

DEPARTMENT OF STATISTICS  
University of Wisconsin  
1210 West Dayton St.  
Madison, WI 53706

TECHNICAL REPORT NO. 1058

June 25, 2002

## Optimal Spline Smoothing of fMRI Time Series by Generalized Cross-Validation <sup>1</sup>

John D. Carew,<sup>2</sup> Grace Wahba,<sup>3</sup> Xianhong Xie,<sup>3</sup> Erik V. Nordheim  
Department of Statistics, University of Wisconsin, Madison, WI

M. Elizabeth Meyerand <sup>4</sup>  
Department of Medical Physics, University of Wisconsin, Madison, WI

**Key Words and Phrases:** functional MRI, temporal autocorrelation, general linear model, smoothing spline, generalized cross-validation

---

<sup>1</sup>Corresponding author address: John Carew, Department of Statistics, University of Wisconsin, 1210 W. Dayton St., Madison, WI 53706, jcarew@stat.wisc.edu.

<sup>2</sup>Research supported in part by NIH grant T32 EY07119.

<sup>3</sup>Research supported in part by NSF grant DMS0072292 and NIH grant EY09946.

<sup>4</sup>Research supported in part by a grant from the Whitaker Foundation.

## ABSTRACT

Linear parametric regression models of fMRI time series have correlated residuals. One approach to address this problem is to condition the autocorrelation structure by temporal smoothing. Smoothing splines with the degree of smoothing selected by generalized cross-validation (GCV-Spline) provide a method to find an optimal smoother for an fMRI time series. The purpose of this study was to determine if GCV-Spline of fMRI time series yields unbiased variance estimates of linear regression model parameters. GCV-Spline was evaluated with a real fMRI data set and bias of the variance estimator was computed for simulated time series with autocorrelation structures derived from fMRI data. The results from the real data suggest that GCV-Spline determines appropriate amounts of smoothing. The simulations show that the variance estimates are, on average, unbiased. This study demonstrates that GCV-Spline is an appropriate method for smoothing fMRI time series.

# 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 paper 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 paper 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 paper 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.

## 2 Materials and Methods

### 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),

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{K}\epsilon, \tag{1}$$

where  $\mathbf{y} = (y_1, \dots, y_n)^T$  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,  $\beta$  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) to give

$$\mathbf{S}\mathbf{y} = \mathbf{S}\mathbf{X}\beta + \mathbf{S}\mathbf{K}\epsilon. \quad (2)$$

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  $\beta$  in equation (2) given by  $\hat{\beta} = (\mathbf{S}\mathbf{X})^+ \mathbf{S}\mathbf{y}$  where  $^+$  denotes the pseudoinverse that satisfies the Moore-Penrose conditions, 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  $\beta$ . 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) 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.

## 2.2 Smoothing Splines

Smoothing splines model the observed time series  $y_i$  as

$$y_i = f(t_i) + \epsilon_i, \quad (3)$$

where  $f$  is a smooth function,  $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ , and  $t_i$  for  $i = 1, \dots, n$  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

$$f(\hat{t}_i) = \arg \min_{f \in C^2[t_1, t_n]} \left( \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda \int_{t_1}^{t_n} (f''(x))^2 dx \right). \quad (4)$$

The unique solution of (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 (4). When  $\lambda = \infty$ , (4) is a linear approximation. When  $\lambda = 0$ , i.e., no smoothing, (4) interpolates the  $y_i$  with a piecewise cubic polynomial.

The solution of (4) is computed with linear algebra. The penalty function in equation (4), namely

$$S(f) = \sum_{i=1}^n (y_i - f(t_i))^2 + \lambda \int_{t_1}^{t_n} (f''(x))^2 dx, \quad (5)$$

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  $i = 1, \dots, n$  and  $j = 2, \dots, n - 1$ , 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

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

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 (5):

$$\int_{t_1}^{t_n} f''(x)^2 dx = \mathbf{f}^T \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{f}. \quad (6)$$

Substitution simplifies equation (5) to

$$S(f) = (\mathbf{y} - \mathbf{f})^T (\mathbf{y} - \mathbf{f}) + \lambda \mathbf{f}^T \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T \mathbf{f}. \quad (7)$$

It is now easy to show that the minimum of (7) is

$$\hat{\mathbf{f}} = (\mathbf{I} + \lambda \mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T)^{-1} \mathbf{y}. \quad (8)$$

The estimator in equation (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

$$\mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T = \mathbf{\Gamma} \mathbf{D} \mathbf{\Gamma}^T, \quad (9)$$

where the orthogonal matrix  $\mathbf{\Gamma}$  contains the eigenvectors and

$$\mathbf{D} = \begin{bmatrix} 0 & & \cdots & 0 \\ & 0 & & \vdots \\ & & l_1 & \\ \vdots & & & \ddots \\ 0 & \cdots & & & l_{n-2} \end{bmatrix} \quad (10)$$

contains the  $n - 2$  positive eigenvalues. Substituting this factorization for  $\mathbf{Q} \mathbf{R}^{-1} \mathbf{Q}^T$  in equation (8) gives

$$\hat{\mathbf{f}} = (\mathbf{I} + \lambda \mathbf{\Gamma} \mathbf{D} \mathbf{\Gamma}^T)^{-1} \mathbf{y}. \quad (11)$$

Since  $\Gamma$  is orthogonal (i.e.,  $\Gamma^T = \Gamma^{-1}$ ), equation (11) can be simplified to

$$\hat{\mathbf{f}} = \Gamma(\mathbf{I} + \lambda\mathbf{D})^{-1}\Gamma^T\mathbf{y}. \quad (12)$$

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

$$(\mathbf{I} + \lambda\mathbf{D})^{-1} = \begin{bmatrix} 1 & & \dots & 0 \\ & 1 & & \vdots \\ & & \frac{1}{1+\lambda l_1} & \\ \vdots & & & \ddots \\ 0 & \dots & & \frac{1}{1+\lambda l_{n-2}} \end{bmatrix}. \quad (13)$$

Thus, one need only perform the eigenvector-eigenvalue decomposition in equation (9) once for an entire fMRI data set since  $\mathbf{QR}^{-1}\mathbf{Q}^T$  depends only on the time between scans (i.e., the sample rate). Then, for each spline fit, only  $\lambda$  need be changed in equation (13). This allows for very fast spline fits for different  $\lambda$ . This feature is important for GCV since a spline must be fit for many different degrees of smoothing.

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

$$GCV(\lambda) = \frac{1}{n} \cdot \frac{\sum_{i=1}^n (y_i - f(\hat{t}_i))^2}{(1 - n^{-1}\text{tr}\mathbf{A}(\lambda))^2}. \quad (14)$$

The matrix  $\mathbf{A}(\lambda)$  is the hat matrix for a given  $\lambda$ . 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.

### 2.3 Bias of the Variance Estimator for Spline Smoothing

The smoothing spline from (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 (2). The representation of the smoothing matrix is given directly from the solution to the penalty function. Moreover, the smoothing matrix is

$$\mathbf{S} = \Gamma(\mathbf{I} + \lambda\mathbf{D})^{-1}\Gamma^T. \quad (15)$$

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):

$$\text{var}(\mathbf{c}^T\hat{\beta}) = \sigma^2\mathbf{c}^T(\mathbf{SX})^+\mathbf{SVS}^T(\mathbf{SX})^{+T}\mathbf{c} \quad (16)$$

and

$$\begin{aligned} \text{bias}(\mathbf{S}, \mathbf{V}) &= \frac{\text{var}(\mathbf{c}^T\hat{\beta}) - \widehat{\text{E}[\text{var}(\mathbf{c}^T\hat{\beta})]}}{\text{var}(\mathbf{c}^T\hat{\beta})} \\ &= 1 - \frac{\text{tr}[\mathbf{LSVS}^T]\mathbf{c}^T(\mathbf{SX})^+\mathbf{SV}_a\mathbf{S}^T(\mathbf{SX})^{+T}\mathbf{c}}{\text{tr}[\mathbf{LSV}_a\mathbf{S}^T]\mathbf{c}^T(\mathbf{SX})^+\mathbf{SVS}^T(\mathbf{SX})^{+T}\mathbf{c}}, \end{aligned} \quad (17)$$

where  $\mathbf{L} = \mathbf{I} - \mathbf{S}\mathbf{X}(\mathbf{S}\mathbf{X})^+$  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  $\sigma^2$  with its estimate

$$\hat{\sigma}^2 = \frac{(\mathbf{L}\mathbf{S}\mathbf{y})^T \mathbf{L}\mathbf{S}\mathbf{y}}{\text{tr}(\mathbf{L}\mathbf{V}_a)}, \quad (18)$$

given in Worsley and Friston (1995).

## 2.4 Relationship to Partial Spline Model

In this paper  $\mathbf{A}(\lambda)$ , the smoother matrix associated with the minimization problem of (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:

$$y_i = \sum_{\nu=1}^p \phi_{\nu}(t_i) \beta_{\nu} + f(t_i) + \epsilon_i \quad (19)$$

where the  $\phi_{\nu}$  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 (5), or, alternatively as a zero mean Gaussian stochastic process with some multiple of a particular covariance that is related to  $(\mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}^T)^+$ , (Wahba, 1983, Wahba, 1990, Chiang *et al.*, 1999, and elsewhere). In the partial spline paradigm, one finds  $\beta$  and  $f$  to minimize

$$\tilde{S}(f, \beta) = \sum_{i=1}^n \left( y_i - \left[ \sum_{\nu=1}^p \phi_{\nu}(t_i) \beta_{\nu} + f(t_i) \right] \right)^2 + \lambda \int_{t_1}^{t_n} (f''(x))^2 dx \quad (20)$$

and  $\lambda$  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  $\mathbf{X}\tilde{\beta} + \tilde{\mathbf{f}} = \tilde{\mathbf{A}}(\lambda)\mathbf{y}$ . In this paradigm the  $\hat{\beta}$  which minimizes (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  $\sigma^2$  and hypothesis tests for  $\beta$  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.

## 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<sup>2</sup> pixel matrix with in-plane resolution of 3.75mm<sup>2</sup>, 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 (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  $\mathbf{t} = (t_1, t_2, \dots, t_n)^T$ . An optimal spline smoothing matrix was computed for each time series with equation (12). The optimal  $\lambda =: \lambda_{\text{opt}}$  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

$$t = \frac{\mathbf{c}^T \hat{\beta}}{\sqrt{\text{var}(\mathbf{c}^T \hat{\beta})}}. \quad (21)$$

## 2.6 Bias Computations with Simulated Data

Bias of the variance estimator can be computed with equation (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

$$\mathbf{r} = \mathbf{L}\mathbf{y}. \quad (22)$$

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

$$r_i = b_1 r_{i-1} + \dots + b_8 r_{i-8} + \zeta_i, \quad (23)$$

where  $b_1, \dots, b_8$  are the AR coefficients and  $\zeta_i \sim \mathcal{N}(0, \sigma_r^2)$ . The more convenient matrix representation of equation (23) is, following Friston *et al.* (2000)

$$\mathbf{r} = (\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\zeta}, \quad (24)$$

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

$$\hat{\mathbf{B}} = \begin{bmatrix} 0 & 0 & 0 & 0 & \cdots & 0 \\ \hat{b}_1 & 0 & 0 & 0 & \cdots & 0 \\ \hat{b}_2 & \hat{b}_1 & 0 & 0 & \cdots & 0 \\ \hat{b}_3 & \hat{b}_2 & \hat{b}_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \hat{b}_8 & \hat{b}_7 & \hat{b}_6 & \hat{b}_5 & \cdots & 0 \\ 0 & \hat{b}_8 & \hat{b}_7 & \hat{b}_6 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (25)$$

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

$$\tilde{\mathbf{K}} = (\mathbf{I} - \hat{\mathbf{B}})^{-1}. \quad (26)$$

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

$$\tilde{\mathbf{y}} = 0.15 \cdot \mathbf{s} + \tilde{\mathbf{K}}\mathbf{e}, \quad (27)$$

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 Lange *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 (27). The simulated time series were assigned to the spatial location where the  $\tilde{\mathbf{K}}$  was estimated. This preserves the 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  $\lambda$  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 (16) and (17), respectively. The known variance-covariance matrix  $\mathbf{V} = \tilde{\mathbf{K}} \tilde{\mathbf{K}}^T$  and the assumed variance-covariance matrix  $\mathbf{V}_a = \mathbf{S} \mathbf{S}^T$  with  $\mathbf{S}$  given by the particular smoothing matrix under investigation.

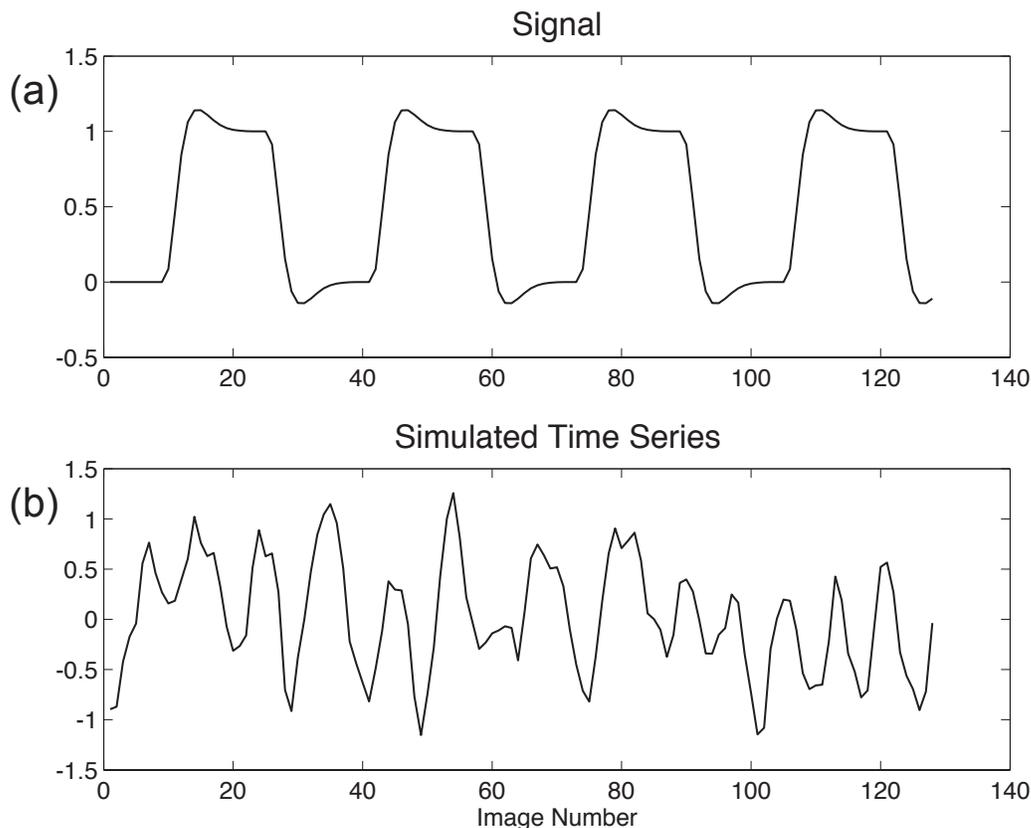


Figure 1: The hypothesized BOLD response to the photic stimulation in the fMRI experiment (a). A simulated time series used in the bias computations (b).

## 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](http://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.

## 3 Results

### 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 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

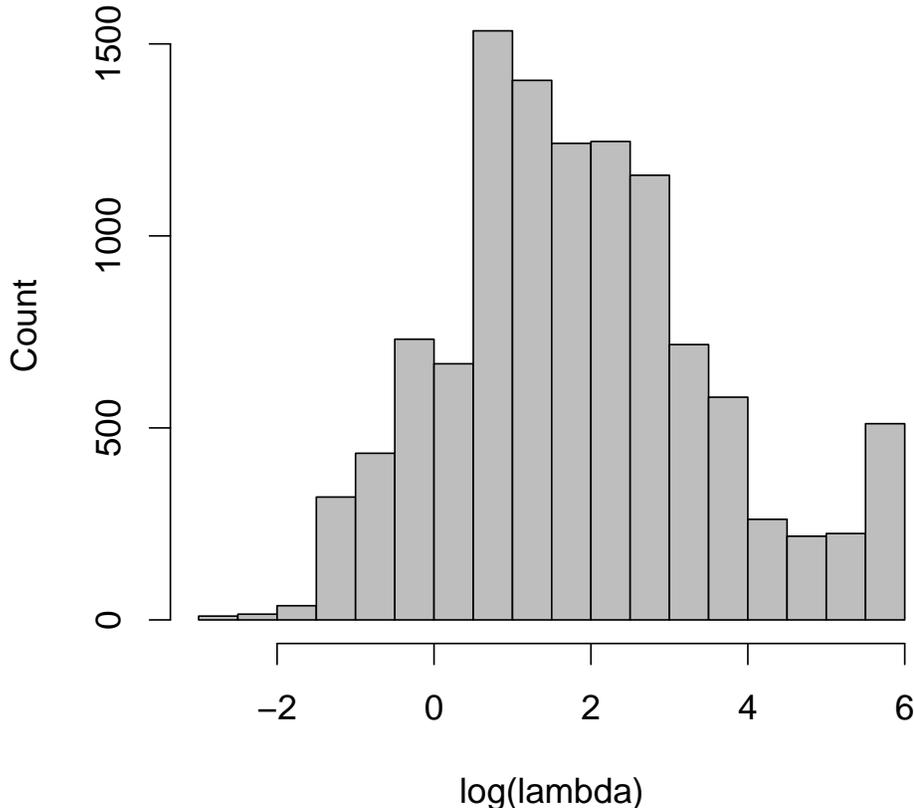


Figure 2: Histogram of  $\log_{10}(\lambda_{\text{opt}})$  from the visual stimulus experiment.

search. The amount of optimal smoothing with the GCV-spline method is illustrated in Figure 2. The distribution of the optimal  $\lambda$  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  $\text{median}(\lambda) = 50.01$ ). 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.

### 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  $\lambda$  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  $\text{median}(\lambda) = 398.1$ ) 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 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).

**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	$-1.800 \times 10^{-5}$	0.1801	0.5964
Median Bias	-0.0332	0.2312	0.6627

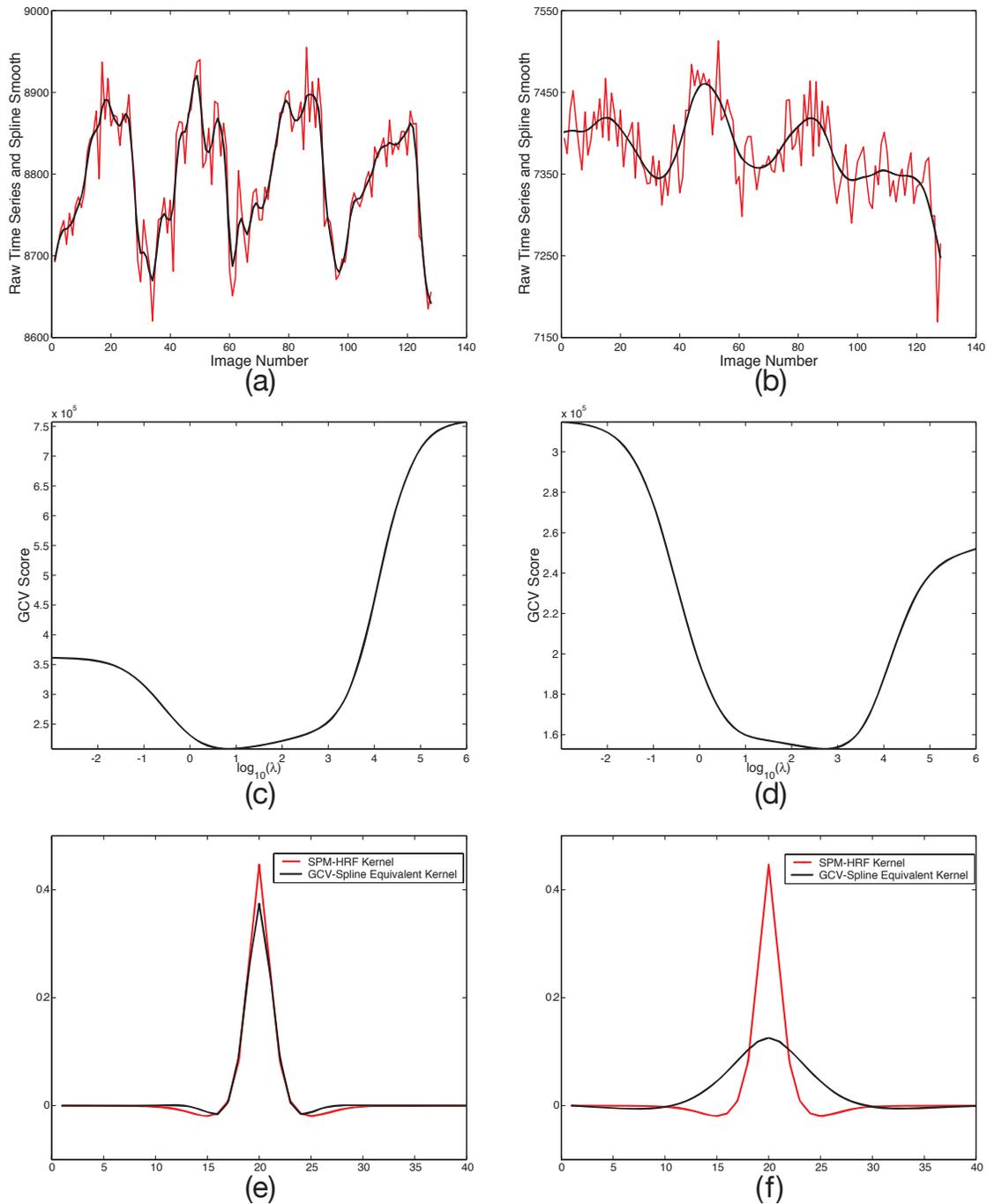


Figure 3: Spline smoothing of fMRI time series with the degree of smoothing selected by generalized cross-validation. The time series in (a) required approximately the same amount of smoothing as is provided by the SPM-HRF smoothing kernel. The plot of corresponding GCV score (c) is well-behaved with 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.

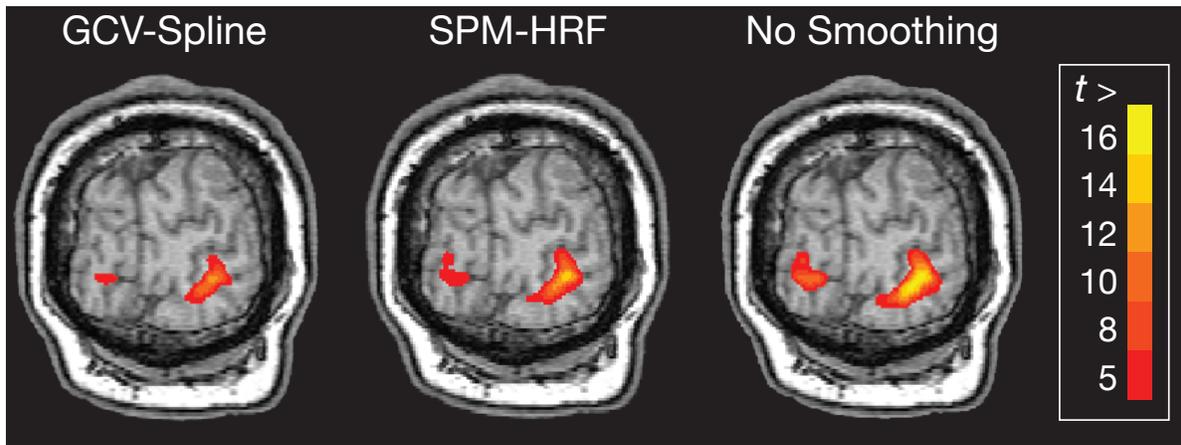


Figure 4: Images of the  $t$ -statistic under the null hypothesis for GCV-Spline smoothing, SPM-HRF smoothing, and no smoothing.

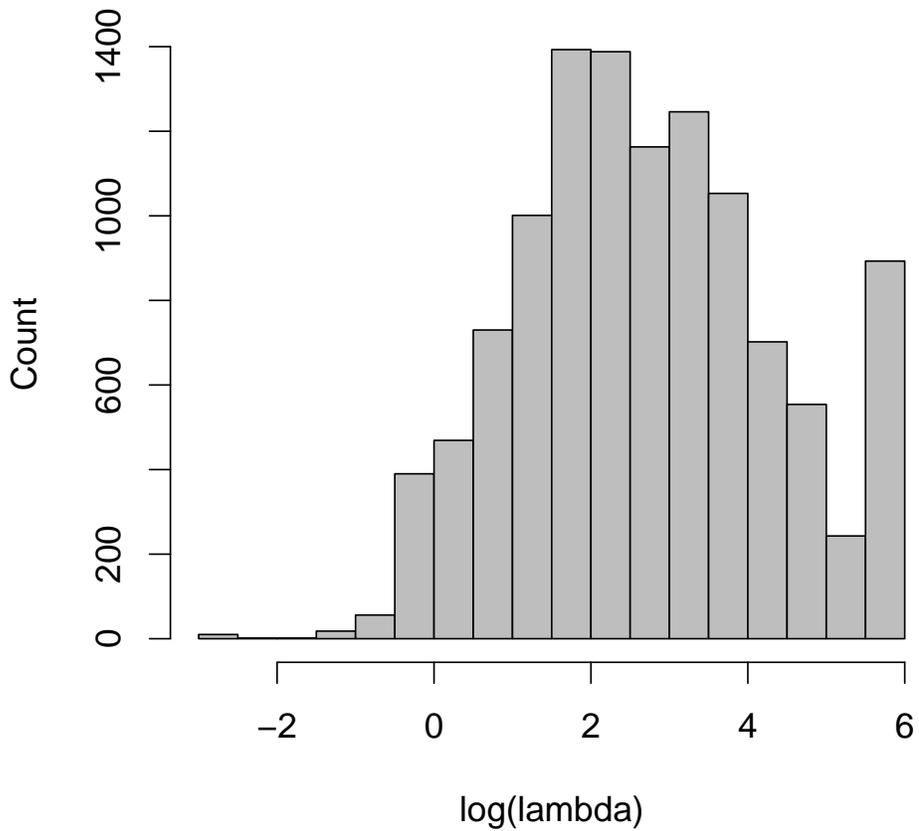


Figure 5: Histogram of  $\log_{10}(\lambda_{\text{opt}})$  from the simulated data set.

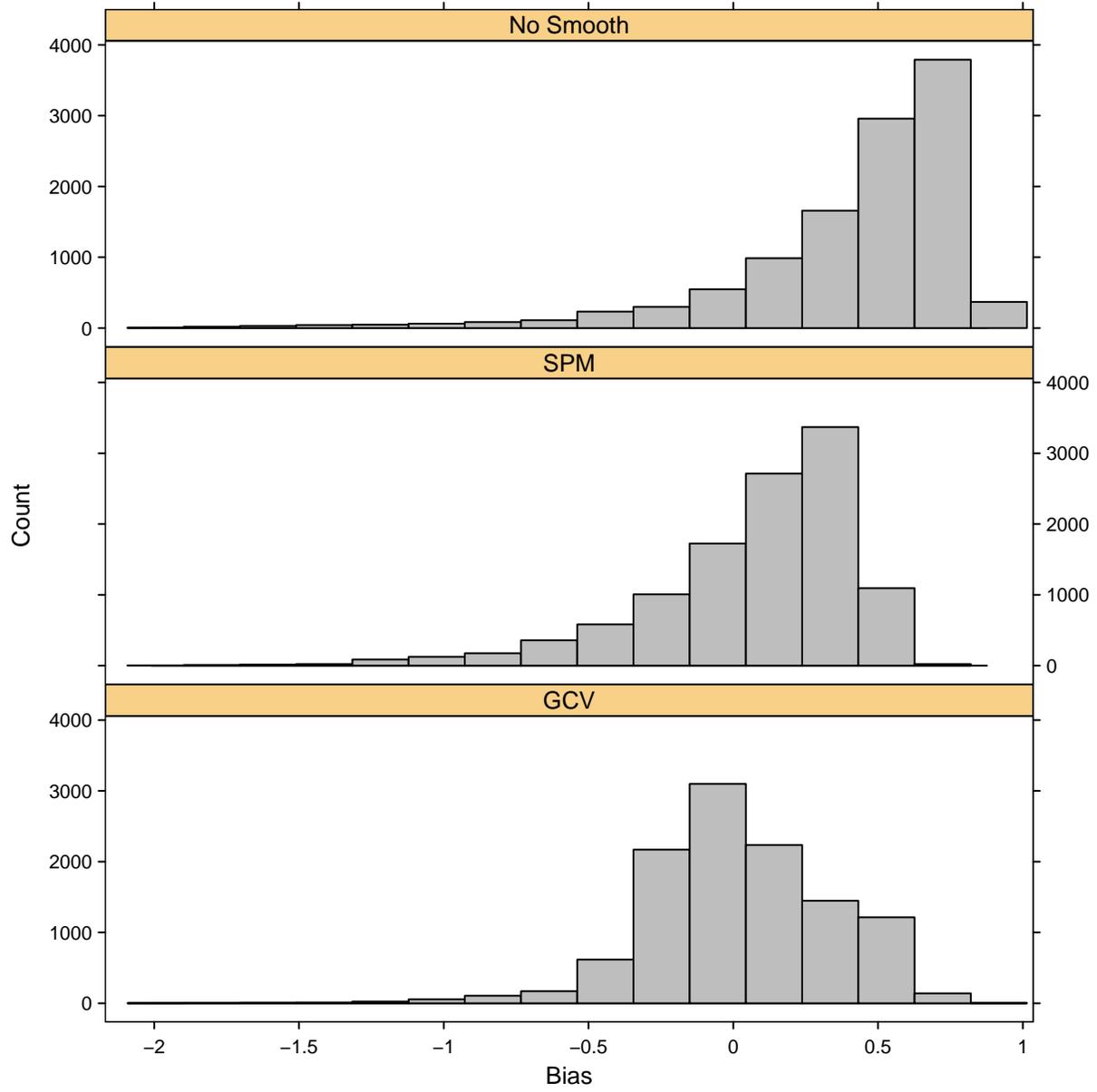


Figure 6: Histograms of  $\text{bias}(\hat{\text{var}}(\mathbf{c}^T \hat{\beta}))$  from the simulated data with no smoothing (top), SPM-HRF smoothing (middle), and GCV-Spline smoothing (bottom). Smoothing with the GCV-Spline method produces  $\hat{\text{var}}(\mathbf{c}^T \hat{\beta})$  estimates that are, on average, unbiased.

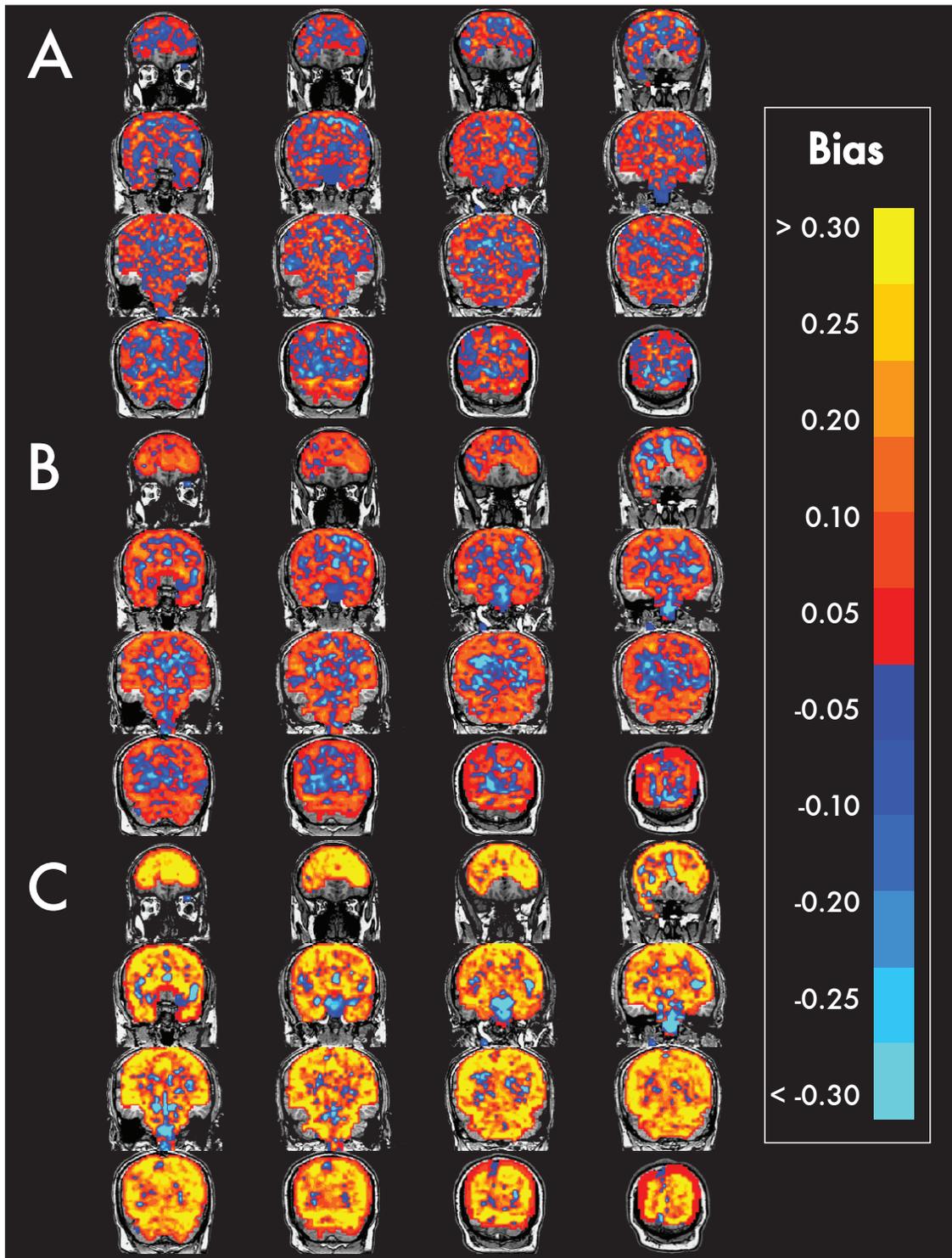


Figure 7: Images of bias of the bias of  $\hat{\text{var}}(\mathbf{c}^T \hat{\beta})$  for GCV-Spline (A), SPM-HRF (B), and no smoothing (C). The bias in each voxel was computed for simulated time series with autocorrelation structures estimated from the corresponding voxel. Voxels with positive bias 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.

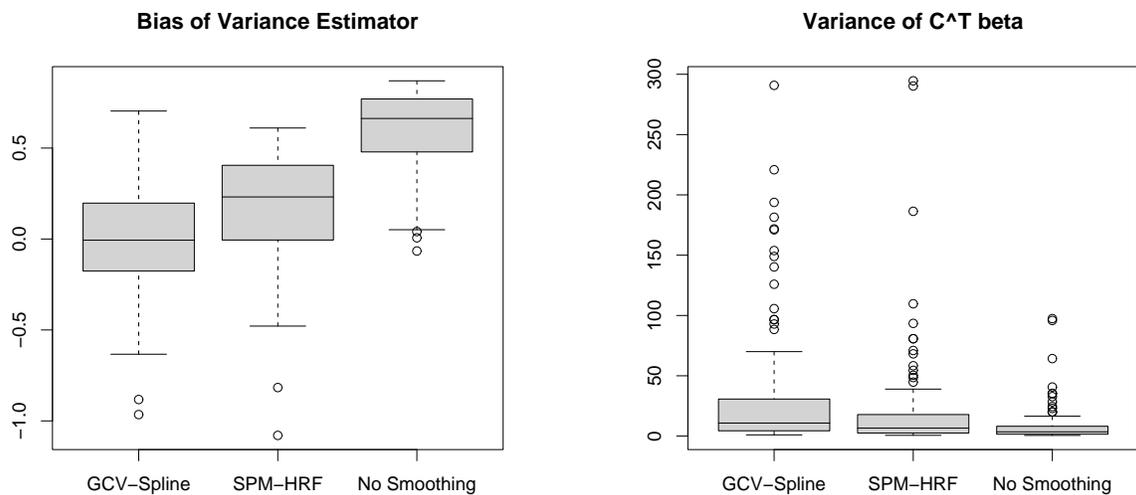


Figure 8: Bias of the variance estimator and estimated variance of  $\mathbf{c}^T \hat{\beta}$  with GCV-Spline, SPM-HRF, and no smoothing from the simulated data set with autocorrelation structures estimated from the 100 most significantly task-related time series.

## 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  $\lambda$  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  $\text{var}(\mathbf{c}^T \hat{\beta})$  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](http://www.netlib.org/gcv). To find optimal smoothers with RKPACk for 12,000 time series takes about six minutes on computer hardware 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 assumptions 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 (Worsley and Friston, 1995; Friston *et al.*, 2000; Woolrich *et al.*, 2001).

## 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.

## 6 Acknowledgments

We thank the National Institutes of Health for support of J.D. Carew under grant T32 EY07119, the Whitaker foundation for a grant to M.E. Meyerand and the University of Wisconsin, and the National Science Foundation and the National Institutes of Health for partial support of G. Wahba and X. Xie under grants DMS0072292 and EY09946.

## References

- [1] Bullmore, E., Brammer, M., Williams, S., Rabe-Hesketh, S., Janot, N., David, A., Mellers, J., Howard, R., and Sham, P. 1996. Statistical methods of estimation and inference for functional MR image analysis. *Magn. Reson. Med.* **35**:261-277.
- [2] Chatfield, C. 1996. *The analysis of time series*. Chapman & Hall, London.
- [3] Chiang, A., Wahba, G., Tribbia, J., and Johnson, D. 1999. A quantitative study of smoothing spline-ANOVA based fingerprint methods for attribution of global warming. Technical Report 1010, Department of Statistics, University of Wisconsin, Madison, WI (available from [www.stat.wisc.edu/~wahba/trindex.html](http://www.stat.wisc.edu/~wahba/trindex.html)).
- [4] Cox, R.W. 1996. AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comp. Biomed. Res.* **29**:162-173.
- [5] Craven, P., and Wahba, G. 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.
- [6] de Boor, C. 1978. *A practical guide to splines*. Springer-Verlag, New York.
- [7] Friston, K.J., Jezzard, P., and Turner, R. 1994. Analysis of functional MRI time-series. *Hum. Brain Mapp.* **1**:153-171.
- [8] Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., and Turner, R. 1995. Analysis of fMRI time-series revisited. *NeuroImage* **2**:45-53.
- [9] Friston, K.J., Josephs, O., Zarah, E., Holmes, A.P., Rouquette, S., and Poline, J.-B. 2000. To smooth or not to smooth? *NeuroImage* **12**:196-208.
- [10] Green, P.J., and Silverman, B.W. 1994. *Nonparametric regression and generalized linear models. A roughness penalty approach*. Chapman & Hall, London.
- [11] Gu, C. 1989. Rkpack and its applications: fitting smoothing spline models. Technical Report 857, Department of Statistics, University of Wisconsin, Madison, WI.
- [12] Lange, N., Strother, S.C., Anderson, J.R., Nielsen, F. Å., Holmes, A.P., Kolenda, T., Savoy, R., Hansen, L.K. 1999. Plurality and resemblance in fMRI data analysis. *NeuroImage* **10**:282-303.

- [13] Lowe, M.J., and Sorenson, J.A. 1997. Spatially filtering functional magnetic resonance imaging data. *Magn. Reson. Med.* **37**:723-729.
- [14] Wahba, G. 1983. Bayesian “confidence intervals” for the cross-validated smoothing spline. *J. Roy. Stat. Soc. Ser. B.* **45**:133-150.
- [15] Wahba, G. 1990. *Spline Models for Observational Data*. SIAM, Philadelphia.
- [16] Woolrich, M.W., Ripley, B.D., Brady, M., and Smith, S.M. 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. *NeuroImage* **14**:1370-1386.
- [17] Worsley, K.J., and Friston, K.J. 1995. Analysis of fMRI time-series revisited—again. *NeuroImage* **2**:173-181.
- [18] Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., Evans, A.C. 2002. A general statistical analysis for fMRI data. *NeuroImage* **15**:1-15.
- [19] Zarahn, E., Aguirre, G.K., and D’Esposito, M. 1997. Empirical analysis of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **5**:179-197.