defines the noise level. These two parameters, however, are not the mean and variance of the Rice distribution; rather, its first two moments are given by \( E(r_t|x'_t \beta_R, \sigma^2_R) = \sqrt{\pi \sigma^2_R / 2} L_{1/2}(-(x'_t \beta_R)^2/2\sigma^2_R) \) and \( E(r_t^2|x'_t \beta_R, \sigma^2_R) = (x'_t \beta_R)^2 + 2\sigma^2_R \) (Zhu et al., 2009), where the Laguerre polynomial \( L_{1/2}(x) = \exp(-x/2)[(1-x)I_0(-x/2) - xI_1(-x/2)] \), \( I_1(\cdot) \) being the first order modified Bessel function of the first kind (Abramowitz and Stegun, 1965).

### 4.2.1 Models for magnitude fMRI time series

Next, we present the models and associated likelihood ratio tests (LRTs) for activation which we compare in Section 4.3, beginning with some general considerations. All models assume temporal independence of the magnitude time series, perhaps after a prewhitening step. To differentiate the signal and noise parameters, \( \beta \) and \( \sigma^2 \), respectively, and the LRT statistics
A for the different models, we attach identifying subscripts; in contrast, the design matrix $X$

is the same under each model. The test for activation is generally posed as $H_0 : C\beta = 0$ vs. $H_a : C\beta \neq 0$, and the LRT statistics follow asymptotic $\chi^2_m$ null distributions under all models, where $m = \text{rank}(C)$. In deriving LRT statistics under each model, we illustrate calculation of restricted and unrestricted MLEs, in which the likelihood function is maximized under $H_0$ and $H_a$, respectively.

We begin with the Gaussian model, which is the most widely used of all the presented models due its ease of application and the fact that the Ricean-distributed magnitudes are approximately Gaussian-distributed at high SNR. It is given by $r = X\beta_G + \epsilon$, where the error term $\epsilon \sim N(0, \sigma^2_G I_n)$, $I_n$ being the identity matrix of order $n$. Unrestricted MLEs for the parameters $\beta_G$ and $\sigma^2_G$ are given by $\hat{\beta}_G = (X'X)^{-1}X'r$ and $\hat{\sigma}^2_G = (r - X\hat{\beta}_G)'(r - X\hat{\beta}_G)/n$, and restricted MLEs are $\hat{\beta}_G = \Psi \hat{\beta}_G$, where $\Psi = I_q - (X'X)^{-1}C'(C(X'X)^{-1}C)'^{-1}C$, and $\hat{\sigma}^2 = (r - X\hat{\beta}_G)'(r - X\hat{\beta}_G)/n$ (Rowe, 2005b). The LRT statistic is given by $A_G = n \log(\hat{\sigma}^2/\hat{\sigma}^2_G)$.

The Ricean model is given by $r_t \sim \text{indep Rice}(x_t\beta_R, \sigma_R^2)$, $t = 1, \ldots, n$, and, following (4.1), has log-likelihood function

$$\log L(\beta_R, \sigma_R^2 | r) = \sum_{t=1}^n \left[ \log(r_t/\sigma_R^2) - \frac{r_t^2}{2\sigma_R^2} + \log I_0 \left( \frac{r_t(x_t\beta_R)}{\sigma_R^2} \right) \right].$$

(4.2)

Using the Gaussian-model estimates as starting values, we propose hybrid schemes utilizing both EM and Newton-Raphson (NR) iterates to find MLEs to capitalize on the former’s stability and the latter’s convergence speed (McLachlan and Krishnan, 2008). Under unrestricted maximization, EM iterates update the $k$th step estimates $\hat{\beta}_R^{(k)}$ and $\hat{\sigma}_R^{2(k)}$ by $\hat{\beta}_R^{(k+1)} = (X'X)^{-1}X'\hat{u}_R^{(k)}$ and $\hat{\sigma}_R^{2(k+1)} = [r'r - (X'\hat{u}_R^{(k)})'(X'X)^{-1}(X'\hat{u}_R^{(k)})]/2n$, where $\hat{u}_R^{(k)}$ is an $n$-vector with $t$th entry $\hat{u}_R^{(k)} = r_tA(x_t'\hat{\beta}_R^{(k)} r_t/\hat{\sigma}_R^{2(k)})$, $t = 1, \ldots, n$, $A(\cdot)$ being $I_1(\cdot)/I_0(\cdot)$ (Solo and Noh, 2007). Under restricted maximization, EM updates are given by $\hat{\beta}_R^{(k+1)} = \Psi (X'X)^{-1}X'\tilde{u}_R^{(k)}$ and $\hat{\sigma}_R^{2(k+1)} = [r'r - (X'\tilde{u}_R^{(k)})'(X'X)^{-1}(X'\tilde{u}_R^{(k)})]/2n$, where $\Psi$ is as before and $\tilde{u}_R^{(k)}$ is an $n$-vector with $t$th entry $\tilde{u}_R^{(k)} = r_tA(x_t'\hat{\beta}_R^{(k)} r_t/\hat{\sigma}_R^{2(k)})$, $t = 1, \ldots, n$ (Solo and Noh, 2007). The NR iterations can be derived from (4.2) noting the derivative forms $I_0'(\cdot) = I_1(\cdot)$ and $A'(x) = 1 - A(x)/x - A^2(x)$, for $x \neq 0$, $A'(0) = 0.5$ (Schou, 1978). In practice, we applied the hybrid scheme as follows: we began with a large number of EM iterations (say, 1000) due to their simple form and stability.
These EM iterations brought about convergence – as measured by the change in (4.2) – in most cases, with the exception of very low-SNR data, which potentially required thousands of EM iterations. In this case, NR iterations are mixed with EM iterations to bring about convergence much more quickly. Another issue is that the MLEs of Ricean signal parameters are subject to the (parameter space) constraints $x_i'\beta_R \geq 0, \ t = 1, \ldots, n$, which can effect estimation for low-SNR data. The LRT statistic is given by $\Lambda_R = 2[\ell_R(\tilde{\beta}_R, \tilde{\sigma}_R^2) - \ell_R(\hat{\beta}_R, \hat{\sigma}_R^2)]$, where $\ell_R(\cdot, \cdot)$ is shorthand for the log-likelihood function (4.2).

As mentioned in Section 4.1, den Dekker and Sijbers (2005) derives Gaussian- and Ricean-model-based LRT statistics under the assumption of known noise parameters. Notationally, we add asterisks to parameters and LRT statistics under this assumption to distinguish them from their counterparts in the previously discussed models where the noise variance is estimated. For the Gaussian model, the MLEs are the same under the assumption of known noise variance as when the variance is estimated; that is, $\hat{\beta}_G^* = \hat{\beta}_G$ and $\tilde{\beta}_G^* = \tilde{\beta}_G$. The Gaussian-model LRT statistic under the known variance assumption is given by $\Lambda^*_G = [(r - X\hat{\beta}_G^*)'(r - X\hat{\beta}_G^*) - (r - X\tilde{\beta}_G^*)'(r - X\tilde{\beta}_G^*)]/\sigma^2_G$, where $\sigma^2_G$ is the assumed variance. For the Ricean model (with assumed known variance), we improve on the “nonstandard numerical optimization” (Solo and Noh, 2007) of den Dekker and Sijbers (2005) by calculating MLEs via similar EM-NR hybrid schemes as before, except that $\sigma^2_R$, the assumed value of the Ricean noise parameter, should be substituted for all iterates $\hat{\sigma}^2_R$ and $\tilde{\sigma}^2_R$. The Ricean-model LRT statistic under known noise variance is given by $\Lambda^*_R = 2[\ell_R(\tilde{\beta}_R^*, \tilde{\sigma}_R^2) - \ell_R(\hat{\beta}_R^*, \hat{\sigma}_R^2)]$.

Continuing, we present the “Taylor model” approach introduced in Rowe (2005b), where the Rice distribution is approximated by replacing the cosine term in (4.1) by the first two terms of its Taylor series expansion. The paper illustrates maximizing the resulting log-likelihood through an iterative approach, but we find it fails to produce exact MLEs. Therefore we utilize NR iterations instead. In addition, we find that the Taylor-model “PDF” does not integrate to one for low-SNR parameter values, as shown in Figure 4.1. Though this is cause for concern, for consistency, we do not correct for this in calculating the LRT statistic $\Lambda_T$.

We also note that the Gaussian distribution does not integrate to one (over positive support) at low SNRs. To correct for this, we examine a Gaussian model which is truncated at zero and
integral over positive support

\[ \text{Taylor} \]

\[ \text{Gaussian} \]

\[ \text{Rice} / \text{Tr. Norm.} \]

Figure 4.1 Integrals of Taylor, Gaussian, Ricean, and truncated normal PDFs over positive support for different signal parameters \( \mu \) and noise parameter \( \sigma^2 = 1.0 \).

<table>
<thead>
<tr>
<th>LRT Statistic</th>
<th>Model Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Lambda_G )</td>
<td>Gaussian model with estimated variance</td>
</tr>
<tr>
<td>( \Lambda_R )</td>
<td>Ricean model with estimated noise parameter</td>
</tr>
<tr>
<td>( \Lambda_G^* )</td>
<td>Gaussian model with assumed variance</td>
</tr>
<tr>
<td>( \Lambda_R^* )</td>
<td>Ricean model with assumed noise parameter</td>
</tr>
<tr>
<td>( \Lambda_T )</td>
<td>Taylor model</td>
</tr>
<tr>
<td>( \Lambda_{TG} )</td>
<td>Truncated Gaussian model</td>
</tr>
</tbody>
</table>

Table 4.1 Summary of the models and LRT statistics presented in Section 4.2.1.

normalized to integrate to one, which has PDF

\[
f(r_t|\beta_{TG}, \sigma_{TG}^2) = (2\pi\sigma_{TG}^2)^{-1/2} \exp \left[ -\frac{1}{2\sigma_{TG}^2}(r_t - x_t'\beta_{TG})^2 \right] \left[ 1 - \Phi \left( \frac{x_t'\beta_{TG}}{\sigma_{TG}} \right) \right]^{-1}, \quad (4.3)
\]

for \( r_t \geq 0 \), where \( \Phi(\cdot) \) is the standard normal cumulative distribution function (CDF). The LRT statistic under this model, \( \Lambda_{TG} \), can be computed using NR iterations.

Table 4.1 summarizes the models and LRT statistics presented. In the next section, we illustrate methods of evaluating these statistics.

4.2.2 Methods for evaluating activation statistics

We evaluate the previous tests for their ability to discriminate between activated and non-activated voxel time series. That is, in the test of \( H_0 : C\beta = \mathbf{0} \) (not activated) vs. \( H_a : C\beta \neq \mathbf{0} \) (activated), we want these tests to have large probabilities of true activation detection, which is rejecting \( H_0 \) when \( H_0 \) is false, and small probabilities of false detection, which is rejecting \( H_0 \) when \( H_0 \) is in fact true. In simulation experiments, we can control whether \( H_0 \) is true and thus
compute these true and false detection probabilities as the proportions of statistics that reject \( H_0 \). Following standard practice, we reject \( H_0 \) when test statistics are greater than a cutoff value, the \((1 - \alpha)\)th quantile of the \( \chi^2_m \) null distribution, where \( \alpha \) is the specified significance level.

Before a test statistic is considered a candidate for evaluation, we first must check that its false detection rate is equal (or approximately equal) to the specified significance level and constant over the desired range of parameter values. If this is not the case, the false detection rate cannot be controlled in practice and the test statistic is “of little practical use” (den Dekker and Sijbers, 2005). Given that tests satisfy the above criterion, comparing their true detection rates for a (somewhat arbitrary) significance level provides a rough evaluation of them. However, this practice can overlook small differences in false detection rates.

A more informative evaluation tool, the receiver operating characteristic (ROC) curve, considers true and false detection rates simultaneously. To define ROC curves formally, denote the \( r \)th-method test statistics computed under \( H_0 \) and \( H_a \) as \( \{T^{(r)}_{0_i}\}_{i=1}^{n_0} \) and \( \{T^{(r)}_{aj}\}_{j=1}^{n_a} \), respectively, \( r = 1, \ldots, k \), which we assume to be continuous random variables. For any cut-off value \( z \), the true and false detection rates are \( \text{TDR}^{(r)}(z) = \frac{1}{n_a} \sum_{j=1}^{n_a} I(T^{(r)}_{aj} > z) \) and \( \text{FDR}^{(r)}(z) = \frac{1}{n_0} \sum_{i=1}^{n_0} I(T^{(r)}_{0i} > z) \), respectively, where the indicator function \( I(B) \) is 1 if \( B \) is true and 0 otherwise. Then the ROC curve \( \{(\text{FDR}^{(r)}(z), \text{TDR}^{(r)}(z)) : z \in \mathbb{R}\} \) passes through \((0,0)\) when \( z \) is larger than all the test statistics and monotonically increases to \((1,1)\) as \( z \) decreases to be smaller than all the test statistics. Test statistics that discriminate best between activated and non-activated voxel time series will have ROC curves that are closest to the top and left, indicating higher true detection rates and lower false detection rates, respectively.

For more formal comparison of ROC curves, we report results in terms of the area under the ROC curve, or AUC, which summarizes the graphical information in an ROC curve by a single number. Continuing with the notation above, the trapezoidal-rule-based AUC, which we denote \( \hat{\tau}^{(r)} \), can be shown to be equal to the Mann-Whitney \( U \)-statistic \( \sum_{i=1}^{n_0} \sum_{j=1}^{n_a} I(T^{(r)}_{0i} < T^{(r)}_{aj}) \) \( (n_0n_a) \) (Bamber, 1975; Mann and Whitney, 1947). In other words, the AUC \( \hat{\tau}^{(r)} \) is the proportion of null-alternative statistic pairs in which the statistic computed under \( H_a \) is greater than the one computed under \( H_0 \) – i.e. the null and alternative statistics would be correctly
assigned. Denote the sample AUCs as $\tilde{\tau} = (\tilde{\tau}^{(1)}, \ldots, \tilde{\tau}^{(k)})$ and their population versions as $\tau$. Using the general theory of $U$-statistics, DeLong et al. (1988) shows that $\tilde{\tau}$ is asymptotically normal, unbiased for $\tau$, and has covariance matrix $S = S_0/n_0 + S_a/n_a$, where $S_0$ and $S_a$ are $k \times k$ matrices with $(r,s)$-elements

$$S_0^{(r,s)} = \frac{1}{n_0 - 1} \sum_{i=1}^{n_0} [V_0^r(T_{0i}^{(r)}) - \tilde{\tau}^{(r)}][V_0^s(T_{0i}^{(s)}) - \tilde{\tau}^{(s)}],$$

$$S_a^{(r,s)} = \frac{1}{n_a - 1} \sum_{j=1}^{n_a} [V_a^r(T_{aj}^{(r)}) - \tilde{\tau}^{(r)}][V_a^s(T_{aj}^{(s)}) - \tilde{\tau}^{(s)}],$$

for $1 \leq r, s \leq k$, with $V_0^r(\cdot) = (1/n_a) \sum_{j=1}^{n_a} I(\cdot < T_{aj}^{(r)})$ and $V_a^r(\cdot) = (1/n_0) \sum_{i=1}^{n_0} I(T_{0i}^{(r)} < \cdot)$ (DeLong et al., 1988). As a result, the test comparing the AUCs of the $r$th and $s$th test, $H_0 : \tau^{(r)} = \tau^{(s)}$ vs. $H_a : \tau^{(r)} \neq \tau^{(s)}$, has test statistic $z = (\tilde{\tau}^{(r)} - \tilde{\tau}^{(s)})/\sqrt{e_{rs}^T S e_{rs}}$ with a standard normal asymptotic null distribution, where $e_{rs}$ is a vector of length $k$ with 1 and -1 at the $r$th and $s$th positions, respectively, and zeros elsewhere.

### 4.3 Experimental Evaluations

We generated fMRI magnitude time series according to $r_t \sim \text{indep Rice}(x_t^T \beta, \sigma^2)$, $t = 1, \ldots, 256$, which can be efficiently simulated as the magnitude of bivariate normal components as described in Section 4.2. The design matrix $X$ included an intercept to model the baseline MR signal level and a $\pm 1$ square wave, alternating every 16 time points, to model the expected BOLD response of a block-design experiment. In addition, we modeled linear drift in the signal with an arithmetic sequence from -1 to 1; separate simulations included and excluded this effect. Thus, $X$ contained $q = 2$ or 3 columns, and $\beta$ correspondingly consisted of $(\beta_0, \beta_1)$ or $(\beta_0, \beta_1, \beta_2)$, representing the size of the baseline, activation, and (when included) drift effects. Since only $\beta_1$ is activation-related, the activation test is $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ and the LRT statistics have $\chi^2_1$ null distributions. When we simulated the drift effect, before testing for activation, we first performed an LRT on $\beta_2$ to determine whether it should be included in the model.

In specifying simulation parameters, we fix the noise parameter $\sigma^2 = 1.0$ for easy interpretation of the signal-to-noise ratio, which we define as $\text{SNR} = \beta_0/\sigma$ because the baseline signal,
in practice, is much larger than the other effects. We varied $\beta_0$ from 0.2 to 5.0, which represent low-SNR values, set $\beta_1 = 0.2$ and 0.0 for activated and non-activated time series, respectively, and set $\beta_2 = 0.2$ (when used). For consistency, we follow den Dekker and Sijbers (2005) in assuming that the known noise parameters are equal to the true Ricean noise parameter – i.e. $\sigma^2_R = \sigma^2_G = \sigma^2 = 1.0$. We computed 100,000 LRT statistics for each of the six models and set of parameter values. In Section 4.3.1, we examine the properties of the parameter estimates and the computation times under each model. Then, we evaluate the activation statistics in Section 4.3.2.

4.3.1 Parameter estimation and computation times

We examine the properties of the MLEs under each model by computing bias, standard error (SE), and root mean squared error (RMSE). Recall that $\text{RMSE}^2(\cdot) = \text{Bias}^2(\cdot) + \text{SE}^2(\cdot)$ for any estimator, so the RMSE encompasses both bias and variation in measuring the quality of an estimator. These quantities are plotted against SNR in Figure 4.2 for the simulations without drift. At low SNR, the Ricean-model MLEs show less bias than the other models, but they also show larger standard errors. When the two are combined, the RMSE results are mixed: while Ricean-model MLEs show the lowest RMSEs for $\hat{\beta}_0$ and $\hat{\sigma}^2$, they show the highest RMSEs for $\hat{\beta}_1$.

We next examine the computation times under each model, which, considering that fMRI datasets contain tens of thousands of voxel time series, can potentially be restrictive for fMRI applications. Table 4.2 displays the number of seconds required to compute 10,000 LRT statistics under each model on a 2.13 GHz Intel® Core™ 2 CPU. The LRTs $\Lambda_R$ and $\Lambda_{TG}$, whose calculations involve numerous iterations, require the largest amount of computation time while the Gaussian-model LRTs, which only involve closed-form calculations, require the least. As a side note, we also show that changing the implementation of the EM-NR hybrid Ricean parameter estimation scheme does not dramatically effect computation time; in Table 4.2, $\Lambda_{R,NR}$ refers to computing $\Lambda_R$ with mostly NR iterations, beginning with only 10 EM steps, while the $\Lambda_R$-column refers the time involved for the scheme as described in Section 4.2.1, which begins with 1000 EM iterations.
Figure 4.2 (a)-(c) Biases, (d)-(f) standard errors (SE), and (g)-(i) root mean squared errors (RMSE) of the unrestricted MLEs under each model plotted against SNR(= β0). The models are labeled in (a) as in Table 4.1.

4.3.2 Evaluation of activation tests

The true and false detection rates of the LRT statistics, according to an α = 0.05 significance level, are displayed for different SNRs in Table 4.3 and Figure 4.3, respectively. Only the “no-drift” simulations are shown as the other results are very similar. Before examining the results as a whole, we confirm the results of den Dekker and Sijbers (2005), which compares Λ* R and Λ* G, and Rowe (2005b), which compares Λ T and Λ G. We agree with the former’s assessment that the true detection rates of Λ* R are higher than those of Λ* G and note that Λ T, as suggested by Rowe (2005b), has a higher true detection rate than Λ G. Recall that, in both papers, the authors’ used such comparisons as evidence that Ricean-model-based activation tests outperformed their
Table 4.2 Computation times (in seconds) for 10,000 LRT statistics under each model.

<table>
<thead>
<tr>
<th>β₀</th>
<th>Λ₉</th>
<th>Λ₉,R,ΝR</th>
<th>Λ₉,G</th>
<th>Λ₉,R⁺</th>
<th>Λ₉,G⁺</th>
<th>Λ₉,T</th>
<th>Λ₉,TG</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>152.5</td>
<td>108.7</td>
<td>3.3</td>
<td>105.0</td>
<td>3.3</td>
<td>20.0</td>
<td>186.5</td>
</tr>
<tr>
<td>1.0</td>
<td>119.8</td>
<td>90.7</td>
<td>3.2</td>
<td>31.2</td>
<td>3.3</td>
<td>19.2</td>
<td>171.2</td>
</tr>
<tr>
<td>1.5</td>
<td>41.0</td>
<td>62.5</td>
<td>3.2</td>
<td>22.1</td>
<td>3.3</td>
<td>17.7</td>
<td>145.7</td>
</tr>
<tr>
<td>2.0</td>
<td>26.0</td>
<td>49.7</td>
<td>3.2</td>
<td>22.1</td>
<td>3.3</td>
<td>17.1</td>
<td>119.7</td>
</tr>
<tr>
<td>2.5</td>
<td>21.7</td>
<td>41.1</td>
<td>3.3</td>
<td>16.0</td>
<td>3.3</td>
<td>17.2</td>
<td>95.8</td>
</tr>
<tr>
<td>3.0</td>
<td>18.4</td>
<td>41.1</td>
<td>3.3</td>
<td>15.2</td>
<td>3.2</td>
<td>17.3</td>
<td>75.2</td>
</tr>
</tbody>
</table>

Table 4.3 True detection rates of the different LRT statistics, according to an α = 0.05 significance level, at different β₀ (or, equivalently, SNR) values.

<table>
<thead>
<tr>
<th>β₀</th>
<th>Λ₉</th>
<th>Λ₉,G</th>
<th>Λ₉,R⁺</th>
<th>Λ₉,G⁺</th>
<th>Λ₉,T</th>
<th>Λ₉,TG</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.10</td>
<td>0.09</td>
<td>0.05</td>
<td>0.01</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>0.4</td>
<td>0.22</td>
<td>0.21</td>
<td>0.18</td>
<td>0.04</td>
<td>0.25</td>
<td>0.21</td>
</tr>
<tr>
<td>0.6</td>
<td>0.38</td>
<td>0.37</td>
<td>0.37</td>
<td>0.13</td>
<td>0.42</td>
<td>0.37</td>
</tr>
<tr>
<td>0.8</td>
<td>0.53</td>
<td>0.51</td>
<td>0.52</td>
<td>0.26</td>
<td>0.57</td>
<td>0.51</td>
</tr>
<tr>
<td>1.0</td>
<td>0.64</td>
<td>0.63</td>
<td>0.63</td>
<td>0.41</td>
<td>0.67</td>
<td>0.63</td>
</tr>
<tr>
<td>1.2</td>
<td>0.71</td>
<td>0.71</td>
<td>0.71</td>
<td>0.54</td>
<td>0.74</td>
<td>0.71</td>
</tr>
<tr>
<td>1.4</td>
<td>0.76</td>
<td>0.76</td>
<td>0.76</td>
<td>0.64</td>
<td>0.79</td>
<td>0.76</td>
</tr>
<tr>
<td>2.0</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.79</td>
<td>0.85</td>
<td>0.84</td>
</tr>
<tr>
<td>3.0</td>
<td>0.87</td>
<td>0.87</td>
<td>0.87</td>
<td>0.86</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>4.0</td>
<td>0.88</td>
<td>0.88</td>
<td>0.88</td>
<td>0.87</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td>5.0</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.88</td>
<td>0.89</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Gaussian-based counterparts. However, we argue that these comparisons are invalid, because the false detection rates of Λ̂₉,G⁺ and Λ₉,T differ from α = 0.05 (and from the tests they are compared to). Note also that, for this reason, neither of these tests would be used in practice.

For the most part, the other four tests have similar true and false detection rates. As discussed in Section 4.2.2, we “combine” this information in calculating the areas under the ROC curve (AUCs), which allow clearer evaluation of the tests. Using Λ₉,G (the standard test) as a baseline for comparison, Table 4.4 displays the z-statistics obtained, as illustrated in Section 4.2.2, when the other activation tests’ AUCs are compared with it. Very few of these statistics represent significantly different test performance, as measured at α = 0.05 with a Bonferroni correction, which is surprising in that it differs from the conclusions of den Dekker and Sijbers (2005) and Rowe (2005b). These few significant differences are seen for the Ricean-based tests Λ₉,R and Λ₉,R⁺, mostly the latter, at SNRs below one, SNRs not likely present for in vivo fMRI data.
Figure 4.3 False detection rates of the different LRT statistics, according to an $\alpha = 0.05$ significance level, are plotted against SNR. The legend follows Table 4.1.

Furthermore, the “significant” differences are quite small in reality, noting that the standard errors of the AUCs tested, as shown in Table 4.4, are on the order of $10^{-3}$. Because these results seemingly run counter to those in den Dekker and Sijbers (2005) and Rowe (2005b), which show advantages of Ricean-based tests for SNRs as high as 5, we offer explanations in the following paragraphs.

First, we argue that the comparison of $\Lambda^*_R$ and $\Lambda^*_G$ in den Dekker and Sijbers (2005) unfairly favors the former due to a faulty assumption made when calculating the latter. Namely, the paper assumes that the Gaussian-model variance parameter is equal to the true Ricean noise parameter; that is, $\sigma^2_G^* = \sigma^2_R$. We argue that when the Gaussian model is applied to the simulated Rice-distributed data, the Gaussian-model noise parameter $\sigma^2_G$ represents the variance of the Rice-distributed data, which, as discussed in Section 4.2, is different from the Ricean parameter $\sigma^2_R$. To illustrate, we plot the (theoretical) variance of the Rice($\mu, 1$) distribution and the middle 95% of the estimates $\hat{\sigma}^2_G$ for simulated Rice($\mu, 1$) data for different $\mu$ in Figure 4.4. At low SNR, the estimates $\hat{\sigma}^2_G$, which are centered around the Ricean variance as in Figure 4.4, are smaller than the assumed value $\sigma^2_G^*$ (again, which is equal to $\sigma^2_R = 1$ under the assumption). Thus, we argue that the assumption results in $\sigma^2_G^*$ being over-specified at low SNR, which, due to the expression for $\Lambda^*_G$, results in $\Lambda^*_G$ taking lower values than it otherwise would. There are no such problems in the calculation of $\Lambda^*_R$: the specification of $\sigma^2_R^* = \sigma^2_R$ is a perfect one (although, on another topic, it unrealistically neglects estimation error). As
Figure 4.4 The (theoretical) variance of the Rice($\mu, 1.0$) distribution plotted against $\mu$ (or alternatively, SNR), with estimates of the middle 95% of the distributions of $\hat{\sigma}_G^2$ (obtained from simulation) at $\mu = 0.0, 0.5, \ldots, 6.0$. A horizontal line at $\sigma_G^2 = \sigma_R^2 = 1.0$ is given for comparison.

a result, as we have seen, $\Lambda^*_G$ has much lower true and false detection rates than $\Lambda^*_R$, which den Dekker and Sijbers (2005) use as evidence of the superiority of the Ricean-model LRT. However, we have shown that if the Gaussian-model variance is estimated correctly, as in $\Lambda_G$, the Gaussian-model test performs as well as the Ricean model tests $\Lambda_R$ and $\Lambda^*_R$ at all but the lowest SNRs.

The explanation involved for Rowe (2005b) is simpler. The paper focuses on parameter estimation under different models and thus, as is done for the parameter estimates, only displays the mean and variance of the activation statistics. Although the mean values and, as a result, the true detection rates of $\Lambda_T$ are higher than $\Lambda_G$, we have seen the false detection rate of $\Lambda_T$ is also higher. When the two are considered together, the performances of $\Lambda_T$ and $\Lambda_G$ as measured by AUC are not significantly different. In fact, because the Taylor-model LRT has problems maintaining the nominal $\alpha$-level false detection rate, perhaps due to its PDF not integrating to one, we prefer $\Lambda_G$. As a side note, the properness of the Gaussian PDF over positive support does not seem to effect $\Lambda_G$, as the detection rates and AUCs of $\Lambda_G$ and $\Lambda_{TG}$ are nearly identical.
4.4 Discussion

In this paper, we have performed an expanded study on Gaussian and Ricean modeling of low-SNR fMRI magnitude time series. Noting that previous work showing improved performance of Ricean-based activation tests was based on assumptions and approximations, we removed them by applying an EM-NR hybrid algorithm to Ricean parameter estimation. Through a simulation study, we found that the performances of Ricean- and Gaussian-model activation tests, as measured by AUC, are only significantly different at very low SNRs, SNRs most likely below the range of real fMRI data. We explained how the previous work could give seemingly different results. Based on the Gaussian model’s simple implementation and low computational expense, we recommend it over the Ricean model at all SNR for activation tests based on fMRI magnitude time series.
Table 4.4 Comparisons of $\hat{\tau}_G$, the AUCs of $\Lambda_G$, to those of the other LRT statistics, as measured by the $z$-statistics described in Section 4.2.2. The notation follows that of Table 4.1 – e.g. $z_R$ refers to the $z$-statistic calculated from comparing the AUCs of $\Lambda_G$ and $\Lambda_R$. Bold statistics represent significant differences at the $\alpha = 0.05$ level, after a Bonferroni adjustment.

(a) with drift term

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>$\hat{\tau}_G$</th>
<th>SE($\hat{\tau}_G$)</th>
<th>$z_R$</th>
<th>$z^*_R$</th>
<th>$z^*_G$</th>
<th>$z_T$</th>
<th>$z_{TG}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>0.668</td>
<td>0.0012</td>
<td>2.7</td>
<td>$5.5$</td>
<td>0.6</td>
<td>-2.6</td>
<td>0.1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.776</td>
<td>0.0011</td>
<td>2.6</td>
<td>$3.6$</td>
<td>1.6</td>
<td>-0.5</td>
<td>1.1</td>
</tr>
<tr>
<td>0.8</td>
<td>0.854</td>
<td>0.0009</td>
<td>0.1</td>
<td>3.3</td>
<td>0.6</td>
<td>-3.5</td>
<td>-1.1</td>
</tr>
<tr>
<td>1.0</td>
<td>0.899</td>
<td>0.0007</td>
<td>-0.2</td>
<td>0.5</td>
<td>-0.7</td>
<td>-0.6</td>
<td>-0.4</td>
</tr>
<tr>
<td>1.2</td>
<td>0.926</td>
<td>0.0006</td>
<td>1.0</td>
<td>0.5</td>
<td>0.4</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>1.4</td>
<td>0.943</td>
<td>0.0005</td>
<td>0.3</td>
<td>0.3</td>
<td>1.8</td>
<td>-0.8</td>
<td>0.9</td>
</tr>
<tr>
<td>2.0</td>
<td>0.964</td>
<td>0.0004</td>
<td>-1.1</td>
<td>-0.3</td>
<td>1.4</td>
<td>-1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>3.0</td>
<td>0.972</td>
<td>0.0003</td>
<td>0.7</td>
<td>0.2</td>
<td>0.1</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>4.0</td>
<td>0.974</td>
<td>0.0003</td>
<td>0.5</td>
<td>1.2</td>
<td>2.6</td>
<td>0.9</td>
<td>1.1</td>
</tr>
<tr>
<td>5.0</td>
<td>0.975</td>
<td>0.0003</td>
<td>-0.5</td>
<td>2.6</td>
<td>-0.1</td>
<td>-0.2</td>
<td>-0.5</td>
</tr>
</tbody>
</table>

(b) without drift term

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>$\hat{\tau}_G$</th>
<th>SE($\hat{\tau}_G$)</th>
<th>$z_R$</th>
<th>$z^*_R$</th>
<th>$z^*_G$</th>
<th>$z_T$</th>
<th>$z_{TG}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.555</td>
<td>0.0013</td>
<td>0.3</td>
<td>17.7</td>
<td>1.1</td>
<td>-2.5</td>
<td>-1.5</td>
</tr>
<tr>
<td>0.4</td>
<td>0.670</td>
<td>0.0012</td>
<td>2.3</td>
<td>$7.5$</td>
<td>0.2</td>
<td>-3.1</td>
<td>-0.1</td>
</tr>
<tr>
<td>0.6</td>
<td>0.779</td>
<td>0.0010</td>
<td>4.1</td>
<td>$4.0$</td>
<td>1.5</td>
<td>-2.1</td>
<td>-0.5</td>
</tr>
<tr>
<td>0.8</td>
<td>0.853</td>
<td>0.0008</td>
<td>1.9</td>
<td>$4.6$</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>1.0</td>
<td>0.900</td>
<td>0.0007</td>
<td>1.4</td>
<td>1.4</td>
<td>0.9</td>
<td>0.6</td>
<td>-0.2</td>
</tr>
<tr>
<td>1.2</td>
<td>0.927</td>
<td>0.0006</td>
<td>-0.6</td>
<td>0.4</td>
<td>0.7</td>
<td>-0.7</td>
<td>-1.0</td>
</tr>
<tr>
<td>1.4</td>
<td>0.943</td>
<td>0.0005</td>
<td>-0.3</td>
<td>0.0</td>
<td>-0.4</td>
<td>0.7</td>
<td>-0.3</td>
</tr>
<tr>
<td>2.0</td>
<td>0.963</td>
<td>0.0004</td>
<td>0.1</td>
<td>0.9</td>
<td>1.2</td>
<td>-0.3</td>
<td>-0.1</td>
</tr>
<tr>
<td>3.0</td>
<td>0.971</td>
<td>0.0003</td>
<td>0.6</td>
<td>1.9</td>
<td>1.9</td>
<td>0.0</td>
<td>0.2</td>
</tr>
<tr>
<td>4.0</td>
<td>0.974</td>
<td>0.0003</td>
<td>0.3</td>
<td>2.6</td>
<td>0.8</td>
<td>-0.6</td>
<td>0.2</td>
</tr>
<tr>
<td>5.0</td>
<td>0.975</td>
<td>0.0003</td>
<td>-0.6</td>
<td>0.0</td>
<td>1.3</td>
<td>0.9</td>
<td>0.1</td>
</tr>
</tbody>
</table>
CHAPTER 5. ESTIMATING PARAMETERS FOR RICE-DISTRIBUTED TIME SERIES OBSERVATIONS WITH APPLICATIONS TO FMRI DATA

A paper in preparation
Ranjan Maitra, Daniel B. Rowe, Daniel W. Adrian

Abstract

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

5.1 Introduction

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

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

Most statistical analyses focus on the magnitude data computed from the complex-valued measurements resulting from Fourier reconstruction (Jezzard and Clare, 2001) and discard the phase information. Because the real and imaginary measurements are well-modeled as two independent normal random variables with the same variance (Wang and Lei, 1994), these magnitude measurements follow the Rice distribution (Rice, 1944; Gudbjartsson and Patz, 1995). However, standard analyses assume that magnitude data are Gaussian-distributed, even though the Gaussian approximation of the Rice distribution is only valid at high signal-to-noise ratios (SNRs). This factor is increasingly important because the SNR is proportional to voxel volume (Lazar, 2008); thus an increase in the fMRI spatial resolution will correspond to a lowering of the SNR, making the Gaussian distributional approximation for the magnitude
In its simplest form, analysis of magnitude fMRI time series assumes no autocorrelation; however the naïveté of this assumption is widely recognized. There are many reasons for this: one is that the hemodynamic response disperses (or “smears,” in fMRI jargon) neural activation. The hemodynamic (or BOLD) response to a single neural activation takes 15 to 20 seconds (Lazar, 2008), which is much longer than the sampling intervals of many fMRI techniques – 100 ms-5 s for echo-planar imaging (EPI) techniques (Friston et al., 1994). Further, the neuronal response, which can be modeled as a point response or a delta function (Friston et al., 1994), is itself very fast when compared to the BOLD response. Since fMRI experiments measure the BOLD response over time, the above discussion means that the observed time series within each voxel are correlated. Friston et al. (1994) also contend that the neuronal process is composed of “intrinsic” neuronal activities in addition to the stimulus-related response. Consequently, the authors say, autocorrelations in the observed time series arise from two neural components, both measured through the hemodynamic response: one that is experimentally induced owing to the stimulus and another that is due to intrinsic neuronal activity. The first component is modeled by convolution of the stimulus time course with the HRF, as discussed previously, while the second is modeled with autocorrelation. Additional sources of autocorrelation are also provided by the subject’s cardiac and respiratory cycles (Friston et al., 2000).

Precise modeling of this temporal correlation is essential to maintaining assumed significance levels in tests for activation (Purdon and Weisskoff, 1998). Many analyses extend the linear model by introducing autocorrelated errors (Lazar, 2008). Prewhitening these errors is a common procedure, based on estimated autoregressive (AR) (Bullmore et al., 1996; Marchini and Ripley, 2000) or autoregressive moving average (ARMA) (Locascio et al., 1997) models, which produces the most efficient estimators. However, this approach can bias significance levels (Friston et al., 2000; Woolrich et al., 2001), so temporal (Worsley and Friston, 1995) and spatial (Worsley et al., 2002) smoothing have been recommended for more robustness. Likelihood-based activation statistics, based on incorporating an AR temporal correlation structure into the likelihood function, have also been proposed as a less-biased alternative to prewhitening approaches (den Dekker et al., 2009).
The above approaches all make Gaussian distributional assumptions for the observed magnitude time series, which as discussed before, is not appropriate, even approximately, at low SNR. This has led to the development of Rice-distributed magnitude-data models (den Dekker and Sijbers, 2005; Rowe, 2005b; Solo and Noh, 2007; Zhu et al., 2009) which have, understandably, shown improved power of detection over their Gaussian counterparts at low SNR. These models, however, assume independence in the time series; incorporating autocorrelation in Rice-based models is impeded by the fact that the above approaches, such as ARMA modeling and prewhitening, are based on the Gaussian distribution.

In this paper, we develop a Ricean model for fMRI magnitude time series which incorporates AR\((p)\) dependence. Due to the previously discussed “mismatch” between Gaussian-based time series techniques and Ricean-distributed magnitude data, we do not model the magnitudes directly and instead utilize the fMRI data acquisition process as follows. Because the Ricean-distributed magnitude observations are computed from Gaussian-distributed real and imaginary components, we apply the AR\((p)\) dependence to this latent complex-valued data. In Section 5.2, because this complex-valued data is composed of observed magnitudes and “missing” phase data, we present the model through an EM algorithm (Dempster et al., 1977) framework and illustrate its use in parameter estimation; we also illustrate computation of approximate standard errors of these parameter estimates and tests for activation and AR order detection through extensions of the EM algorithm. We compare these AR\((p)\)-Ricean-model parameter estimates and test statistics to those based upon a Gaussian AR\((p)\) model for simulated and real fMRI data in Sections 5.3 and 5.4, respectively, and discuss our results in Section 5.5.

5.2 Methodology

We focus on the magnitude time series at a voxel, which we denote as \(r = (r_1, \ldots, r_n)\), where \(n\) is the number of scans. As discussed in Section 5.1, we incorporate autocorrelation into a Ricean-distributed model for \(r\) by applying AR\((p)\) errors to the real and imaginary (Gaussian) time series, denoted \(y_R = (y_{R1}, \ldots, y_{Rn})\) and \(y_I = (y_{I1}, \ldots, y_{In})\), respectively. After transforming the distribution of \((y_R, y_I)\) to the magnitude-phase variables \((r, \phi)\), where \(\phi = (\phi_1, \ldots, \phi_n)\), we apply the EM algorithm to obtain maximum likelihood estimates (MLEs) of
model parameters, which we denote by $\tau$. Because the phase data $\phi$ is discarded in “magnitude-only” data analysis, the observed, missing, and complete data of EM algorithm terminology are represented by $r$, $\phi$, and $(r, \phi)$, respectively.

### 5.2.1 Parameter estimation via the EM algorithm

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

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

$$
\begin{pmatrix}
y_R \\
y_I
\end{pmatrix} = 
\begin{pmatrix}
X & 0 \\
0 & X
\end{pmatrix}
\begin{pmatrix}
\beta \cos \theta \\
\beta \sin \theta
\end{pmatrix} + 
\begin{pmatrix}
\eta_R \\
\eta_I
\end{pmatrix},
$$

(5.1)

where the $n \times q$ design matrix $X$ models effects such as the baseline signal level, signal drift, and the expected BOLD response, and $\theta$ represents the constant mean of the phase time series $\phi$. (Note that since magnitude time series contain no phase information, $\theta$ is neither known nor estimated.) The error terms $\eta_R = (\eta_{R1}, \ldots, \eta_{Rn})$ and $\eta_I = (\eta_{I1}, \ldots, \eta_{In})$ are independent AR$(p)$ time series parameterized by AR coefficients $\alpha = (\alpha_1, \ldots, \alpha_p)$ and white noise variance $\sigma^2$; that is, for $t = 1, \ldots, n$, $\eta_{Rt} = \sum_{i=1}^{p} \alpha_i \eta_{R,t-i} + \epsilon_{Rt}$ and $\eta_{It} = \sum_{i=1}^{p} \alpha_i \eta_{I,t-i} + \epsilon_{It}$, with $\epsilon_{Rt}, \epsilon_{It} \sim \text{iid } \mathcal{N}(0, \sigma^2)$. We also denote $\gamma_j$ as the lag-$j$ autocovariance, $j = 0, \ldots, p$, and use $(\gamma_0, \ldots, \gamma_p)$ as an alternative parameterization of $(\alpha, \sigma^2)$, obtained via the Yule-Walker equations (Shumway and Stoffer, 2006). Additionally, we define $R_n$ such that $\text{Cov}(\eta_R) = \text{Cov}(\eta_I) = \sigma^2 R_n$. The AR order $p$, which is assumed known in this model formulation, can be chosen using the order selection procedure described in Section 5.2.2. Thus, the complete-data
log-likelihood is
\[ \log f(y_R, y_I; \tau) = -n \log \sigma^2 - \log |R_n| - h/2\sigma^2, \tag{5.2} \]
where \( h = \alpha' D \alpha \), \( \alpha \) being the \((p+1)\)-vector \((1, -\alpha_1, \ldots, -\alpha_p)\) and \( D \) the \((p+1) \times (p+1)\) symmetric matrix with \((i, j)\)th entry \( d_{ij} = \sum_{t=1}^{n-i-j} [\eta_{R,t+i}\eta_{R,t+j} + \eta_{I,t+i}\eta_{I,t+j}] \), \( 0 \leq i, j \leq p \) (Pourahmadi, 2001). To transform the complete log-likelihood function from real-imaginary to magnitude-phase variables, we apply the relations \( y_{Rt} = r_t \cos \phi_t \) and \( y_{It} = r_t \sin \phi_t \), \( t = 1, \ldots, n \), to \( d_{ij} \), which produces \( d_{ij} = \sum_{t=1}^{n-i-j} r_{t+i}r_{t+j} \cos(\phi_{t+i} - \phi_{t+j}) - \mu_{t+i}r_{t+j} \cos(\phi_{t+j} - \theta) - \mu_{t+j}r_{t+i} \cos(\phi_{t+i} - \theta) + \mu_{t+i}\mu_{t+j} \), where \( \mu_t = x_t' \beta \), \( x_t' \) being the \( t \)th row of \( X \). In view of (5.2) and this expression for \( d_{ij} \), we note that the E-step involves two categories of expectations: the univariate expectations \( E[\cos(\phi_t - \theta)|r_t, \tau^{(k)}] \), \( t = 1, \ldots, n \), and the bivariate expectations \( E[\cos(\phi_t - \phi_{t+j})|r_t, r_{t+j}, \tau^{(k)}] \), \( j = 1, \ldots, p \), \( t = 1, \ldots, n - j \).

The former expectations are with respect to the von-Mises distribution \( VM(\cdot, \cdot) \), which is defined in Appendix 5.6.1. It can be shown that \( \phi_t|r_t, \tau^{(k)} \) is von-Mises by transforming \((y_{Rt}, y_{It})\), which are independent and follow \( N(\mu_t^{(k)} \cos \theta, \gamma_0^{(k)}) \) and \( N(\mu_t^{(k)} \sin \theta, \gamma_0^{(k)}) \) distributions, to the variables \((r_t, \phi_t)\). This gives \( f(\phi_t|r_t, \tau^{(k)}) \propto \exp\{[\mu_t^{(k)} r_t/\gamma_0^{(k)}] \cos(\phi_t - \theta)\} \), where \( \mu_t^{(k)} = x_t' \beta^{(k)} \), which, as seen in Appendix 5.6.1, is the probability density function (PDF) of the \( VM(\theta, \mu_t^{(k)} r_t/\gamma_0^{(k)}) \) distribution. Thus, again following from Appendix 5.6.1, the univariate expectations \( E[\cos(\phi_t - \theta)|r_t, \tau^{(k)}] = A(\mu_t^{(k)} r_t/\gamma_0^{(k)}) \), \( t = 1, \ldots, n \), where \( A(\cdot) = I_1(\cdot)/I_0(\cdot) \), \( I_j(\cdot) \) being the \( j \)th order modified Bessel function of the first kind (Abramowitz and Stegun, 1965).

The bivariate expectations can be approximated via univariate Monte Carlo integration as follows. First, we generate \( \psi^{(1)}, \ldots, \psi^{(m)} \sim iid VM(0, \mu_t^{(k)} r_t/\gamma_0^{(k)}) \), which can be efficiently accomplished through the rejection sampling algorithm of Best and Fisher (1979). The expectation \( E[\cos(\phi_t - \phi_{t+j})|r_t, r_{t+j}, \tau^{(k)}] \) is then approximated as
\[ \frac{1}{m} \sum_{i=1}^{m} \frac{A(K^{(i)})}{K^{(i)}} \kappa^* \cos \psi^{(i)} + \delta, \tag{5.3} \]
where \( K^{(i)} = \sqrt{\kappa^*^2 + \delta^2 + 2\kappa^*\delta \cos \psi^{(i)}} \), \( \kappa^* = r_{t+j}(\gamma_0^{(k)} \mu_{t+j}^{(k)} - \gamma_j^{(k)} \mu_j^{(k)})/(\gamma_0^{(k)} - \gamma_j^{(k)}) \), and \( \delta = \gamma_j^{(k)} r_{t+j}(\gamma_0^{(k)} - \gamma_j^{(k)})/2(\gamma_0^{(k)} - \gamma_j^{(k)}) \). See Appendix 5.6.2 for a derivation of (5.3).

To summarize, the E-step replaces the \( h \)-term in (5.2) by its expectation, which we denote
by $h^{(k)} = \hat{\alpha}'D^{(k)}\hat{\alpha}$, where $D^{(k)}$ is a $(p + 1) \times (p + 1)$ symmetric matrix with $(i, j)$th entry

$$d^{(k)}_{ij} = \sum_{t=1}^{n-i-j} \left\{ r_{t+i} r_{t+j} E[\cos(\phi_{t+i} - \phi_{t+j})|r_{t+i}, r_{t+j}, \tau^{(k)}] - \mu_{t+i} r_{t+j} A(r_{t+j} \mu_{t+j}/\gamma^{(k)}_0) - \mu_{t+j} r_{t+i} A(r_{t+i} \mu_{t+i}/\gamma^{(k)}_0) + \mu_{t+i} \mu_{t+j} \right\}. \tag{5.4}$$

Also, to prepare for the M-step, we represent the $\beta$-dependent portion of $h^{(k)}$ as $h^{(k)}(\beta) = \beta'X'R^{-1}nX\beta - 2\beta'X'R^{-1}nXu^{(k)}$, where $u^{(k)}$ is a vector of length $n$ with $t$th entry $u_t^{(k)} = r_t A(\mu_t^{(k)} r_t/\gamma^{(k)}_0)$, $t = 1, \ldots, n$.

Because the M-step does not have a closed form, we obtain $\tau^{(k+1)}$ through three conditional maximization steps as described in the ECM algorithm of Meng and Rubin (1993). First, given $\beta^{(k)}$, we calculate $\alpha^{(k+1)}$ from the equations (a modification of Miller, 1995)

$$\sum_{j=1}^{p} \left( d^{(kk)}_{ij} + 2 j \gamma^{(kk)}_{j-i} \right) \alpha_j = d^{(kk)}_{i0}, \quad i = 1, \ldots, p, \tag{5.5}$$

where $d^{(kk)}_{ij}$ substitutes $\mu^{(k)}_t$ for $\mu_t$, $t = 1, \ldots, n$, in (5.4) and $\gamma^{(kk)}_j = d^{(kk)}_{0j}/(2n)$ estimates the lag-$j$ autocovariance, $j = 0, \ldots, p$. Second, conditioned on $\alpha^{(k+1)}$, we calculate $\beta^{(k+1)}$ as $\beta^{(k+1)} = (X'R^{-1}(k+1)nX)^{-1}X'R^{-1}(k+1)nXu^{(k)}$, which follows from minimizing the expression for $h^{(k)}(\beta)$ in the previous paragraph, where $R^{-1}(k+1)n$ is calculated from $\alpha^{(k+1)}$ (Pourahmadi, 2001). Finally, we calculate $\sigma^{2(k+1)} = h^{(k,k+1)}/(2n)$, where $h^{(k,k+1)}$ substitutes $(\alpha^{(k+1)}, \beta^{(k+1)})$ for $(\alpha, \beta)$ in $h^{(k)}$. To complete the EM algorithm, we calculate $\gamma^{(k+1)}_j$, $j = 0, \ldots, p$, from $(\alpha^{(k+1)}, \sigma^{2(k+1)})$ and the Yule-Walker equations for use in the next E-step.

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

We utilize the Fisher information matrix for calculation of approximate standard errors for the MLEs. However, because the observed-data likelihood function is intractable in this case, we use the empirical information matrix (Meilijson, 1989), an estimate of the Fisher information matrix which can be calculated from the complete-data likelihood and its expectation under the E-step. More specifically, the empirical information matrix is the sum over independent observations of outer products of the score statistics, where each (observed-data) score statistic can be calculated as the expectation of the corresponding complete-data score statistic with respect to the distribution of the missing data conditioned on the observed data. In our context, we do not have independence, but we can exploit the fact that AR($p$) processes have a conditional independence structure which provides similar factoring of the likelihood: when conditioned on the first $p$ observations, the complete data consists of $n - p$ conditionally independent (complex-valued) observations; that is, $(r_t, \phi_t)|(r_{t-1}, \phi_{t-1}), \ldots, (r_{t-p}, \phi_{t-p})$, which we abbreviate as $[r, \phi]|t=p$, are independent for $t = p+1, \ldots, n$. Similarly denoting $[r]|t=p$ as $r|_{t-1}, \ldots, r_{t-p}$, the empirical information matrix is given by $I_e(\hat{\tau}; r) = \sum_{t=p+1}^n s([r]|t=p; \hat{\tau})s'(([r]|t=p; \hat{\tau})\tau=\hat{\tau}$, the (E-step) expectation of the complete-data score statistic $s([r, \phi]|t=p; \tau) = \partial/\partial\tau \log L([r, \phi]|t=p; \tau)$. More details on this calculation are given in Appendix 5.6.3. For verification, we also estimated the Fisher information matrix using Louis’s method (Louis, 1982). Following standard practice, approximate standard errors for $\hat{\tau}_i, i = 1, \ldots, q+p+1$, are given by $\{I_e^{-1}(\hat{\tau}; r)_{\tau_i}\tau_i\}^{1/2}$, the square-root of the diagonal entry of $I_e^{-1}(\hat{\tau}; r)$ corresponding to $\tau_i$.

Wald statistics for activation and order detection follow from the information matrix. We generally pose the test for activation as $H_0 : C\beta = 0$ vs. $H_a : C\beta \neq 0$, which has corresponding Wald statistic $(C\hat{\beta})'[CI_e^{-1}(\hat{\tau}; r)\beta\beta]^{-1}(C\hat{\beta})$, where $I_e^{-1}(\hat{\tau}; r)\beta\beta$ refers to the $q \times q$ block of $I_e^{-1}(\hat{\tau}; r)$ corresponding to $\beta$. This statistic has an asymptotic $\chi^2$ null distribution with degrees of freedom equal to the rank of $C$. We utilize a sequential hypothesis testing procedure (similar to forward model selection in regression modeling) for order detection, in which the detected order $\hat{p} = i’ - 1$, where $i’$ is the smallest $i, i \geq 1$, such that a test of $H_0 : \alpha_i = 0$ vs. $H_a : \alpha_i \neq 0$
does not reject $H_0$ (see Adrian et al., for more detail). A Wald statistic for this test is given by $\hat{\alpha}_i^2 / I_e^{-1}(\hat{\tau}; \hat{r})_{\alpha_0, \alpha_1}$, where $\hat{\alpha}_i$ and $I_e^{-1}(\hat{\tau}; \hat{r})_{\alpha_0, \alpha_1}$ are computed under the AR(i) Ricean model, which follows an asymptotic $\chi^2_1$ null distribution.

Before proceeding, we caution that the use of such Wald statistics relies on two assumptions: first, that the asymptotic result that the MLE is Gaussian-distributed with covariance equal to the inverse Fisher information matrix, and second, that the empirical estimate equals the Fisher information matrix. As a result, in Section 5.3, we examine the null distributions of the Wald statistics before evaluating their detection rates.

5.2.3 Gaussian Autoregressive model

We compare the parameter estimates and test statistics derived under the AR(p) Ricean model to those based on a standard Gaussian AR(p) model. In this Gaussian model, $r = X\beta + \epsilon$, where $X$ is the same as before and $\epsilon$ follows an AR(p) dependence structure parameterized by $\alpha$ and $\sigma^2$. The MLEs of $\beta$, $\alpha$, and $\sigma^2$ can be obtained according to Cochrane and Orcutt (1949). For comparison with the test statistics derived in Section 5.2.2, the Gaussian AR(p) model test statistics are also calculated as Wald statistics which utilize the empirical information matrix.

5.3 Experimental Evaluations

We generated Ricean-distributed magnitude time series of length $n = 256$ by simulating from (5.1) and computing $r_t = \sqrt{y_{tR}^2 + y_{tI}^2}$ for $t = 1, \ldots, 256$. The design matrix $X$ contained $q = 2$ columns, which included an intercept term to model the baseline signal and a $\pm 1$ square wave alternating every 16 time points to model the expected BOLD response. Thus, in $\beta = (\beta_0, \beta_1)$, only $\beta_1$ was activation-related, so the test for activation was posed as $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 \neq 0$ and the corresponding activation test statistic follows an asymptotic $\chi^2_1$ null distribution. We maintained $\sigma^2 = 1.0$ over all simulations for simple interpretation of the signal-to-noise ratio $\text{SNR} = \beta_0 / \sigma$. We varied $\beta_0$ from 0.4 to 10 to examine low SNR values (most fMRI data has SNR above 10) and used $\beta_1 = 0.2$ and 0.0 to represent activated and non-activated voxel time series, respectively. We applied AR(1) dependence with AR coefficient values $\alpha_1 = 0.2, 0.4,$ and 0.6 and fit AR(1) models to all time series. We simulated
100,000 magnitude time series at each collection of parameter values and computed MLEs, standard error estimates, and (activation and AR order detection) test statistics under Ricean and Gaussian models as described in Section 5.2. Implementation of the AR(1) Ricean-model EM algorithm included ten preliminary iterations generating \( m = 2 \) Monte Carlo samples per expectation, followed by iterations utilizing \( m = 10 \) samples until convergence was reached.

First, we examine the biases of the Ricean and Gaussian-model MLEs, which are plotted against \( \beta_0 \) (or alternatively, SNR) in Figure 5.1. Of the two models, the Ricean-model estimates show less bias; the Gaussian-model estimates become increasingly biased as the SNR decreases due to the worsening Gaussian approximation of the Rice distribution discussed in Section 5.1. Further, we note that these (Gaussian-model) biases increase with \( \alpha_1 \) and attribute this to decreasing SNR as well: as \( \alpha_1 \) increases, the SNR decreases due to the increasing variance of AR(1) processes, which is given by \( \gamma_0 = \sigma^2/(1 - \alpha_1^2) \). Though the opposite is true for the Gaussian-model estimate of \( \sigma^2 \), we argue that this effect is “artificial” and due to the bias of \( \hat{\alpha}_1 \) increasing with \( \alpha_1 \).

Next, we look at the standard error estimates and focus on \( SE(\hat{\beta}_1) \) and \( SE(\hat{\alpha}_1) \) because the activation and order detection test statistics are calculated as \( \hat{\beta}_1^2/SE^2(\hat{\beta}_1) \) and \( \hat{\alpha}_1^2/SE^2(\hat{\alpha}_1) \), respectively. In Table 5.1, we display these standard errors for simulated data under the null hypothesis values of \( \beta_1 = 0.0 \) and \( \alpha_1 = 0.0 \) as a way of evaluating the null distributions of these statistics. (A more direct evaluation comes later.) We report the 10th and 90th percentiles of the distributions of the empirical-information-based standard errors, denoted by \( SE_{\text{emp},0.10} \) and \( SE_{\text{emp},0.90} \), for different \( \beta_0 \) values and compare them to the standard deviation of the MLEs themselves, which we denoted by \( SE_{\text{boot}} \) since this is bootstrap-based estimate (Efron, 1981). Unfortunately, it is evident that the empirical-information-based standard error estimates are unstable for low SNRs under the Ricean model. The estimates for both \( SE(\hat{\beta}_1) \) and \( SE(\hat{\alpha}_1) \) show values that are too large: in the former case, the 90th percentiles reveal that some empirical-information-based standard error estimates are (wildly) excessive for \( \beta_0 \leq 1.0 \); in the latter, the middle 80% of the empirical-information-based standard errors fails to capture the (reliable) bootstrap-based standard error for \( \beta_0 < 3 \). The standard errors estimated using Louis’s method yielded similar results. The standard errors under the Gaussian AR(1) model
showed no such issues (so are not displayed).

Due to these unstable standard errors, the Ricean-model Wald test statistics for activation and order detection have difficulty following their null distributions at low SNRs. We demonstrate this through computing the false detection rate of each statistic, which is the proportion of time series generated under $H_0$ in which $H_0$ is rejected. Specifically, we generated activation and order detection statistics under $\beta_1 = 0$ and $\alpha_1 = 0$, respectively, and, specifying a significance level of 0.05, rejected either null hypothesis if the test statistic was greater than the $\chi^2_{1,0.95}$ quantile. Figure 5.2 shows that the false detection rates of both Ricean-model test statistics fail to adhere to the nominal 0.05 level at low SNRs. The Gaussian-model statistics show no such problems. This result follows simply from Table 5.1: the Ricean-model empirical-information-based standard error estimates are too large, so the Wald statistics based on them
(a) Standard errors of $\hat{\beta}_1$

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>$SE_{\text{boot}}$</th>
<th>$SE_{\text{emp},0.10}$</th>
<th>$SE_{\text{emp},0.90}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>0.144</td>
<td>0.067</td>
<td>31.7</td>
</tr>
<tr>
<td>0.6</td>
<td>0.143</td>
<td>0.069</td>
<td>27.8</td>
</tr>
<tr>
<td>0.8</td>
<td>0.145</td>
<td>0.068</td>
<td>18.8</td>
</tr>
<tr>
<td>1.0</td>
<td>0.137</td>
<td>0.067</td>
<td>5.49</td>
</tr>
<tr>
<td>1.2</td>
<td>0.118</td>
<td>0.065</td>
<td>0.121</td>
</tr>
<tr>
<td>1.4</td>
<td>0.094</td>
<td>0.064</td>
<td>0.093</td>
</tr>
<tr>
<td>1.6</td>
<td>0.076</td>
<td>0.063</td>
<td>0.084</td>
</tr>
<tr>
<td>1.8</td>
<td>0.071</td>
<td>0.061</td>
<td>0.079</td>
</tr>
<tr>
<td>2.0</td>
<td>0.068</td>
<td>0.061</td>
<td>0.076</td>
</tr>
<tr>
<td>3.0</td>
<td>0.065</td>
<td>0.058</td>
<td>0.071</td>
</tr>
<tr>
<td>4.0</td>
<td>0.064</td>
<td>0.058</td>
<td>0.070</td>
</tr>
<tr>
<td>5.0</td>
<td>0.064</td>
<td>0.058</td>
<td>0.070</td>
</tr>
</tbody>
</table>

(b) Standard errors of $\hat{\alpha}_1$

<table>
<thead>
<tr>
<th>$\beta_0$</th>
<th>$SE_{\text{boot}}$</th>
<th>$SE_{\text{emp},0.10}$</th>
<th>$SE_{\text{emp},0.90}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>0.067</td>
<td>0.105</td>
<td>0.924</td>
</tr>
<tr>
<td>0.6</td>
<td>0.069</td>
<td>0.102</td>
<td>0.841</td>
</tr>
<tr>
<td>0.8</td>
<td>0.067</td>
<td>0.097</td>
<td>0.631</td>
</tr>
<tr>
<td>1.0</td>
<td>0.066</td>
<td>0.090</td>
<td>0.373</td>
</tr>
<tr>
<td>1.2</td>
<td>0.065</td>
<td>0.084</td>
<td>0.154</td>
</tr>
<tr>
<td>1.4</td>
<td>0.063</td>
<td>0.079</td>
<td>0.111</td>
</tr>
<tr>
<td>1.6</td>
<td>0.063</td>
<td>0.075</td>
<td>0.094</td>
</tr>
<tr>
<td>1.8</td>
<td>0.063</td>
<td>0.071</td>
<td>0.086</td>
</tr>
<tr>
<td>2.0</td>
<td>0.062</td>
<td>0.069</td>
<td>0.081</td>
</tr>
<tr>
<td>3.0</td>
<td>0.063</td>
<td>0.063</td>
<td>0.073</td>
</tr>
<tr>
<td>4.0</td>
<td>0.063</td>
<td>0.061</td>
<td>0.071</td>
</tr>
<tr>
<td>5.0</td>
<td>0.063</td>
<td>0.060</td>
<td>0.070</td>
</tr>
</tbody>
</table>

Table 5.1 Standard error estimates of the AR(1) Ricean estimates (a) $\hat{\beta}_1$ and (b) $\hat{\alpha}_1$ calculated from magnitude time series generated with $\beta_1 = 0.0$, $\alpha_1 = 0.0$, and various values of $\beta_0$. The standard error $SE_{\text{boot}}$ is the standard deviation of the MLEs obtained from simulation (a bootstrap estimate), and $SE_{\text{emp},0.10}$ and $SE_{\text{emp},0.90}$ refer to the 10th and 90th percentiles of the distribution of the empirical-information-based standard errors.

are too small (except for the activation statistic for $\alpha_1 = 0.6$, where the opposite seems to be true). We do not recommend use of the AR($p$) Ricean-model activation and order detection tests below SNRs of 2 and 4, respectively. In the next section, we apply the AR($p$) Ricean and Gaussian models to a real fMRI dataset with SNR above 20 (so do not encounter such issues).

To evaluate the performance of the activation and order detection statistics under each model, we utilize receiver operating characteristic (ROC) curves and the area under the ROC curve, or AUC. An ROC curve plots true detection rate against false detection rate so better-performing test statistics will have ROC curves that are closer to the top and left of the plot, indicating higher true detection rates and lower false detection rates, respectively. By considering true and false detection rates simultaneously, ROC curves can be used to compare tests with different false positive rates. The AUC summarizes the graphical information in an ROC curve by a single number between 0 and 1, with higher AUC indicating better performance. Denoting test statistics computed under $H_0$ and $H_a$ as $\{T_{0i}\}_{i=1}^{n_0}$ and $\{T_{aj}\}_{j=1}^{n_a}$, respectively, the AUC is computed as $\sum_{i=1}^{n_0} \sum_{j=1}^{n_a} I(T_{0i}^{(r)} < T_{aj}^{(r)})/(n_0 n_a)$ (Bamber, 1975), where $I(\cdot)$ is the
Figure 5.2 False detection rates at a 0.05 significance level for the (a) activation and (b) order detection statistics are plotted against $\beta_0$.

indicator function; it is the proportion of null-alternative statistic pairs in which the statistic computed under $H_a$ is greater than the one computed under $H_0$ – i.e. the null and alternative statistics would be correctly assigned. Based on the asymptotic normality of AUCs, DeLong et al. (1988) develops statistical tests for comparing AUCs of different test statistics. Figure 5.3 shows the AUCs of the activation and order detection statistics under both models computed from simulations with alternative hypothesis values of $\beta_1$ and $\alpha_1$ both equal to 0.2. At low SNRs, the Gaussian-model activation statistics have higher AUCs; the opposite is true for the order detection statistics.

Figure 5.3 Areas under the ROC curve (AUCs) for the (a) activation and (b) order detection statistics are plotted against $\beta_0$, where “∗” denotes a statistically significant difference at the 0.05 significance level after a Bonferroni adjustment.
5.4 Detecting Activation in a Finger-Tapping Experiment

Our application for this paper comes from a commonly-performed bilateral sequential finger-tapping experiment, as studied in Rowe and Logan (2004). In this case, the MR images were acquired while the (normal healthy male) volunteer subject was instructed to either lie at rest or to rapidly tap fingers of both hands (hence bilateral) at the same time. The fingers were tapped sequentially in the order of index, middle, ring and little fingers. The experiment consisted of a block design with 16 s of rest followed by eight “epochs” of 16 s tapping alternating with 16 s of rest. MR scans were acquired once every second, resulting in 272 images. Figure 5.4 shows images of the magnitude data at time points 5, 9, 13, 17, 21, 25, 29, and 33 (moving left to right), which represent the first complete 32-s cycle of the finger-tapping experiment, containing 16-s periods of tapping and rest.

Figure 5.4  Images of the magnitude data for time points 5, 9, 13, 17, 21, 25, 29, and 33 (moving left to right), which represent the first complete 32-s cycle of the finger-tapping experiment, containing 16-s periods of tapping and rest.

images of the magnitude data at time points $t = 5, 9, \ldots, 33$, which constitute the first 32-s cycle containing 16-second time periods of tapping and rest, on a single axial slice through the motor cortex consisting of $128 \times 128$ voxels. For simplicity, we restrict attention in this paper to this two-dimensional slice of the dataset. These images appear not to change much in time because, as explained in Section 5.1, the BOLD stimulus response is very small compared to the overall MR signal in all fMRI experiments. A dataset on a well-studied paradigm such as this provides us with as close to a “known” detected activation area as is possible in fMRI: numerous studies have confirmed activation in the sensori-motor finger area cortex in the central sulcus. Thus, this dataset provides us with an ideal case study for both developing and evaluating new methodology.

We modified the design matrix of Section 5.3 by adding a linear trend term to model signal drift and shifting the $\pm 1$ square wave to best correlate with the BOLD response. Our computation of functional activation had three steps: order detection, computation of activation statistics, and thresholding. First, we detected the AR order for each voxel time series using both Ricean and Gaussian test statistics and a maximum order of five, which as shown in Figure
5.5, produced almost identical orders (which, as a side note, conform interestingly with brain anatomy). Based on these detected orders, Wald activation statistics were calculated for both (a) Ricean and (b) Gaussian models. The voxel-wise \( p \)-values, computed from the \( \chi^2 \) null distribution, were converted to \( q \)-values, their analog under false discovery rate (FDR) thresholding (Benjamini and Hochberg, 1995), a criteria for multiple testing which controls the proportion of tests rejecting \( H_0 \) which do so falsely. Images of the \( q \)-values are shown in Figure 5.6, which again are quite similar.

5.5 Discussion

In this paper, we developed the first Ricean model for fMRI magnitude time series that incorporates time dependence. We used an indirect (but natural) approach, applying AR(\( p \)) errors to the Gaussian-distributed real and imaginary components from which the magnitudes are computed. We modeled the latent complex-valued data by augmenting the observed magnitude data by missing phase data according to an EM algorithm framework, which we used to calculate parameter estimates, standard errors, and test statistics for activation and AR order detection. We showed that this AR(\( p \)) Ricean model produces less-biased parameter estimates than its standard AR(\( p \)) Gaussian counterpart and demonstrates comparable activation detec-
Figure 5.6 Images of the $q$-values associated with the voxelwise activation statistics under (a) Ricean and (b) Gaussian models overlayed on a contour plot of brain anatomy.

tion performance on a high-SNR fMRI dataset. However, the instability of the standard error calculation prevents useful AR($p$) Ricean-model activation detection at low SNRs.

5.6 Appendix

5.6.1 The von-Mises distribution

If $\varphi$ follows the von-Mises distribution $VM(\theta_0, \kappa)$, where $\theta_0$ is the mean direction and $\kappa$ is the concentration parameter, the PDF of $\varphi$ is given by $f(\varphi|\theta_0, \kappa) = [2\pi I_0(\kappa)]^{-1} \exp[\kappa \cos(\varphi - \theta_0)]$, for $\varphi \in (-\pi, \pi)$, $\theta_0 \in (-\pi, \pi)$, and $\kappa > 0$ (Mardia and Jupp, 2000), $I_j(\cdot)$ being the $j$th order modified Bessel function of the first kind (Abramowitz and Stegun, 1965). In the E-step, we use the von-Mises expectations $E(\cos \varphi) = A(\kappa) \cos \theta_0$ and $E(\sin \varphi) = A(\kappa) \sin \theta_0$, where $A(\cdot) = I_1(\cdot)/I_0(\cdot)$, as well as the location-family property that $\varphi \sim VM(\theta_0, \kappa) \implies (\varphi - \theta_0) \sim VM(0, \kappa)$. 
5.6.2 Derivation of Monte Carlo approximation

The Monte Carlo approximation (5.3) for $E[\cos(\phi_t - \phi_{t+j})|r_t, r_{t+j}, \tau^{(k)}]$, $j = 1, \ldots, p$, $t = 1, \ldots, n - j$, follows from the expansion

$$E_{\phi_t|r_t, \tau^{(k)}}\{\cos(\phi_t - \theta)E_\theta[\cos(\phi_{t+j} - \theta)] + \sin(\phi_t - \theta)E_\theta[\sin(\phi_{t+j} - \theta)]\},$$

(5.6)

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

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

(5.7)

where $\kappa^*$ and $\delta$ are as in Section 5.2.1. After combining the bracketed portion of (5.7) into a single cosine term, it is evident that $\phi_{t+j}|\phi_t, r_t, r_{t+j}, \tau^{(k)} \sim V M(\Psi(\phi_t), K(\phi_t))$, where $\Psi(\phi_t) = \arctan(\delta \sin(\phi_t - \theta)/[\kappa^* + \delta \cos(\phi_t - \theta)])$ and $K(\phi_t) = \{\kappa^* + \delta^2 + 2\kappa^* \delta \cos(\phi_t - \theta)\}^{1/2}$. Using this distribution to compute the expectations $E_\theta[\cdot]$ in (5.6) and simplifying, we get

$$E_{\phi_t|r_t, \tau^{(s)}}\left\{\frac{A(K(\phi_t))}{K(\phi_t)}[\kappa^* \cos(\phi_t - \theta) + \delta]\right\},$$

(5.8)

which has the same form as (5.3). Because this expectation does not have a closed form, we approximate it by Monte Carlo integration, simulating from $(\phi_t - \theta)|r_t, \tau^{(k)} \sim V M(0, \mu_t^{(s)} r_t/\gamma_0^{(k)})$ and averaging as is in (5.3).

5.6.3 Calculation of empirical information matrix

To illustrate the calculation of $s([r]_{t:t-p}; \tau) = E_{\phi_t|r_t, \tau}[\partial / \partial \tau \log f([r, \phi]_{t:t-p}; \tau)]_{\tau = \hat{r}}$, $t = p + 1, \ldots, n$, we begin with deriving $\log f([r, \phi]_{t:t-p}; \tau)$. By transforming the distributions of $[y_R]_{t:t-p}$ and $[y_I]_{t:t-p}$ (following the notation in Section 5.2.2), which are independent and normal with respective means $\mu_t \cos \theta + \sum_{i=1}^p \alpha_i (y_{R,t-i} - \mu_{t-i} \cos \theta)$ and $\mu_t \sin \theta + \sum_{i=1}^p \alpha_i (y_{I,t-i} - \mu_{t-i} \sin \theta)$ and variances $\sigma^2$, to magnitude-phase variables, it can be shown that $\log f([r, \phi]_{t:t-p}; \tau) = -\log \sigma^2 - h_t/(2\sigma^2)$, where $h_t = \alpha' D_{t} \alpha$, $D_t$ being a $(p+1) \times (p+1)$ matrix with $(i,j)$th-entry
\[ d_t(i, j) = r_{t-i} r_{t-j} \cos(\phi_{t-i} - \phi_{t-j}) - \mu_{t-i} r_{t-j} \cos(\phi_{t-j} - \theta) - \mu_{t-j} r_{t-i} \cos(\phi_{t-i} - \theta) + \mu_{t-i} \mu_{t-j}, \]

\[ 0 \leq i, j \leq p, \text{ and } \hat{\alpha} \text{ as before. After computing derivatives with respect to the parameters, the expectation involves the two types of expectations described in Section 5.2.1.} \]
BIBLIOGRAPHY


Adrian, D. W., Maitra, R., and Rowe, D. B. Improved activation detection via complex-valued ar(p) modelling of fmri voxel time series.


