Published in final edited form as:

J Appl Stat. 2014 ; 41(1): 26–45. doi:10.1080/02664763.2013.830085.

A Method for Choosing the Smoothing Parameter in a Semi-parametric Model for Detecting Change-points in Blood Flow

Sung Wan Han^a, Rickson C. Mesquita^b, Theresa M. Busch^c, and Mary E. Putt^{a,*}

^aDepartment of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

^bInstitute of Physics, University of Campinas, Campinas, SP, 13083-859, Brazil

^cDepartment of Radiation Oncology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

In a smoothing spline model with unknown change-points, the choice of the smoothing parameter strongly influences the estimation of the change-point locations, and the function at the change-points. In a tumor biology example, where change-points in blood flow in response to treatment were of interest, choosing the smoothing parameter based on minimizing generalized cross validation, GCV, gave unsatisfactory estimates of the change-points. We propose a new method, aGCV, that re-weights the residual sum of squares and generalized degrees of freedom terms from GCV. The weight is chosen to maximize the decrease in the generalized degrees of freedom as a function of the weight value, while simultaneously minimizing aGCV as a function of the smoothing parameter and the change-points. Compared to GCV, simulation studies suggest that the aGCV method yields improved estimates of the change-point and the value of the function at the change-point.

Keywords

Change-points; Smoothing spline; Partial spline; Generalized Cross Validation; Generalized Degrees of Freedom

1. Introduction

Our work was motivated by a tumor biology study of blood flow in mice treated with photodynamic therapy (PDT) [15, 27]. Blood flow in this model system typically exhibits two change-points, the first as blood flow declines in response to treatment and the second as it subsequently reaches a plateau. A change-point is defined as discontinuity at one or more of the derivatives of the function at that point [19]. In the mouse tumors, the duration and magnitude of blood flow decline are of interest; i.e., a long decline of large magnitude appears to be associated with better tumor response [27]. Complicating the problem is the fact that the pattern of blood flow in different animals, in the absence of treatment, is irregular and not well fit by a single parametric model. Some animals exhibit substantial curvature and regular sinusoidal patterns; others demonstrate relatively small variation. Fitting these data using a parametric approach requires substantial modelling of the baseline function. Against this background of irregular patterns of baseline blood flow, we sought an

approach to estimate the change-points, those times when blood flow begins to decline and reaches a plateau, as well as the magnitude of blood flow at the change-points.

Spline functions can be a useful alternative to a parametric model in this situation. Wahba [21] described “partial splines”, models that include both a smoothing spline component and a parametric component. Partial splines would be immediately applicable to the blood flow problem if the change-points were known. O’Connell [16] appears to be the first to have explored a smoothing spline to model the baseline function in a change-point problem, where the baseline function was clearly non-linear and the change-points were unknown. O’Connell [16] implemented a Bayesian approach to the estimation and focused largely on identifying change-points where there were “jumps” in the function. More recently, Wang [25] incorporated several change-point examples into his seminal text on smoothing splines. Wang [25] used a smoothing spline for the baseline and implemented a grid-search algorithm, selecting the smoothing parameter and the change-points that minimize generalized cross-validation (*GCV*) or the Akaike Information Criteria (*AIC*). This approach reflects Wahba’s work with partial spline models where the changepoint is known, and where choosing the smoothing parameter based on *GCV* yields minimum mean square error (MSE) [22].

For fitting models without change-points, *GCV* is asymptotically optimal with respect to choosing a smoothing parameter that minimizes MSE [13]. However, in moderate or small samples, fits based on *GCV* may have high variance as very small values of the smoothing parameter are sometimes chosen [11, 13, 23]. While the smoothing based on *GCV* gave very reasonable model fit in the examples given by Wang [25], the propensity of *GCV* to undersmooth has potential to yield substantial bias and variability in change-point estimation. As Speckman [19] suggests for a kernel-based model with change-points, the choice of the bandwidth is critical to accurate estimation of change-points, and the order of the optimal smoothing parameter differs from that used for comparable models without change-points. He suggested choosing the bandwidth for the kernel based on a modified form of *GCV*. Gijbels and Goderniaux [5, 6] later developed a bootstrap-based data-driven approach for estimating the smoothing parameter for the kernel-based estimation procedure.

Here we investigate the choice of the smoothing parameter for a smoothing spline model with unknown change-points. We confirm that the change-points selected by using *GCV* can be biased and highly variable. We propose an adaptive, data-driven approach, *aGCV*, based on changes in the generalized degrees of freedom [26], that better uses the data to inform the selection of the smoothing penalty. We demonstrate in simulation studies that using *aGCV* can substantially reduce bias and improve the precision of the estimated change-points as well as the function at the change-points. We further investigate a method for generating point-wise confidence intervals of the function at the change-points, and evaluate the coverage of these intervals.

Section 2 introduces the motivating example in detail. In Section 3, we review literature describing smoothing splines and their extension to the change-point problem in the context of a partial spline model. Section 4 describes the approach to estimation of the change-points and an algorithm to implement the approach. Section 5 describes the performance of these methods using a simulation study. Throughout the paper, we illustrate the development of our methods using the motivating example. Section 6 concludes with a discussion.

2. Blood flow in PDT-treated mice

Our motivating example comes from observations of blood flow during photodynamic therapy (PDT) in solid tumors grown in mice. PDT is a novel therapy in which subjects

receive a photosensitizing drug followed by light at or near a wavelength specific to the optimal absorption of the photosensitizer [12]. Illumination is directed toward the tumor or tissue of interest, and the subsequent oxygen dependent photo-toxic reactions leads to cellular and/or vascular damage. The ‘dose’ of light is the product of the intensity of the illumination and the duration of its exposure. In tumors, the subsequent oxygen-dependent photo-toxic reaction leads to cellular and/or vascular damage [15, 27]. Patients treated with PDT have variable outcomes in terms of tumor response, survival, and overall toxicity [4]. A number of parameters, including concentration of the photosensitizer, the light dose, and the mechanism of light delivery, *e.g.*, continuous versus intermittent light dosing, can be altered as part of PDT treatment. By varying these parameters, eventually on an individual basis, our research group hopes to ultimately improve the overall success of PDT for treating cancer.

In previous work, Yu *et al.* [27] showed that blood flow is dramatically altered by PDT, and that changes in blood flow are reflected in the rate of tumor regrowth after treatment. If blood flow decreases too rapidly, severe vascular damage may occur, there may be insufficient oxygen for the photo-toxic reaction, and tumor regrowth may be rapid. On the other hand, blood flow which is little changed may also indicate ineffective treatment. A gradual, sustained reduction in blood flow is commonly optimal. In conjunction with PDT, diffuse correlation spectroscopy (DCS), a novel optical technique, is being developed to monitor blood flow in real-time [14]. Flexible and accurate statistical methods are needed to detect change-points and estimate blood flow at the change-points, with the ultimate goal of providing information that physicians may use to evaluate treatment efficacy in real or near-real time.

Figure 1 (upper row) shows examples of blood flow during PDT in tumors grown in two individual mice that underwent PDT using the photosensitizer Photofrin. The data are available at <https://dbe.med.upenn.edu/biostat-research/MaryEPutt>. The full details of this study appear in Mesquita *et al.* [15], where we describe substantial variation in the baseline blood flow of the animals even in the absence of PDT. Typically, in animals undergoing Photofrin-based PDT, blood flow is approximately constant or even rises for an initial time, undergoes a sustained decline, and reaches a lower plateau. Figure 1 displays substantial variation in the blood flow patterns in two mice. In Mouse A, blood flow is initially constant, and at around 15 minutes declines to a lower plateau. In Mouse B, blood flow initially shows a pronounced initial curvature, followed by a sharp decline between 20 and 25 minutes. In a similar data set, Yu *et al.* [27] identified change-points in blood flow visually, and determined that both the slope of the blood flow change in the interval between the change-points and the duration of the change were associated with tumor response. In Mesquita *et al.* [15], we used an *ad-hoc* approach to identify change-points, taking first order differences of the data and applying scan statistics with variable time windows based on likelihood ratios [8]. We then smoothed the data and estimated blood flow at the change-points. The smoothing spline approach is a flexible model-based approach for analyzing these data.

3. Statistical models

We first describe the change-point model, and review the partial spline model described by Wahba [22] and Wang [25].

3.1 Change-point model

Let $y_1, \dots, y_n$ be the outcomes with $t_1, \dots, t_n$, the corresponding observation times in the domain of $t \in [0, 1]$ for n , the number of time points. For the j^{th} observation ($j = 1, \dots, n$), our model is:

$$y_j = f(t_j) + \sum_{k=1}^K \alpha_k (t_j - \tau_k)_+ + e_j \quad (1)$$

where $f(t_j)$ is an unspecified baseline function, τ_k is the k^{th} change-point ($k = 1, \dots, K$) for a model with up to K changepoints, α_k is the coefficient corresponding to a parameter designating the added slope between τ_k and τ_{k+1} , and $(t_j - \tau_k)_+$ is the function, $(t_j - \tau_k)I(t_j > \tau_k)$. The error terms, e_j , are independent and mean-zero with $\text{Var}(e_j) = \sigma^2$. The baseline function $f(t_j)$ in Equation (1) is flexible. One simple case is the linear model such that

$$f(t_j) = \beta_0 + \beta_1 t_j. \quad (2)$$

Alternatively, $f(t_j)$ may be more complex with curvature introduced by adding polynomial terms or through a non-linear parametric model. However, it may be challenging to specify an appropriate baseline function, particularly when, as in our tumor model, individuals show substantial heterogeneity in this baseline function. Thus, we let $f(t_j)$ be unspecified.

3.2 Partial spline model

Wahba [22] and Wang [25] show that the smoothing-spline model for $f(t_j)$ may be based on the reproducing kernel Hilbert space (RKHS). Thus, the cubic spline is written as:

$$f(t_j) = \beta_0 + \beta_1 t_j + \sum_{l=1}^n c_l \xi_l(t_j), \quad (3)$$

where $\xi_l(t) = R_1(t_j, t_l)$, and

$$R_1(t_j, t_l) = \int_0^1 (t_j - \nu)_+ (t_l - \nu)_+ d\nu,$$

where $(t_j - \nu)_+$ is the function, $(t_j - \nu)I(t_j > \nu)$. In Equation (3), the baseline is decomposed into a linear term $\beta_0 + \beta_1 t$ and the remainder term $\sum_{l=1}^n c_l \xi_l(t_j)$. The form in Equation (3) allows flexibility in the curvature of a baseline within a unified framework, which includes parametric and nonparametric models. If all $c_j = 0$, $f(t_j)$ is linear, and the model in Equation (1) is piecewise linear. If $c_l > 0$, $f(t_j)$ allows curvature in the baseline with the extent of the curvature dependent on the magnitude of the c_l .

When the change-points are known, substituting Equation (3) for $f(t_j)$ in Equation (1) yields the partial spline model for the mean function, $g(t_j) = E[y_j]$, where

$$g(t_j) = \beta_0 + \beta_1 t_j + \sum_{l=1}^n c_l \xi_l(t_j) + \sum_{k=1}^K \alpha_k (t_j - \tau_k)_+. \quad (4)$$

When the change-points, τ_k , are known, estimates of $g(t_j)$, $\hat{g}(t_j)$, are constructed using estimates of the parameters in Equation (4) determined by minimizing the residual sum of squares subject to a penalty, *i.e.*,

$$\frac{1}{n} \sum_{j=1}^n \left\{ y_j - \sum_{k=1}^K \alpha_k (t_j - \tau_k)_+ - f(t_j) \right\}^2 + \lambda \int_0^1 f^{(2)}(t) dt, \quad (5)$$

where $f^{(2)}(t)$ is the second order derivative with respect to time t , $\int_0^1 f^{(2)}(t) dt$ is a penalty function, and λ is the smoothing parameter.

4. Estimation

We first consider estimation when the change-points are known, extend the discussion to the case of unknown change-points, and illustrate the estimation procedure with our example data.

4.1 Estimation with known change-points

When the change-points, τ_k , are known, and $f(t_j)$ is modelled using a cubic spline, the matrix form of Equation (1) is:

$$\mathbf{Y} = \mathbf{Z}_\tau \alpha + \mathbf{T} \beta + \sum \mathbf{c} + \mathbf{e}, \quad (6)$$

where $\mathbf{Y} = [y_1, y_2, \dots, y_n]^T$, $\mathbf{Z}_\tau$ is an $n \times K$ matrix with j^{th} row $\{(t_j - \tau_1)_+, (t_j - \tau_2)_+, \dots, (t_j - \tau_K)_+\}$, and $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_K]^T$, $\mathbf{T}$ is an $n \times 2$ matrix with j^{th} row $\{1, t_j\}$, $\beta = [\beta_0, \beta_1]^T$, Σ is an $n \times n$ matrix with elements $\{R_1(t_j, t_l)\}_{j=1, l=1}^n$, $\mathbf{c} = [c_1, c_2, \dots, c_n]^T$, and $\mathbf{e} = [e_1, e_2, \dots, e_n]^T$. Here $\text{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}_{n \times n}$, where $\mathbf{I}_{n \times n}$ is the $n \times n$ identity matrix. Based on Equation (6) and the partial spline derivation described in Wahba [22] and Wang [25], the penalized least square for Equation (6) is

$$\frac{1}{n} \left\| \mathbf{Y} - \mathbf{Z}_\tau \alpha - \mathbf{T} \beta - \sum \mathbf{c} \right\|^2 + \lambda \mathbf{c}^T \Sigma \mathbf{c} \quad (7)$$

where $\|\cdot\|$ is a Euclidian norm. Let $\mathbf{V} = [\mathbf{Z}_\tau, \mathbf{T}]$, Wahba [22] showed that the solution to the penalized least square for Equation (7) may be implemented using the QR decomposition. Let $\mathbf{V} = [\mathbf{Q}_1, \mathbf{Q}_2][\mathbf{R}^T, \mathbf{O}^T]^T$, where $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{R}$ are of dimension $n \times (K+2)$, $n \times (n-K-2)$, and $(K+2) \times (K+2)$, respectively. $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2]$ is an orthogonal matrix, $\mathbf{R}$ is upper triangular and invertible, and $\mathbf{O}$ is an $(n-K-2) \times (K+2)$ matrix. Then for $\mathbf{M} = \Sigma + n\lambda \mathbf{I}_{n \times n}$ and $\mathbf{d} = [\alpha^T, \beta^T]^T$, the estimate of $\mathbf{d}$ is:

$$\hat{\mathbf{d}} = \mathbf{R}^{-1} \mathbf{Q}_1^T (\mathbf{Y} - \mathbf{M} \mathbf{c}),$$

where the elements of $\hat{\mathbf{d}}$ are the first K entries of $\hat{\mathbf{d}}$. The estimate of $\mathbf{c}$ is

$$\hat{\mathbf{c}} = \mathbf{Q}_2 (\mathbf{Q}_2^T \mathbf{M} \mathbf{Q}_2)^{-1} \mathbf{Q}_2^T \mathbf{Y},$$

The smoother matrix $\mathbf{H}_{\tau, \lambda}$ such that the predicted value of $\mathbf{Y}$, $\hat{\mathbf{Y}} = \mathbf{H}_{\tau, \lambda} \mathbf{Y}$ can be represented by

$$\mathbf{H}_{\tau, \lambda} = \mathbf{I}_{n \times n} - n\lambda \mathbf{Q}_2 (\mathbf{Q}_2^T \mathbf{M} \mathbf{Q}_2)^{-1} \mathbf{Q}_2^T. \quad (8)$$

4.2 Estimation when change-points are unknown

Fitting the partial spline assumes that τ is known, and additionally requires specification of the smoothing parameter, λ . For several unknown change-point examples, Wang [25] used a grid-search for combinations of (λ, τ) . Here, for a model with a single change-point, λ and τ were chosen across a grid of possible values to minimize an information criterion such as the generalized cross validation (GCV) criterion where

$$\log(GCV) = \log(RSS_{\tau, \lambda}/n) - 2\log\left(1 - \frac{DF_{\tau, \lambda}}{n}\right), \quad (9)$$

and

$$RSS_{\tau, \lambda} = \|(\mathbf{I}_{n \times n} - \mathbf{H}_{\tau, \lambda})\mathbf{Y}\|^2 = \sum_{j=1}^n (y_j - \hat{y}_j)^2,$$

where $\hat{y}_j = \hat{g}(t_j)$ is the predicted value of y_j . $RSS_{\tau, \lambda}$ measures the fit of the model while

$$DF_{\tau, \lambda} = \text{tr}(\mathbf{H}_{\tau, \lambda})$$

where $\text{tr}(\cdot)$ indicates the trace of a matrix. $DF_{\tau, \lambda}$ is a “generalized degrees of freedom”, or model complexity term representing the sum of the sensitivity of each fitted value to perturbations in the corresponding value in the observed data [26].

4.3 Estimation based on GCV in the blood flow data

Figure 1 (middle row) shows the results of fitting this model based on the minimum GCV, as described in Wang [25]. In each case, the estimated change-points consistently lie outside of the regions where common sense suggests the change-points are located. For example, the estimated change-points in the rBF of Mouse A are between 10 minutes and 15 minutes, while the primary decline in rBF appears to occur after 15 minutes. In Mouse B, the estimated change-points are at 25.5 and 26.1 minutes, while the apparent initial decline in rBF is at around 20 minutes. In contrast, Figure 1 (lower row) where a larger smoothing parameter was used, the change-points appear to be much more reasonable. The procedure used to select the smoothing parameter and the change-points in the lower row of the figure is described below.

4.4 Selection of the smoothing parameter

For the partial spline model, where τ is known, Wahba [22] showed that using GCV to choose λ minimizes the mean square error (MSE) of the estimated function $\hat{\mathbf{g}}$ defined as

$$\begin{aligned} MSE(\hat{\mathbf{g}}) &= E\left(\frac{1}{n} \|\hat{\mathbf{g}} - \mathbf{g}\|^2\right), \\ &= \text{bias}^2(\hat{\mathbf{g}}) + \text{Var}(\hat{\mathbf{g}}), \end{aligned}$$

where $\mathbf{g} = [g(t_1), \dots, g(t_n)]^T$ and $\hat{\mathbf{g}} = [\hat{g}(t_1), \dots, \hat{g}(t_n)]^T$, and where $\text{bias}^2(\hat{\mathbf{g}}) = \frac{1}{n} \|(\mathbf{I} - \mathbf{H}_{\tau, \lambda})\mathbf{g}\|^2$, and $\text{Var}(\hat{\mathbf{g}}) = \frac{1}{n} \sigma^2 \text{tr}(\mathbf{H}_{\tau, \lambda}^2)$. Our model differs from the partial spline models investigated by

Wahba [22], in having unknown change-points. Moreover, our primary interests lie in estimating specific components of the curve related to the change-points, *i.e.* the duration of the change period, and the magnitude of the associated change in blood flow during the change period.

Results in Figure 1 (middle row) suggest that choosing the smoothing parameter, λ , based on minimizing GCV yields a reasonably well-fit model and smooth curve for Mouse B and a somewhat under-smoothed curve for Mouse A. In initial simulation studies, we found that this procedure typically yields a model where the observed MSE is very near the MSE achieved when the change-point is known (results not shown). However, Figure 1 (middle row) suggests that the estimated change-points are highly inaccurate for both animals. We considered several other information criteria such as the Bayesian Information Criteria (BIC), but did not find consistently satisfactory results, leading us to develop our own approach.

The log of GCV in Equation (9) can be seen as a weighted sum of two terms, the residual sum of squares, $RSS_{\tau,\lambda}$, reflecting model fit, and a term relevant to model complexity, $DF_{\tau,\lambda}$ [26]. For GCV , the first and second terms have weights of 1 and -2 . Our data suggest that re-weighting to give more importance to the model complexity term, and thus increase smoothing, could improve change-point estimation. Heuristically, we reasoned that a change-point in our model represents a sharp change in derivative relative to the baseline function. In an under-smoothed model, the fit will include many sharp changes in derivatives. If the magnitude of the derivative change at the change-point is similar to the derivative changes in the fitted function, the change-point will be difficult to locate. On the other hand, if the data are over-smoothed, spurious change-points may be found [19].

Our reweighted form of $\log(GCV)$, is $\log(aGCV)$ where

$$\log(aGCV) = \log(RSS_{\tau,\lambda}/n) - w \log\left(1 - \frac{DF_{\tau,\lambda}}{n}\right), \quad (10)$$

for $w \geq 2$, to allow at least as much smoothing as GCV . Here it is easy to see that as w becomes very large, $aGCV$ will be dominated by the model complexity term, and minimized for smaller $DF_{\tau,\lambda}$, corresponding to a smoother fit. The question is which w to choose. For small values of w , accurate assessment of the change-points has little impact on $aGCV$, because here $aGCV$ places low weight on model complexity and relatively higher weight on model fit, favoring under-smoothed models, where the derivative changes at the change-point and in the baseline function are similar magnitude. However as w increases, and $aGCV$ begins to favor smoother models, accurate assessment of the change-points potentially has a large impact on model complexity. In other words for a smoother model, accurate assessment of the change-points contributes information that allows the spline function to find a fit to the curve with a simpler model. Thus we propose choosing w to maximize the criterion:

$$R(w) = \frac{DF'_k(w, \tau, \lambda)}{DF'_0(w, \lambda)} \quad (11)$$

where $DF_k(w, \tau, \lambda)$ is the generalized degrees of freedom, or model complexity term for the smoothing spline model with k change-points that minimizes $aGCV(w)$ across change-points, τ , and smoothing term λ . Similarly, $DF_0(w, \lambda)$ is the model complexity term for the smoothing spline model without change-points that minimizes $aGCV(w)$ across smoothing terms λ . $DF'_k(w, \tau, \lambda)$ and $DF'_0(w, \lambda)$ are the first derivatives with respect to w of these

terms. Both the numerator and the denominator of Equation (11) are negative as model complexity is a decreasing function of w . Without proof, we hypothesize that the derivatives reflecting the change in model complexity should be similar in the numerator and the denominator in the region where w is small and the estimated change-points are far from the truth. As w increases and the estimated change-points approach the true change-points, model complexity decreases sharply and the derivative for the model with change-points (numerator) becomes large and negative. In contrast the derivative for the model without change-points (denominator) continues to decrease in a smooth fashion. The point where $R(w)$ becomes large and positive suggests a region where the change-points have been accurately estimated.

4.5 Description of algorithm for selecting smoothing parameter

The following grid-search technique can be used to estimate the w that minimizes Equation (11).

1. Chose a grid of possible change-points and smoothing parameters. For each unique change-point/smoothing parameter combination, $(\tau_1, \tau_2, \lambda)$, fit the model in Equation (1) using penalized least squares as described in Equation (5). Compute $(RSS_{\tau,\lambda}, DF_{\tau,\lambda})$.
2. Choose a grid of penalties, $W = \{w_i, i = 1, \dots, I\}$ for $2 \leq w_i \leq 10$.
3. For each change-point/smoothing parameter combination, $(\tau_1, \tau_2, \lambda)$, and each w_i in W , compute $aGCV(w_i)$ from Equation (10).
4. For each w_i in W , determine the value of the change-points that minimizes $aGCV(w_i)$ i.e., $(\tau_1, \tau_2, \lambda)^{(i)} = \underset{\tau_1, \tau_2, \lambda}{\operatorname{argmin}} [aGCV(w_i)]$.
5. Determine $DF_{\tau,\lambda}^{(i)}$, the generalized degrees of freedom or model complexity term, for the model corresponding to $(\tau_1, \tau_2, \lambda)^{(i)}$ obtained in Step 4.
6. Compute $\delta_k(w_i)$, the difference between $DF_{\tau,\lambda}^{(i)}$ for adjacent values of w_i for a model with k changepoints. Thus for $i = 1$

$$\delta_k(w_i) = DF_{\tau,\lambda}^{(i)} - DF_{\tau,\lambda}^{(i+1)}.$$

7. Repeat Steps 1 through 6, but for the model *without change-points*. Now the differences between adjacent values of w_i are denoted as $\delta_0(w_i)$.
8. For the model without change-points, discard all $\delta_0(w_k) = 0$. Fit a smooth function to the plot of the non-zero $\delta_0(w_i)$ versus w_i , and use the curve to interpolate $\delta_0(w_i)$ at each w_i , where $\delta_2(w_i)$ is non-zero. These interpolated points form the final $\delta_0(w_i)$.
9. Compute the ratio of differences,

$$R(w_i) = \frac{\delta_2(w_i) I[\delta_2(w_i) > 0]}{\delta_0(w_i) I[\delta_2(w_i) > 0]}.$$

10. Select the maximum $R(w_i)$ among values larger than a pre-determined threshold of at least 1.0. We used 1.1. The estimated change-points, $\hat{\tau}_1$ and $\hat{\tau}_2$, are the values of (τ_1, τ_2) that minimize $aGCV$ at the value of $w = w_{i+1}$ that maximizes $R(w_i)$. If all

values are smaller than the threshold, use *GCV* as the selection criterion, choosing $w = 2$.

4.6 Point estimates and confidence intervals

Point estimates for $g(\tau_1)$ and $g(\tau_2)$ as well as $g(\tau_1) - g(\tau_2)$ are readily obtained by substituting the parameters with their estimates in Equation (4). Confidence intervals were based on a method, described by Wahba [22] [20] and summarized in Wang [25], for a smoothing spline. Conditional on known τ_1 , τ_2 , and λ , our model is a partial spline. Based on the Hilbert space construction, terms involving the change-points, $\sum_{k=1}^K \alpha_k (t_j - \tau_k)_+$ in Equation (4) and the linear terms $\beta_0 + \beta_1 t_j$ in Equation (3) can be treated as parametric terms and the approach to finding confidence intervals for the smoothing spline applies to the partial spline [22], *i.e.*, a two-sided α -level confidence interval on $g(t_j)$ is based on the Wald-type interval

$$\hat{g}(t_j) \pm z_{\alpha/2} \sqrt{\widehat{Var}(\hat{g}(t_j))}, \quad (12)$$

where the estimated variance of the function is

$$\widehat{Var}(\hat{g}(t_j)) = \hat{\sigma}^2 h_{jj}$$

for

$$\hat{\sigma}^2 = \frac{RSS_{\tau, \lambda}}{tr(\mathbf{I} - \mathbf{H}_{\tau, \lambda})}, \quad (13)$$

and h_{jj} , the j^{th} diagonal entry in $\mathbf{H}_{\tau, \lambda}$. These intervals are confidence intervals on the function rather than point-wise confidence intervals, although we treated them as such. Additionally, this interval does not account for the fact that the change-points are unknown. Thus, similar to Hall *et al.* [7], we incorporated uncertainty in our estimate of the change-points into the variance estimate. Specifically, we considered for any time t_j , that

$$Var(\hat{g}(t_j)) = E[Var(\hat{g}(t_j)|\tau)] + Var[E(\hat{g}(t_j)|\tau)].$$

where we consider $\tau = (\tau_1, \tau_2)$ as random variables, represented by the grid of possibilities for the unknown values of the changepoints. To estimate $Var(\hat{g}(t_j))$, we condition on the λ chosen to minimize *aGCV*. We then estimate $Var(\hat{g}(t_j)|\tau)$ and $E(\hat{g}(t_j)|\tau)$ from the model fit for each combination of change-point(s) in the grid search. Since the distribution of τ is unknown, we simply constructed the moments using the $RSS_{\tau, \lambda}$ as weights. Our interest is in $\widehat{Var}(\hat{g}(t_j))$ at the specific time $t_j = \tau$, where, since τ is unknown, we substitute our *aGCV*-based estimate(s), τ , for τ .

For the two change-point model, we also constructed confidence intervals for $g(\tau_1) - g(\tau_2)$, assuming that the point estimates were independent, *i.e.*,

$$\text{Var}[\hat{g}(t_1) - \hat{g}(t_2)] = \text{Var}(\hat{g}(t_1)) + \text{Var}(\hat{g}(t_2)).$$

4.7 Illustration of aGCV algorithm in the PDT data

For Mouse B, and across a range of values of w , Figure 2 shows minimum values of $aGCV$ decomposed into the component terms, $RSS_{\tau, \lambda}$ (top panel) and $DF_{\tau, \lambda}$ (middle panel) for the model with two change-points (black) and no change-points (blue). Larger values of w imply increased smoothing. Thus for both models, $RSS_{\tau, \lambda}$ increases as a function of w while $DF_{\tau, \lambda}$ decreases. For $w < 2.8$ both the model complexity and the residual sum of squares are just slightly smaller for the model with change-points compared to the model without change-points. For the change-point model, at $w = 2.8$, model complexity decreases sharply while the residual sum of squares shows a sharp increase. For the model without change-points, no such shift occurs. For $w < 2.8$, the curves for the two models diverge. The lower panel shows the values of τ that minimize $aGCV$ for each w . For $w < 2.8$, the change-points in the lower panel of Figure 2 are grossly mis-specified, as shown in Figure 1 (middle row, right panel). At $w = 2.8$, there is a shift in the change-points that minimize $aGCV$, corresponding to the more reasonable estimates seen in Figure 1 (lower row, right panel).

Figure 1 (bottom row) shows the results of fitting this model based on the procedure described above. Table 1 summarizes the results. The two animals had similar weighting factors, w , but substantially different values of $\log_{10}(\lambda)$, the smoothing parameter, indicating that, as expected, the curve for Mouse A involved more smoothing than that for Mouse B. Of interest to the investigators, the results suggested that blood flow underwent a longer decline for Mouse A compared to Mouse B (5.7 versus 3.1 minutes). The magnitude of the decline in blood flow may have been smaller for Mouse A compared to Mouse B (.56 versus 0.69), although there was overlap in the 95% confidence intervals.

5. Simulation study

We investigated the performance of the $aGCV$ algorithm using simulation. Our model is $f(t_j) + \varepsilon_j$, where the ε_j 's are independent with $\varepsilon_j \sim N(0, 1)$, $j = 1, \dots, 100$ where we generated the mean baseline function using a logistic function, *i.e.*,

$$f(t_j) = \frac{\theta_1}{1 + \exp[-(t_j - \theta_2)/\theta_3]}, \quad (14)$$

indexed by θ_1 which determines an asymptotic upper limit of the function, θ_2 , the location parameter, and θ_3 , the scale parameter. The top row of Figure 3 illustrates the baseline function corresponding to two cases: “small curvature” with $\theta = \{\theta_1, \theta_2, \theta_3\} = \{100, -0.25, 0.15\}$ and “large curvature” with $\theta = \{100, 0.0, 0.1\}$.

We considered a model with either one or two change-points constructed on the range $t_j \in \{0, 1\}$. Change-points were set at either or both 0.2 or 0.6, corresponding to the region of greater or less curvature (Figure 3 (middle panel)). We consider $a = -50, -100$, or -200 for the one change-point model or $a = a_1 = -a_2$ for the two change-point model. Results are discussed in terms of absolute values *i.e.*, $|a|$. These a values bound or exceed the maximum $f'(t_j)$ of 89 in the low curvature simulation, while they are somewhat smaller than the maximum $f'(t_j)$ of 250 in the large curvature simulation. Figure 3 illustrates the model with the change-points.

We carried out 100 simulations, calculating the mean and variance across iterations. Bias, the difference between the sample mean and the parameter of interest, and %bias ($\text{bias}/\text{parameter} \times 100$) are reported along with the sample standard deviation (SD) and the *MSE*, the sum of the squared bias and the variance. For each simulation, we computed a 95% confidence interval and coverage of the parameter of interest across all simulations. With 100 iterations, the bounds on the coverage, for a two-sided exact 95% confidence interval is (.888,.984).

5.1 Results for single change-point model

This section compares the performance of the *GCV* and *aGCV* -based estimates for the simulations with one change-point.

- **Choice of w and λ :** Table 2 shows that median w tends to approach 2.0, or equivalently *GCV*, when the degree of curvature in the baseline is smaller or when the absolute value of the change size, $|a|$, is larger. Even when the median of the *aGCV* distribution was equivalent to *GCV*, however, there were typically cases, where larger values of w were selected. With a larger degree of baseline curvature or a smaller change size, *aGCV* tends to suggest larger values of w , moving farther away from the *GCV* criterion.

As expected, for any one change-point/change size combination (e.g., $\tau = 0.2$, $|a| = 100$), the large curvature simulation consistently yielded more negative values of $\log_{10}(\lambda)$ than the small curvature simulation. This corresponds to less smoothing and more curvature in the fitted function for the large curvature simulations. We were initially surprised that, for each case of baseline curvature, the degree of smoothing selected for $\tau = 0.2$, in the region of higher curvature, was higher than for $\tau = 0.6$. We speculate that when the change-point is in a region of higher curvature, the higher degree of smoothing allows the model to more accurately differentiate the derivative changes in the function compared to the change-point.

- **Estimation of the change-point:** Bias and SD in the change-point estimate, $\hat{\tau}$, can be interpreted in the context of the range, (0,1), of τ ; e.g., bias of 0.05 can be interpreted as 5% bias in the change-point estimate. Table 3 shows that when τ , the change-point, is at 0.2, *aGCV* consistently yields estimates that outperform those chosen by *GCV*, both in terms of bias and SD. For *aGCV*, the absolute levels of bias for $\tau = 0.2$ were consistently small, less than 5%.

When the change-point occurred at $\tau = 0.6$, the pattern was less clear. For larger values of $|a|$ of 100 or 200, *aGCV* either outperformed or was similar to *GCV* for both bias and SD; absolute levels of bias were small, consistently less than 10% and often $< 1\%$. For the small curvature baseline, when the change-point occurred at $\tau = 0.6$ and $|a|$ was 50, both *aGCV* and *GCV* yielded change-point estimates with small bias ($< 5\%$) and similar MSE. In contrast, for the large curvature baseline for $\tau = 0.6$ and $|a| = 50$, both *aGCV* and *GCV* yielded estimates that were highly biased. Further examination of the results suggested that the distribution of the estimates of τ for this case was roughly bimodal, with one peak occurring near the true change-point of 0.6, and a second peak occurring below 0.2. This bimodal distribution was particularly pronounced for *aGCV* (not shown).

- **Estimation of the value of the function at the change-point:** Table 4 shows that the performance of the estimates of $g(\tau)$ follows a pattern that is similar to that described for estimation of the change-points. For $\tau = 0.2$, *aGCV* consistently outperformed *GCV*, both in terms of %bias and SD. Levels of bias in the estimate of $g(\tau)$ for *aGCV* were small, on the order of 1 to 2% for $\tau = 0.2$. When the

change-point occurred at $\tau = 0.6$, $aGCV$ outperformed GCV for both bias and SD except for $|a|$ of 50 in the large curvature case. Here the bias and SD were similar for $aGCV$ and GCV .

- **Estimation of change, a :** In contrast to the estimation of τ and $g(\tau)$, Table 5 shows that the bias in the estimate of a from $aGCV$ is often larger than for GCV . However, $aGCV$ yielded estimates with substantially smaller SD and MSE compared to GCV . Bias for $aGCV$ was at least 18% for the change-point at 0.2; smaller bias was achieved for $\tau = 0.6$. Because neither GCV nor $aGCV$ yielded reliable estimates of a , we did not proceed further to develop confidence intervals for these parameters.
- **Coverage of $g(\tau)$:** Table 6 shows that confidence intervals produced with $aGCV$ generally showed better coverage than those produced with GCV . The left panel of Figure 4 shows the relationship between the log (base 10) of the absolute value of the bias in the change-point estimate (Table 3) and coverage. This figure suggests that even small bias in the change-point estimate, roughly $> 1\%$, yields a sharp reduction in coverage.

5.2 Results for two change-point model

The relative performance of GCV and $aGCV$ for the two change-point model was similar to the single change-point model, with $aGCV$ generally outperforming or having a similar performance to GCV . Thus, our discussion focusses exclusively on results for $aGCV$.

- **Estimation of change-point duration and change-points:** Table 7 shows that the overall magnitude of the bias for the difference in change-points was larger than for the single change-point model; in particular for the large curvature case, only the change-point difference at $|a| = 200$, was estimated with $< 5\%$ bias. The duration of the change, equivalently the difference in change-points, is consistently estimated with less bias and SD when the degree of curvature in the baseline is small.
- **Estimation of difference in outcome at the change-point:** Table 7 shows that the overall magnitude of the bias for the difference in change-points was somewhat larger than for the single change-point model; in particular, for the large curvature case, only the difference in outcome at $|a| = 200$, was estimated with $< 5\%$ bias. The difference in outcome at the change-points is consistently estimated with less bias and SD when the degree of curvature in the baseline is small.
- **Estimation of a_1 and a_2 :** Similar to the single change-point model, estimates of the two slopes were highly biased and are not discussed further here.
- **Coverage of $g(\tau_1) - g(\tau_2)$:** Table 6 shows that the confidence intervals for the differences in $g(\tau_1)$ and $g(\tau_2)$ generally showed somewhat lower coverage than for the individual change-points. The right panel of Figure 4 shows the relationship between the absolute value of the bias in the estimated change-point difference (Table 7) and coverage. This figure suggests that, as with the individual change-points, even small bias in the change-point estimate yields a sharp reduction in coverage.
- **Coverage of $g(\tau_1)$ and $g(\tau_2)$:** Nominal coverage rates were satisfactory for the small baseline curvature but fell below the expected lower bound of 0.888 for the large curvature baseline except when the change size was large, $|a| = 200$.

6. Discussion

Our research was motivated by a tumor biology study where we wished to detect change-points in blood flow, and estimate mean blood flow at the change-points, against the background of a model that allows flexibility in the baseline blood flow. Following Wang [25] we developed an approach where the baseline function is modeled using a smoothing spline. The basic profile-likelihood approach used here was used in several examples by Wang [25], who suggested *GCV* or *AIC* to select the smoothing parameter. In partial spline models, Wahba showed that asymptotically, under mild regularity conditions, choosing the smoothing parameter by minimizing *GCV* yields the minimum *MSE* for the overall function [22]. In our problem, several initial simulations suggested that minimizing *GCV* tended to yield an *MSE* that was close to that obtained when the change-points were known. However, for the case of an unknown change-point location, we found that the change-points were often poorly fit by this approach: in essence, minimum, or close to minimum, *MSE* could be obtained with inaccurate change-points. More generally, it is known that choosing the smoothing parameter based on minimizing *GCV* may yield under-smoothed models in models without change-points [11, 13, 23].

We are unaware of analytical work involving *GCV* in smoothing splines where the change-point is unknown. However, the general problem of choosing the smoothing parameter has been considered in some detail in the context of change-point approaches using kernel-based methods of smoothing. Speckman [19] considered the choice of smoothing parameter when the number of change-points is unknown, pointing to the need to balance the dual issues of false positives, when the data are over-smoothed, and the detection problem, if the data are under-smoothed. The potential for *GCV* to yield under-smoothed estimates in a range of non-linear models outside of generalized linear models suggests that *GCV*, and the related *AIC*, may not be good criteria to use when detecting change-points. For the smoothing spline model we also explored *BIC* to choose the smoothing parameter [18]. For a sample size of 100, *BIC* is approximately $aGCV$ with $w = 4.6$ and should thus yield larger smoothing parameters than *GCV*. However *BIC* yielded change-point estimates with substantial bias, particularly for the larger curvature simulations (not shown). This suggested an adaptive approach using the data to select the smoothing parameter. Similarly, Gijbels and Goderniaux [5, 6] used a data-driven approach to both estimate the locations of the discontinuities in derivatives as well as determine the smoothing parameter. This earlier work used kernel smoothing to fit the background function.

Our approach allows the penalty for the smoothing term to be larger than *GCV*, and as we showed in the simulations, potentially penalizes models with large model complexity more than those with small model complexity. Our $aGCV$ statistic has similarities to a ‘modified’ *GCV* criterion [2, 13], which has the form

$$\log(mGCV) = \log(RSS/n) - 2\log\left(1 - a\frac{DF}{n}\right).$$

For models without change-points there is some consensus about good choices of a , the weights in $mGCV$. As with $aGCV$, for $a > 1$, $mGCV$ will tend to be minimized for models with increased smoothing compared to *GCV* due to the additional weight on the complexity term. Work with $mGCV$ in the context of smoothing splines for models without change-points has suggested somewhat arbitrary values of a , e.g., $a = 1.4$. It is not clear that these are applicable to models where change-point detection is the primary interest, and where it may be necessary to undersmooth somewhat for optimal detection of change-points.

In our work the number of change-points was known, and accurate detection of the change-points was the primary focus. We note that the determination of the number of change-points is an additional question that may be a key issue in many applications [10]. Our data-driven approach is based on a comparison of $aGCV$ in models with and without a fixed number of change-points. For this approach, there does not appear to be a single appropriate value of w . For a lower degree of baseline curvature, values of w near 2 seem appropriate, whereas a higher degree of baseline curvature required higher values of w . Not surprisingly, detection of the change-points was most challenging when the size of the change was small and the degree of curvature in the baseline was larger. For the simulations, the problem was more pronounced for the one change-point model when the change occurred in the asymptote of the curve at $t = 0.6$, than when the change-point occurred in the region of higher curvature at $t = 0.2$. Indeed, for the large curvature baseline, we found a bimodal distribution of the change-point estimates when $|a|=50$. We suspect this occurs because the magnitude of the curvature in the baseline at places other than the change-point, and the magnitude of the additional curvature at the change-point, are similar; thus the algorithm cannot reliably discriminate between the additional curvature in the function due to the change-point and baseline curvature.

One of the limitations of our method is that the grid-search does not provide any estimate of the precision of the change-points. Thus, the reliability of any one set of estimates is unknown. We are currently exploring bootstrap as well as Bayesian methods with the goal of addressing these problems. Our work additionally suggests that the estimated slope at the change-point has substantial bias. This finding has been noted before for partial spline models [1, 3, 17] where higher degrees of smoothing are associated with higher levels of bias. O'Connell [16] further noted the difficulty inherent in interpreting deviations from the baseline function modelled using a smoothing spline. In contrast to a parametric model, the baseline function is not clearly defined for a smoothing spline. A reviewer suggested a two step procedure to overcome the bias in the slope, combining the semi-parametric approach with a subsequent parametric model. For example, first, we might estimate the baseline and change-points by the semi-parametric approach described here, and then, fit the data near the estimated change point by, say, a piecewise linear model. This could yield an estimate of the slope with less bias. However, the implementation of such a two-step approach might have some issues, such as how large the neighborhood of the estimated change point should be for the piecewise linear model. In addition, if there were large curvature in the region of the change-point, the approach might still yield substantial bias.

In the presence of substantial bias in the estimate of the slope, as well as these problems with interpretation, we focused our attention on estimating the function at the change-points and developing point-wise confidence intervals for the function at the change-points. Not surprisingly, our results indicate the critical importance of bias in the estimate of the change-points in determining adequate coverage of the function at the change-point. Obtaining adequate coverage when the size of the change is small relative to the derivatives in the baseline function is not possible using our method. We are currently exploring refinements of our approach that may yield better estimation of the change-points which in turn should translate to better coverage. Additionally, during the development of the confidence interval approach several issues arose. First since the change-points are unknown, we need to account for uncertainty in their positions in the estimation of the variance. Hall *et al.* [7] used iterated expectation to adjust the variance; we followed a similar procedure, conditioning on the smoothing parameter. Since the distribution of the change-points is unknown, we simply weighted our estimates by the residual sums of squares. When the size of the change was large, and the change-point estimate was highly accurate, this procedure had little effect. However, for a number of our simulations, including uncertainty in the change-point estimate led to substantial improvement in the coverage compared to our initial

efforts (not shown). Nevertheless, this procedure highlights the fact that our method provides no estimate of the uncertainty in the change-point estimates. Re-sampling the data, as described above, is one approach that we are now considering to address this problem. A further issue arises from the fact that the confidence intervals described by Wahba [20] are intended to have the desired coverage across the function, rather than at individual points. Locally, coverage can be very poor particularly in regions of high curvature. To address this problem when point-wise estimates are of interest, Cummins *et al.* [2] suggested determining the variance estimate for a curve that is locally smoothed. This incorporation of local smoothing appears to improve coverage in regions of high curvature. For the change-point model, this would imply a two-step model where the change-points are estimated in the first step and then conditional local smoothing is applied to improve coverage at the change-point.

Lastly, the aGCV approach presented here, while a large improvement over GCV, is computationally intensive, particularly for larger numbers of change-points. We have extended the research described here, developing a more computationally efficient approach, that has comparable or better MSE to aGCV [9].

Acknowledgments

We are grateful to two anonymous reviewers for their helpful, thought-provoking comments on our original manuscript. We thank Professor Wensheng Guo for initially suggesting the smoothing spline approach, Professor Eli Glatstein for sharing his wealth of knowledge regarding photodynamic therapy with us, and Professor Arjun Yodh for his insights into DCS. Dr. Andrew Pole contributed valuable insights into the original analyses of these data. We thank Joann Miller and Steven S. Schenkel for help with data collection. Research supported by NIH-NCI 5-P01-CA-087971; R01 CA85831; and CA087971-S1 and CA129554. Code for implementing the programs is available from the first author upon request.

References

1. Chen H, Shiau JH. A two-stage spline smoothing method for partially linear models. *Journal of Statistical Planning and Inference*. 1991; 27:187–201.
2. Cummins DJ, Filloon TG, Nychka D. Confidence intervals for nonparametric curve estimates: toward more uniform pointwise coverage. *Journal of the American Statistical Association*. 2001; 96:233–246.
3. Engle RF, Granger CWJ, Rice J, Weiss A. Semi-parametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association*. 1986; 81:310–320.
4. Friedberg JS, Mick R, Stevenson JP, Zhu T, Busch TM, Shin D, Smith D, Culligan M, Dimofte A, Glatstein E, Hahn SM. Phase II trial of pleural photodynamic therapy and surgery for patients with non-small-cell lung cancer with pleural spread. *Journal of Clinical Oncology*. 2004; 22:2192–2201.
5. Gijbels I, Goderniaux A-C. Bootstrap test for change-points in nonparametric regression. *Journal of Nonparametric Statistics*. 2004; 16:591–611.
6. Gijbels I, Goderniaux A-C. Data-driven discontinuity detection in derivatives of a regression function. *Communications in Statistics - Theory and Methods*. 2005; 33:851–871.
7. Hall CB, Lipton RB, Sliwinski M, Stewart WF. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in Medicine*. 2000; 19:1555–1566.
8. Han SW, Tsui K-L, Ariyajunyah B, Kim SB. A comparison of CUSUM, EWMA, and temporal scan statistics for detection of increases in Poisson rates. *Quality and Reliability Engineering International*. 2010; 26:279–289.
9. Han SW, Zhong H, Putt M. An efficient operator for the change point estimation in partial spline models *Communications in Statistics. Computation and Simulation*. In press.
10. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for join-point regression with applications to cancer rates. *Statistics in Medicine*. 2000; 19:335–351. [PubMed: 10649300]
11. Kou SC, Efron B. Smoothers and the C_p , generalized maximum likelihood, and extended exponential criteria. *Journal of the American Statistical Association*. 2002; 97:766–782.

12. Lane N. New light on medicine. *Scientific American*. 288:38–45. [PubMed: 12506423]
13. Lukas M, De Hoog FR, Anderssen R. Performance of robust *GCV* and modified *GCV* for spline smoothing. *Scandinavian Journal of Statistics*. 2012; 39:97–115.
14. Mesquita RC, Durduran T, Yu G, Buckley EM, Kim MN, Zhou C, Choe R, Sunar U, Yodh AG. Direct measurement of tissue blood flow and metabolism with diffuse optics. *Philosophical Transactions of the Royal Society A*. 2011; 369:4358–4379.
15. Mesquita RC, Putt M, Pole A, Han SW, Busch TM. Mouse strain affects the dynamics of tumor vascular response during PDT. *PLoS ONE*. 2012:e37322. [PubMed: 22624014]
16. O'Connell, J. BCM Honours Dissertation. Faculty of Engineering, Computing and Mathematics, University of Western Australia; 2005. Penalised splines with change points.
17. Rice J. Convergence rates for partially splined models. *Statistics and Probability Letters*. 1986; 4:203–208.
18. Ruppert, D.; Wand, MP.; Carroll, RJ. *Semiparametric regression*. Cambridge University Press; Cambridge, UK: 2003.
19. Speckman P. Fitting curves with features: semiparametric change-point methods. *Computing science and statistics*. 1995; 26:257–264.
20. Wahba G. Bayesian 'Confidence Intervals' for the cross-validated smoothing spline. *Journal of the Royal Statistical Society Series B (Methodological)*. 1983; 45:133–150.
21. Wahba, G. Partial spline models for the semiparametric estimation of functions of several variables. *Statistical Analysis of Time Series, Proceedings of the Japan U.S. Joint Seminar*; Tokyo. 1984. p. 319-329.
22. Wahba, G. *Spline models for observational data*. Society for Industrial and Applied Mathematics; Philadelphia, PA: 1990.
23. Wahba G, Wang Y. Behavior near zero of the distribution of *GCV* smoothing parameter estimates. *Statistics & Probability Letters*. 1995; 25:105–111.
24. Wang H, Li R, Tsai CL. Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*. 2007; 94:553–568. [PubMed: 19343105]
25. Wang, Y. *Monographs on Statistics and Applied Probability*. Vol. 121. CRC Press. Taylor & Francis; Boca Raton, Florida, USA: 2011. *Smoothing Splines: Methods and Applications*.
26. Ye J. On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*. 1998; 93:120–131.
27. Yu G, Durduran T, Zhou C, Wang HW, Putt ME, Saunders HM, Sehgal CM, Glatstein E, Yodh AG, Busch TM. Noninvasive monitoring of murine tumor blood flow during and after photodynamic therapy provides early assessment of therapeutic efficacy. *Clinical Cancer Research*. 2005; 11:3543–3552. [PubMed: 15867258]

Figure 1. Relative blood flow (rBF) in Mouse A and Mouse B over time. Data alone are shown in top row; fitted curves (blue lines) with estimated change-points (red lines) are shown for the fits based on *GCV* (middle row) or with increased smoothing as determined by *aGCV* (bottom row).

Figure 2. Components of $aGCV$ (top and middle panels) for Mouse B as a function of the weighting factor, w , for the model with two change-points (black) or no change-points (blue). The bottom panel shows the change-points that minimize $aGCV$ at each value of w .

Figure 3.

Mean function used in the simulations. Top panel shows baseline alone; middle panels show baseline with single change-points at $\tau = 0.2$ or $\tau = 0.6$; bottom panel shows two change-points. Black, medium gray and light gray lines represent $\alpha = \{-50, -100, -200\}$ respectively.

Figure 4. Coverage of 95% confidence interval as a function of bias in the estimated change-point(s). Left panel shows results for the single change-point model for $g(\tau_1)$. Right panel shows results for the two change-point model for $g(\tau_1) - g(\tau_2)$.

Table 1Results for blood flow data using $aGCV$.

Estimates	Mouse A $w=2.7, \log_{10}(\lambda)=-3.5$	Mouse B $w=2.8, \log_{10}(\lambda)=-6$
Change-points (minutes)	(18.7, 24.4)	(21.8, 24.9)
Duration of change (minutes)	5.7	3.1
rBF at τ_1 (95% CI)	0.88 (0.86, 0.90)	1.19 (1.15, 1.24)
rBF at τ_2 (95% CI)	0.32 (0.30, 0.35)	0.50 (0.46, 0.55)
Change in rBF (95% CI)	-0.56 (-0.66, -0.46)	-0.69 (-0.75, -0.62)

Values of the weighting factor (w) and smoothing parameter $\log(\lambda)$ for $aGCV$ for a model with one change-point. Values shown are medians (interquartile range).

Table 2

τ	Baseline Curvature	w chosen for $ \alpha $ of:			$\log_0(\lambda)$ for $ \alpha $ of:		
		50	100	200	50	100	200
0.2	Small	2.25 (2.00,2.93)	2.20 (2.00,3.13)	2.10 (2.00,2.70)	-4.20 (-4.50, -3.90)	-4.20 (-4.50, -3.90)	-4.25 (-4.50, -3.90)
0.2	Large	5.25 (2.90,7.40)	4.45 (2.08,7.05)	3.50 (2.00,6.65)	-5.00 (-5.40, -4.90)	-5.30 (-6.00, -4.90)	-5.65 (-6.00, -4.90)
0.6	Small	2.00 (2.00,2.20)	2.00 (2.00,2.20)	2.00 (2.00,2.00)	-5.00 (-5.20, -4.90)	-5.10 (-5.30, -4.90)	-5.10 (-5.30, -4.90)
0.6	Large	3.15 (2.00,7.10)	2.00 (2.00,2.13)	2.00 (2.00,2.00)	-5.40 (-5.70, -4.80)	-5.75 (-5.90, -5.60)	-5.80 (-5.90, -5.70)

Table 3

Performance of estimates of τ based on minimizing GCV or aGCV for a model with one change-point

τ	Baseline Curvature	Metric	With GCV for $ \alpha $ of:			With aGCV for $ \alpha $ of:		
			50	100	200	50	100	200
0.2	Small	Bias	0.1055	0.0663	0.0060	-0.0023	-0.0022	-0.0005
		SD	0.2519	0.1986	0.0545	0.0417	0.0095	0.0043
		MSE	0.0746	0.0439	0.0030	0.0017	<0.0001	<0.0001
0.2	Large	Bias	0.2048	0.1429	0.0211	0.0250	0.0215	-0.0037
		SD	0.3056	0.2585	0.1172	0.1570	0.1243	0.0059
		MSE	0.1353	0.0872	0.0142	0.0253	0.0159	<0.0001
0.6	Small	Bias	-0.0597	-0.0057	-0.0015	-0.0556	-0.0008	-0.0006
		SD	0.2200	0.1273	0.0791	0.1598	0.0091	0.0044
		MSE	0.0520	0.0162	0.0063	0.0286	<0.0001	<0.0001
0.6	Large	Bias	-0.1699	-0.0683	-0.0170	-0.3443	-0.0836	0.0006
		SD	0.2812	0.2012	0.0910	0.2032	0.1805	0.0042
		MSE	0.1079	0.0451	0.0086	0.1598	0.0396	<0.0001

Table 4

Performance of estimates of $g(\tau)$ based on minimizing GCV or aGCV for a model with one change-point.

τ	Case	Metric	With GCV for $ \mathbf{d} $ of:			With aGCV for $ \mathbf{d} $ of:		
			50	100	200	50	100	200
0.2	Small	Bias (%bias)	-5.3 (-5.5)	-6.5 (-6.8)	-1.3 (-1.4)	0.5 (0.5)	0.7 (0.8)	0.7 (0.8)
		SD	10.3	18.2	9.7	1.5	0.4	0.5
		MSE	134.6	373.3	96.6	2.7	0.7	0.7
0.2	Large	Bias (%bias)	-11.0 (-12.5)	-14.5 (-16.4)	-5.4 (-6.1)	-2.1 (-2.4)	-1.7 (-2.0)	1.3 (1.5)
		SD	11.9	19.7	20.3	6.1	9.9	0.9
		MSE	262.0	598.4	441.7	41.8	101.4	2.4
0.6	Small	Bias (%bias)	-2.6 (-2.6)	-1.9 (-1.9)	-1.4 (-1.4)	-0.8 (-0.8)	0.1 (0.1)	0.2 (0.2)
		SD	5.2	7.7	10.7	2.6	0.5	0.5
		MSE	34.1	62.9	116.5	7.4	0.3	0.3
0.6	Large	Bias (%bias)	-11.4 (-11.4)	-5.5 (-5.5)	-1.3 (-1.3)	-13.5 (-13.6)	-3.5 (-3.5)	0.0 (0.0)
		SD	13.6	11.2	7.2	9.1	8.3	0.6
		MSE	312.9	154.7	54.2	265.2	81.1	0.4

Table 5

Performance of estimates of α using *GCV* or *aGCV* for a model with one change-point.

τ	Baseline Curvature	Metric	With GCV for			d of:			With aGCV for d of:		
			50	100	200	50	100	200	50	100	200
0.2	Small	Bias (%Bias)	-16.8 (33.7)	-3.9 (3.9)	-12.8 (6.4)	-38.3 (76.5)	-37.2 (37.2)	-36.6 (18.3)			
		SD	90.1	93.3	87.0	9.0	7.8	8.9			
		MSE	8403	8720	7727	1545	1447	1420			
0.2	Large	Bias (%Bias)	-5.4 (10.7)	6.5 (-6.5)	-34.1 (17.1)	-105.1 (210.1)	-97.1 (97.1)	-99.5 (49.8)			
		SD	132.5	128.0	119.4	49.0	53.0	31.9			
		MSE	17596	16427	15413	13444	12238	10928			
0.6	Small	Bias (%Bias)	0.0 (0.0)	4.8 (-4.8)	1.6 (-0.8)	-5.6 (11.1)	-4.0 (4.0)	-3.4 (1.7)			
		SD	71.2	72.1	60.5	12.4	6.8	6.7			
		MSE	5065	5227	3665	186	62	57			
0.6	Large	Bias (%Bias)	-9.4 (18.8)	15.9 (-15.9)	19.3 (-9.6)	-60.3 (120.6)	-8.9 (8.9)	-2.8 (1.4)			
		SD	120.2	95.1	104.4	38.1	19.4	12.9			
		MSE	14530.6	9288.0	11263.8	5091.5	456.3	174.0			

Table 6

Coverage of 95% CI for $g(\tau_1)$, $g(\tau_2)$ or $g(\tau_2) - g(\tau_1)$ for one change-point or two change-point model.

Change-points in Model	Change-point		with GCV for $ d $ of:				with aGCV for $ d $ of:			
	Parameter	Value	Baseline Curvature		50	100	200	50	100	200
1	τ_1	0.2	Small		0.52	0.78	0.94	0.98	1.00	1.00
1	τ_1	0.2	Large		0.08	0.24	0.72	0.52	0.76	0.96
1	τ_1	0.6	Small		0.52	0.7	0.92	0.62	0.92	1.00
1	τ_1	0.6	Large		0.36	0.62	0.86	0.18	0.64	0.94
2	τ_1	0.2	Small		0.56	0.74	0.93	0.94	1.00	0.99
2	τ_1	0.2	Large		0.04	0.16	0.71	0.35	0.56	0.90
2	τ_2	0.6	Small		0.56	0.74	0.93	0.93	0.99	0.99
2	τ_2	0.6	Large		0.21	0.43	0.71	0.13	0.60	0.90
2	$\tau_2 - \tau_1$	0.4	Small		0.49	0.70	0.90	0.94	1.00	0.98
2	$\tau_2 - \tau_1$	0.4	Large		0.03	0.18	0.64	0.11	0.57	0.86

Table 7

Performance of estimates of $\tau_2 - \tau_1$ and $g(\tau_2) - g(\tau_1)$ based on minimizing GCV or aGCV for a two change-point model.

Estimate	Baseline Curvature	Metric	With GCV for $ d $ of:			With aGCV for $ d $ of:		
			50	100	200	50	100	200
$\hat{\tau}_2 - \hat{\tau}_1$	Small	Bias (%bias)	-0.1368 (-34.2)	-0.0527 (-13.2)	-0.0043 (-1.1)	-0.0041 (-1.0)	0.0046 (1.2)	0.0023 (0.6)
		SD	0.1818	0.1447	0.0530	0.0899	0.0108	0.0065
		MSE	0.0518	0.0237	0.0028	0.0081	0.0001	<.0001
	Large	Bias (%bias)	-0.2644 (-66.1)	-0.1559 (-39.0)	-0.0449 (-11.2)	-0.2537 (-63.4)	-0.0906 (-22.7)	-0.0022 (-0.6)
		SD	0.1686	0.2019	0.1327	0.1053	0.1499	0.0530
		MSE	0.0983	0.0651	0.0196	0.0754	0.0307	0.0028
$\hat{g}(\hat{\tau}_2) - \hat{g}(\hat{\tau}_1)$	Small	Bias (%bias)	7.2 (-46.2)	7.2 (-20.2)	1.8 (-2.4)	0.1 (-0.9)	-0.8 (2.2)	-0.8 (1.0)
		SD	8.5	14.4	11.9	4.2	0.5	0.6
		MSE	124.8	258.6	145.0	17.6	0.9	1.0
	Large	Bias (%bias)	14.7 (-177.1)	19.5 (-68.8)	13.7 (-20.0)	15.3 (-183.7)	11.3 (-39.8)	2.9 (-4.2)
		SD	9.8	16.3	26.2	8.0	18.1	16.0
		MSE	314.1	644.4	871.4	298.3	456.6	263.2