DEPARTMENT OF STATISTICS
University of Wisconsin
1300 University Ave.
Madison, WI 53706

TECHNICAL REPORT NO. 1133

December 18, 2006

Statistical Methods for Magnetic Resonance Images

John D. Carew[1]

[1]Department of Statistics, Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison (jcarew@stat.wisc.edu).

STATISTICAL METHODS FOR MAGNETIC RESONANCE IMAGES

By

John D. Carew

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(STATISTICS)

at the

UNIVERSITY OF WISCONSIN – MADISON

2006

Abstract

Magnetic resonance imaging (MRI) has revolutionized modern medical practice. MRI is an essentially noninvasive tool for eliciting information about both the function and structure of the human body. Investigators who use MRI are faced with many unique challenges in the design, analysis, and interpretation of their studies. I present novel statistical methodology for three types of MRI data: functional, phase contrast, and diffusion tensor.

Functional MRI (fMRI) measures the blood oxygen level dependent signal in tissue. When applied to the brain, the goal is to measure brain function where changes in blood oxygenation are surrogates for changes in neuronal function. One method for analyzing fMRI time series involves smoothing the data to induce a known autocorrelation structure. The goal of smoothing is to reduce the bias in estimates of linear model parameters. I present a method for spline smoothing of fMRI time series where the amount of smoothing is selected by generalized cross-validation. I show that this procedure substantially reduces bias.

Phase contrast MRI (PC-MRI) is sensitive to flow of fluids. PC-MRI is often applied to measure flow of blood. Information about blood flow and shear is particularly important since these properties are related to formation and progression of atherosclerotic plaque. I developed a nonparametric method for estimating blood flow and shear (proportional to the gradient of the velocity function) in a Matern reproducing kernel Hilbert space.

Diffusion tensor imaging (DTI) is a quantitative magnetic resonance imaging

method that is widely used to study the microstructural properties of white matter in the brain. Tensor-derived quantities such as trace and fractional anisotropy (FA) are important for characterizing the normal, diseased, and developing brain. I derive asymptotic properties of the nonlinear least squares estimator of the diffusion tensor, trace, and FA. I show, with simulations, experimental designs where the asymptotic distributions are very close to the empirical distributions. The asymptotic methods are applied to estimate variances in a healthy volunteer. Variances of trace and FA are found to vary significantly throughout the brain. This renders many popular tests used in group analysis invalid. Unequal variances for tests with tensor-derived quantities is discussed.

Acknowledgements

I am profoundly thankful for the support and guidance of my advisors, Grace Wahba and Beth Meyerand. I thank Grace, Beth, and Rick Nordheim for encouraging me to stay at UW for graduate school. Looking back on the last five years, I can now say that it was a very wise decision. Grace is a brilliant statistician and an outstanding, caring mentor. It has been an honor to have her as my preceptor. I am particularly grateful for her encouragement and always having her door open to listen to my latest ideas and results. Beth provided me with unprecedented and outstanding support both as an undergraduate and as a graduate student. Through Beth's support, I was able to attend multiple meetings each year and always have whatever resources I needed to conduct my research. Attending these meetings was critical to identify and understand important issues in MRI. Beth made sure that I always had the best possible computer hardware to carryout my research. Currently, I am on my sixth computer. Both of my advisors gave me tremendous freedom. I could not ask for better mentors. I am indebted to them.

Several other people contributed to my education and research. I especially thank Andy Alexander, Konstantinos Arfanakis, Peter Basser, Doug Bates, Sean Fain, Vic Haughton, Guan Koay, Yi Lin, Michael Newton, Rick Nordheim, and Orhan Unal. I am particularly grateful to Dave DeMets, Dori Kalish, and Michael Newton for their support of me with an NIH traineeship through the Department of Biostatistics and Medical Informatics. Grace's current and former students have been very helpful throughout my time at UW. I thank Sang-Hoon Cho, Hector Corrada Bravo, Hyonho

Chun, Fan Lu, Chenlei Leng, Weiliang Shi, Ming Yuan, and Xianhong Xie. I also thank Beth's current and former students. Many thanks go to the members of my defense committee: Andy Alexander, Doug Bates, Sunduz Keles, Beth Meyerand, Rick Nordheim, and Grace Wahba.

I thank my family, particularly Raymond C. Carew, and friends for their support. Finally, I express my deepest gratitude to Jack Heil for teaching me how to be a scientist.

List of Figures

List of Tables

Contents

Chapter 1

Spline smoothing of fMRI time series by generalized cross-validation

1.1 Introduction

Linear parametric regression models of fMRI time series have autocorrelated residual errors (Friston et al. 1994). Two general approaches to deal with these autocorrelated residuals are temporal smoothing (Friston et al. 1995; Worsley and Friston 1995) and whitening (Bullmore et al. 1996). Data are whitened by first modeling the intrinsic autocorrelations and then removing them from the data. Provided that the model of the autocorrelations is correct, whitening yields the minimum variance estimate among all unbiased estimates of the linear regression model parameters (Bullmore et al. 1996). Smoothing conditions the autocorrelation structure of an fMRI time series. Appropriate smoothing can minimize the bias in variance estimators for a contrast of a linear model parameter and, thus, the difference between the applied autocorrelation structure and the intrinsic autocorrelations (Friston et al. 2000). Friston et al. (2000) argue that it is preferable to smooth with the goal of minimizing

bias rather than whiten data since it is difficult to obtain an accurate estimate of the intrinsic autocorrelations. Thus, it is prudent to investigate the appropriateness of various smoothing methods for fMRI time series.

This chapter focuses on the use of spline smoothing in the context of fMRI analysis as described by Worsley and Friston (1995). Smoothing splines with automatic optimal smoothing parameter selection via generalized cross-validation (GCV) have a number of desirable properties (Wahba 1990). GCV provides an objective method to determine the correct degree of smoothing for optimally separating a smooth function from white noise. The purpose of this work is to describe and validate spline smoothing of fMRI time series with GCV smoothing parameter selection (GCV-spline). GCV-spline is validated with respect to the minimum bias criteria proposed by Friston et al. (2000). They study the properties of a relevant contrast for a regression model parameter by examining the bias of the variance estimator of this contrast. We compare the variance and the bias of the variance estimator for a contrast of a regression model parameter under GCV-spline smoothing to those under the low pass filter (SPM-HRF) implemented in SPM99 (Wellcome Department of Cognitive Neurology, London) and no smoothing. This chapter describes smoothing splines combined with a parametric regression model. This is related to, but different than, the well known partial spline paradigm for signal detection. The difference will be briefly described.

1.2   Materials and Methods

1.2.1   The General Linear Model

One popular model of fMRI time series is a general linear model (GLM) (Friston et al. 1995; Worsley and Friston 1995),

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{K}\epsilon, \tag{1.1}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$ is an $n \times 1$ matrix of equally-spaced samples of the time series, $\mathbf{X}$ is the model (design) matrix with columns that contain signals of interest and nuisance signals, β is an unknown parameter, $\mathbf{K}$ is an unknown convolution matrix which characterizes the autocorrelations, and $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. The autocorrelation matrix is given by $\mathbf{V} \propto \mathbf{K}\mathbf{K}^T$. Let $\mathbf{S}$ be a linear transformation. The matrix $\mathbf{S}$ is applied to the model (1.1) to give

$$\mathbf{S}\mathbf{y} = \mathbf{S}\mathbf{X}\beta + \mathbf{S}\mathbf{K}\epsilon. \tag{1.2}$$

If $\mathbf{K}$ is known, the approach of whitening is to choose $\mathbf{S} = \mathbf{K}^{-1}$. This transformation will allow a minimum variance, unbiased ordinary least squares (OLS) estimate of β in equation (1.2) given by $\hat{\beta} = (\mathbf{S}\mathbf{X})^+ \mathbf{S}\mathbf{y}$ where $^+$ denotes the pseudoinverse, i.e., $\mathbf{X}^+ = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$. Suppose that $\mathbf{S} \neq \mathbf{K}^{-1}$. Then the assumed autocorrelation $\mathbf{V}_a \propto \mathbf{S}^{-1}(\mathbf{S}^{-1})^T$ will differ from the actual autocorrelation $\mathbf{V}$ and result in biased OLS estimates of the variance of a contrast of β. The amount of bias depends on the accuracy of the approximation of $\mathbf{K}^{-1}$. Friston et al. (2000) found that computing $\mathbf{V}_a$ with a low order autoregressive (AR) model or a $1/f$ model (Zarahn et al. 1997) results in unacceptable bias for fMRI inference. An alternative approach to whitening

is to smooth model (1.1) with $\mathbf{S}$ such that the assumed autocorrelation $\mathbf{V}_a \propto \mathbf{S}\mathbf{S}^T \approx \mathbf{S}\mathbf{V}\mathbf{S}^T$, the true autocorrelation under smoothing. Even when $\mathbf{V}$ is unknown, bias can be minimized (Friston et al. 2000). Since the bias as a function of $\mathbf{S}$ is difficult to directly minimize it is more practical to determine if a method for computing $\mathbf{S}$ gives acceptable levels of bias. One method for computing $\mathbf{S}$ is spline smoothing.

1.2.2 Smoothing Splines

Smoothing splines model the observed time series y_i as

$$y_i = f(t_i) + \epsilon_i, \tag{1.3}$$

where f is a smooth function, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$, and t_i for $i = 1, \ldots, n$ are equally-spaced times when fMR images are acquired. Green and Silverman (1994) give an elementary introduction to smoothing splines. A general reproducing kernel Hilbert space approach to smoothing splines is found in Wahba (1990).

An estimator of $f(t_i)$ is obtained from

$$\hat{f(t_i)} = \arg \min_{f \in C^2[t_1, t_n]} \left(\sum_{i=1}^{n} (y_i - f(t_i))^2 + \lambda \int_{t_1}^{t_n} (f''(x))^2 \, dx \right). \tag{1.4}$$

The unique solution of (1.4) is a natural cubic spline (NCS). A comprehensive introduction to cubic splines can be found in de Boor (1978). The amount of smoothing is controlled by the parameter $\lambda \geq 0$ through weighting the contribution of the second derivative to the penalty function in (1.4). When $\lambda = \infty$, (1.4) is a linear approximation. When $\lambda = 0$, i.e., no smoothing, (1.4) interpolates the y_i with a piecewise cubic polynomial.

The solution of (1.4) is computed with linear algebra. The objective function in

equation (1.4), namely

$$S(f) = \sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \int_{t_1}^{t_n} (f''(x))^2 \, dx, \tag{1.5}$$

can be simplified if the NCS f is represented with its value-second derivative form. The NCS representation from Green and Silverman (1994) is simplified for the case of equally-spaced time points. Following Green and Silverman (1994) with simplifications, let $\mathbf{Q}$ and $\mathbf{R}$ be matrices. $\mathbf{Q}$ has size $n \times (n-2)$ with entries $\mathbf{Q}_{ij}$ for $i = 1, \ldots, n$ and $j = 2, \ldots, n-1$, where

$$\mathbf{Q}_{j-1,j} = \mathbf{Q}_{j+1,j} = (\Delta t)^{-1} \quad \text{and} \quad \mathbf{Q}_{jj} = -2(\Delta t)^{-1},$$

with $\mathbf{Q}_{ij} = 0$ for $|i - j| \geq 2$. Note that the columns are indexed with an unusual convention starting with $j = 2$. The $(n-2) \times (n-2)$ matrix $\mathbf{R}$ is given by

$$
\begin{aligned}
\mathbf{R}_{ii} &= \frac{2\Delta t}{3}, \text{ for } i = 2, \ldots, n-1, \\
\mathbf{R}_{i,i+1} &= \mathbf{R}_{i+1,i} = \frac{\Delta t}{6}, \text{ for } i = 2, \ldots, n-2, \quad \text{and} \\
\mathbf{R}_{i,j} &= 0, \text{ for } |i - j| \geq 2.
\end{aligned}
$$

Let $\mathbf{f}$ be a $1 \times n$ matrix containing the values of f at the t_i. The NCS conditions set the second and third derivatives of f equal to zero at the boundary points t_1 and t_n. The following relationship exists on the smoothness penalty term of equation (1.5):

$$\int_{t_1}^{t_n} f''(x)^2 \, dx = \mathbf{f}^T \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{f}. \tag{1.6}$$

Substitution simplifies equation (1.5) to

$$S(f) = (\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}) + \lambda \mathbf{f}^T \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{f}. \tag{1.7}$$

It is now easy to show that the minimum of (1.7) is

$$\hat{\mathbf{f}} = (\mathbf{I} + \lambda \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T)^{-1}\mathbf{y}. \tag{1.8}$$

The estimator in equation (1.8) is valid, but its form is not computationally efficient. Consider an eigenvector-eigenvalue decomposition of the symmetric, positive semi-definite $n \times n$ matrix $\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T$, namely

$$\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T = \mathbf{\Gamma}\mathbf{D}\mathbf{\Gamma}^T, \tag{1.9}$$

where the orthogonal matrix Γ contains the eigenvectors and

$$\mathbf{D} = \begin{bmatrix} 0 & & & \cdots & & 0 \\ & 0 & & & & \vdots \\ & & l_1 & & & \\ \vdots & & & \ddots & \\ 0 & \cdots & & & l_{n-2} \end{bmatrix} \tag{1.10}$$

contains the $n-2$ positive eigenvalues. Substituting this factorization for $\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T$ in equation (1.8) gives

$$\hat{\mathbf{f}} = (\mathbf{I} + \lambda \mathbf{\Gamma}\mathbf{D}\mathbf{\Gamma}^T)^{-1}\mathbf{y}. \tag{1.11}$$

Since Γ is orthogonal (i.e., $\Gamma^T = \Gamma^{-1}$), equation (1.11) can be simplified to

$$\hat{\mathbf{f}} = \mathbf{\Gamma}(\mathbf{I} + \lambda \mathbf{D})^{-1}\mathbf{\Gamma}^T\mathbf{y}. \tag{1.12}$$

Observing that $\mathbf{I} + \lambda\mathbf{D}$ is diagonal,

$$(\mathbf{I}+\lambda\mathbf{D})^{-1} = \begin{bmatrix} 1 & & & \cdots & & 0 \\ & 1 & & & & \vdots \\ & & \frac{1}{1+\lambda l_1} & & & \\ \vdots & & & \ddots & \\ 0 & \cdots & & & & \frac{1}{1+\lambda l_{n-2}} \end{bmatrix}. \tag{1.13}$$

Thus, one need only perform the eigenvector-eigenvalue decomposition in equation (1.9) once for an entire fMRI data set since $\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T$ depends only on the time between scans (i.e., the sample rate). Then, for each spline fit, only λ need be changed in equation (1.13). This allows for very fast spline fits for different λ. This feature is important for GCV since a spline must be fit for many different degrees of smoothing.

The optimal smoothing parameter λ is determined with GCV (Craven and Wahba 1979). For a given λ, the GCV score is

$$GCV(\lambda) = \frac{1}{n} \cdot \frac{\sum_{i=1}^{n}(y_i - f(\hat{t}_i))^2}{(1 - n^{-1}\mathrm{tr}\mathbf{A}(\lambda))^2}. \tag{1.14}$$

The matrix $\mathbf{A}(\lambda)$ is the hat matrix for a given λ. By definition, $\mathbf{A}(\lambda) = \Gamma(\mathbf{I}+\lambda\mathbf{D})^{-1}\Gamma^T$ since it is the linear transformation that maps the data to their fitted values.

1.2.3 Bias of the Variance Estimator for Spline Smoothing

The smoothing spline from (1.4) can be represented in the form of a smoothing matrix. This allows the spline smoother to be incorporated into the smoothed GLM in equation (1.2). The representation of the smoothing matrix is given directly from

the solution to the penalty function. Moreover, the smoothing matrix is

$$\mathbf{S} = \Gamma(\mathbf{I} + \lambda\mathbf{D})^{-1}\Gamma^T. \tag{1.15}$$

Given $\mathbf{S}$ computed with GCV-spline, the variance of a contrast of $\hat{\beta}$ and the bias of the variance estimator can be computed with equations given in Friston et al. (2000):

$$\text{var}(\mathbf{c}^T\hat{\beta}) = \sigma^2\mathbf{c}^T(\mathbf{SX})^+\mathbf{SVS}^T(\mathbf{SX})^{+T}\mathbf{c} \tag{1.16}$$

and

$$
\begin{aligned}
\text{bias}(\mathbf{S}, \mathbf{V}) &= \frac{\text{var}(\mathbf{c}^T\hat{\beta}) - \text{E}[\widehat{\text{var}(\mathbf{c}^T\hat{\beta})}]}{\text{var}(\mathbf{c}^T\hat{\beta})} \\
&= 1 - \frac{\text{tr}[\mathbf{LSVS}^T]\mathbf{c}^T(\mathbf{SX})^+\mathbf{SV}_a\mathbf{S}^T(\mathbf{SX})^{+T}\mathbf{c}}{\text{tr}[\mathbf{LSV}_a\mathbf{S}^T]\mathbf{c}^T(\mathbf{SX})^+\mathbf{SVS}^T(\mathbf{SX})^{+T}\mathbf{c}},
\end{aligned} \tag{1.17}
$$

where $\mathbf{L} = \mathbf{I} - \mathbf{SX}(\mathbf{SX})^+$ is the residual forming matrix and $\mathbf{c}$ is a contrast vector for hypothesis testing of the components of $\hat{\beta}$. An estimate of $\text{var}(\mathbf{c}^T\hat{\beta})$ is obtained by replacing $\mathbf{V}$ with its assumed value, $\mathbf{V}_a$, and σ^2 with its estimate

$$\hat{\sigma}^2 = \frac{(\mathbf{LSy})^T\mathbf{LSy}}{\text{tr}(\mathbf{LV}_a)}, \tag{1.18}$$

given in Worsley and Friston (1995).

1.2.4   Relationship to Partial Spline Model

In this paper $\mathbf{A}(\lambda)$, the smoother matrix associated with the minimization problem of (1.4) is taken as the matrix $\mathbf{S}$ in the Worsley and Friston (1995) paradigm. A different smoothing spline paradigm, which is also designed for signal detection, is based on the the partial spline model. The partial spline model is:

$$y_i = \sum_{\nu=1}^{p} \phi_\nu(t_i)\beta_\nu + f(t_i) + \epsilon_i \tag{1.19}$$

where the ϕ_ν are specified signal functions whose values $\phi_\nu(t_i)$ provide the entries of $\mathbf{X}$, $\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ and f may be considered as a smooth function, as in (1.5), or, alternatively as a zero mean Gaussian stochastic process with some multiple of a particular covariance that is related to $(\mathbf{QR}^{-1}\mathbf{Q}^T)^+$, (Chiang et al. 1999; Wahba 1983; Wahba 1990, and elsewhere). In the partial spline paradigm, one finds β and f to minimize

$$\tilde{S}(f, \beta) = \sum_{i=1}^{n} \left(y_i - \left[\sum_{\nu=1}^{p} \phi_\nu(t_i)\beta_\nu + f(t_i) \right] \right)^2 + \lambda \int_{t_1}^{t_n} (f''(x))^2 dx \qquad (1.20)$$

and λ is chosen to minimize the corresponding GCV score which now has $\sum_{i=1}^{n}(y_i - [\sum_{\nu=1}^{p} \phi_\nu(t_i)\beta_\nu + f(t_i)])^2$ in the numerator of (14), and $\tilde{\mathbf{A}}(\lambda)$ replacing $\mathbf{A}(\lambda)$ where $\tilde{\mathbf{A}}(\lambda)$ is the matrix satisfying $\mathbf{X}\hat{\beta} + \hat{\mathbf{f}} = \tilde{\mathbf{A}}(\lambda)\mathbf{y}$. In this paradigm the $\hat{\beta}$ which minimizes (1.20) is $\hat{\beta} = (\mathbf{X}^T(\mathbf{I} - \mathbf{S})\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{I} - \mathbf{S})\mathbf{y}$. Estimates for σ^2 and hypothesis tests for β when f is treated as a stochastic process are given in the references. At this time it is not known how the partial spline paradigm might compare, in practice, with the Worsley and Friston (1995) paradigm studied here.

1.2.5   fMRI Experiment

One hundred thirty-two brain volumes were acquired from a healthy volunteer with a 1.5T scanner running a gradient echo EPI pulse sequence for BOLD contrast. The specific parameters were: 22 coronal slices, 7mm thick, 1mm gap, 64^2 pixel matrix with in-plane resolution of 3.75mm^2, TR 2000ms, TE 40ms, and flip angle $85°$. During the fMRI experiment, four symmetric blocks of photic stimulation and darkness were presented to the volunteer. This stimulus was designed to activate the visual cortex. The raw data were spatially smoothed in the frequency domain with a

Hamming filter to increase SNR (Lowe and Sorenson 1997). The first four time points were discarded to prevent known signal instabilities from confounding the analysis. An intensity mask was applied to the images to exclude voxels located outside of the brain. Head motion was corrected with the algorithm in SPM99.

The data were analyzed with the GLM in equation (1.2) with $\mathbf{S}$ determined by GCV-spline. The design matrix contained one column with a boxcar function that was convolved with the SPM99 HRF to model the expected BOLD response to the fMRI experiment (Figure 1a). Polynomial terms up to order three were also included to achieve high pass filtering following the motivation of Worsley et al. (2002). Moreover, the design matrix $\mathbf{X} = [\mathbf{s}\ \mathbf{1}\ \mathbf{t}\ \mathbf{t}^2\ \mathbf{t}^3]$, where $\mathbf{s}$ is the convolved boxcar, $\mathbf{1}$ is a $n \times 1$ column of 1s, and $\mathbf{t} = (t_1, t_2, \cdots, t_n)^T$. An optimal spline smoothing matrix was computed for each time series with equation (1.12). The optimal $\lambda =: \lambda_{\mathrm{opt}}$ was determined by a grid search for the minimum of the GCV score over $\lambda \in [10^{-3}, 10^6]$ on a $\log_{10}$ scale with steps of size 0.1. The data were also analyzed with $\mathbf{S}$ given by the SPM-HRF and $\mathbf{S} = \mathbf{I}$ (no smoothing). Images of voxel t-statistics from a test of the null hypothesis $\mathbf{c}^T\beta = 0$ for the contrast $\mathbf{c}^T = [1\ 0\ 0\ 0\ 0]$ were created for each of the three smoothing strategies with the statistic given by

$$t = \frac{\mathbf{c}^T\hat{\beta}}{\sqrt{\widehat{\mathrm{var}(\mathbf{c}^T\hat{\beta})}}}. \tag{1.21}$$

1.2.6   Bias Computations with Simulated Data

Bias of the variance estimator can be computed with equation (1.12). However, this demands knowledge of the true variance. For an observed fMRI time series, the true variance is unknown. It is possible to generate simulated a time series with a

known autocorrelation structure to allow direct calculation of bias. To generate a set of reasonable simulated fMRI time series, autocorrelation structures can be estimated from real data. The autocorrelation estimates can be used to induce correlations in pseudo-random numbers independently-sampled from a Gaussian density with a known variance. Finally, a signal can be added to the correlated Gaussian samples. Figure 1b shows a simulated time series.

Autocorrelation structures were estimated from the residuals of the fit of the time series from the fMRI experiment with the smoothing matrix $\mathbf{S} = \mathbf{I}$ (i.e., no smoothing) with an AR(8) model. The residuals for a single time series $\mathbf{y}$ are

$$\mathbf{r} = \mathbf{L}\mathbf{y}. \tag{1.22}$$

The AR(8) model of the i-th residual is

$$r_i = b_1 r_{i-1} + \cdots + b_8 r_{i-8} + \zeta_i, \tag{1.23}$$

where $b_1, \cdots, b_8$ are the AR coefficients and $\zeta_i \sim \mathcal{N}(0, \sigma_r^2)$. The more convenient matrix representation of equation (1.23) is, following Friston et al. (2000),

$$\mathbf{r} = (\mathbf{I} - \mathbf{B})^{-1}\zeta, \tag{1.24}$$

where $\mathbf{B}$ is a matrix of AR model coefficients and $\zeta \sim \mathcal{N}(\mathbf{0}, \sigma_r^2 \mathbf{I})$. The AR coefficients

were estimated with the least squares procedure (Chatfield 1996) and organized into

$$\hat{\mathbf{B}} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ \hat{b}_1 & 0 & 0 & 0 & \cdots & 0 \\ \hat{b}_2 & \hat{b}_1 & 0 & 0 & \cdots & 0 \\ \hat{b}_3 & \hat{b}_2 & \hat{b}_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \hat{b}_8 & \hat{b}_7 & \hat{b}_6 & \hat{b}_5 & \cdots & 0 \\ 0 & \hat{b}_8 & \hat{b}_7 & \hat{b}_6 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{1.25}$$

where $\hat{b}_1, \cdots, \hat{b}_8$ are the estimated AR coefficients. Then, the estimated convolution matrix is

$$\tilde{\mathbf{K}} = (\mathbf{I} - \hat{\mathbf{B}})^{-1}. \tag{1.26}$$

A simulated time series $\tilde{\mathbf{y}}$ is constructed by

$$\tilde{\mathbf{y}} = 0.15 \cdot \mathbf{s} + \tilde{\mathbf{K}}\mathbf{e}, \tag{1.27}$$

where $\mathbf{s}$ is the signal in Figure 1(a) and $\mathbf{e}$ is a sample from $\mathcal{N}(\mathbf{0}, \mathbf{I})$. The coefficient of the signal, namely 0.15, and the unit variance error term $\mathbf{e}$ were selected to match the simulated data parameters used by Lang et al. (1999).

For each voxel in the masked data set, a separate $\tilde{\mathbf{K}}$ was estimated and a separate $\mathbf{e}$ was generated with the `randn()` function in MATLAB (Mathworks, Nattick, MA) to produce a simulated time series with equation (1.27). The simulated time series were assigned to the spatial location where the $\tilde{\mathbf{K}}$ was estimated. This preserves the
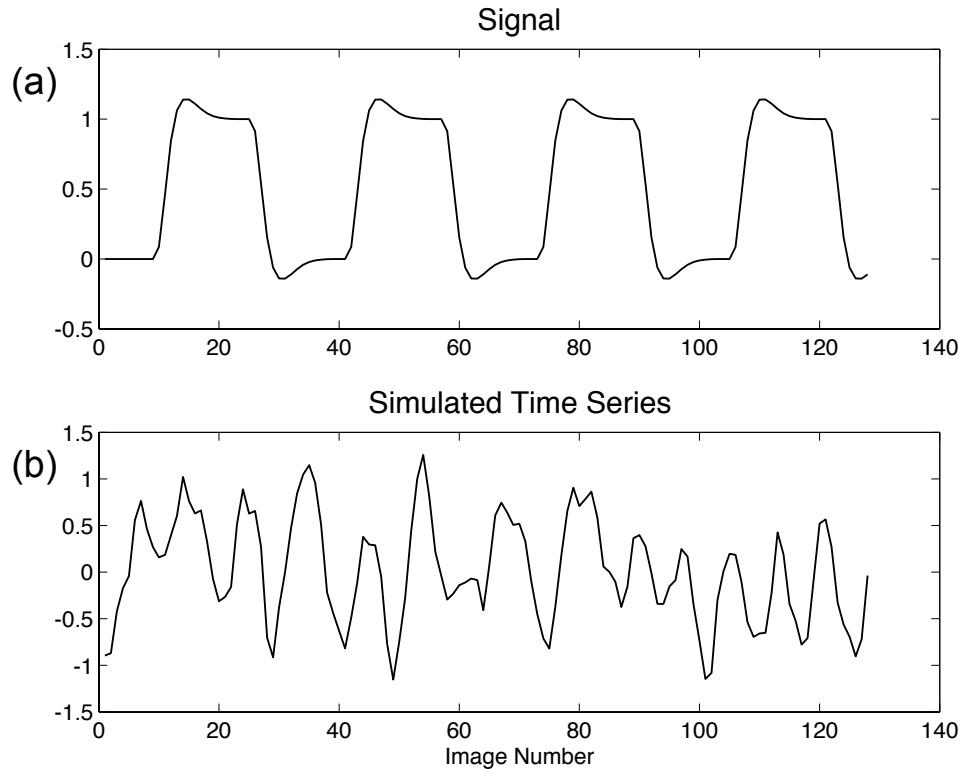
Figure 1: The hypothesized BOLD response to the photic stimulation in the fMRI experiment (a). A simulated time series used in the bias computations (b).

spatial variability of autocorrelation structures in the simulated data. A subset of the simulated data was selected based on 100 voxels with the largest t-statistics of a test of the null hypothesis $\mathbf{c}^T \beta = 0$ for the contrast $\mathbf{c}^T = [1\ 0\ 0\ 0\ 0]$ in the real data under $\mathbf{S} = \mathbf{I}$. This subset contained simulated time series with autocorrelation structures from regions that were presumably related to the stimulus in the fMRI experiment.

The variance and bias of the variance estimator were calculated for each time series in the entire simulated data set with three different smoothing matrices: $\mathbf{S} = \mathbf{I}$, $\mathbf{S}$ equal to the SPM-HRF filter matrix, and the spline smoothing matrix $\mathbf{S} = (\mathbf{I} + \lambda \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T)^{-1}$. The SPM-HRF smoothing matrix is generated with the `spm_make_filter` function with the option for no high pass filtering and "hrf" for the low pass filter. This function is part of the SPM99 package. For the spline smoothing, the optimal λ was determined with the same parameters as used with the real data, i.e., for each

time series a GCV search over $\lambda \in [10^{-3}, 10^6]$ on a $\log_{10}$ scale with steps of size 0.1. The variance of $\mathbf{c}^T\beta$ with $\mathbf{c}^T = [1\ 0\ 0\ 0\ 0]$ and the bias of the variance estimator were computed for each of the time series under each smoothing method with equations (1.16) and (1.17), respectively. The known variance-covariance matrix $\mathbf{V} = \tilde{\mathbf{K}}\tilde{\mathbf{K}}^T$ and the assumed variance-covariance matrix $\mathbf{V}_a = \mathbf{SS}^T$ with $\mathbf{S}$ given by the particular smoothing matrix under investigation.

1.2.7 Computer Software and Hardware

The algorithms were written and carried out in the MATLAB technical computing package. Some of the plots were created with the R statistical environment (www.r-project.org). The t-statistic images and bias images were co-registered with anatomic images with AFNI (Cox 1996). A linux workstation with dual 1.2 GHz AMD Athlon MP processors was used for the computations in this study. The algorithms were not designed to utilize both processors simultaneously.

1.3 Results

1.3.1 FMRI Experiment

The results of the fMRI experiment and motion correction were technically adequate for further analysis. After the intensity mask was applied, approximately 12,000 voxels were included in the analysis. The computational time for GCV-spline analysis of the entire brain volume was four hours and fifteen minutes. Generalized cross-validation performed well on the fMRI time series; sensible amounts of smoothing
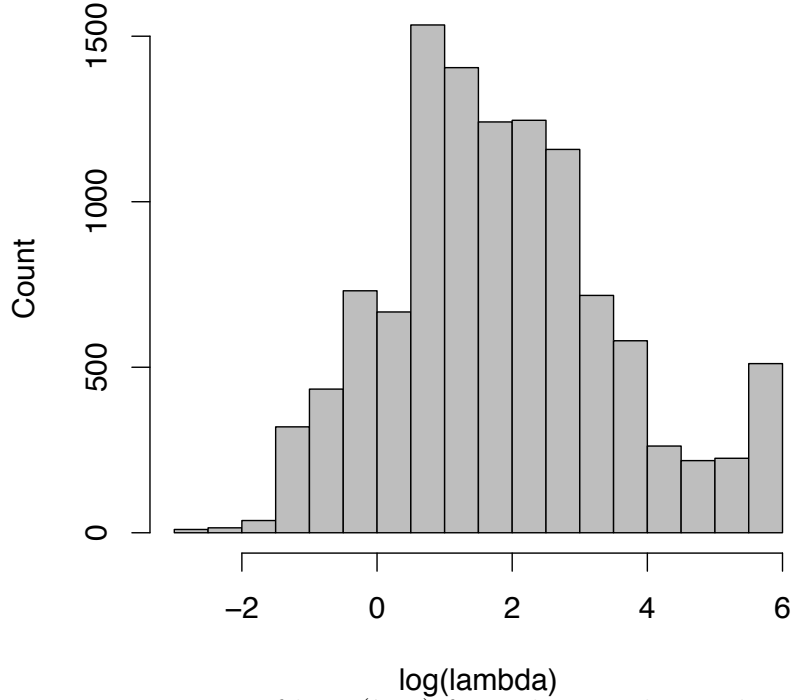
Figure 2: Histogram of $\log_{10}(\lambda_{\mathrm{opt}})$ from the visual stimulus experiment.

were determined. High frequency fluctuations of the observed BOLD signal received more smoothing than lower frequency fluctuations. The GCV score was numerically well-behaved over the region of the parameter search. The amount of optimal smoothing with the GCV-spline method is illustrated in Figure 2. The distribution of the optimal λ is skewed toward the smaller values with only a few time series that required large amounts of smoothing ($\bar{\lambda} = 4.790 \times 10^4$ and median(λ) = 50.01). An example of a time series that required little smoothing and one that required more smoothing are included in Figure 3 (a) and (b). The corresponding plots of the GCV score demonstrate the numerical stability of GCV for these two time series (Figure 3 (c) and (d)). These plots are similar in the relative amount of curvature to the GCV scores from the other time series. The GCV-spline method performs, on average, more smoothing than the SPM-HRF smoothing kernel. Figure 3 (e) and (f) shows

the equivalent smoothing spline kernels and the SPM-HRF kernel for the time series in Figure 3 (a) and (b), respectively. For $\log_{10}(\lambda) \approx 0.78$, both methods provide the same amount of smoothing (Figure 3(e)).

Inference from the model fits with GCV-spline, SPM-HRF, and no smoothing give qualitatively similar results. Images of t-statistics from a test of the null hypothesis $\mathbf{c}^T\beta = 0$ for the contrast $\mathbf{c}^T = [1\ 0\ 0\ 0\ 0]$ are given for a coronal slice through the primary visual cortex (V1) in Figure 4. The clusters of activation are centered at approximately the same location for each of the models. The spatial extent of the clusters and the level of significance was different with each method. When no smoothing is performed (i.e., no accounting for residual autocorrelations), the t-statistics are highest and the spatial extent of the clusters is greatest. Conversely, with GCV-spline smoothing, the level of significance was lowest and the activated clusters are the smallest. Smoothing with the SPM-HRF gave t-maps that were somewhere in between the other two methods.

1.3.2 Simulated Data

The amount of bias in the variance estimates for the simulated data differed between the three smoothing strategies. The computational time for GCV-spline fitting of the simulated time series was the same as with the real fMRI data set. The distribution of the optimal λ was skewed towards the lower values (Figure 5). On average, the simulated time series required more smoothing ($\bar{\lambda} = 8.154 \times 10^4$ and median(λ) = 398.1) than the real fMRI data. The mean and median bias with the simulated data for the three methods are shown in Table 1. The mean and median

Table 1: Bias for Three Smoothing Strategies (Whole Brain)

	GCV-Spline	SPM-HRF	No Smoothing
Mean Bias	0.0200	0.0701	0.4019
Median Bias	0.0037	0.1608	0.5399

Table 2: Bias for Three Smoothing Strategies (100 Significant Voxels)

	GCV-Spline	SPM-HRF	No Smoothing
Mean Bias	-1.800×10^{-5}	0.1801	0.5964
Median Bias	-0.0332	0.2312	0.6627

bias were positive for all methods. GCV-spline smoothing had bias that was closest to zero, whereas the bias was greatest for no smoothing. Histograms of the bias show that, on average, GCV-spline is unbiased and that SPM-HRF and no smoothing are biased (Figure 6). Images of bias are included in Figure 7. These images show how the bias of each method varies over different regions of the brain. There are slightly more voxels with positive bias in the grey matter regions and more voxels with negative bias in the ventricles and white matter regions with GCV-spline smoothing (Figure 7A). With SPM-HRF smoothing (Figure 7B) and no smoothing (Figure 7C), bias is nearly systematically positive in the grey matter regions. The 100 voxel subset of the simulated data show similar trends in the bias that are summarized in Table 2. Boxplots of the bias and variance for the three methods show that the reduction in bias comes at the cost of only a small increase in variance (Figure 8).

1.4   Discussion

The novel contribution of this paper is to show that spline smoothing with generalized cross-validation provides a method to determine the optimal amount of smoothing for an fMRI time series. This method not only conditions the autocorrelation structure on the data, but it does so in a way to optimally separate the underlying signal from noise. By selecting the λ that minimizes the GCV score, the smoothing spline estimator for the signal will minimize the predictive mean-square error (Craven and Wahba 1979). The empirical results from the spline smoothing of the fMRI data show that GCV is a sufficient method to automatically choose the appropriate amount of smoothing. On average, the spline method determined that a greater amount of smoothing was necessary than the amount provided by the SPM-HRF kernel. This suggests that if one chooses to use a single smoother for all time series the underlying signal might, on average, be better estimated if a smoothing kernel with a greater bandwidth than the SPM-HRF kernel is used. However, if computational time is not a major issue it is preferable to find the optimal degree of smoothing for each time series. It must be emphasized that the premise for using a fixed smoothing kernel such as the SPM-HRF was a computational constraint (Friston et al. 2000).

The results from the comparison of GCV-spline smoothing with the SPM-HRF and no smoothing of the simulated data show that optimal spline smoothing of each time series is, on average, significantly less biased than smoothing all time series with an identical SPM-HRF kernel or ignoring residual autocorrelations. The mean bias reported in Table 1 for the SPM-HRF is deceptive in the context of fMRI studies since the majority of voxels with negative bias are located in regions other than grey

matter. These negative bias voxels shift the mean bias closer to zero. A more complete study where grey matter voxels are segmented from the rest of the brain would likely show that the mean bias for SPM-HRF smoothing is higher. This is also true for GCV-spline smoothing. However, the effect of studying only gray matter voxels on the bias is expected to be lower since the distribution of bias is more symmetric about zero for GCV-spline smoothing (Figure 6).

The bias improvement is attributed to the ability of the spline method to select appropriate smoothing for each time series. The gain in bias reduction comes at the cost of a slight increase in variance (e.g., Figure 8). This increase in variance simply reflects a bias-variance tradeoff in the variance estimator that is controlled by the amount of smoothing. A large reduction in statistical efficiency is not expected from the greater amount of smoothing with the GCV-spline than the SPM-HRF since $\mathrm{var}(\mathbf{c}^T\hat{\beta})$ was not orders of magnitude greater with GCV-spline smoothing. The t-maps in Figure 4 also reflect how greater smoothing causes greater variances which lead to lower values of the t-statistic.

One limitation of the GCV-spline method is the additional computational expense. Fitting a GLM with SPM-HRF smoothing is on the order of minutes compared to a few hours for the GCV-spline method. The algorithms used in this study for the GCV-spline method were not numerically optimal. A simple grid search for the minimum of the GCV score is inefficient. A fixed step size of $\log_{10}(\lambda) = 0.1$ is a particularly poor choice since the GCV score is very well-behaved for fMRI time series (Figure 3c,d). Both improved algorithms and better optimization could dramatically reduce the amount of computer time needed for spline smoothing and generalized cross-validation. Interpreted languages such as MATLAB are often slower than compiled

code. FORTRAN routines for minimizing the GCV score are available in RKPACK (Gu 1989) which is freely available on www.netlib.org/gcv. To find optimal smoothers with RKPACK for 12,000 time series takes about six minutes on computer harware comparable to that used in this study. Thus, the use of RKPACK or other compiled code is encouraged for researchers adopting the methods of this paper.

Generalization of the bias analysis of simulated data to real data depends on how realistic the assumtions made when the simulated time series were constructed. The first assumption is that the signal $\mathbf{s}$ has a given amplitude and is generated by the convolution of the hemodynamic response function with a boxcar function. The second assumption is that the noise is additive and specified by $\mathbf{K}\epsilon$. The third assumption is that the AR(8) model provides an accurate estimate to construct an estimate of $\mathbf{K}$, namely, $\tilde{\mathbf{K}}$. The first assumption is not critical in the context of this study since $\mathbf{s}$ is modeled exactly in the design matrix. The structure and additivity of the noise model is generally accepted with strong evidence to suggest its validity from the null data studies of Woolrich et al. (2001). Finally, the assumption that the AR(8) model is sufficient is the critical assumption to establish the validity of the simulated time series. An examination of the necessary AR order from six null data sets by (Woolrich et al. 2001) concluded that AR(6) was sufficient for their data. Thus, the AR(8) is conservative with enough freedom to accommodate even more complex AR processes than expected.

This study did not consider the performance or applicability of spline smoothing to a random event design fMRI experiment. A Tukey taper applied to a spectral density estimate and nonlinear spatial smoothing can be used to estimate the autocorrelation for the purposes of prewhitening that yields acceptable levels of bias (Woolrich et al.

2001). This is likely to be a more efficient method for handling autocorrelations in random event data than spline smoothing. However, there is likely to be no gain in efficiency with this method over spline smoothing since smoothing and prewhitening have similar efficiencies for block design experiments (Friston et al. 2000; Woolrich et al. 2001; Worsley and Friston 1995).

1.5   Conclusion

Spline smoothing with the optimal degree of smoothing selected with generalized cross-validation is a method for smoothing fMRI time series that may be used to separate a smooth signal from white noise. In this study, we use the implied spline smoother to select an appropriate smoothing matrix for a GLM of an fMRI time series. For fMRI experiments with block designs, there is a significant reduction in bias over smoothing with the SPM-HRF kernel or simply ignoring residual autocorrelations. Since the appropriate degree of smoothing is selected for each time series, spline smoothing (with compiled code such as RKPACK) is slightly computationally more expensive than applying a single smoothing kernel to all time series. Nonetheless, the bias advantage of the GCV-spline smoothing suggests that it is an appropriate smoothing method for regression analysis of fMRI time series.
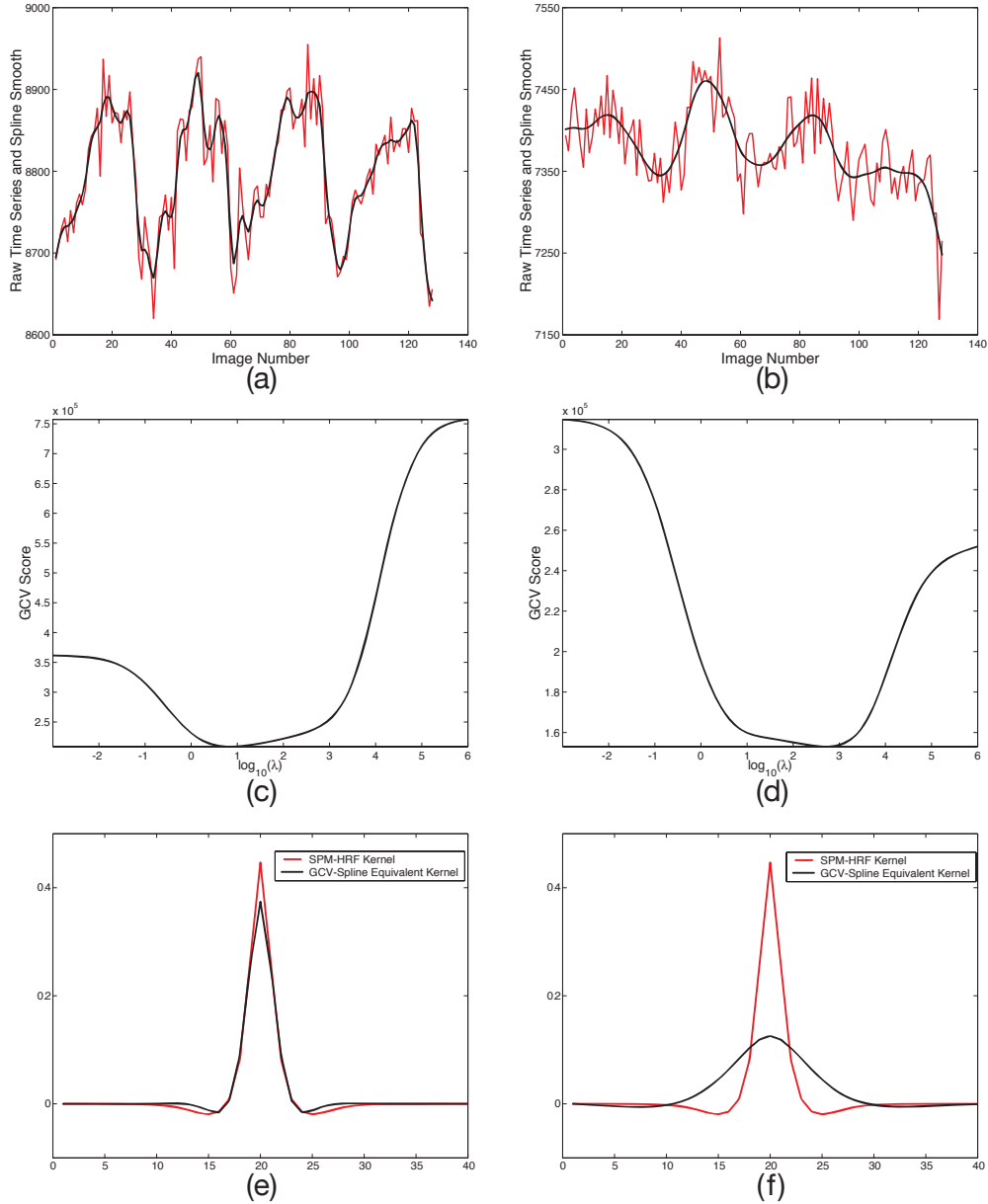
Figure 3: Spline smoothing of fMRI time series with the degree of smoothing selected by generalized cross-validation. The time series in (a) required approximately the same amount of smoothing as is provided by the SPM-HRF smoothing kernel. The plot of corresponding GCV score (c) is well-behaved with a minimum at $\lambda = 6.31$. The SPM-HRF kernel is very similar to the equivalent smoothing spline kernel (e). The other time series (b) was smoothed more with the spline method than by the SPM-HRF kernel. The underlying signal is estimated well at the minimum of the GCV score (d) where $\lambda = 501$. The equivalent smoothing spline kernel (f) has a noticeably greater bandwidth than the SPM-HRF smoothing kernel.
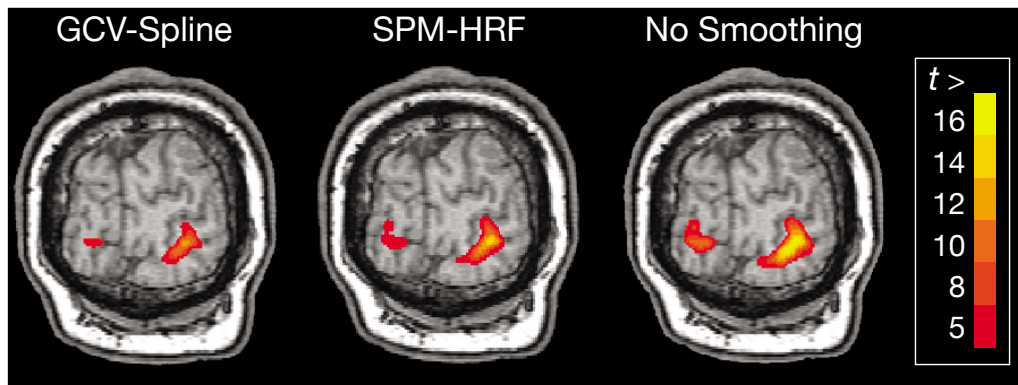
Figure 4: Images of the t-statistic under the null hypothesis for GCV-Spline smoothing, SPM-HRF smoothing, and no smoothing.
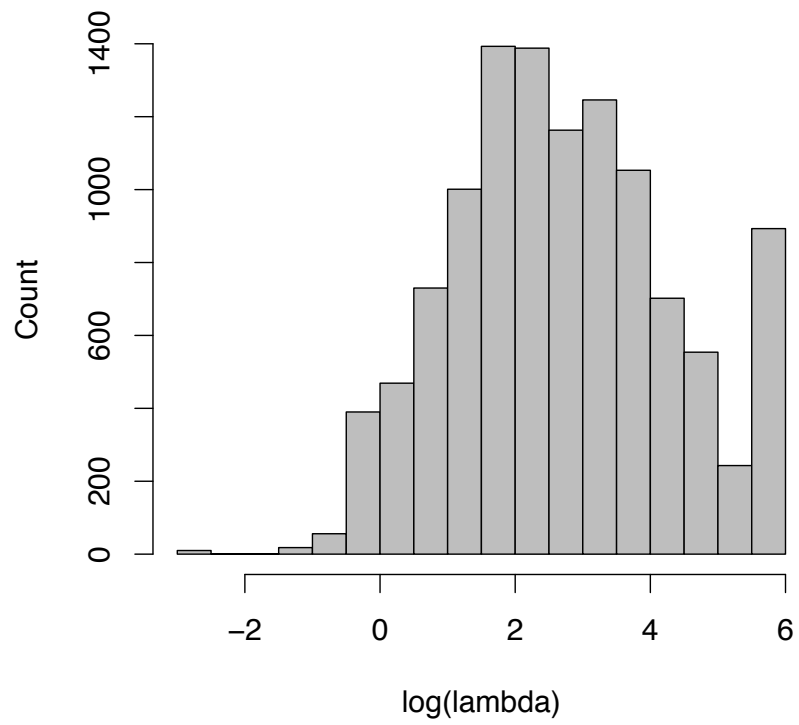
Figure 5: Histogram of $\log_{10}(\lambda_{\mathrm{opt}})$ from the simulated data set.
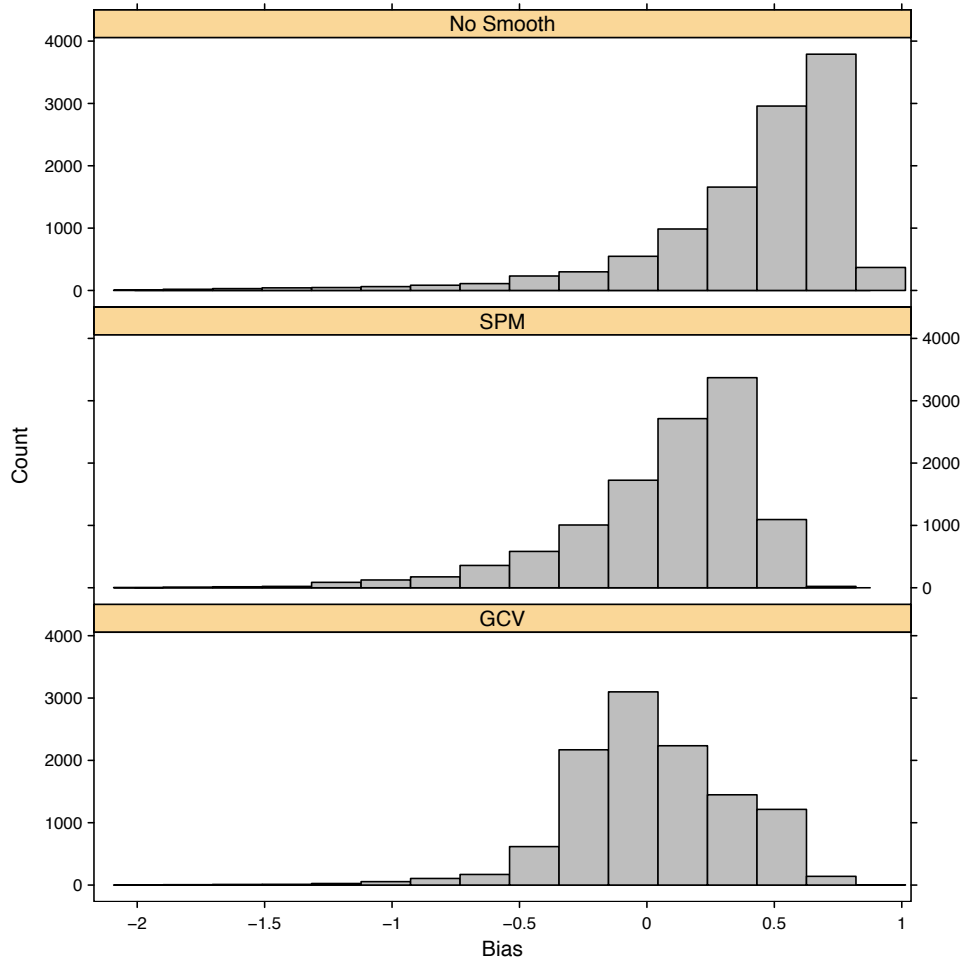
Figure 6: Histograms of bias(vâr($\mathbf{c}^T\hat{\beta}$)) from the simulated data with no smoothing (top), SPM-HRF smoothing (middle), and GCV-Spline smoothing (bottom). Smoothing with the GCV-Spline method produces vâr($\mathbf{c}^T\hat{\beta}$) estimates that are, on average, unbiased.
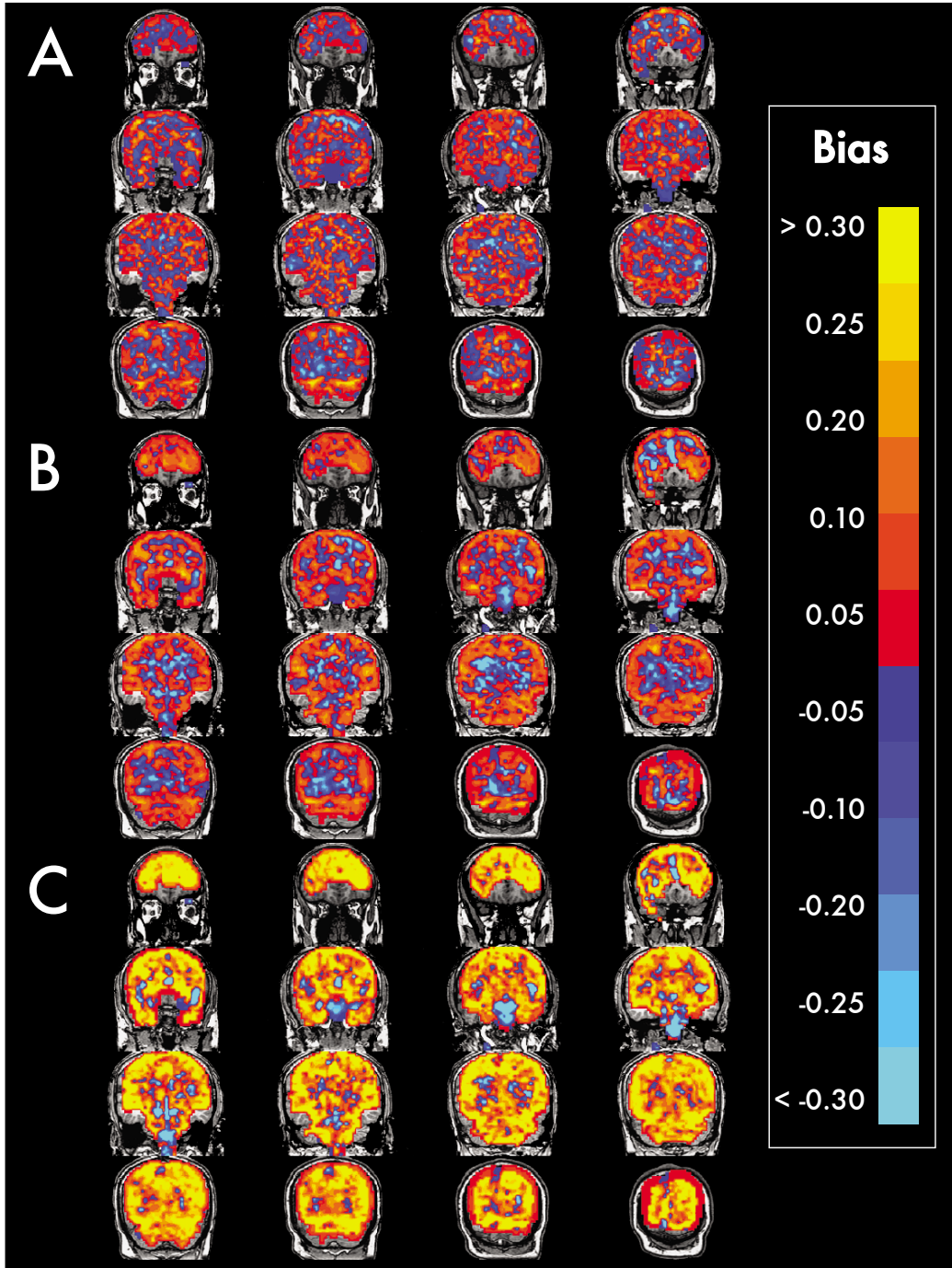
Figure 7: Images of bias of the bias of $\hat{\text{var}}(\mathbf{c}^T\hat{\beta})$ for GCV-Spline (A), SPM-HRF (B), and no smoothing (C). The bias in each voxel was computed for simulated time series with autocorrelation structures estimated from the corresponding voxel. Voxels with positive bias underestimate the true variance of its regression parameter estimate. The inference in these regions is anticonservative. The converse is true for voxels with negative bias.
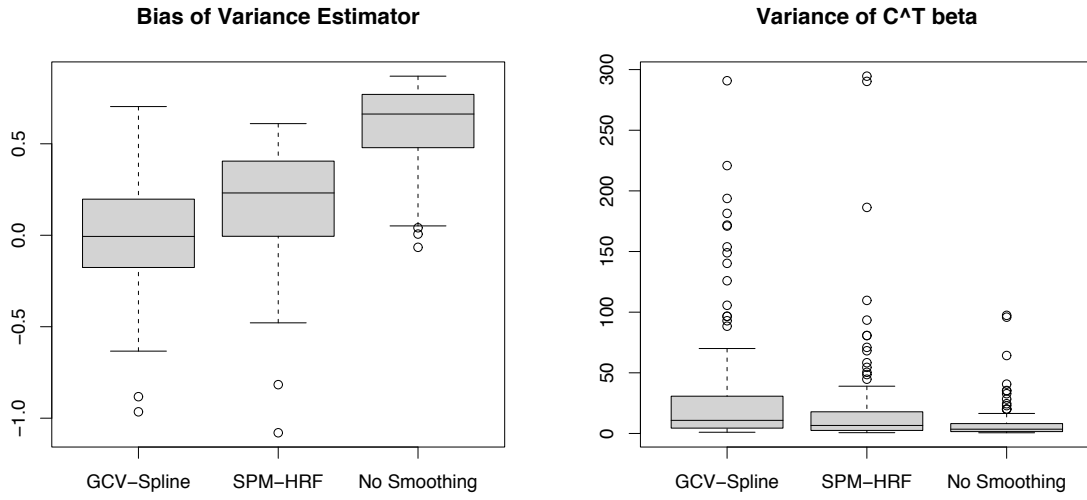
Figure 8: Bias of the variance estimator and estimated variance of $\mathbf{c}^T\hat{\beta}$ with GCV-Spline, SPM-HRF, and no smoothing from the simulated data set with autocorrelation structures estimated from the 100 most significantly task-related time series.

Chapter 2

Nonparametric estimation of fluid velocity and shear in the Matern reproducing kernel Hilbert space

2.1 Introduction

The level of cardiovascular plaque formations is related to the risk of stroke and heart disease. Atherosclerosis preferentially forms in areas of reduced wall shear stress (WSS) (Friedman et al. 1987; Ku et al. 1985; Zarins et al. 1983). Thus, arterial WSS measurements may provide a diagnostic indicator for atherosclerosis. Most of the evidence supporting this connection is from studies of arteries that were dissected from cadavers. In vivo methods are beneficial for studying the formation and progression of the disease. One method for noninvasively estimating WSS is through post-processing of phase contrast magnetic resonance images (PC-MRI). PC-MRI is capable of measuring blood velocity along the direction of arterial blood flow. The WSS is proportional to the component normal to the wall of the gradient of the blood velocity at the arterial wall. The post-processing step involves fitting a surface to the PC-MRI velocity images over the cross-section of the artery and computing

the derivatives at the boundary. It is difficult to find a parametric surface that fits the velocity data well. In fact, only for the simplest vessel geometries are there closed-form analytical descriptions of fluid flow (Barna 1969). The general flow of fluid is governed by solutions to the Navier-Stokes partial differential equations. A "sectored" method which fits several parabolic surfaces to overlapping parts of the image has been used to estimate WSS (Oyre et al. 1998). This method is limited to straight vessels with circular cross-sections. The Oyre method cannot handle complex flow near vessel bifurcations and cannot model bidirectional flow, which can occur at vessel bifurcations. A more direct approach is to fit a nonparametric function to the velocity data.

In this chapter, we describe flexible, nonparametric estimators for fluid velocity and shear. The estimation is based on penalized least squares over a reproducing kernel Hilbert space (RKHS). A Bayesian confidence interval is derived for the shear estimator. Our methods are applied to three PC-MRI data sets that were acquired from vascular "phantoms." The phantoms are idealized models of human arteries.

2.2 Theory and Methods

2.2.1 Wall Shear Stress

When a fluid flows along one axis, WSS[1] at a point $\mathbf{t} := (x, y)' \in \mathbb{R}^2$ along the wall with unit normal vector $\mathbf{n}(\mathbf{t}) := (n_x, n_y)'$ is

$$
\begin{aligned}
W(\mathbf{t}) &:= \mu \, \nabla f(\mathbf{t}) \cdot \mathbf{n}(\mathbf{t}) && (2.1) \\
&= \mu \left(\frac{\partial f(\mathbf{t})}{\partial x} \, n_x + \frac{\partial f(\mathbf{t})}{\partial y} \, n_y \right). && (2.2)
\end{aligned}
$$

The function $f(\mathbf{t})$ is the spatially dependent fluid velocity where $\{\mathbf{t}\}$ lie in a plane that is perpendicular to the axis of fluid flow. The constant μ is the dynamic viscosity of the fluid. We can estimate $W(\mathbf{t})$ by first estimating $f(\mathbf{t})$ and then differentiate $\hat{f}(\mathbf{t})$. Substituting derivatives of $\hat{f}(\mathbf{t})$ for the derivatives of the velocity function in equation (2.2) gives an estimator for $W(\mathbf{t})$ that is

$$
\hat{W}(\mathbf{t}) = \mu \left(\frac{\partial \hat{f}(\mathbf{t})}{\partial x} \, n_x + \frac{\partial \hat{f}(\mathbf{t})}{\partial y} \, n_y \right). \tag{2.3}
$$

2.2.2 Estimating Fluid Velocity in a RKHS

The objective is to estimate a function from a set of fluid velocity observations that are contaminated by noise. Consider data $\{(\mathbf{t}_i, z_i)\}$, $i = 1, \ldots, n$, where $\mathbf{t}_i = (x_i, y_i)' \in \mathbb{R}^2$ is the point where velocity z_i is observed[2]. In our particular application, the $\mathbf{t}_i$ are pixel locations in a PC-MR image, and the z_i are the corresponding pixel intensities,

---

[1]Shear can be calculated at any point within the vessel and not only at the wall. It is because of the connection to plaque formation that we are particularly interested in shear at the wall.

[2]The methods that follow extend to data from higher dimensional spaces, e.g., 3D images, contiguous 2D images, and sequences of 2D or 3D images acquired over time.

which represent fluid velocity. The data are modeled as

$$z_i = f(\mathbf{t}_i) + \epsilon_i, \tag{2.4}$$

where f is assumed to be a smooth function and the errors $\epsilon_i \sim_{iid} \mathcal{N}(0, \sigma^2)$.

Model (2.4) is fit from data by selecting an appropriate f in a RKHS. Let $\mathcal{H}_R$ be a RKHS with reproducing kernel R. The reproducing kernel (RK) is a symmetric, positive definite function $R(\mathbf{s}, \mathbf{t})$ such that $R : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^+$. Any such R defines a unique RKHS (Aronszajn 1950). An estimator for $f \in \mathcal{H}_R$ is given from the solution to the variational problem

$$\hat{f}(\mathbf{t}) := \arg\min_{f \in \mathcal{H}_R} \left(\sum_{i=1}^{n} (z_i - f(\mathbf{t}_i))^2 + \lambda ||f||_{\mathcal{H}_R}^2 \right). \tag{2.5}$$

The regularization parameter $\lambda \geq 0$ in equation (2.5) controls the smoothness of $\hat{f}(\mathbf{t})$ through weighting the penalty $||f||_{\mathcal{H}_R}^2$, where $|| \cdot ||_{\mathcal{H}_R}$ is the norm in $\mathcal{H}_R$. The solution to (2.5) is a special case of a solution to a more general variational problem that was discovered by Kimeldorf and Wahba (1971). The solution is

$$\hat{f}(\mathbf{t}) = \sum_{i=1}^{n} \mathbf{c}_i R(\mathbf{t}, \mathbf{t}_i), \tag{2.6}$$

where,

$$\mathbf{c} := (\mathbf{\Sigma} + \lambda \mathbf{I})^{-1} (z_1, \cdots, z_n)' \tag{2.7}$$

$$\mathbf{\Sigma} := [R(\mathbf{t}_i, \mathbf{t}_j)]_{ij}, \quad i, j = 1, \cdots, n. \tag{2.8}$$

The important thing to note about equation (2.6) is that the solution to (2.5) at any point $\mathbf{t}$ in the domain is simply a linear combination of the RK evaluated at $\mathbf{t}$ and the $\mathbf{t}_i$, with coefficients depending on the $\mathbf{t}_i$, z_i, and λ. As the notation suggests,

the symmetric, positive definite matrix $\mathbf{\Sigma}$ can be thought of as the covariance of the $f(\mathbf{t}_i)$.

An explicit solution to (2.5) has been given. However, two related issues remain: what are appropriate choices for R and λ? These choices are best made with consideration given to the characteristics of the specific problem and to the data. First, consider selection of the RK R. Since R specifies the RKHS, an R that gives rise to a sufficiently rich RKHS such that it contains the true, unknown function (that is to be estimated) is required. Thus, consider selecting a RK from a family of functions such as the Matérn family of radial basis functions (RBF) (Matérn 1986; Stein 1999). Let $\tau := ||\mathbf{s} - \mathbf{t}||$ be the Euclidean distance between points $\mathbf{s}, \mathbf{t} \in \mathbb{R}^d$. Then, the ν-th order[3] Matérn function is:

$$R_\nu(\tau) := e^{-\tau}\pi_\nu(\tau), \quad \nu = 0, 1, 2, \ldots, \tag{2.9}$$

where $\pi_\nu(\cdot)$ is a ν-th order polynomial of a particular form. See Table 3 for $R_1, \ldots, R_7$ and a recursion formula for higher orders. It should be noted that the Matérn RBFs are defined for general, positive orders. However, the general form of the Matérn RBF involves a modified Bessel function for which there is no closed form for orders not defined here. Moreover, $R_\nu(\tau)$ can be thought of as an isotropic covariance, which gives rotational invariance of the solution. This is important for imaging applications.

2.2.3 Estimating Shear Stress

The estimator for WSS is obtained by differentiating the velocity estimator. Let $\mathbf{z} := (z_1, \ldots, z_n)'$ and $\mathbf{A}$ be the influence matrix, which maps to data to the fitted

---

[3]What is referred to here as a ν-th order Matérn RBF is sometimes called order $m = \nu + 1/2$.

Table 3: Matérn Radial Basis Functions R_ν of Order ν.

ν	$R_\nu(\tau)$, $\tau := \|s - t\|$
0	$e^{-\tau}$
1	$e^{-\tau}(1 + \tau)$
2	$e^{-\tau}(3 + 3\tau + \tau^2)$
3	$e^{-\tau}(15 + 15\tau + 6\tau^2 + \tau^3)$
4	$e^{-\tau}(105 + 105\tau + 45\tau^2 + 10\tau^3 + \tau^4)$
5	$e^{-\tau}(945 + 945\tau + 420\tau^2 + 105\tau^3 + 15\tau^4 + \tau^5)$
6	$e^{-\tau}(12285 + 12285\tau + 5565\tau^2 + 1470\tau^3 + 240\tau^4 + 23\tau^5 + \tau^6)$
7	$e^{-\tau}(184275 + 184275\tau + 71345\tau^2 + 23205\tau^3 + 4110\tau^4 + 470\tau^5 + 32\tau^6 + \tau^7)$
$\vdots$	$\vdots$
$m+1$	$e^{-\tau} \sum_{i=0}^{m+1} a_{m+1,i}\tau^i$, where $\quad a_{m+1,0} = (2m+1)a_{m,0},$ $\quad a_{m+1,i} = (2m+1-i)a_{m,i} + a_{m,i-1},\ i = 1, \cdots, m,$ and $\quad a_{m+1,m+1} = 1$

values. Let $\mathbf{A}(\mathbf{t})$ be the influence vector for a generic point $\mathbf{t}$. Note that if $\mathbf{t}$ is a particular $\mathbf{t}_i$ then $\mathbf{A}(\mathbf{t})$ is the ith row of the influence matrix. Using this notation we can express the estimator of f at point $\mathbf{t}$ in matrix form for simplicity

$$\hat{f}(\mathbf{t}) = \mathbf{A}(\mathbf{t})\,\mathbf{z} \tag{2.10}$$

$$= \mathbf{r}'(\mathbf{\Sigma} + \lambda\mathbf{I})^{-1}\mathbf{z}, \tag{2.11}$$

where $\mathbf{r} := (R(\mathbf{t}, \mathbf{t}_1), \ldots, R(\mathbf{t}, \mathbf{t}_n))'$. Differentiating $\hat{f}$ with respect to x and y provides the necessary derivatives to estimate WSS. Estimator $\hat{f}(\mathbf{t})$ depends on x and y only through $\mathbf{r}$, which depends on R. Thus,

$$\frac{\partial \hat{f}(\mathbf{t})}{\partial x} = \mathbf{r}'_x(\mathbf{\Sigma} + \lambda\mathbf{I})^{-1}\mathbf{z} \tag{2.12}$$

$$\frac{\partial \hat{f}(\mathbf{t})}{\partial y} = \mathbf{r}'_y(\mathbf{\Sigma} + \lambda\mathbf{I})^{-1}\mathbf{z}, \tag{2.13}$$

where

$$\mathbf{r}_x \quad := \quad \left(\frac{\partial R_\nu(\mathbf{t}, \mathbf{t}_1)}{\partial x}, \dots, \frac{\partial R_\nu(\mathbf{t}, \mathbf{t}_n)}{\partial x} \right)' \tag{2.14}$$

$$\mathbf{r}_y \quad := \quad \left(\frac{\partial R_\nu(\mathbf{t}, \mathbf{t}_1)}{\partial y}, \dots, \frac{\partial R_\nu(\mathbf{t}, \mathbf{t}_n)}{\partial y} \right)'. \tag{2.15}$$

Substitution gives

$$\hat{W}(\mathbf{t}) \quad = \quad \mu \left(n_x\, \mathbf{r}_x'(\mathbf{\Sigma} + \lambda\mathbf{I})^{-1}\mathbf{z} + n_y\, \mathbf{r}_y'(\mathbf{\Sigma} + \lambda\mathbf{I})^{-1}\mathbf{z} \right) \tag{2.16}$$

$$= \quad \mu \left(n_x\, \mathbf{r}_x' + n_y\, \mathbf{r}_y' \right) (\mathbf{\Sigma} + \lambda\mathbf{I})^{-1}\mathbf{z}. \tag{2.17}$$

From equation (2.17), we see that to estimate WSS, we need only differentiate the reproducing kernel. In general, to estimate any bounded linear functional we apply the functional to the reproducing kernel in (2.6).

2.2.4   Model Selection: Choosing ν and λ

Generalized cross-validation (GCV) (Craven and Wahba 1979) is one objective method for automatically selecting both the order of the RK R_ν and the smoothing parameter λ. Let $\hat{\mathbf{f}}_{\nu,\lambda} := (\hat{f}_\nu(\mathbf{t}_1, \lambda), \dots, \hat{f}_\nu(\mathbf{t}_n, \lambda))'$ be the $n \times 1$ matrix of fitted values under R_ν and λ. Also, let $\mathbf{\Sigma}_\nu$ be defined as in (2.8) where ν denotes the use of $R_\nu(\mathbf{s}, \mathbf{t})$. Since $\mathbf{\Sigma}_\nu$ is symmetric, positive definite, it has an eigenvalue-eigenvector decomposition $\mathbf{\Sigma}_\nu = \mathbf{\Gamma D \Gamma}'$, where $\mathbf{\Gamma}$ is an orthogonal matrix of eigenvectors and $\mathbf{D}$ is a diagonal

matrix of the corresponding eigenvalues. Then, it follows from (2.6) that

$$\hat{\mathbf{f}}_{\nu,\lambda} = \mathbf{\Sigma}'_\nu \mathbf{c} \tag{2.18}$$

$$= \mathbf{\Sigma}'_\nu (\mathbf{\Sigma}_\nu + \lambda \mathbf{I})^{-1} \mathbf{z} \tag{2.19}$$

$$= \mathbf{\Gamma}' \mathbf{D} \mathbf{\Gamma} (\mathbf{\Gamma} \mathbf{D} \mathbf{\Gamma}' + \lambda \mathbf{I})^{-1} \mathbf{z} \tag{2.20}$$

$$= \mathbf{\Gamma} \mathbf{D} (\mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{\Gamma}' \mathbf{z}. \tag{2.21}$$

Let $\mathbf{A}_{\nu,\lambda} := \mathbf{\Gamma} \mathbf{D} (\mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{\Gamma}'$. Then,

$$\hat{\mathbf{f}}_{\nu,\lambda} = \mathbf{A}_{\nu,\lambda} \mathbf{z}. \tag{2.22}$$

The GCV score for ν and λ is

$$V(\nu, \lambda) := \frac{\mathbf{z}'(\mathbf{I} - \mathbf{A}_{\nu,\lambda})'(\mathbf{I} - \mathbf{A}_{\nu,\lambda})\mathbf{z}}{(1 - n^{-1}\mathrm{tr}(\mathbf{A}_{\nu,\lambda}))^2}. \tag{2.23}$$

The numerator of (2.23) is the residual sum of squares. To recognize the denominator, first note that the degrees of freedom for the signal is $\mathrm{df}_{sig} = \mathrm{tr}(\mathbf{A}_{\nu,\lambda}) = n - \mathrm{df}_{res}$, where df_{res} is the residual degrees of freedom. Thus, the denominator of (2.23) is $(\mathrm{df}_{res}/n)^2$. The GCV estimates of ν and λ, denoted $\hat{\nu}$ and $\hat{\lambda}$, respectively, are the values that minimize (2.23). In practice, $\mathbf{A}_{\nu,\lambda}$ is computed for several values of λ, typically on a log scale, and for several orders of the Matérn RBFs. For, say, $\nu = 1, \ldots, 7$, $\hat{\nu} := \arg\min_\nu \left\{ V(\nu = 1, \hat{\lambda}_1), \ldots, V(\nu = 7, \hat{\lambda}_7) \right\}$, where $\hat{\lambda}_i$ is the minimizer of $V(\nu = i, \lambda)$. The GCV estimates have a number of desirable properties that are discussed in Craven and Wahba (1979), Wahba (1990), and the references therein. One such property is that GCV is consistent for estimating the true, unknown function. This means that as the number of observations increases, $\hat{\mathbf{f}}_{\hat{\nu},\hat{\lambda}}$ approaches the true $\mathbf{f}$, where $\hat{\nu}, \hat{\lambda}$ minimize $V(\nu, \lambda)$.

2.2.5 Bayesian Confidence Intervals for Shear

In this section we consider inference for WSS. One disadvantage of our nonparametric estimator for W is that it is biased due to the penalty on the norm of f. This complicates inference for estimates of shear. By interpreting our estimators as Bayes estimators with respect to a (conjugate) normal prior on f, we can construct a Bayesian confidence interval. The original Bayesian confidence interval (Wahba 1983) was constructed for the smoothing spline. Here, we extend Wahba's interval to WSS.

A Bayesian confidence interval for $W(\mathbf{t})$ is constructed from the mean squared error (MSE) of its estimator $\hat{W}(\mathbf{t})$. Recall that

$$\text{MSE}[\hat{W}(\mathbf{t})] = \text{Var}[\hat{W}(\mathbf{t})] + \text{Bias}^2[\hat{W}(\mathbf{t})]. \tag{2.24}$$

A $(1-\alpha)100\%$ approximate confidence interval has the form

$$\hat{W}(\mathbf{t}) \pm Z_{1-\alpha/2}\sqrt{\text{MSE}[\hat{W}(\mathbf{t})]}, \tag{2.25}$$

where $Z_{1-\alpha/2}$ denotes the $1-\alpha/2$ quantile of the standard normal distribution. To estimate the MSE, we need to estimate both the variance and the bias. We first consider the variance.

The variance of the estimator is a routine calculation. To simplify the expressions, define

$$\mathbf{h} = \left(n_x\,\mathbf{r}'_x + n_y\,\mathbf{r}'_y\right)(\boldsymbol{\Sigma} + \lambda\mathbf{I})^{-1} \tag{2.26}$$

so that

$$\hat{W}(\mathbf{t}) = \mu\mathbf{h}\mathbf{z}. \tag{2.27}$$

The variance of the WSS estimator is

$$\mathrm{Var}[\hat{W}(\mathbf{t})] = \mathrm{Var}[\mu\mathbf{hz}] \qquad (2.28)$$

$$= \mu^2\mathbf{h}\mathrm{Var}[\mathbf{z}]\mathbf{h}' \qquad (2.29)$$

$$= \mu^2\mathbf{h}\sigma^2\mathbf{I}\mathbf{h}' \qquad (2.30)$$

$$= \mu^2\sigma^2\mathbf{h}\mathbf{h}'. \qquad (2.31)$$

The error variance, σ^2, is estimated from the residuals to the fit of the velocity function f. The appropriate degrees of freedom is $n - \mathrm{tr}(\mathbf{A}_{\nu,\lambda})$.

The squared bias is more difficult since we do not know the true velocity function f. It is possible to compute the expected bias from assuming that f is well-approximated in the RKHS and from a Bayesian argument. The true velocity can be represented by its minimum norm interpolant plus error $\rho(\mathbf{t})$, i.e.,

$$f(\mathbf{t}) = \mathbf{r}'\mathbf{\Sigma}^{-1}\mathbf{f} + \rho(\mathbf{t}), \qquad (2.32)$$

where

$$\mathbf{f} := (f(\mathbf{t}_1), \ldots, f(\mathbf{t}_n))'. \qquad (2.33)$$

Notice that $\mathbf{r}'\mathbf{\Sigma}^{-1}\mathbf{f}$ is $\hat{f}$ when $\lambda = 0$ and the data $\mathbf{z}$ are replaced with the values of the velocity function at the $\mathbf{t}_i$. Assume that f is well-approximated by its minimum norm interpolant at all points. This means that $\rho(\mathbf{t})$ is negligible and we let $\rho(\mathbf{t}) = 0$.

The minimum norm interpolant of WSS is obtained from the minimum norm interpolant of the velocity function by similar calculations to obtain $\hat{W}(\mathbf{t})$ from $\hat{f}(\mathbf{t})$. A minimum norm interpolant of WSS is (ignoring approximation errors)

$$W(\mathbf{t}) = \mu\left(n_x\,\mathbf{r}'_x + n_y\,\mathbf{r}'_y\right)\mathbf{\Sigma}^{-1}\mathbf{f}. \qquad (2.34)$$

Equation (2.34) lets us express the bias of $\hat{W}(\mathbf{t})$ as

$$\text{Bias}[\hat{W}(\mathbf{t})] = \text{E}[\hat{W}(\mathbf{t})] - W(\mathbf{t}) \tag{2.35}$$

$$= \text{E}[\mu \mathbf{hz}] - \mu \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \boldsymbol{\Sigma}^{-1} \mathbf{f} \tag{2.36}$$

$$= \mu \left(\mathbf{h}\text{E}[\mathbf{z}] - \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \boldsymbol{\Sigma}^{-1} \mathbf{f}\right) \tag{2.37}$$

$$= \mu \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \left((\boldsymbol{\Sigma} + \lambda \mathbf{I})^{-1} - \boldsymbol{\Sigma}^{-1}\right) \mathbf{f}. \tag{2.38}$$

The only unknown in (2.38) is $\mathbf{f}$. By placing a prior distribution on $\mathbf{f}$, it is possible to compute the expected square bias.

To estimate the bias, a Bayesian argument that places a prior distribution on $\mathbf{f}$ yields an estimate of the expected square bias. The conjugate prior distribution placed on $\mathbf{f}$ is $\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \sigma^2/\lambda \boldsymbol{\Sigma})$. This prior corresponds to a posterior estimate of $\hat{f}(\mathbf{t}) = \mathbf{r}'(\boldsymbol{\Sigma} + \lambda \mathbf{I})^{-1} \mathbf{z}$. The expected square bias is

$$
\begin{aligned}
\text{E}\left[\text{Bias}^2[\hat{W}(\mathbf{t})]\right] &= \text{E}\left[\mu \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \left(\boldsymbol{\Sigma} + \lambda \mathbf{I}\right)^{-1} - \boldsymbol{\Sigma}^{-1}\right) \mathbf{f}\right] \\
&\quad \cdot \left(\mu \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \left(\boldsymbol{\Sigma} + \lambda \mathbf{I}\right)^{-1} - \boldsymbol{\Sigma}^{-1}\right) \mathbf{f}\right)' \\
&= \mu^2 \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \left(\boldsymbol{\Sigma} + \lambda \mathbf{I}\right)^{-1} - \boldsymbol{\Sigma}^{-1}\right) \text{Var}[\mathbf{f}] \\
&\quad \cdot \left(\boldsymbol{\Sigma} + \lambda \mathbf{I}\right)^{-1} - \boldsymbol{\Sigma}^{-1}\right)' \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right)' \\
&= \frac{\mu^2 \sigma^2}{\lambda} \left[\mathbf{h} \boldsymbol{\Sigma} \mathbf{h}' - 2\mathbf{h} \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right)' \right. \\
&\quad \left. + \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \boldsymbol{\Sigma}^{-1} \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right)'\right].
\end{aligned}
\tag{2.39}
$$

Collecting the expressions for the variance and the expected square bias, we compute the MSE (or more appropriately named the expected MSE due to the prior)

$$
\begin{aligned}
\text{MSE}[\hat{W}(\mathbf{t})] &= \mu^2 \sigma^2 \mathbf{h}\mathbf{h}' + \frac{\mu^2 \sigma^2}{\lambda} \left[\mathbf{h} \boldsymbol{\Sigma} \mathbf{h}' - 2\mathbf{h} \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right)' \right. \\
&\quad \left. + \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right) \boldsymbol{\Sigma}^{-1} \left(n_x \, \mathbf{r}'_x + n_y \, \mathbf{r}'_y\right)'\right].
\end{aligned}
\tag{2.40}
$$

Equation (2.40) can be simplified by substituting the an eigenvalue-eigenvector decomposition for Σ and factoring the eigenvalues (somewhat tedious) to get

$$\text{MSE}[\hat{W}(\mathbf{t})] = \mu^2 \sigma^2 \mathbf{a}'(\Sigma + \lambda \mathbf{I})^{-1} \Sigma^{-1} \mathbf{a} \qquad (2.41)$$

where

$$\mathbf{a} = n_x\, \mathbf{r}'_x + n_y\, \mathbf{r}'_y. \qquad (2.42)$$

By replacing σ^2 in (2.41) with an estimate, we can have a Bayesian estimator for MSE that can be used to construct confidence intervals for WSS.

2.2.6   Experimental Evaluation and Data

The methods of this paper were applied to three phantom data sets. The goal of the phantom studies was to determine if our method is capable of estimating velocity and shear in situations where we have an approximate knowledge of the truth. With this goal in mind we selected λ and ν to match the expected WSS.

The data consist of PC-MR images of constant flow of a blood substitute, which was matched to the viscosity of human blood. The flow rate was set to 6.9 ml/s for the first two data sets and 6.5 ml/s for the third data set. These are typical flow rates during diastole in humans. With each PC-MR image comes a perfectly co-registered "magnitude image" which has good contrast for resolving the walls of the phantom vessels. The magnitude images were used to identify the vessel wall pixels and segment the interior pixels in the PC-MR images. Vessel wall pixels were identified by the Canny edge detection algorithm (Canny 1986), which is part of the Matlab (Mathworks, Inc.) image processing toolbox. A periodic cubic smoothing spline was fit to each set of wall pixel locations to compute unit normal vectors along

the wall of each of the phantom vessels (Carew et al. 2003). The detected vessel wall was eroded by a distance of one pixel to eliminate observations from partially-volumed pixels. The partially-volumed pixels intersect both the vessel wall and the interior of the vessel. The software and algorithms used to carry out this analysis were developed for the R statistical environment (www.r-project.org). Some of the 3D plots and images were produced with Matlab.

The first data set is from a straight glass tube phantom that has a circular cross-section with a 3.9mm inner diameter radius. For a straight vessel with a circular cross-section, the flow can be expressed analytically as a paraboloid (Barna 1969). This allows us to compute that the theoretically-expected WSS is approximately 0.13 N/m^2. The image field of view was 100mm×100mm and the image matrix was 512×512 pixels. The pixel intensities outside of the tube were set to zero and the PC-MR images were reduced to a rectangular region of approximately 50×50 pixels that covered the entire glass tube. The boundary points (i.e., points in the image at the tube wall) were automatically identified.

The second data set is from a modified version of the straight glass tube phantom where the tube was laterally indented on one side to yield a non-convex cross section. For this vessel geometry there is no closed-form parametric model describing fluid flow. The indentation is not so extreme that we expect the WSS to be on the order of 0.13 N/m^2, particularly on the side opposite the indentation where the wall is circular. The image size and pre-processing were identical to that of the first data set.

A replica of a human carotid bifurcation (Shelley Medical, Toronto) was selected as the third phantom. Estimating WSS in the internal carotid artery (ICA) near the

bifurcation is of great interest since in many patients stroke is caused by stenosis or obstruction of the ICA. The image matrix was 256×256 pixels with a 0.39 mm pixel size. The image plane was selected to intersect the ICA approximately 1 cm superior to the bifurcation and oriented to be perpendicular to the anterior wall of the vessel. The perpendicularity ensures that nearly all flow is into or out of the image plane.

2.3   Results

The nonparametric model fits the segmented PC-MR velocity images well for each of the three phantom data sets. We first consider the straight glass tube phantom with a circular cross-section. A plot of the raw velocity measurements after segmentation is in Figure 9A. The measurements appear slightly more variable in the interior of the glass tube than near the edges. A fit with $\nu = 4$ and $\lambda = 300$ is in Figure 9B. As expected, the nonparametric fit appears parabolic. The parameters were tuned to best match the expected WSS (0.13 N/m^2). Figure 9C is a diagram of how the WSS estimates (Figure 9D) are ordered. Starting at the wall point that is at the largest value along the x-axis, we move counter clockwise. The shear stress estimates in Figure 9D are accompanied by 95% Bayesian confidence intervals. The solid horizontal line denotes the expected value of WSS. Based on the circular symmetry of the vessel, we expect WSS to be constant at all wall points. On average the estimates of WSS are very close to the expected value. The estimates of WSS appear to have correlated errors. Of the 35 equally-spaced points along the vessel wall where WSS was estimated, 5 of the intervals do not cover the expected value. This gives roughly 86% coverage of the theoretically expected value.
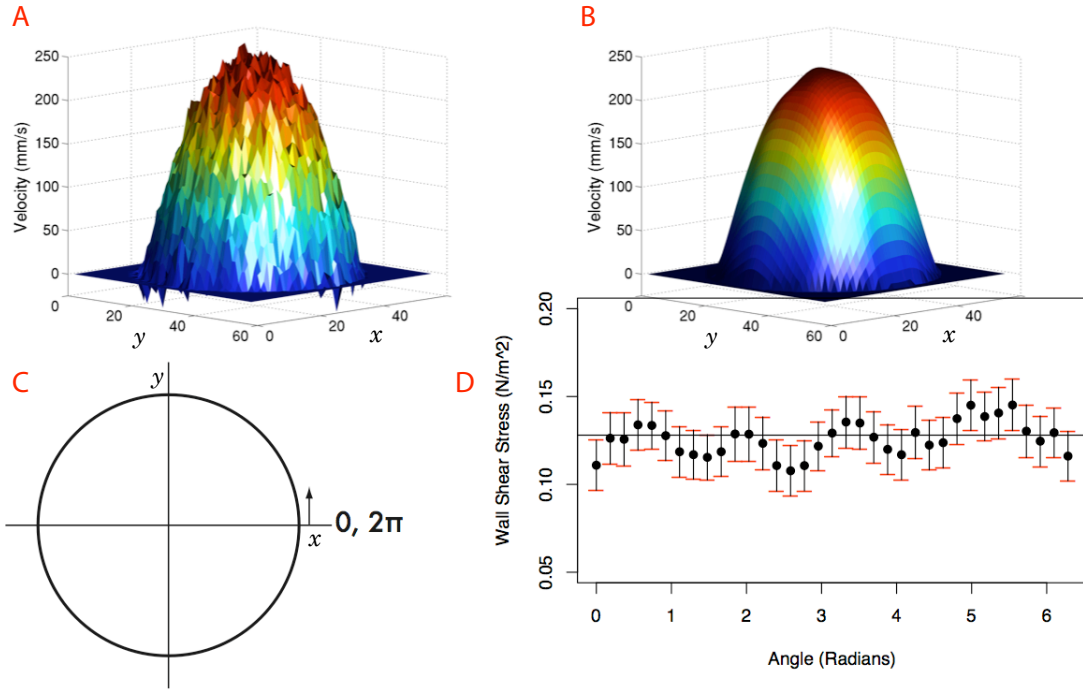
Figure 9: Fluid velocity and wall shear stress estimates in a straight glass tube phantom with a circular cross-section. The raw velocity measurements are in A. The nonparametric fit with $\nu = 4$ and $\lambda = 300$ is in B. The fitted velocity function appears parabolic, which is predicted by physical principles. In D are the WSS estimates and 95% Bayesian confidence intervals. The diagram in C shows how the measurements are ordered in D. Starting at the point with the largest value along the x-axis the order of the WSS estimates increase counter clockwise, indexed by the angle.
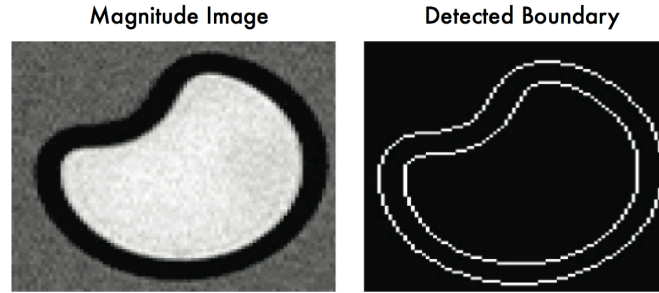
Figure 10: Magnitude and detected boundary for the straight glass tube phantom with non-convex cross-section.

The straight glass tube phantom with a non-convex cross section presents a case where there is no parametric form describing fluid velocity. Figure 10 shows the magnitude image and the vessel wall detected by the Canny method. The magnitude image has excellent contrast for detecting the vessel wall. Figure 11 shows an image of the nonparametric fit with $\nu = 4$ and $\lambda = 150$ and WSS estimates with 95% Bayesian confidence intervals. The shear estimates are ordered as in Figure 9. WSS is roughly constant around the vessel wall except for the region of the indentation where it sharply increases. The horizontal line in Figure 11 denotes WSS=0.13 N/m^2, which is the expected shear stress for a cylindrical glass tube of 3.9 mm radius. Since there is no closed-form description of velocity, this line serves as an approximate reference. In the regions of the vessel wall away from the indentation, the WSS estimates are very close, on average, to 0.13 N/m^2. Figure 12 plots the sample quantiles of the residuals versus the quantiles of a standard normal distribution. That the points lie closely on a 45 degree line indicates that the residuals closely follow a normal distribution.

The data from the carotid bifurcation demonstrate that the nonparametric method can model complex flow and shear in a situation that is of direct clinical interest.
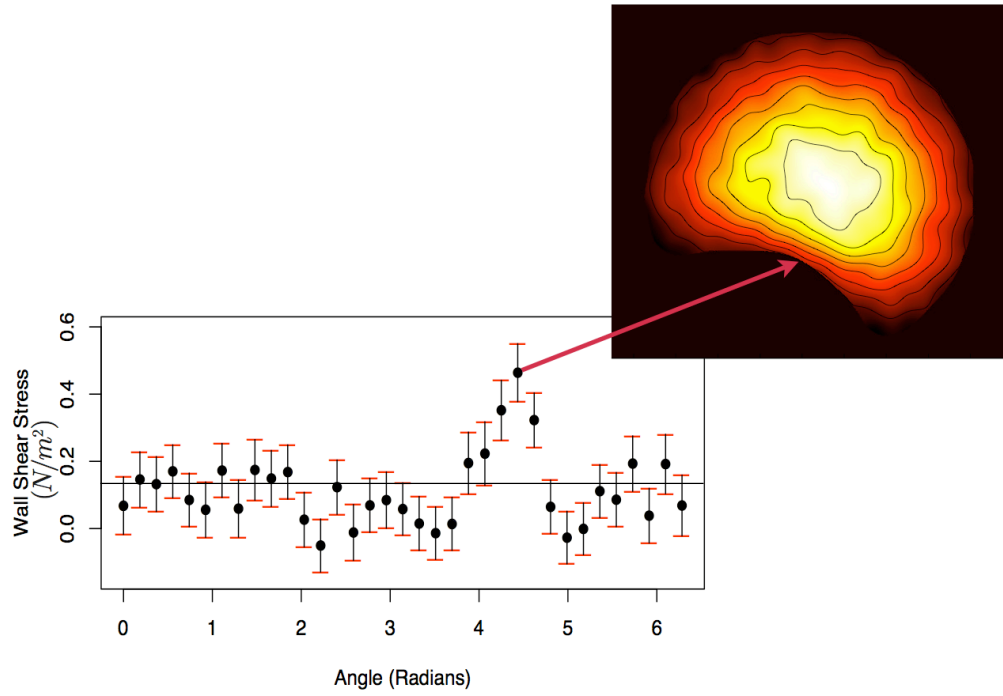
Figure 11: Fluid velocity and wall shear stress estimates in a straight glass tube phantom with non-convex cross section. The nonparametric fit uses $\nu = 4$ and $\lambda = 150$. The shear stress estimates are ordered as described in Figure 9C. The WSS is nearly constant except for the region near the indentation where the shear stress increases. The increased shear can be seen through the more closely-spaced contour lines on the velocity image near the center of the indentation. The horizontal line shows the expected value of WSS calculated for a vessel with a circular cross-section.
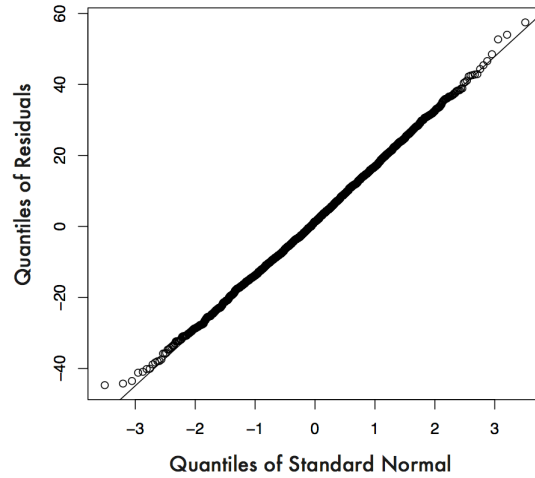
Figure 12: Residuals from the fit to the non-convex phantom. The quantile-quantile plot indicates that the residuals closely follow a normal distribution.

Figure 13A shows the geometry of the bifurcation phantom as well as the orientation of the image slice (green). The slice goes through the internal carotid artery, which is a major source of blood to the brain. The common carotid artery is located just inferior to the bifurcation. The other branching artery is the external carotid artery, which supplies blood to non-brain tissues in the head. The image in Figure 13B shows the nonparametric fit with $\nu = 4$ and $\lambda = 100$. The parameters were selected by visual inspection of the fit and residuals for several values of ν and λ. The boundary of the vessel in Figure 13B shows less circular symmetry than in the other two phantoms. This is partly due to the smaller diameter of the vessel (approximately 5 mm) and the lower spatial resolution of the image (256×256 pixels versus 512×512 pixels). Figure 13B shows that the velocity and rate of change of velocity are much higher along the anterior wall of the vessel than on the posterior wall. This suggests that the shear stress is higher against the anterior wall than against the posterior wall. This

is confirmed in the plot of the WSS estimates in Figure 13C. The point estimates are accompanied by 95% Bayesian confidence intervals and are ordered as described in Figure 9.

2.4   Discussion

The preliminary results from this study demonstrate the feasibility of our nonparametric estimator of fluid velocity and shear from phase-contrast MR images, which measure fluid velocity. The nonparametric estimator is based on penalized least squares in a reproducing kernel Hilbert space generated by the Matérn family of radial basis functions. The main strength of our method is that it does not require restrictive assumptions about vessel geometry (e.g., convex) or flow (e.g., parabolic). Unlike the Oyre et al. (1998) method, our method is not constrained to arterial cross-sections and velocity data that can be modeled by parabolic sectors. Another advantage of the nonparametric method is that it can be easily extended to higher dimensions for 3D images, 2D images acquired at regularly (or even irregularly) spaced intervals, and multiple images acquired over time. This is important for estimating shear stress over a cardiac cycle since the force driving the blood is not constant. This causes the WSS to vary over the cardiac cycle.

The results from the phantom models show promise for the clinical utility of our method. The straight glass tube phantom demonstrates that, for very simple vessel geometries, the nonparametric method produces results that are quite consistent with the parabolic model of flow. This is not to suggest that the nonparametric model is better than a parametric model for these simple situations. The power of the
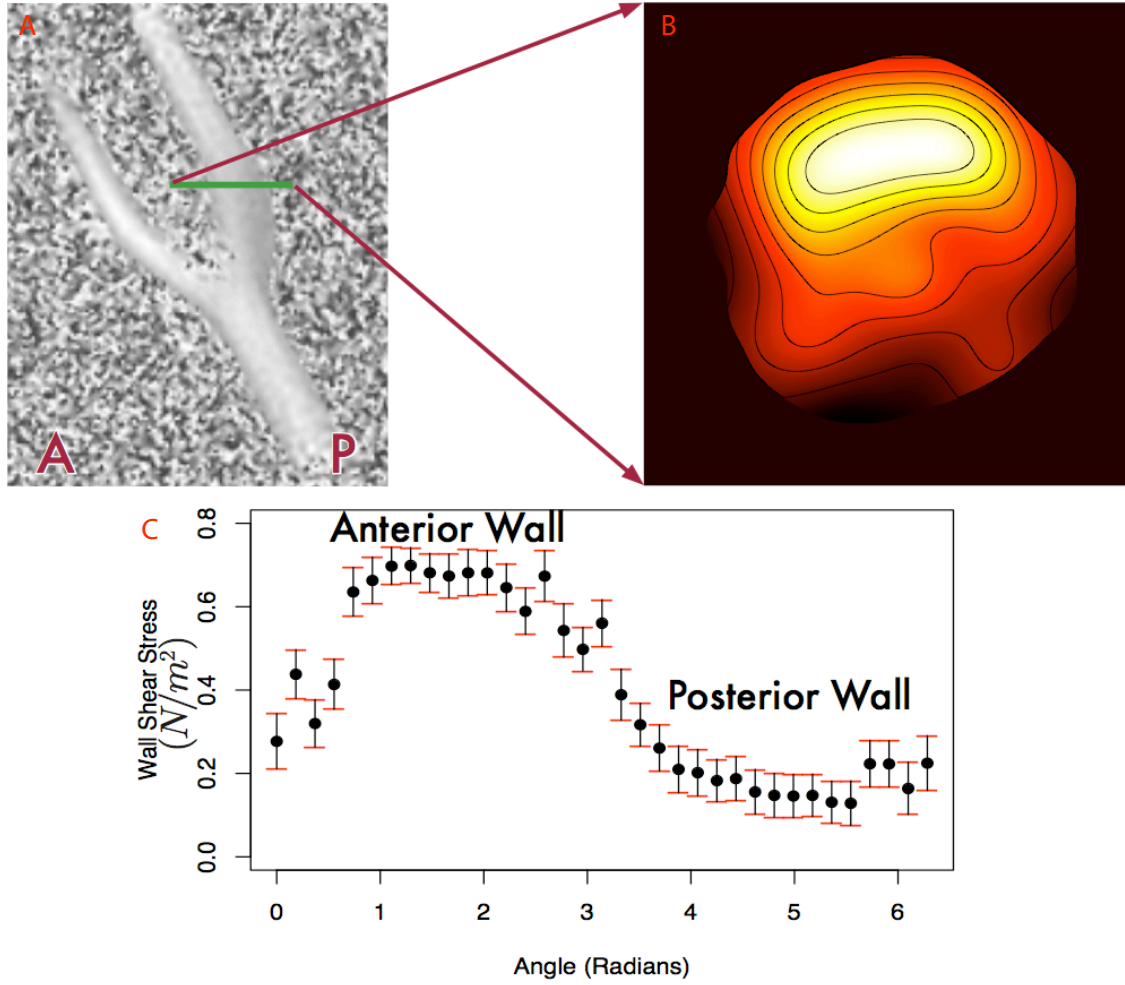
Figure 13: Velocity and WSS in a human carotid bifurcation phantom. The green line segment that bisects the internal carotid artery in A indicates the orientation of the imaging plane. The fitted velocity function with $\nu = 4$ and $\lambda = 100$ is in B. The velocity and rate of change of velocity is much higher along the anterior wall of the vessel than along the posterior wall. The effect of this flow regime is seen in the WSS estimates in C. The point estimates are ordered as in Figure 9 and the intervals are 95% Bayesian confidence intervals.

nonparametric method is best seen in the fits to the indented tube phantom data and the carotid bifurcation model. The indented tube phantom provided an example of a simple flow regime where parametric methods would typically fail. The estimated flow patterns in the bifurcation phantom (Figure 13B) are consistent with the bifurcation model studies of Motomiya and Karino (1984). The low shear stress on the posterior wall of the ICA is consistent with several findings (Friedman et al. 1987; Ku et al. 1985; Zarins et al. 1983) which link low shear stress to formation of plaque on the posterior wall.

Our method has a number of limitations. We require the velocity to be a smooth function. Our method may not work well in very turbulant flow. An accurate definition of the glass tube boundary or arterial wall is needed. It is not clear that the magnitude images would provide as good wall contrast in vivo as in the phantoms. It may be necessary to acquire e.g., a black-blood image to achieve better contrast. If this is the case, then registration between the wall image and the phase image may become an issue. The Oyre et al. (1998) method has the benefit of estimating the boundary points from the data. However, this depends on the model fitting the data well, which does not appear to be the case for vessels with noncircular cross-sections and nonparabolic flow. We have not yet addressed automatic selection of ν and λ. We propose model selection with GCV. This needs further study. If our method is to be used in a clinical setting, automatic selection will become necessary. In the three phantom data sets considered here, $\nu = 4$ seems to work quite well. It may not be necessary to tune this parameter for each data set; ν may depend more on the class of data than on the specific set of data. The smoothing parameter λ took values 300, 150, and 100 on our phantom data sets. Less smoothing was required

for the carotid phantom data set. This can partly be explained by the larger pixel sizes in the image. The larger pixel sizes give a higher signal to noise ratio in the measurements. Overall, we did not find the estimates to be sensitive to λ for the same order of magnitude. Less smoothing generally means wider Bayesian confidence intervals for the estimates.

More investigation is needed to study the frequentist properties of the Bayesian confidence intervals. For the straight glass tube phantom with circular cross-section we have an approximate expectation of shear stress based on a parabolic model. The 95% Bayesian confidence interval covered this value by roughly 86% (Figure 9D). In an experiment, there is no guarantee that the theoretically predicted value was the true value. Experimental variation may partly explain why the coverage is less than 95%. That this Bayesian interval does not explicitly control the coverage probability may also partly explain the 86% coverage. We propose to simulate data by taking the nonparametric fits to the three phantom data sets as the ground truth. Simulated data sets will be generated by adding independent normal errors. The simulated data sets will be fit and the coverage probability of the Bayesian confidence intervals will be estimated by counting the fraction of intervals that cover the ground truth.

Chapter 3

The asymptotic behavior of the nonlinear estimators of the diffusion tensor and tensor-derived quantities

3.1   Introduction

Diffusion tensor imaging (DTI) (Basser et al. 1994b; Basser et al. 1994a) is a quantitative magnetic resonance imaging (MRI) method that is widely used to study the microstructural properties of white matter in the brain. The diffusion tensor, which describes the diffusion properties of the imaged tissue, is known to be proportional to the covariance matrix for Brownian motion of water molecules. Since DTI can provide microstructural information, it can reveal disease-correlated tissue changes that are not evident on conventional MRI. This sensitivity to microstructure and tissue organization makes DTI an important tool for studying neurologic diseases. Several functions of the diffusion tensor (e.g., trace and fractional anisotropy (FA)) are important for characterizing neurologic diseases (Lebihan et al. 2001).

Determining the statistical properties of the diffusion tensor estimator is necessary to determine if the method of estimation is appropriate, and to describe how noise in diffusion-weighted images leads to variability in estimates of the tensor and functions of the tensor estimate. Estimators of the variance of the tensor estimator and functions of the tensor estimator are needed for statistical inference, comparing experimental designs, and tractography. Studying the statistical properties of the nonlinear least squares estimator (NLSE) of the diffusion tensor is particularly timely since it was recently shown that the NLSE outperforms, in terms of mean square error, linear tensor estimators (Koay, Carew, Alexander, Basser, and Meyerand 2006).

In this chapter, we derive asymptotic properties of the NLSE of the diffusion tensor. The asymptotic distributions of the tensor estimator and functions of the tensor estimator are used to obtain estimators for their variances. We show that the NLSE of the diffusion tensor is a maximum likelihood estimator (MLE). This connection allows us to directly apply the theory of maximum likelihood estimation to obtain asymptotic properties. For diffusion tensor estimation, the asymptotic properties are achieved in the limit as the signal to noise ratio (SNR) and the number of samples of a set of diffusion directions go to infinity. Asymptotic properties are commonly used in many statistical procedures and can be very accurate well before the appropriate quantities approach infinity. We assess the utility of the asymptotic approximations with a series of simulations.

We show that the NLSE is consistent and asymptotically normal. Consistency is a desirable property of estimators, which implies that the estimates approach their true, unknown values in a probabilistic sense. Furthermore, certain functions of the tensor estimator are also consistent and asymptotically normal. To illustrate and

validate the theory we derive the asymptotic distributions of trace and FA and show, with simulations, experimental designs that have asymptotic distributions that are very close to the empirical distributions. The methods introduced in this chapter are applied to estimate the variance of trace and FA from a DTI data set acquired from a healthy human volunteer. Finally, we discuss implications of unequal variances for statistical tests with tensor-derived quantities.

3.2 Diffusion Tensor Model

The diffusion tensor model for a signal $S := (S_1, \ldots, S_n)'$ from a single voxel with n diffusion measurements is

$$S = S_0 \exp(-\mathbf{X}\boldsymbol{\beta}) + \boldsymbol{\epsilon}. \tag{3.1}$$

The symbol $'$ denotes matrix transposition and $\exp()$ denotes the exponential function applied to each element of an array. The signal with no diffusion weighting is denoted by S_0. The diffusion encoding matrix $\mathbf{X}$ is given by

$$\mathbf{X} := \begin{bmatrix} b_1 g_{x1}^2 & b_1 g_{y1}^2 & b_1 g_{z1}^2 & 2b_1 g_{x1} g_{y1} & 2b_1 g_{y1} g_{z1} & 2b_1 g_{x1} g_{z1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_n g_{xn}^2 & b_n g_{yn}^2 & b_n g_{zn}^2 & 2b_n g_{xn} g_{yn} & 2b_n g_{yn} g_{zn} & 2b_n g_{xn} g_{zn} \end{bmatrix},$$

where the b_j, $j = 1, \ldots, n$, are the diffusion weightings and the g_{ij}, $i \in \{x, y, z\}$, are the components of the gradient encoding unit vectors, which specify the direction of the diffusion weighting. The parameter $\boldsymbol{\beta}$ contains (at most) 6 unique elements of the symmetric, positive-definite diffusion tensor, namely

$$\boldsymbol{\beta} := (D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{yz}, D_{xz})',$$

where the diffusion tensor $\mathbf{D}$ is

$$\mathbf{D} := \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}.$$

We assume that the errors ϵ in (3.1) are independent and normal with constant variance, i.e., $\epsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, where $\mathbf{I}_n$ is an $n \times n$ identity matrix. Noise in magnitude MR images formed with quadrature detection follows a Rician distribution (Henkelman 1985; Koay and Basser 2006). As the SNR goes to infinity, the noise in magnitude images follows a normal distribution with variance equal to the variance of the normally-distributed noise in the quadrature channels (Gudbjartsson and Patz 1995). Yet, surprisingly, when the SNR is greater than three, the noise in the SE-EPI MR images is well-approximated by a normal distribution (Gudbjartsson and Patz 1995). With routine diffusion imaging parameters (e.g., $b \approx 1000$ s/mm^2 or less and the SNR is at least 20). Consequently, the normality and constant variance assumptions are quite sensible, and achieved at low SNR and number of images.

3.3   Nonlinear Least Squares Tensor Estimator

In this section, we show that the nonlinear least squares estimator (NLSE) of the diffusion tensor is the same as the maximum likelihood estimator (MLE). We start with the maximum likelihood estimator. The diffusion measurements S are viewed as a single sample drawn from a multivariate normal distribution. Under the distributional assumption $\epsilon \sim \mathcal{N}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, the signal is distributed as $S \sim \mathcal{N}_n(S_0 \exp(-\mathbf{X}\boldsymbol{\beta}), \sigma^2 \mathbf{I}_n)$. Define the parameter to be $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_8)' := (\boldsymbol{\beta}, S_0, \sigma^2)'$. The likelihood, $L(\boldsymbol{\theta}|S)$,

is given by

$$L(\boldsymbol{\theta}|S) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[S - S_0\exp(-\mathbf{X}\boldsymbol{\beta})\right]'\left[S - S_0\exp(-\mathbf{X}\boldsymbol{\beta})\right]\right\}.$$

Taking the logarithm gives the loglikelihood, $l(\boldsymbol{\theta}|S)$,

$$l(\boldsymbol{\theta}|S) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[S - S_0\exp(-\mathbf{X}\boldsymbol{\beta})\right]'\left[S - S_0\exp(-\mathbf{X}\boldsymbol{\beta})\right]. \qquad (3.2)$$

Note that the NLSE for $\boldsymbol{\beta}$ and S_0 do not depend on σ^2. In these cases, maximizing (3.2) with respect to $\boldsymbol{\beta}$ and S_0 is the same as minimizing the nonlinear least squares objective function, $h(\boldsymbol{\beta}, S_0)$, given by

$$h(\boldsymbol{\beta}, S_0) := \frac{1}{2}\left[S - S_0\exp(-\mathbf{X}\boldsymbol{\beta})\right]'\left[S - S_0\exp(-\mathbf{X}\boldsymbol{\beta})\right]. \qquad (3.3)$$

Therefore, the minimizers of (3.3), call them $\hat{\boldsymbol{\beta}}$ and $\hat{S}_0$, are the maximum likelihood estimators of the components $\boldsymbol{\beta}$ of the tensor and S_0. To get the MLE for σ^2, we substitute $\hat{\boldsymbol{\beta}}$ and $\hat{S}_0$ for $\boldsymbol{\beta}$ and S_0, respectively, in equation (3.2) and find the σ^2 that maximizes $l(\boldsymbol{\theta}|S)$. It is easy to show that the MLE for σ^2 is

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n}\left[S - \hat{S}_0\exp(-\mathbf{X}\hat{\boldsymbol{\beta}})\right]'\left[S - \hat{S}_0\exp(-\mathbf{X}\hat{\boldsymbol{\beta}})\right] = \frac{\text{RSS}}{n}. \qquad (3.4)$$

This estimator is the residual sum of squares (RSS) divided by the number of diffusion measurements. The MLE for the error variance σ^2 is not the best estimator, particularly for small n. In the next section we will provide a better estimator that relies on the asymptotic properties of $\hat{\boldsymbol{\beta}}$ and $\hat{S}_0$.

3.4 Asymptotic Properties of the Tensor Estimator

The NLS estimators of $\boldsymbol{\beta}$, the components of the tensor, and S_0 are the same as the maximum likelihood estimator under the assumption of independent normal errors with constant variance. This fact allows us to use the asymptotic properties of the MLE to obtain the asymptotic properties of the NLS estimators. The two main properties of the MLE are consistency and asymptotic normality. Consistency, also called convergence in probability, is denoted by $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}$. This means that for all $a > 0$, $\lim_{N,\text{SNR}\to\infty} P(|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}| > a) = 0$, where N is the number of samples of the diffusion directions. In words, this means that for large enough N and SNR, the probability that the tensor estimator is more than an arbitrarily small distance away from the true value goes to zero. By theorem 5.1 on page 463 of Lehmann and Casella (1998), we have the following asymptotic properties:

1. $\hat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}$, and

2. $\hat{\boldsymbol{\theta}} \xrightarrow{d} \mathcal{N}_8(\boldsymbol{\theta}, I^{-1}(\boldsymbol{\theta}))$.

The first property states that as the number of samples of the diffusion directions and the SNR increase, the NLSE of the tensor components converges in probability to the true tensor components. The second property implies the asymptotic normality of the NLSE, i.e., $\hat{\boldsymbol{\theta}}$ converges in distribution to a multivariate normal. Convergence in distribution, denoted by $\xrightarrow{d}$, means that as N and SNR go to infinity, the distribution of the tensor estimator (at all continuity points of its distribution) is normal with mean equal to $\boldsymbol{\theta}$. Both of the asymptotic properties of the MLE hold under certain

regularity conditions. Under our normality assumption, these regularity conditions are satisfied.

Note that $I^{-1}(\boldsymbol{\theta})$, the covariance matrix of the asymptotic distribution, is the inverse of the Fisher information matrix, which is proportional to the Hessian of the expected loglikelihood. Specifically, from equation 6.11 on page 125 of Lehmann and Casella (1998),

$$I_{ij}(\boldsymbol{\theta}) := -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}l(\boldsymbol{\theta}|S)\right], \tag{3.5}$$

where $\mathrm{E}[\cdot]$ denotes the expectation operator, i.e., for random variable X, $\mathrm{E}[X] := \int X\, dP$, where P is a probability measure. In the case of equation (3.5), the expectation is taken with respect to the multivariate normal distribution that is parameterized by $\boldsymbol{\theta}$.

Since we are primarily interested in the properties of the components of the tensor, we can apply a simple transformation $\mathbf{C}$ to $\hat{\boldsymbol{\theta}}$ to get the asymptotic distribution of $\boldsymbol{\beta}$. Let

$$\mathbf{C} := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Then, $\mathbf{C}\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\beta}}$ and

$$\hat{\boldsymbol{\beta}} \xrightarrow{d} \mathcal{N}_6(\boldsymbol{\beta}, \mathbf{C}I^{-1}(\boldsymbol{\theta})\mathbf{C}'). \tag{3.6}$$

From equation (3.6), we see that $\hat{\boldsymbol{\beta}}$ is asymptotically normal with mean $\boldsymbol{\beta}$ and covariance matrix $\mathbf{C}I^{-1}(\boldsymbol{\theta})\mathbf{C}'$.

We will derive the Fisher information matrix for each component, first considering the derivatives with respect to the components of $\boldsymbol{\beta}$. It follows from the Hessian of (3.3), which was derived by Koay, Chang, Carew, Pierpaoli, and Basser (2006), that

$$\nabla_{\boldsymbol{\beta}}^2 l(\boldsymbol{\theta}|S) = -\frac{1}{\sigma^2}\nabla_{\boldsymbol{\beta}}^2 h(\boldsymbol{\beta}, S_0)$$

$$= -\frac{1}{\sigma^2}\mathbf{X}'(2\tilde{\mathbf{S}}^2 - \mathbf{S}\tilde{\mathbf{S}})\mathbf{X},$$

where

$$\mathbf{S} := \begin{bmatrix} S_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & S_n \end{bmatrix}, \text{and } \tilde{\mathbf{S}} := S_0 \begin{bmatrix} \exp(-\mathbf{X}_1\boldsymbol{\beta}) & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & & \cdots & 0 & \exp(-\mathbf{X}_n\boldsymbol{\beta}) \end{bmatrix},$$

with $\mathbf{X}_i$ denoting the ith row of the diffusion gradient matrix. The notation $\nabla_{\boldsymbol{\beta}}^2 h$ denotes a 6×6 matrix with ijth element equal to $\partial^2 h/\partial\beta_i\partial\beta_j$. Thus, the upper left 6×6 block of $I(\boldsymbol{\theta})$ is

$$\frac{1}{\sigma^2}\mathrm{E}[\mathbf{X}'(2\tilde{\mathbf{S}}^2 - \mathbf{S}\tilde{\mathbf{S}})\mathbf{X}] = \frac{1}{\sigma^2}\mathbf{X}'\tilde{\mathbf{S}}^2\mathbf{X}.$$

In the left hand side of the previous equation, the only random quantity is $\mathbf{S}$. Since $S \sim \mathcal{N}_n(S_0\exp(-\mathbf{X}\boldsymbol{\beta}), \sigma^2\mathbf{I}_n)$ and $\mathrm{E}[\mathbf{S}] = \tilde{\mathbf{S}}$, the right hand side follows. For the terms

involving S_0 and σ^2, the following can be shown:

$$-\mathrm{E}\left[\frac{\partial^2}{\partial S_0^2}\, l(\boldsymbol{\theta}|S)\right] = \frac{1}{\sigma^2}\exp(-\mathbf{X}\boldsymbol{\beta})'\exp(-\mathbf{X}\boldsymbol{\beta})$$

$$-\mathrm{E}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\, l(\boldsymbol{\theta}|S)\right] = \frac{n}{2(\sigma^2)^2}$$

$$-\mathrm{E}\left[\frac{\partial^2}{\partial\sigma^2\partial S_0}\, l(\boldsymbol{\theta}|S)\right] = 0$$

$$-\mathrm{E}\left[\frac{\partial^2}{\partial\sigma^2\partial\beta_i}\, l(\boldsymbol{\theta}|S)\right] = 0$$

$$-\mathrm{E}\left[\frac{\partial^2}{\partial\beta_i\partial S_0}\, l(\boldsymbol{\theta}|S)\right] = \frac{S_0}{\sigma^2}\sum_{j=1}^{n}\mathbf{X}_{ji}[\exp(-\mathbf{X}_j\boldsymbol{\beta})]^2.$$

From the last equation, we see that estimates of S_0 are correlated with estimates of $\boldsymbol{\beta}$. From a practical standpoint, this means that experimental designs that yield poor estimates of S_0 will adversely affect estimates of $\boldsymbol{\beta}$. For this reason, experimental designs that include only one image with no diffusion weighting should be avoided.

The MLE for the error variance in equation (3.4) has a large bias for small n. This causes under estimation of the variance. The consistency and asymptotic normality of the NLSE for S_0 and $\boldsymbol{\beta}$ allow us to get a better estimator. In particular,

$$\frac{\mathrm{RSS}}{\sigma^2} \xrightarrow{d} \chi^2_{n-7},$$

where χ^2_{n-7} denotes a χ^2 distribution with $n-7$ degrees of freedom. This suggests that the appropriate estimator of the error variance is

$$\hat{\sigma^2} := \frac{\mathrm{RSS}}{n-7} \tag{3.7}$$

rather than the MLE given in (3.4).

3.5 Distribution of a Linear Function of the Tensor

The distribution of a linear function of the tensor (e.g., trace) is asymptotically normal. This follows from the well-know property that linear functions of normal random variables are normally-distributed. To obtain the asymptotic distribution of the trace of the tensor estimator, let $\mathbf{c} := [1\ 1\ 1\ 0\ 0\ 0]$. Then, $\widehat{\mathrm{tr}\mathbf{D}} = \mathbf{c}\hat{\boldsymbol{\beta}}$, where tr denotes the trace of a matrix. From equation [**3.6**] it follows that

$$\widehat{\mathrm{tr}\mathbf{D}} \xrightarrow{d} \mathcal{N}_1(\mathrm{tr}\mathbf{D}, \mathbf{c}\mathbf{C}I^{-1}(\boldsymbol{\theta})\mathbf{C}'\mathbf{c}'). \tag{3.8}$$

The general form of the normal distribution for the trace of the diffusion tensor was previously given in Basser and Pajevic (2003). The difference in our paper is that we prove asymptotic normality and show how the variance of the asymptotic distribution comes from the Fisher information matrix.

To estimate $\mathrm{Var}(\widehat{\mathrm{tr}\mathbf{D}})$, we use the variance in (3.8) evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$, with σ^2 estimated according to (3.7), i.e.,

$$\widehat{\mathrm{Var}}(\widehat{\mathrm{tr}\mathbf{D}}) = \mathbf{c}\mathbf{C}I^{-1}(\hat{\boldsymbol{\theta}})\mathbf{C}'\mathbf{c}'. \tag{3.9}$$

The consistency of the estimator for the variance of trace follows directly from the consistency of the estimator for $\boldsymbol{\theta}$. The variance of the mean apparent diffusion coefficient (ADC), which is defined as $1/3\ \mathrm{tr}\mathbf{D}$, is the variance of the trace of the tensor scaled by $1/9$.

3.6 Distribution of a Nonlinear Function of the Tensor

In this section the multivariate delta method is used to show that a nonlinear function of the tensor estimate is asymptotically normal. In particular, we are interested in scalar functions of the tensor such as the fractional anisotropy. It is also possible that a vector-valued function of the tensor is of interest, e.g., the primary eigenvector of the tensor. To include this possibility, let

$$f(\hat{\boldsymbol{\beta}}) := (f_1(\hat{\boldsymbol{\beta}}), \ldots, f_r(\hat{\boldsymbol{\beta}}))$$

denote an r-dimensional vector-valued function of the tensor estimate. The multivariate delta method (Theorem 8.22, page 61 (Lehmann and Casella 1998)) yields

$$\left[(f_1(\hat{\boldsymbol{\beta}}) - f_1(\boldsymbol{\beta})), \ldots, (f_r(\hat{\boldsymbol{\beta}}) - f_r(\boldsymbol{\beta})) \right] \xrightarrow{d} \mathcal{N}_r(\mathbf{0}, \mathbf{B}\mathbf{C}I^{-1}(\boldsymbol{\theta})\mathbf{C}'\mathbf{B}'), \qquad (3.10)$$

where $\mathbf{B}_{ij} = \partial f_i / \partial \beta_j$. The matrix $\mathbf{B}$ of partial derivatives must be nonsingular in a neighborhood ω of $\boldsymbol{\beta}$. For scalar-valued f, the derivatives cannot be zero at $\boldsymbol{\beta}$. We also require that $f_1, \ldots, f_r$ are continuously differentiable in ω.

To compute the asymptotic distribution of FA, which is a normalized standard deviation of the eigenvalues of the diffusion tensor, we need to compute the partial derivatives of FA with respect to each of the six components of $\boldsymbol{\beta}$. FA takes values between 0 and 1, where 0 indicates isotropic diffusion and 1 indicates completely anisotropic diffusion. FA is written in terms of the eigenvalues λ_1, λ_2, and λ_3 of the diffusion tensor $\mathbf{D}$, as follows

$$\text{FA}(\lambda_1, \lambda_2, \lambda_3) := \left[\frac{3[(\lambda_1 - \bar{\lambda})^2 + (\lambda_2 - \bar{\lambda})^2 + (\lambda_3 - \bar{\lambda})^2]}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)} \right]^{\frac{1}{2}},$$

where $\bar{\lambda} := (1/3)(\lambda_1 + \lambda_2 + \lambda_3)$. We can express FA directly in terms of $\boldsymbol{\beta}$, which is more convenient for computing partial derivatives with respect to the components of $\boldsymbol{\beta}$. In this form FA is given by

$$\text{FA}(\boldsymbol{\beta}) = \left[\frac{3}{2} \left(1 - \frac{(\text{tr}\mathbf{D})^2}{3\,\text{tr}(\mathbf{D}^2)} \right) \right]^{\frac{1}{2}}. \tag{3.11}$$

Differentiating [**3.11**] gives the following:

$$\frac{\partial \text{FA}}{\partial D_i} = -\frac{1}{2\text{FA}} \left[\frac{\text{tr}\mathbf{D}\,\text{tr}(\mathbf{D}^2) - (\text{tr}\mathbf{D})^2 D_i}{(\text{tr}(\mathbf{D}^2))^2} \right], \quad i \in \{xx, yy, zz\}$$

$$\frac{\partial \text{FA}}{\partial D_j} = \frac{(\text{tr}\mathbf{D})^2 D_j}{\text{FA}(\text{tr}(\mathbf{D}^2))^2}, \quad j \in \{xy, yz, xz\}.$$

Let

$$\mathbf{B} := \left(\frac{\partial \text{FA}}{\partial D_{xx}}, \frac{\partial \text{FA}}{\partial D_{yy}}, \frac{\partial \text{FA}}{\partial D_{zz}}, \frac{\partial \text{FA}}{\partial D_{xy}}, \frac{\partial \text{FA}}{\partial D_{yz}}, \frac{\partial \text{FA}}{\partial D_{xz}} \right)$$

and apply (3.10) to get the asymptotic distribution of FA. To use this result to estimate the variance of an FA estimate, evaluate the Fisher information at the value of $\hat{\boldsymbol{\theta}}$, with σ^2 estimated according to (3.7), to get

$$\widehat{\text{Var}}(\text{FA}(\hat{\boldsymbol{\beta}})) = \mathbf{B}CI^{-1}(\hat{\boldsymbol{\theta}})\mathbf{C}'\mathbf{B}'. \tag{3.12}$$

It is important to note that when FA is zero (completely isotropic diffusion), the asymptotic variance estimator in equation (3.12) is singular. This is due to the dependence of the derivatives of FA on the value of FA. When FA goes to zero, the factor 1/FA in the expressions for the derivatives goes to infinity.

3.7 Validation and Application

3.7.1 Simulation Study

To validate the range of utility of the asymptotic distributions of trace and FA for finite SNR and N, we simulated diffusion measurements for three cylindrically symmetric tensors with FA= 0.3578, 0.7840, and 0.9623. The trace of each simulated tensor was held constant to 2.189×10^{-3} mm^2/s. The trace and FA values are representative of brain white matter (Pierpaoli et al. 1996). The SNR, defined as S_0/σ was set to 20 and $S_0 = 1000$. Three acquisition designs with different numbers of directions and diffusion weightings were selected to examine the performance of the asymptotic approximations for typical experimental designs. The designs are as follows.

1. Six directions at each $b = 0, 300, 650, 1000$ s/mm^2 (24 total measurements).

2. 16 directions at each $b = 0, 300, 650, 1000$ s/mm^2 (64 total measurements).

3. 46 directions at each $b = 0, 300, 650, 1000$ s/mm^2 (184 total measurements).

The six and 46 directions sets are icosahedral directions and the 16 direction set is constructed with the six icosahedral directions plus the dodecahedral directions (Hasan et al. 2001). The three designs were applied to each of the three tensors for a total of 9 simulation combinations. These 9 simulations were then repeated with a trace of 1.0945×10^{-3} mm^2/s to mimic conditions of acute ischemia, where trace can be reduced by as much as 50% relative to healthy white matter (Moseley et al. 1990). For each of the 18 simulations, there were 50000 simulated data sets. Data were simulated with Rician noise according to the method described by Pierpaoli and Basser (1996).

The NLSE estimates of the simulated tensors were computed with Newton's method (Koay, Chang, Carew, Pierpaoli, and Basser 2006). Trace and FA estimates were computed for each data set from the tensor estimate. The empirical distributions of the trace and FA estimates were graphically compared to a normal distribution. Sample means of the trace and FA estimates were computed and compared to their true values to check consistency. The sample variances of the estimated trace and FA were compared to the variances of their asymptotic distributions.

To investigate the dependence of the variance of trace and FA on the values of trace and FA, we simulated data sets according to Design 2. For each point on a 15×15 grid, 50000 simulated data sets were generated for equally-spaced trace and FA values. See Appendix A for a description of how to generate tensors from pre-specified trace and FA. The trace values ranged from 1.0945×10^{-3} mm^2/s to 3.2835×10^{-3} mm^2/s and the FA values ranged from 0.30 to 0.95. All other simulation parameters are as previously specified. For each trace/FA combination, the sample variances of trace and FA were computed. These variances were then plotted as a function of trace and FA.

3.7.2 Application to Human DTI Data

The methods for estimating the variance of FA were applied to a healthy human male volunteer, age 27, who provided informed consent in accordance with the guidelines of our Institutional Review Board for human subject studies. The images were acquired on a 3 Tesla scanner with a spin echo diffusion weighted imaging (DWI) pulse sequence. The following scan parameters were used: 10 axial slices, 3 mm thick, 5

mm gap, FOV: 240×240 mm, matrix size: 120×120 zero-padded to 256×256 pixels, 16 diffusion directions, TE: 72.3 ms, diffusion weightings $b = 0, 300, 650, 1000$ s/mm². The diffusion weightings and gradient direction set are identical to Design 2 of the simulation studies. The scans were cardiac gated using a pulse oximeter attached to the right index finger. The effective TR was 5 heart beats (approximately 4600 ms). No image averaging was performed, i.e., one image per gradient direction per b-value was obtained. The total scan time was 10 min 40 s. The images were first masked to eliminate pixels outside of the brain. The NLSEs of the tensors were computed at each pixel using Newton's method (Koay, Chang, Carew, Pierpaoli, and Basser 2006). Images of FA and the trace of the tensor estimates were generated. For FA, a second mask based on the trace of the estimated tensors was applied to eliminate pixels that contain isotropic cerebrospinal fluid (CSF). After masking, the variances of the trace and FA estimates were estimated according to (3.9) and (3.12), respectively. Images of $\log_{10}\{\widehat{\text{Var}}(\text{tr}\mathbf{D})\}$ and $\log_{10}\{\widehat{\text{Var}}(\text{FA}(\hat{\boldsymbol{\beta}}))\}$ were generated.

3.8   Results

The results from the simulation studies show that approximate normality is achieved for as few as six directions with four b-values for each gradient direction. Normal probability plots of trace and FA are given in Figures 14 and 15, respectively. These probability plots assess how well the empirical distributions of the trace and FA estimates follow normal distributions. If the empirical distribution perfectly followed a normal distribution, the estimates would fall on the red diagonal line. The plots of trace (Figure 14) show that the empirical distribution of the trace estimates is

very well-approximated by a normal distribution for Design 1 and each of the FA values. Departures from normality are minor and occur only in the distant tails of the distribution (probability < 0.001 and > 0.999). Trace estimates for the other designs, which use more images, are at least as well-approximated by a normal distribution and are, consequently, omitted. Figure 15 shows normal probability plots for each of the three FA values for Designs 1, 2, and 3. The plots show that the empirical distribution of the FA estimates is well-approximated by a normal distribution. Convergence to normality is slower for very high FA (0.9623). The largest departures from normality are seen in Design 1 with FA=0.9623. These departures occur in the tails of the distribution (probability < 0.05 and > 0.95). In this case, we see slightly slower convergence to normality than in other designs and lower values of FA. Moving to Design 2, the discrepancies for FA=0.9623 are in the distant tails. For FA at 0.3578 and 0.7840, the departures from normality are in the distant tails for all designs. The reduced range of the FA estimates when moving from Design 1 to other designs indicates the reduced variability that comes from acquiring more directions.

The asymptotic variances of trace and FA are very accurate. Table 4 summarizes the simulation results when the trace mimics that of healthy white matter. The sample means of the trace FA estimates are very close to the true value in all of the 18 simulations. The sample means of the trace estimates are all very close to the true value (2.189×10^{-3} mm^2/s) with discrepancies only in the third significant digit. The largest difference between the true FA and the sample means is found in Design 1 with low FA (6.7% over-estimation). Other differences are significantly lower. The closeness of the sample means to the true values of trace and FA are expected from the consistency of the NLSE. The sample variance of the trace and FA estimates,

which are good estimates of the true variances, shows that the asymptotic variances are excellent approximations. In all 18 designs, the sample variance of trace differs from the asymptotic variance by 1.61% or less. For low FA (0.3578) the asymptotic variance tends to over-estimate the true FA variance. This bias is highest for Design 1 with FA = 0.3578 where the asymptotic variance is 23.8% higher. Moving to Design 2, the over-estimation is reduced to 5.66%. The over-estimation is not considerable and will not lead to increased type I error rates. For very high FA (0.9623) there is a slight tendency for the asymptotic variance to under-estimate the true variance. This effect is minor with the largest under-estimation at less than 2.36%.

Also seen in Table 4 is that the variances of trace and FA depend of the value of FA. Variance in estimates of trace from tensors with low FA have lower variance than estimates from tensors with higher FA. We found that trace estimates from higher FA tensors are 30% or less variable than trace estimates from lower FA tensors. In the variance of estimates of FA, the opposite effect is seen and is much more dramatic. Estimates of FA from tensors with relatively high FA have lower variance than estimates from tensors with lower FA. In our simulation studies, the variances of FA estimates from tensors with a true FA of 0.3578 are approximately one order of magnitude higher than tensors with a true FA of 0.9623.

The simulation studies for the case of acute ischemia reveal similar behavior as in the case for healthy white matter (Table 5). For the variance of trace, the asymptotic variance is, on average, slightly more accurate than where trace is higher. All of the asymptotic variances have an error of less than 1.36%. For FA, the asymptotic variances are less accurate. Within the same design, tensors with the lowest FA (0.3578) have the highest errors, with the largest error at 39.7% for Design 1. This

overestimation is reduced to 13.2% and 4.32% for Designs 2 and 3, respectively.

Plots of the variance of trace and FA as a function of trace and FA (Figure 16) show graphically how the variance of these tensor-derived quantities depends on their values. These plots reiterate the findings from the simulation studies with three values of FA and two values of trace.

The results from the human data illustrate the anatomical dependence of the variances of trace and FA estimates. The SNR in the diffusion weighted images is approximately 22. The trace and FA maps for four slices are in Figures 17 and 18, respectively. The corresponding variance maps are displayed using a $\log_{10}$ scale since they vary over more than one order of magnitude. The variance of trace estimates depends on the tissue type. Comparing the FA map in Figure 18 to the trace variance image in Figure 17, we see that a region of high FA is where the variance of trace is also relatively high. The variability of FA estimates is not uniform over the whole brain and, like trace, depends on the tissue type. Lowest variances are seen in regions of high anisotropy (e.g., corpus callosum, internal capsule). The opposite effect is seen with trace and is consistent with the simulation studies.

3.9   Discussion

One important result of this study is that the NLSE of the diffusion tensor and tensor-derived quantities are consistent and asymptotically normal. The asymptotic distributions are reached as the SNR and number of DWIs increases. We described consistent estimators for the variances of trace and FA. The simulation results show that the asymptotic approximations for trace and FA are accurate. This suggests the

Table 4: Simulation Results When Trace = 2.189×10^{-3} mm^2/s (Parenthetical Quantities are Percent Error)

Design	True FA	FA Sample Mean	FA Sample Variance	FA Asymptotic Variance	Trace Sample Mean	Trace Sample Variance	Trace Asymptotic Variance
1	0.3578	0.3819	5.819×10^{-3}	7.201×10^{-3} (23.8)	2.185×10^{-3}	1.510×10^{-8}	1.513×10^{-8} (0.199)
1	0.7840	0.7854	2.026×10^{-3}	2.109×10^{-3} (4.10)	2.183×10^{-3}	1.626×10^{-8}	1.651×10^{-8} (1.54)
1	0.9623	0.9578	1.231×10^{-3}	1.202×10^{-3} (-2.36)	2.178×10^{-3}	1.953×10^{-8}	1.984×10^{-8} (1.59)
2	0.3578	0.3667	1.980×10^{-3}	2.092×10^{-3} (5.66)	2.180×10^{-3}	5.652×10^{-9}	5.667×10^{-9} (0.265)
2	0.7840	0.7837	6.008×10^{-4}	5.993×10^{-4} (-0.250)	2.178×10^{-3}	6.060×10^{-9}	6.115×10^{-9} (0.908)
2	0.9623	0.9617	1.778×10^{-4}	1.741×10^{-4} (-2.08)	2.174×10^{-3}	6.669×10^{-9}	6.767×10^{-9} (1.47)
3	0.3578	0.3603	7.007×10^{-4}	7.195×10^{-4} (2.68)	2.179×10^{-3}	1.948×10^{-9}	1.971×10^{-9} (1.18)
3	0.7840	0.7830	2.075×10^{-4}	2.057×10^{-4} (-0.867)	2.177×10^{-3}	2.119×10^{-9}	2.127×10^{-9} (0.378)
3	0.9623	0.9618	5.887×10^{-5}	5.810×10^{-5} (-1.31)	2.173×10^{-3}	2.300×10^{-9}	2.337×10^{-9} (1.61)

Table 5: Simulation Results When Trace = 1.0945×10^{-3} mm^2/s (Parenthetical Quantities are Percent Error)

| Design	True FA	FA Sample Mean	FA Sample Variance	FA Asymptotic Variance	Trace Sample Mean	Trace Sample Variance	Trace Asymptotic Variance
1	0.3578	0.4149	9.9154×10^{-3}	1.3856×10^{-2} (39.7)	1.0946×10^{-3}	1.0035×10^{-8}	9.8990×10^{-9} (-1.36)
1	0.7840	0.7915	4.2829×10^{-3}	4.5564×10^{-3} (6.39)	1.0944×10^{-3}	1.0006×10^{-8}	1.0103×10^{-8} (0.969)
1	0.9623	0.9608	2.1175×10^{-3}	2.1176×10^{-3} (0.00047)	1.0949×10^{-3}	1.0436×10^{-8}	1.0522×10^{-8} (0.824)
2	0.3578	0.3803	4.0397×10^{-3}	4.5717×10^{-3} (13.2)	1.0920×10^{-3}	3.6979×10^{-9}	3.7114×10^{-9} (0.365)
2	0.7840	0.7866	1.5569×10^{-3}	1.5775×10^{-3} (1.31)	1.0920×10^{-3}	3.7720×10^{-9}	3.7845×10^{-9} (0.331)
2	0.9623	0.9626	5.6950×10^{-4}	5.6113×10^{-4} (-1.47)	1.0909×10^{-3}	3.9010×10^{-9}	3.8971×10^{-9} (-0.01)
3	0.3578	0.3657	1.5156×10^{-3}	1.5810×10^{-3} (4.32)	1.0913×10^{-3}	1.2939×10^{-9}	1.2910×10^{-9} (-0.22)
3	0.7840	0.7847	5.4062×10^{-4}	5.4435×10^{-4} (0.67)	1.0980×10^{-3}	1.3048×10^{-9}	1.3164×10^{-9} (0.89)
3	0.9623	0.9621	1.9397×10^{-4}	1.9254×10^{-4} (-0.74)	1.0905×10^{-3}	1.3402×10^{-9}	1.3547×10^{-9} (1.08)
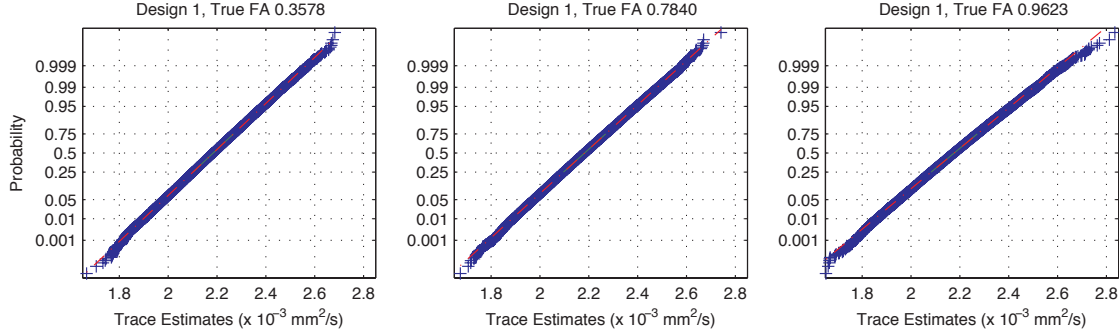
Figure 14: Each plot shows the trace estimates for 50,000 simulated data sets with Design 1 (six directions at each of four b-values). The trace estimates are very well-approximated by a normal distribution. Departures from normality are in the distant tails of the distribution ($P < 0.001$, $0.999 < P$).

utility of the asymptotic variance as the measure of variability in tensor estimates. The asymptotic approximations for trace, which is a linear function of the tensor, perform slightly better than the approximations for FA. This is expected since the use of the delta method for FA adds one additional layer of approximation. The reduction in the variance of trace and FA as the number of unique gradient encoding directions increases has been studied by Jones (2004) with Monte Carlo simulations. Our results on consistency of the estimators provide a theoretical explanation for his findings.

Unlike bootstrap approaches (Pajevic and Basser 2003) that are used to obtain estimates of the variance of FA estimates, our method does not require additional diffusion weighted images or intense computation. Since the asymptotic approximations to the variances of NLSEs of trace and FA are accurate, use of the bootstrap may be unnecessary for data collected under routine scan parameters, provided that the systematic artifacts in the DWIs can be shown to be small. However, it is best to first
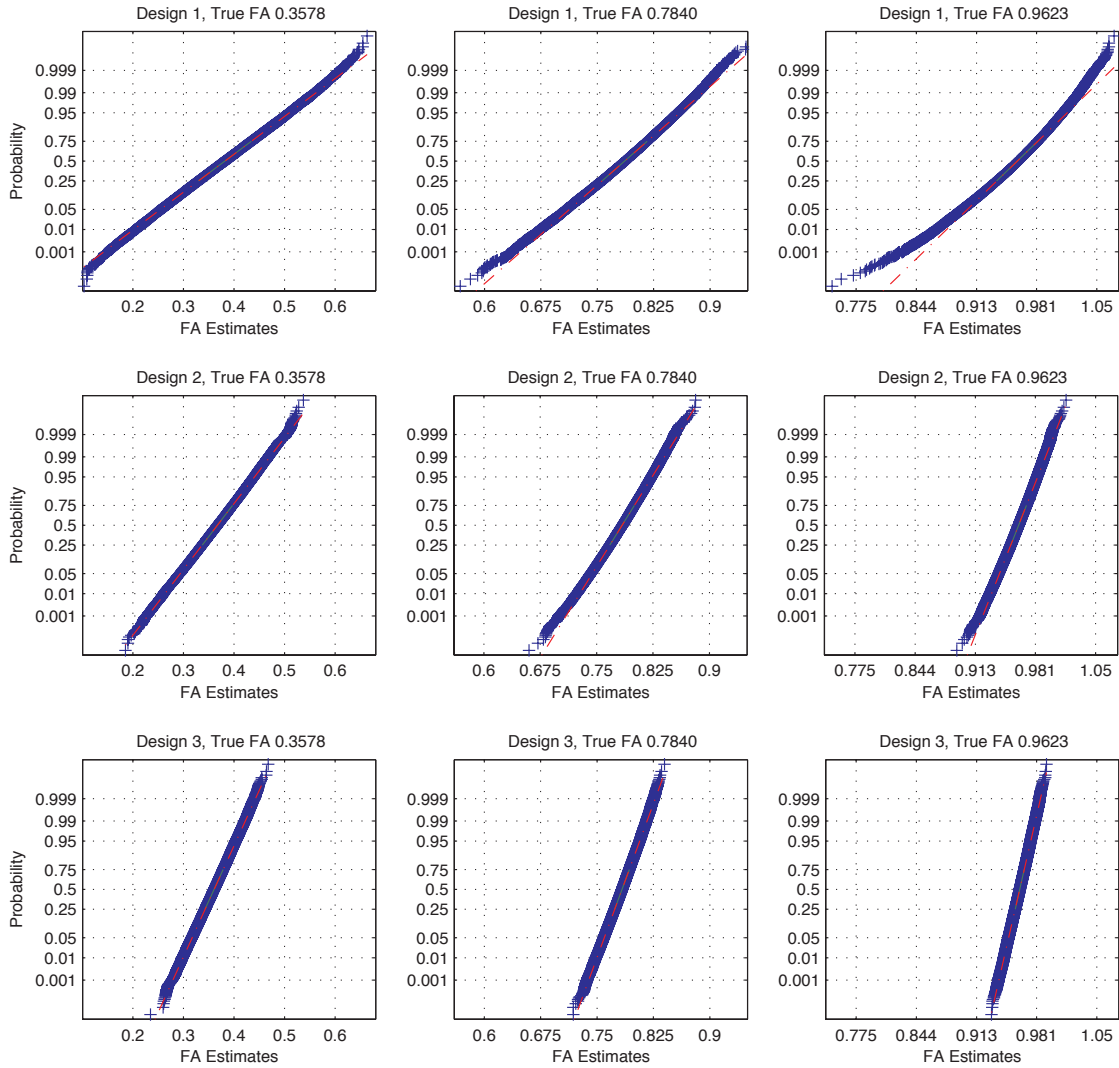
Figure 15: Each plot shows the FA estimates for 50,000 simulated data sets. Departures from normality are minor, with the largest occurring in for Design 1 (six directions at each of four *b*-values) with high FA (0.9623). The discrepancies here occur in the tails of the distribution ($P < 0.05$, $0.95 < P$). Discrepancies for other designs and FA values occur in the far tails of the distribution. Also evident from the decreased range of the FA estimates when moving from Design 1 to Designs 3 and 5 is that the variance of the estimates decreases.
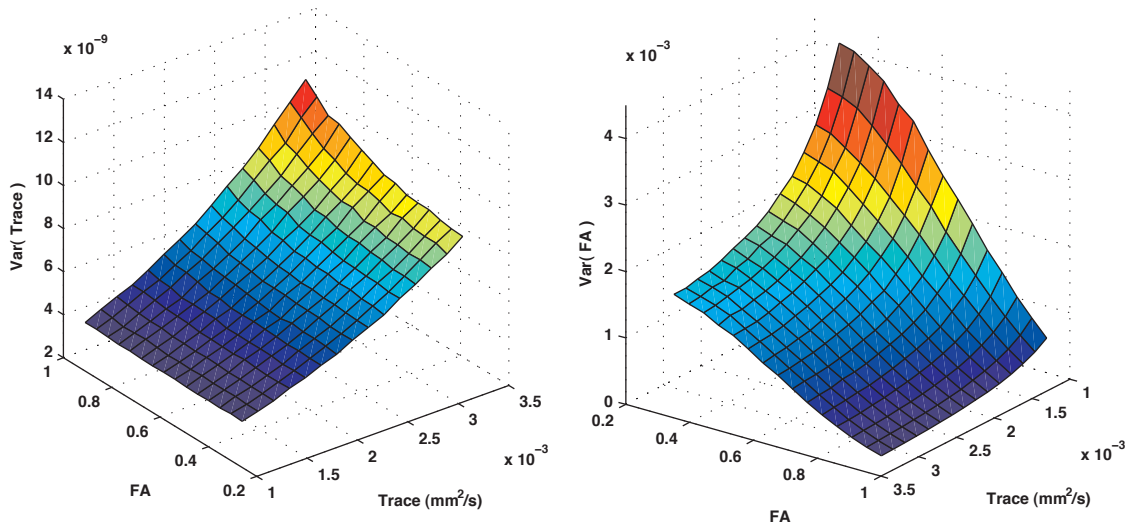
Figure 16: The variances of trace and FA depend on the values of trace and FA. These plots show sample variances of trace and FA for tensors various trace/FA combinations. Each point on the plots is the sample variance of the respective quantity, computed from 50000 data sets simulated with Design 2.
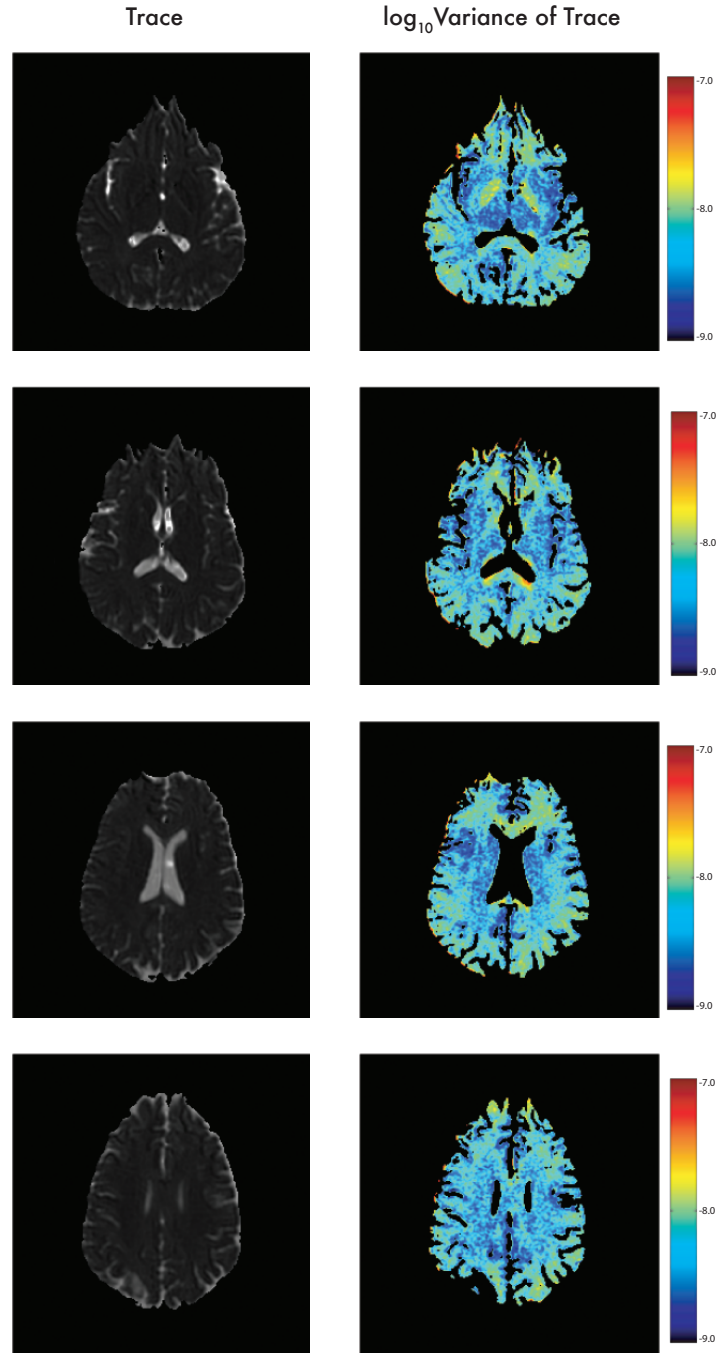
Figure 17: The variances of the trace estimates vary throughout the brain. In some of the highly anisotropic regions (e.g., corpus callosum), trace estimates have higher variance than in more isotropic regions. This is consistent with the simulation results.

Figure 18: The variance of the FA estimates also vary throughout the brain. Regions where FA is relatively high (e.g., the corpus callosum) have relatively low variance. Within white matter, the FA variances vary by approximately one order of magnitude.
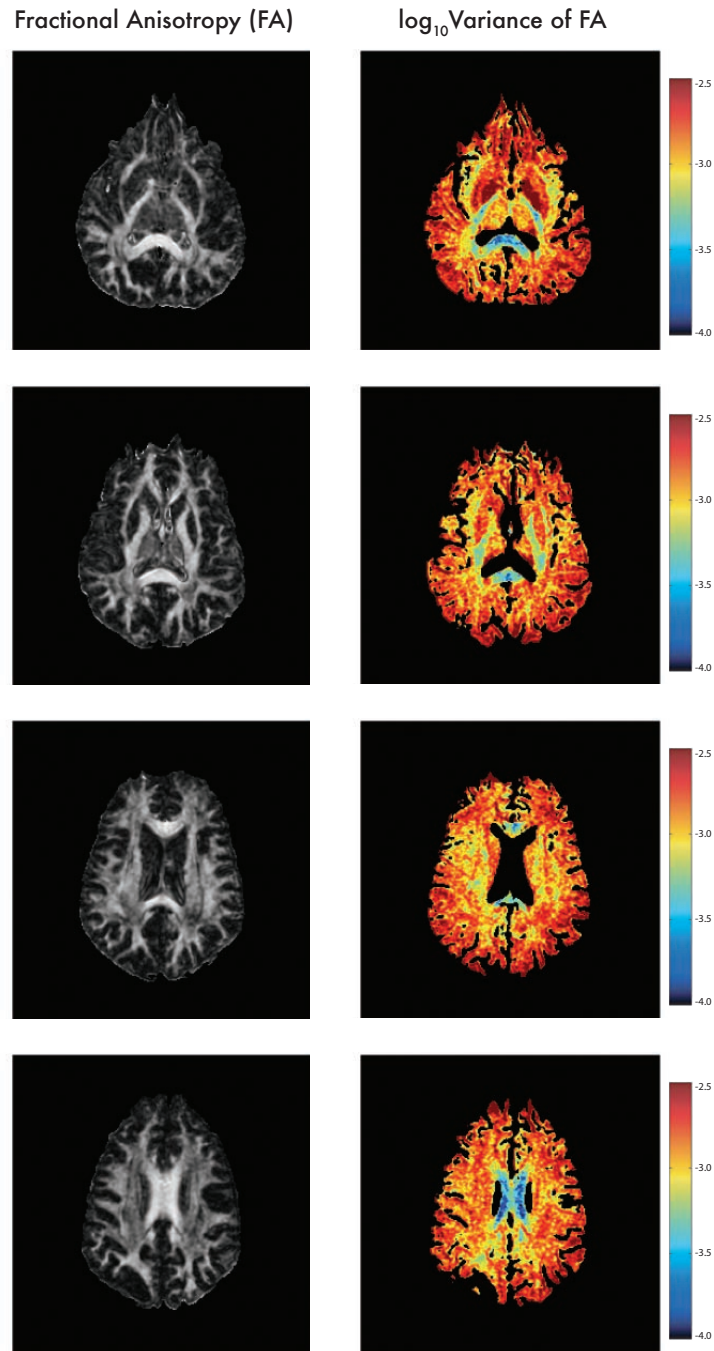
simulate data under a particular design to evaluate the usefulness of the asymptotic approximations.

It is important to note that while the noise in DWIs may be relatively constant throughout the brain, the noise in estimates of the tensor and tensor-derived quantities is not. Our simulation studies show that regions with high FA have less variable FA estimates than regions with low FA. The opposite effect is seen with the trace. When FA is high, the estimates of trace have relatively higher variance. These results are consistent with the findings of Pierpaoli and Basser (1996). The dependence of the tensor-derived quantities on their respective quantities is best illustrated in Figure 16.

The dependence of the variances of trace and FA on the values of trace and FA has important implications for testing for differences in the means of these tensor-derived quantities. The assumption of equal variance that is common to many standard statistical tests used in group analyses could be potentially violated. Suppose that one uses a t-test to test the null hypothesis that the mean of FA in one group is the same as the mean FA in another group. This test is commonly used for both voxel-based morphometry (VBM) and ROI-based group analyses. Jones et al. (2005) cites several VBM studies of tensor-derived quantities in the context of examining the effect of the filter size. When the null hypothesis is true, the variances in each group are equal. When the alternative hypothesis, that the means are unequal, is true, then the variances of the FA measurements will differ in the two groups. This aspect of trace and FA makes estimating the null distribution, which is necessary for comparing the test statistic, impossible. The effect of the variance heterogeneity under the alternative hypothesis may be substantial. Variances of trace and FA differ

considerably over white matter. For example, the variance of FA varies by roughly one order of magnitude (Figure 18). More work is needed to evaluate the impact of unequal variances on the inferences that are made with simple testing procedures.

Combining estimates of the variances of trace or FA measurements with a weighted linear model may provide one solution to the problem of unequal variances. Take, for example, testing the mean of FA between two groups. We have a set of n_1 estimates from group 1, $\mathrm{FA}_{11}, \ldots, \mathrm{FA}_{1n_1}$, and a set of n_2 estimates from group 2, $\mathrm{FA}_{21}, \ldots, \mathrm{FA}_{2n_2}$. Corresponding to these FA measurements are estimates of their variances, denoted $\xi_{11}, \ldots, \xi_{1n_1}, \xi_{21}, \ldots, \xi_{2n_2}$. These variance estimates can be computed with our estimator in equation (3.12) or other methods such as the bootstrap. Then choose the weights $w_1, \ldots, w_{n_1+n_2}$ as $1/\xi_{11}, \ldots, 1/\xi_{1n_1}, 1/\xi_{21}, \ldots, 1/\xi_{2n_2}$, respectively. Define $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$ as the model parameter, where α_0 is the mean FA in group 1 and α_1 is the difference in the means between groups 1 and 2. The parameter $\boldsymbol{\alpha}$ is estimated as

$$\hat{\boldsymbol{\alpha}} := \arg\min_{\boldsymbol{\alpha}} \sum_{i=1}^{n_1+n_2} w_i (\mathrm{FA}_i - \alpha_0 - y_i \alpha_1)^2,$$

where y_i has a value of 1 when the ith observation is from group 2 and otherwise has a value of 0. Then, testing the null hypothesis that the mean FA of group 1 is equal to the mean FA of group 2 is equivalent to testing that $\alpha_1 = 0$. Provided that the variance estimates are good, the residuals from the linear model fit should have zero mean and approximately constant variance. This allows valid testing of the null hypothesis, that $\alpha_1 = 0$, with a t-test. This type of weighted least squares estimation and testing can be performed in many standard statistical packages. More general linear models of FA or other tensor-derived quantities that are fit with weighted least squares are possible. These models may include other factors or covariates.

One limitation of our method is that it assumes the errors in the diffusion weighted images are independent, normal, and have constant variance. The independence assumption is satisfied by the independent acquisition of the DWIs themselves. Ideally, for high SNR, the normality and constant variance assumptions are satisfied. At high values of b ($\gg$ 1000 s/mm^2), which cause lower SNR, these assumptions can break down. For most routine DTI scans, the SNR is at least 20 and, consequently, normality is a good assumption. Head motion, eddy current distortion, susceptibility effects, and other systematic artifacts, as well as subtle physiologic noise may also cause violations in these assumptions.

Rician noise has the unique property that the variance of noise in magnitude MR images depends on the mean of the signal even when the noise in the quadrature channels is constant. This is a source of non-constant variance in diffusion-weighted images since directions along the major axis of axons will have lower mean signal than from directions perpendicular to the major axis. The effect of non-constant variance is only substantial for very low SNR (Koay and Basser 2006) and is not a significant issue for the imaging parameters considered in our study. Nevertheless, our methods can be easily extended to situations where non-constant variance is an issue. The noise covariance matrix, $\sigma^2\mathbf{I}_n$, needs to be replaced with a matrix of the form $\sigma^2\mathbf{w}\mathbf{I}_n$, where $\mathbf{w}$ is a vector of weights, which depend on the SNR of each measurement. Koay and Basser (2006) derived an analytical expression for the weights. The tensor is then estimated iteratively by weighted nonlinear least squares.

Our method can be extended to other functions of the diffusion tensor and to study design of DTI experiments. Other functions of the tensor that are commonly used include the direction of the primary eigenvector of the tensor and relative anisotropy.

Methods for testing for group differences in the direction of the primary eigenvector have been described (Schwartzman et al. 2005). The issue of the affect of experimental design on the variance of tensor-derived quantities has been studied with Monte Carlo methods (Jones 2004; Papadakis et al. 1999; Skare et al. 2000). The asymptotic variances can also be used as a benchmark for comparing designs of DTI experiments.

Finally, Salvador et al. (2005) proposed an estimator for the variance of mean ADC based on the linearized tensor model. They did not consider an estimator for the variance of FA. However, by applying the delta method, an estimator for the variance of FA can also be obtained from the linearized tensor model. By substituting the inverse of the Fisher information matrix in equation (3.12) with the estimator for the variance of the tensor from Salvador et al. (2005), one obtains an estimator for the variance of FA from the linearized tensor model. It is unknown how the performance of such an estimator compares to our estimator in equation (3.12). Since the NLSE outperforms the linearized tensor estimator (Koay, Carew, Alexander, Basser, and Meyerand 2006), we suspect that the FA variance estimator from the NLSE will outperform the FA variance estimator for the linearized model. This requires closer examination.

Appendix A

Construct a tensor with a known FA and trace

In this appendix, we consider the problem of how to construct a cylindrically symmetric tensor that has a pre-specified FA and trace. This is useful for simulation studies where one wants to generate tensors. Trace and FA alone do not uniquely specify a cylindrically symmetric tensor. Additional information about the orientation is needed. Since it is possible to apply a rotation to the tensor to obtain any desired orientation, we will assume that the major axis of the tensor coincides with the x-axis. Denote the three eigenvalues of the tensor in descending order with λ_1, λ_2, and λ_3. Under cylindrical symmetry, $\lambda_2 = \lambda_3$. When the major axis of the tensor $\mathbf{D}$ is coincident with the x-axis, the (at most) six unique elements of the diffusion tensor are written as $(D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{yz}, D_{xz}) = (\lambda_1, \lambda_2, \lambda_2, 0, 0, 0)$. Thus, specifying the cylindrically symmetric tensor oriented along the x-axis from trace and FA is equivalent to specifying λ_1 and λ_2. For the remainder of this appendix, when we refer to the tensor, we are referring to a cylindrically symmetric tensor that is oriented along the x-axis.

We now consider the mathematical relationship between λ_1, λ_2 and trace, FA. By

cylindrical symmetry,

$$\mathrm{tr}\mathbf{D} = \lambda_1 + 2\lambda_2. \tag{A.1}$$

Using the previous equation, we can express FA as follows,

$$\frac{2}{3}\mathrm{FA}^2 = \frac{(\lambda_1 - (1/3)\mathrm{tr}\mathbf{D})^2 + 2[(1/2)(\mathrm{tr}\mathbf{D} - \lambda_1) - (1/3)\mathrm{tr}\mathbf{D}]^2}{\lambda_1^2 + (1/2)(\mathrm{tr}\mathbf{D} - \lambda_1)^2}. \tag{A.2}$$

From the previous two equations, we see that we have a system of two equations and two unknowns, namely λ_1 and λ_2. We can solve the quadratic equation (A.2) for λ_1. Writing (A.2) in the form $a\lambda_1^2 + b\lambda_1 + c = 0$, we see that

$$a = \mathrm{FA}^2 - 3/2,$$

$$b = \mathrm{tr}\mathbf{D}[1 - (2/3)\mathrm{FA}^2], \quad \text{and}$$

$$c = (\mathrm{tr}\mathbf{D})^2[(1/3)\mathrm{FA}^2 - 1/6].$$

Solving for λ_1 gives two roots. One solution is greater than the other. The larger of the two solutions gives a prolate tensor and satisfies the ordering $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The other solution results in an oblate tensor. Since diffusion of non-crossing axons is represented by a prolate tensor, we select the larger solution for λ_1. To get the value of λ_2, substitute the solution for λ_1 in equation (A.1) and solve for λ_2.

Bibliography

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans Amer Math Soc 68*, 337–404.

Barna, P. S. (1969). *Fluid mechanics for engineers*. New York: Plenum Press.

Basser, P. J., J. Mattiello, and D. Lebihan (1994a). Estimation of the effective self-diffusion tensor from the nmr spin echo. *J Magn Reson B 103*, 247–254.

Basser, P. J., J. Mattiello, and D. Lebihan (1994b). Mr diffusion tensor spectroscopy and imaging. *Biophys J 66*, 259–267.

Basser, P. J. and S. Pajevic (2003). Dealing with uncertainty in diffusion tensor mr data. *Israel J Chem 43*, 129–144.

Bullmore, E., M. Brammer, S. Williams, S. Rabe-Hesketh, N. Janot, A. David, J. Mellers, R. Howard, and P. Sham (1996). Statistical methods of estimation and inference for functional mr image analysis. *Magn Reson Med 35*, 261–277.

Canny, J. (1986). A computational approach to edge detection. *IEEE J Pattern Analysis Machine Intelligence 8*, 679–698.

Carew, J. D., R. Dalal, G. Wahba, and S. B. Fain (2003). Estimating arterial wall shear stress. Technical Report 1088, University of Wisconsin Department of Statistics.

Chatfield, C. (1996). *The analysis of time series*. Chapman & Hall.

Chiang, A., G. Wahba, J. Tribbia, and D. Johnson (1999). A quantitative study of smoothing spline-anova based fingerprint methods for attribution of global warming. Technical Report 1010, University of Wisconsin Department of Statistics.

Cox, R. W. (1996). Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Comp Biomed Res 29*, 162–173.

Craven, P. and G. Wahba (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer Math 31*, 377–403.

de Boor, C. (1978). *A practical guide to splines.* New York: Springer-Verlag.

Friedman, M. H., C. B. Bargeron, O. J. Deters, G. M. Hutchins, and F. F. Mark (1987). Correlation between wall shear and intimal thickness at a coronary artery branch. *Atherosclerosis 68*, 27–33.

Friston, K. J., A. P. Holmes, J.-B. Poline, P. J. Grasby, S. C. R. Williams, R. S. J. Frackowiak, and R. Turner (1995). Analysis of fmri time-series revisited. *NeuroImage 2*, 45–53.

Friston, K. J., P. Jezzard, and R. Turner (1994). Analysis of functional mri time-series. *Hum Brain Mapp 1*, 153–171.

Friston, K. J., O. Josephs, E. Zarahn, A. P. Holmes, S. Rouquette, and J.-B. Poline (2000). To smooth or not to smooth? *NeuroImage 12*, 196–208.

Green, P. J. and B. W. Silverman (1994). *Nonparametric regression and generalized linear models. A roughness penalty approach.* London: Chapman & Hall.

Gu, C. (1989). Rkpack and its applications: fitting smoothing spline models. Technical Report 857, University of Wisconsin Department of Statistics.

Gudbjartsson, H. and S. Patz (1995). The rician distribution of noisy mri data. *Magn Reson Med 16*, 87–90.

Hasan, K. M., D. L. Parker, and A. L. Alexander (2001). Comparison of gradient encoding schemes for diffusion-tensor mri. *J Magn Reson Imag 13*, 769–780.

Henkelman, R. M. (1985). Measurement of signal intensities in the presence of noise in mr images. *Med Phys 12*, 232–233.

Jones, D., M. R. Symms, M. Cercignani, and R. J. Howard (2005). The effect of filter size on vbm analyses of dt-mri data. *NeuroImage 26*, 546–554.

Jones, D. K. (2004). The effect of gradient sampling schemes on measures derived from diffusion tensor mri: a monte carlo study. *Magn Reson Med 51*, 807–815.

Kimeldorf, G. and G. Wahba (1971). Some results on tchbycheffian spline functions. *J Math Anal Appl 33*, 82–95.

Koay, C. G. and P. J. Basser (2006). Analytically exact correction scheme for signal extraction from noisy magnitude mr signals. *J Magn Reson 179*, 477–482.

Koay, C. G., J. D. Carew, A. L. Alexander, P. J. Basser, and M. E. Meyerand (2006). Investigation of anomalous estimates of tensor-derived quantities in diffusion tensor imaging. *Magn Reson Med 55*, 930–936.

Koay, C. G., L.-C. Chang, J. D. Carew, C. Pierpaoli, and P. J. Basser (2006).

A unifying theoretical and algorithmic framework for least squares methods of estimation in diffusion tensor imaging. *J Magn Reson 182*, 115–125.

Ku, D. N., D. P. Giddens, C. K. Zarins, and S. Glagov (1985). Pulsatile flow and atherosclerosis in the human carotid bifurcation. positive correlation between plaque location and low oscillating shear stress. *Atherosclerosis 5*, 293–302.

Lang, N., S. C. Strother, J. R. Anderson, F. A. Nielson, A. P. Holmes, T. Kolenda, R. Savoy, and L. K. Hansen (1999). Plurality and resemblance in fmri data analysis. *NeuroImage 10*, 282–303.

Lebihan, D., J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat (2001). Diffusion tensor imaging: concepts and applications. *J Magn Reson Imag 13*, 534–546.

Lehmann, E. L. and G. Casella (1998). *Theory of point estimation* (2nd ed.). New York: Springer-Verlag.

Lowe, M. J. and J. A. Sorenson (1997). Spatially filtering functional magnetic resonance imaging data. *Magn Reson Med 37*, 723–729.

Matérn, B. (1986). *Spatial Variation* (2nd ed.). Berlin: Springer-Verlag.

Moseley, M. E., J. Kucharczyk, J. Mintorovitch, Y. Cohen, J. Kurhanewicz, N. Derugin, H. Asgari, and D. Norman (1990). Diffusion-weighted mr imaging of acute stroke: correlation with t2-weighted and magnetic susceptibility-enhanced mr imaging in cats. *Am J Neuroradiol 11*, 423–429.

Motomiya, M. and T. Karino (1984). Flow patterns in the human carotid artery bifurcation. *Stroke 20*, 119–127.

Oyre, S., S. Ringgaard, S. Kozerke, W. P. Paaske, M. B. Scheidegger, P. Boesiger, and E. M. Pedersen (1998). Quantitation of circumferential subpixel vessel wall position and wall shear stress by multiple sectored three dimensional paraboloid modeling of velocity encoded cine mr. *Magn Reson Med 40*, 645–655.

Pajevic, S. and P. J. Basser (2003). Parametric and non-parametric statistical analysis of dt-mri data. *J Magn Reson 161*, 1–14.

Papadakis, N. G., D. Xing, G. C. Houston, J. M. Smith, M. I. Smith, M. F. James, A. A. Parsons, C. L. Huang, L. D. Hall, and T. A. Carpenter (1999). A study of rotationally invariant and symmetric indices of diffusion anisotropy. *Magn Reson Imaging 17*, 881–892.

Pierpaoli, C. and P. J. Basser (1996). Toward a quantitative assessment of diffusion anisotropy. *Magn Reson Med 36*, 893–906.

Pierpaoli, C., P. Jezzard, P. J. Basser, A. Barnett, and G. Chiro (1996). Diffusion tensor mr imaging of the human brain. *Radiology 201*, 637–648.

Salvador, R., A. Pena, D. K. Menon, T. A. Carpenter, J. D. Pickard, and E. T. Bullmore (2005). Formal characterization and extension of the linearized diffusion tensor model. *Hum Brain Mapp 24*, 144–155.

Schwartzman, A., R. F. Dougherty, and J. E. Taylor (2005). Cross-subject comparison of principal diffusion direction maps. *Magn Reson Med 53*, 1423–1431.

Skare, S., M. Hedehus, M. E. Moseley, and T. Q. Li (2000). Condition number as a measure of noise performance or diffusion tensor data acquisition schemes with mri. *J Magn Reson 147*, 340–352.

Stein, M. L. (1999). *Interpolation of spatial data: some theory for kriging.* New York: Springer-Verlag.

Wahba, G. (1983). Bayesian "confidence intervals" for the cross-validated smoothing spline. *J Roy Stat Soc Ser B 45*, 133–150.

Wahba, G. (1990). *Spline models for observational data.* Philadelphia: SIAM.

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

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

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

Zarahn, E., G. K. Aguirre, and M. D'Esposito (1997). Empirical analysis of bold fmri statistics. i. spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage 5*, 179–197.

Zarins, C. K., D. P. Giddens, B. K. Bharadvaj, V. S. Sottiurai, R. F. Mabon, and S. Glagov (1983). Carotid bifurcation atherosclerosis. quantitative correlation of

plaque localization with flow velocity profiles and wall shear stress. *Circ Res 53*, 502–514.