



A Case Study in Pharmacologic Colon Imaging Using Principal Curves in Single-Photon Emission Computed Tomography

Brian S. Caffo, Ciprian M. Crainiceanu, Lijuan Deng & Craig W. Hendrix


To cite this article: Brian S. Caffo, Ciprian M. Crainiceanu, Lijuan Deng & Craig W. Hendrix (2008) A Case Study in Pharmacologic Colon Imaging Using Principal Curves in Single-Photon Emission Computed Tomography, Journal of the American Statistical Association, 103:484, 1470-1480, DOI: 10.1198/016214508000000832

To link to this article: <https://doi.org/10.1198/016214508000000832>



Published online: 01 Jan 2012.



Submit your article to this journal 



Article views: 77



Citing articles: 11 View citing articles 

A Case Study in Pharmacologic Colon Imaging Using Principal Curves in Single-Photon Emission Computed Tomography

Brian S. CAFFO, Ciprian M. CRAINICEANU, Lijuan DENG, and Craig W. HENDRIX

In this article we are concerned with functional imaging of the colon to assess the kinetics of microbicide lubricants. The overarching goal is to understand the distribution of the lubricants in the colon. Such information is crucial for understanding the potential impact of microbicides on human immunodeficiency virus transmission. The experiment was conducted by imaging a radiolabeled lubricant distributed in the subject's colon. The tracer imaging was conducted via single-photon emission computed tomography, a noninvasive, in vivo functional imaging technique. We have developed a novel principal curve algorithm to construct a three-dimensional curve through the colon images. The algorithm was tested and debugged on several difficult two-dimensional images of familiar curves where the original principal curve algorithm does not apply. The final curve fit to the colon data is compared with experimental sigmoidoscope collection.

KEY WORDS: Curve fitting; Medical imaging; Pharmacology; Smoothing; SPECT.

1. INTRODUCTION

Antimicrobial lubricants could have a profound impact in the spread of sexually transmitted diseases, such as human immunodeficiency virus (HIV). Understanding the efficacy of these treatments requires understanding the anatomic distribution and the kinetics of viral-mixed semen, along with those of the lubricant within the lumen of sexually receptive organs, ideally under typical forces of vaginal or anal intercourse. If the lubricant had a low penetration or its distribution did not comport with the location of the viral-mixed semen, then its efficacy would be in question. But the complexity of the physical systems involved prohibit the analytical study of such questions. Instead, they must be evaluated experimentally.

This article considers data arising from one of the first experiments to empirically evaluate antimicrobial lubricants. As foundational research, its aims are focused, and thus we solely consider the problem of estimating the distributional penetrance of the lubricant after anal intercourse.

We now describe the experiment in greater detail. An over-the-counter lubricant was mixed with a radioactive tracer; this served as a surrogate marker for the antimicrobial treatment. This lubricant was inserted into a subject's colon, and physical forces similar to those of anal coitus were subsequently used to distribute the liquid. The distribution of the tracer/lubricant mixture was then imaged using a single-photon emission computed tomography (SPECT) scanner.

SPECT is a noninvasive, in vivo functional imaging technique. SPECT images arise from the application of computed tomography (CT) techniques to projections obtained by counting emitted photons from a radioactive tracer placed in the body. SPECT images are of lower resolution than those obtained from other modalities, such as X-ray, CT, magnetic resonance imaging (MRI), and positron emission tomography (PET). However, SPECT represents a relatively low-cost

method for obtaining functional imaging, that is, it offers the ability to image the body as it functions through biological interactions with the tracer.

After scanning, a followup procedure was performed in which a sigmoidoscope was used to sample the tracer/lubricant mixture at various positions within the colon. For our purposes, the sigmoidoscope was a lubricated tube with an optical fiber and an additional channel for a mechanical sampling device, which in this experiment was a brush inside a casing.

Before discussing our statistical goals, we briefly describe relevant colon anatomy. A diagram of the colon is given in Figure 1. From the anus, the next structure is the anal canal, then rectum. The sigmoid colon follows by traveling a highly variable course, anteriorly and slightly inferiorly, from right to left, where it transitions into the descending colon, which travels up the left side of the abdomen.

A few sample transverse slices of the raw (reconstructed) image data are shown in Figure 2. (A transverse plane divides the body into upper and lower regions.) As can be seen, the raw data are difficult to interpret or to get any sort of anatomic bearings. Transverse slices of the data after thresholding and overlaying on the accompanying X-ray CT image are given in Figure 3. Here the X-ray CT and SPECT are collected at the same time and registered in the same space by the scanner software. The top four plots, show the hip bones on the X-ray CT image and the tracer distribution in the descending colon. The middle four plots show the tracer distribution around the lower portion of the descending colon and the sigmoid colon, whereas the lower four plots display the tracer near the rectum.

The primary goal of this investigation is to provide a semiautomated procedure to estimate the concentration of the lubricant by distance in the colon using the SPECT image. With reliable information from the imaging data, the sigmoidoscope collection would not be necessary in future studies. In addition, the sigmoidoscope itself displaces the liquid, and thus has limitations for measuring the tracer distribution.

We focus our attention entirely on the SPECT images, because interest lies only in the distributional penetrance of the tracer/lubricant mixture. The accompanying X-ray CT images convey very little relevant additional information, because they

Brian S. Caffo is Associate Professor (E-mail: bcaffo@jhsph.edu) and Ciprian M. Crainiceanu is Assistant Professor, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe Street, Baltimore, MD 21205. Lijuan Deng is Researcher, Boston Scientific Corporation, Natick, MA 21205. Craig W. Hendrix is Director of Drug Development Unit, Associate Professor, Division of Clinical Pharmacology and Infectious Diseases, Johns Hopkins Medicine, 615 N. Wolfe Street, Baltimore, MD 21205. This work was supported by the National Institutes of Health grant K25EB003491 and the Centers for Disease Control and Prevention grant 200-2001-08015.

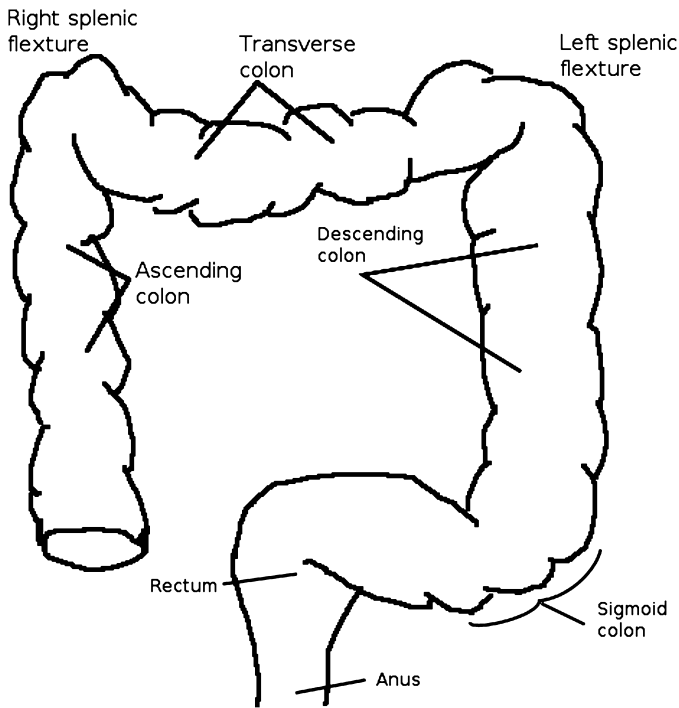


Figure 1. Anatomic diagram of the colon.

were collected primarily as attenuation maps for the SPECT reconstruction. (Attenuation maps are used to perform the reconstruction algorithm that converts the raw data to images.) As such, they display almost no colon anatomy and largely high-light skeletal structures. In principle, other imaging modalities, such as high-resolution MRI, could be used as anatomic complements to the SPECT images. It would be difficult to register these anatomic scans with the SPECT images, however. Moreover, they would have to be collected at a different time, at which point the distribution of the tracer/lubricant mixture, as well as the colon shape itself, may have changed dramatically.

To solve the problem, we develop an algorithm based on principal curves (Hastie and Stuetzle 1989) to perform the fitting. The resulting algorithm incorporates constrained endpoints (of the curve), constrained interior points, and the image intensities. Moreover, a novel “molding” procedure is proposed that greatly improves the algorithm’s ability to fit complex data structures. The algorithm is tested and debugged on a collection of difficult two-dimensional (2-D) images that the original, unmodified principal curve algorithm cannot fit.

With one subject, the scientific contribution of this work is largely a proof of concept. None the less, we show that the al-

gorithm appears to work quite well and requires little user input. The excellent results of the algorithm have led to the potential policy change of eliminating the costly and invasive sigmoidoscopy collection.

The article is organized as follows. Section 2 overviews the data. Section 3 covers curve fitting algorithms, beginning with a literature review. Section 3.1 covers curve characterizations, Section 3.2 discusses principal curves, and Section 3.3 presents our modified algorithm. Section 4 tests these algorithms on constructed 2-D data, whereas Section 5 applies the algorithms to the real SPECT data. Section 6 concludes with a discussion.

2. DATA

Ten milliliters of radiolabeled lubricant [^{99m}Tc -sulfur colloid mixed with K-Y Jelly (Johnson and Johnson, New Brunswick, NJ)] was injected into the subject’s colon. Following rectal administration of the radiolabeled gel, the subject underwent simulated receptive anal intercourse. The experimental paradigm was designed to mimic the typical forces that would influence the lubricant’s distribution.

The patient was imaged on a dual-head VG SPECT-CT imaging system (GE Medical Systems, Waukesha, WI) equipped with a low-end CT unit (Hawkeye). Accompanying each SPECT image, an X-ray CT image also was collected for anatomic reference and reconstruction of the SPECT image. The image was reconstructed using the ordered-subsets EM algorithm (Hudson and Larkin 1994) and filtered as provided with the scanner software (General Electric eNTEGRA workstation, version 1.04; GE Medical Systems).

The SPECT data are represented as a three-dimensional (3-D) array. In our application, the dimension of the array is $128 \times 128 \times 128$. Each voxel (no 3-D pixel) represents a 3.45-mm^3 physical area. The image intensity values are proportional to the concentration of the tracer at that location. The intensity values range from 0 to 187, although we note that the absolute scale is somewhat arbitrary, because the image was rescaled during reconstruction.

After imaging, a sigmoidoscope was used to collect physical concentration measurements. Samples were taken at 5-cm intervals up to 40 cm. The study was approved by the Johns Hopkins Institutional Review Board, and informed written consent was obtained from all subjects.

3. CURVE FITTING ALGORITHMS

Calculating centerlines for such anatomic structures as blood vessels, neurons, and colons has a rich history in the computer vision and medical image processing literature. Much of the

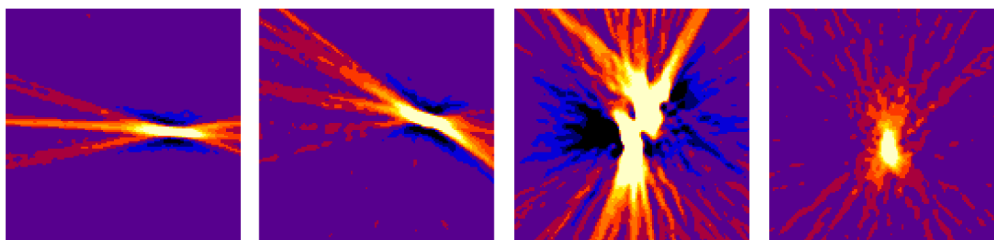


Figure 2. Example raw image data axial slices. The left two show the tracer distribution in axial slices of the descending colon while the third from the left shows distribution around the sigmoid colon. The final plot shows the distribution near the anus.

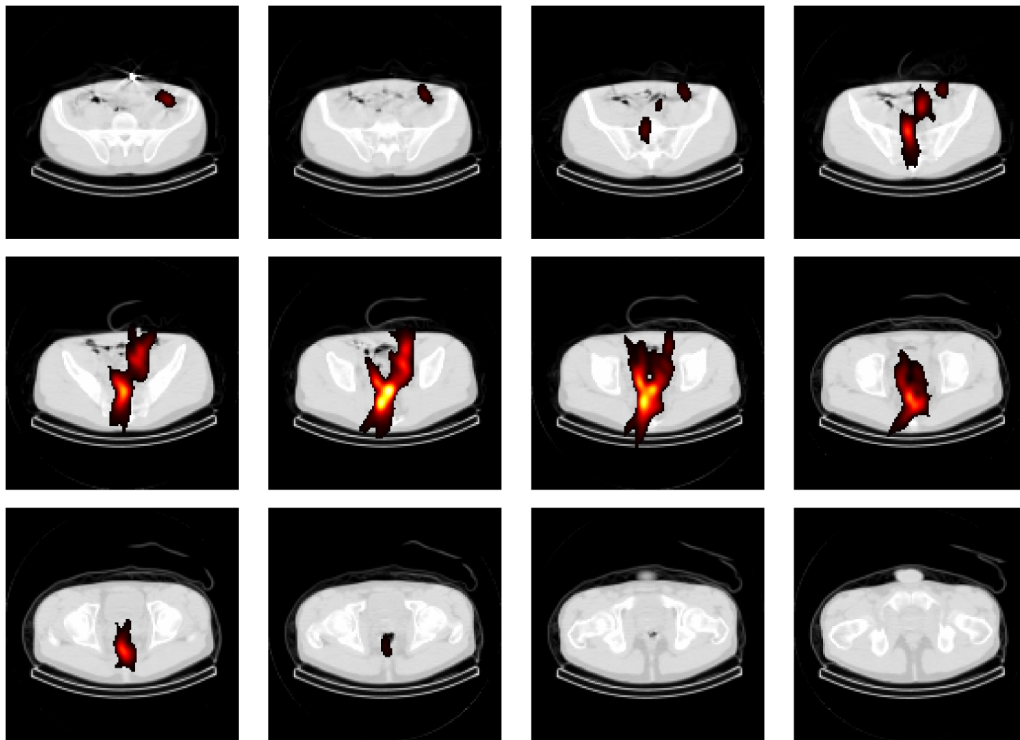


Figure 3. Processed SPECT images by transverse slices. Images proceeding from the upper left to the lower right proceed inferiorly (toward the feet). The selected images are spaced roughly 10 cm apart, with each image representing a 3.45-mm-thick slice. For anatomic reference, the hip bones are clearly visible in the upper-left plots.

research involving the colon has been applied to X-ray CT images for the purpose of finding polyps; for example, McFarland et al. (1997) presented a semiautomated centerline extraction algorithm, and Samara, Fiebich, Dachman, Doi, and Hoffmann (1998) proposed a semiautomated voxel search algorithm for centerline construction.

Another class of methods uses Dijkstra's algorithm (see Dijkstra 1959), in which voxels and intensity values are treated as a networked graph. Search algorithms are used to find minimal distance paths through the graph (Hong, Muraki, Kaufman, Bartz, and He 1997; Chiou, Kaufman, Liang, Hong, and Achiotou 1998; Bitter, Kaufman, and Sato 2001). Wan, Dachille, and Kaufman (2001) used distance fields to compute central paths, whereas Chaudhuri, Khandekar, Sethi, and Kalra (2004) used similar concepts, but with a different distance measure. Cohen and Kimmel (1997) derived a path-tracking routine in 2-D images by calculating a minimal path between two fixed endpoints.

Deschamps and Cohen (2001) extended the so-called "fast marching" algorithm to 3-D objects to extract a minimal path through the colon. Ge, Stels, Wang, and Vining (1999) used a fast topological thinning algorithm to generate a 3-D skeleton of a binary colon volume, which was then pruned. Bouix, Siddiqi, and Tannenbaum (2003) used a technique called medial surface extraction to compute a centerline curve, which was then pruned. Finally, Telea and Vilanova (2003) gave a level-set algorithm for building a colon centerline.

Having mentioned only a subset of the related work on calculating colon centerlines, we emphasize that our problem differs markedly from those approaches in six key as-

pects: (a) This work considers SPECT images rather than high-resolution anatomic X-ray CT images; (b) the image is of the tracer/lubricant mixture, not of colon anatomy; the tracer may be at lower concentrations at different areas of the colon, where the centerline is still desired, so that image intensities are not the primary concern (as opposed to the analysis of X-ray CT images); (c) unlike the colon anatomy, the tracer distribution can be discontinuous, interrupted by stool and gas; therefore, techniques requiring connected graphs will not apply; (d) the scientific application is extremely novel, with no comparable experiments; (e) there is a direct comparison measurement available in the sigmoidoscope data; and (f) our approach and characterization of the problem is more statistically oriented than existing algorithms.

Statistical approaches in the area of curve fitting are few, as opposed to the embarrassment of riches available for fitting proper functions. Here we discuss the principal curve algorithm, a fundamental algorithm in the area of curve fitting. However, we found that when unmodified, this algorithm cannot handle complex images. Moreover, it does not incorporate image intensities. Consequently, we propose a modified algorithm with several notable benefits.

3.1 Characterization of Curves

We characterize the problem as follows. Let $f: \mathbb{R} \rightarrow \mathbb{R}^3$ be defined so that $f(t) = \{f^x(t), f^y(t), f^z(t)\} = \{X(t), Y(t), Z(t)\}$. Here f is the 3-D position of a curve through the colon at a latent argument $t \in [0, 1]$. The value of $f(t)$ represents the coordinate points in the image; thus the curve in 3-D space is then the projection of $f(t)$ over t . As such, this is a standard

representation of a curve in 3-D space (see, e.g., Thorpe 1979). The constraint that t resides in $[0, 1]$ is arbitrary and used for identifiability. Throughout, we conceptualize $f(t)$ as the position of a curve at a “time point” t . We emphasize that this is only for pedagogy; the image has no temporal component. Conceptually considering time highlights the identifiability issue that two functions traveling the same path at different rates yield the same curve.

The general requirements for the fitted curve are that (a) the curve should follow a smooth path through the center of the tracer distribution; (b) the curve must be able to traverse empty spaces, where the continuity of the tracer is interrupted; (c) user-specified starting and ending points should be incorporated; (d) the algorithm should be flexible enough to allow for other user specified points through which the curve must travel; and (e) the curve should prefer traversing higher-intensity points to lower-intensity-points. We note that point (e) must be factored in carefully, because excessively emphasizing the image intensities can result in poor fits. Consider the example data, in which the highest-intensity points are aggregated near the anus. Any algorithm that minimizes the curve integral of the intensities through the image will focus on spending as much time as possible in these areas, contrary to goals (a) and (b).

To achieve these requirements, we build on the principal curve algorithm of Hastie and Stuetzle (1989). Our modifications grow out of difficulty in getting reasonable fits from the original algorithm for challenging imaging problems and the fact that it was not designed to incorporate image intensity values. Our approach does require user-specified endpoints for the curve, mostly because of the nature of the scientific problem. But a novel molding procedure automates the selection of a starting curve, starting at a line connecting the two specified endpoints and gradually capturing finer details of the curve. Moreover, we provide an elegant solution to accounting for differences in image intensity that does not suffer from the problems outlined earlier.

We describe the principal curve algorithm and our modifications in the next two sections. But before discussing the algorithms, we discuss necessary preprocessing and some basic starting points for curve fitting through images.

A first step is to threshold the image. This is done both to remove background noise and to reduce the number of points included in the fitting. Further acceleration is obtained by randomly sampled points that survive the threshold. Points should be sampled uniformly among those that are above the threshold. Sampling points with probabilities relative to image intensities gives poor results, because areas of the colon with low intensities are omitted.

Notationally, let $\{(X_i, Y_i, Z_i)\}_{i=1}^n$ be the sampled locations. To be specific, (X_i, Y_i, Z_i) represents the lattice value of a single sampled point above the original threshold. The index ordering, i , is arbitrary; that is, the points can be selected in any order. Figure 4 shows the sampled colon data. Let $\{t_i\}_{i=1}^n$ represent a collection of (unknown) associated time variables for the sampled points. Throughout, dropping the subscript represents the vector of the collection of variables, such as $t = (t_1, \dots, t_n)'$.

Much of the challenge of this problem is to appropriately assign values for the time points. Tabling this issue for the moment, presume that these points were known. Then a smooth

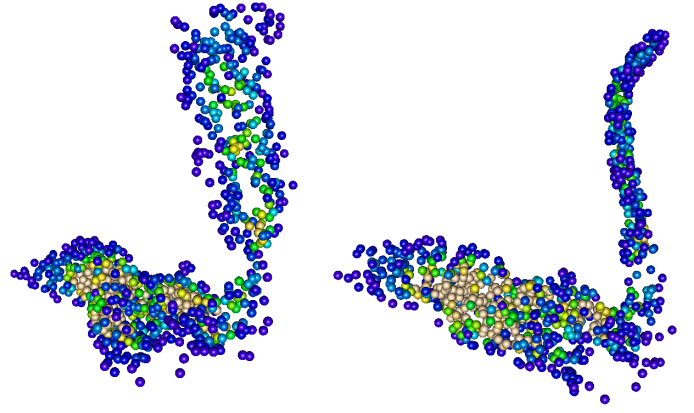


Figure 4. Sampled colon images in two orientations. Colors represent image intensity.

curve through the data can be easily fit. Our approach uses the three spline equations.

$$\begin{aligned}
 E[X_i] &= f^x(t_i) \\
 &= \beta_0^x + \beta_1^x t_i + \beta_2^x t_i^2 + \beta_3^x t_i^3 + \sum_{k=1}^K b_k^x (t_i - \xi_k)_+^3, \\
 E[Y_i] &= f^y(t_i) \\
 &= \beta_0^y + \beta_1^y t_i + \beta_2^y t_i^2 + \beta_3^y t_i^3 + \sum_{k=1}^K b_k^y (t_i - \xi_k)_+^3, \\
 E[Z_i] &= f^z(t_i) \\
 &= \beta_0^z + \beta_1^z t_i + \beta_2^z t_i^2 + \beta_3^z t_i^3 + \sum_{k=1}^K b_k^z (t_i - \xi_k)_+^3,
 \end{aligned} \tag{1}$$

where $\{\xi_k\}_{k=1}^K$ are knots placed at equally spaced quantiles of $\{t_i\}_{i=1}^n$. Here $K + 4$ is the degrees of freedom of the smoother for each dimension. The benefits of using regression splines are many, including easy specification of the basis and easy calculation of derivatives, which are required later.

3.2 Principal Curves

The principal curve algorithm is a general method for fitting a curve through data residing in an arbitrary dimensional space. A curve, $f(t)$, is said to be a principal curve if for each data value, say (X_i, Y_i, Z_i) , the curve at t_i satisfies

$$f(t_i) = E[(X_i, Y_i, Z_i) \mid \text{closest point of } f \text{ to } (X_i, Y_i, Z_i) \text{ occurred at time } = t_i].$$

This recursive form of self-consistency motivates an algorithm. Suppose that a starting collection of time points is given. The algorithm is then as follows:

1. Approximate the principal curve by a scatterplot smoother, regressing $\{(X_i, Y_i, Z_i)\}_{i=1}^n$ on $\{t_i\}_{i=1}^n$, as in eq. (1).
2. Update the time points by redefining t_i as the time point on the curve closest to (X_i, Y_i, Z_i) for $i = 1, \dots, n$; that

is, define

$$t_i = \arg \min_{t \in [0,1]} \{X_i - \tilde{f}^x(t)\}^2 + \{Y_i - \tilde{f}^y(t)\}^2 + \{Z_i - \tilde{f}^z(t)\}^2)^{1/2}$$

for $i = 1, \dots, n$, where $\tilde{f}^x(t)$, for example, represents the current estimate of $f^x(t)$.

These steps are then iterated until convergence. Of course, instead of a starting at a collection of time points, a starting curve could be given, in which case the algorithm simply starts at step 2. For example, the algorithm could be started at the first principal component line through the data. In our setting, we start the algorithm at a line connecting user-specified endpoints.

Conceptually, the steps can be described as follows. First, the data are projected onto the three planes considering time and each spatial dimension; that is, the data (t, X) , (t, Y) , and (t, Z) are considered, where, for example, (t, X) refers to the collection of $\{t_i\}_{i=1}^n$ and $\{X_i\}_{i=1}^n$ data points. Second, a scatter-plot smoother is fit in those three planes to calculate an updated curve. Next, the orthogonal projections of the data points onto the curve are calculated. The time points associated with the projections onto the curve are used to then update the latent time variable. Although this strategy has considerable intuitive appeal, a formal proof of convergence is not available; however, it has been used successfully in (unrelated) image processing settings (Banfield and Raftery 1992).

3.3 A Modified Principal Curve Algorithm

In this section we discuss generalizations of the principal curve algorithm that allow it to be viable in our scientific setting. These modifications include allowing for user-specified endpoints and interior points of the function; incorporating the image intensities; warming up, or molding, the algorithm to achieve better fit; performing a grid search for the minimization in the second step of the algorithm; and incorporating a stopping rule based on relative mean squared error.

Consider the incorporation of user-specified endpoints. Notationally, suppose that (x_0, y_0, z_0) and $t_0 = 0$ and $(x_{n+1}, y_{n+1}, z_{n+1})$ and $t_{n+1} = 1$ are given. A modification of the algorithm that forces the curve to start and end at these points (respectively) simply forces the constraint $0 \leq t_i \leq 1$ for $i = 1, \dots, n$ and adds the relevant Lagrange multiplier terms to eq. (1); that is, the multiplier terms force $f^x(0) = x_0$, $f^x(1) = x_{n+1}$, $f^y(0) = y_0$, $f^y(1) = y_{n+1}$, $f^z(0) = z_0$, and $f^z(1) = z_{n+1}$. Specifically, let W be the basis matrix for eq. (1). (Note that the same basis is used for all three dimensions.) As in eq. (1), we used regression splines with knots at equally spaced points between 0 and 1, although any linear smoother applies. Let \tilde{W} be the basis evaluated at the constrained values of t , and let \tilde{x} , \tilde{y} , and \tilde{z} be vectors of the constrained values. Then the goal is to maximize the least squares equations for the models

$$E[X] = W\beta^x, \quad E[Y] = W\beta^y, \quad \text{and} \quad E[Z] = W\beta^z$$

subject to the constraints

$$\tilde{W}\beta^x = \tilde{x}, \quad \tilde{W}\beta^y = \tilde{y}, \quad \text{and} \quad \tilde{W}\beta^z = \tilde{z}.$$

The fitted values are then (see, e.g., Searle 1971)

$$\begin{aligned} \hat{\beta}_c^x &= \hat{\beta}^x - (W'W)^{-1} \tilde{W} \{ \tilde{W}'(W'W)^{-1} \}^{-1} (\tilde{W}\hat{\beta}^x - \tilde{x}), \\ \hat{\beta}_c^y &= \hat{\beta}^y - (W'W)^{-1} \tilde{W} \{ \tilde{W}'(W'W)^{-1} \}^{-1} (\tilde{W}\hat{\beta}^y - \tilde{y}), \\ \hat{\beta}_c^z &= \hat{\beta}^z - (W'W)^{-1} \tilde{W} \{ \tilde{W}'(W'W)^{-1} \}^{-1} (\tilde{W}\hat{\beta}^z - \tilde{z}), \end{aligned} \quad (2)$$

where $\hat{\beta}^x$, $\hat{\beta}^y$, and $\hat{\beta}^z$ are the unconstrained fitted coefficients.

It is worth noting that constraining the endpoints as such implies that the final fitted curve will not satisfy the self-consistency property. In particular, points near the fixed endpoints will have their values of t updated not by the orthogonal projection to the curve extended in perpetuity, but rather by the closest value to the constrained ends.

This solution can be adapted to incorporate other constrained points along the curve. But unlike the endpoints, the corresponding values of t_i are not known; therefore, these associated time points must be estimated in step 2 of the principal curve algorithm, and the matrix of constrained time points, \tilde{W} , must be updated. We note that this procedure must be used with care, because identifiability problems can occur with constrained interior points. A minimum requirement is that there generally needs to be more parameters than constrained points; for example, one cannot constrain a line to traverse three points (unless those points fall exactly on a line).

Consider incorporating the image intensities. Specifically, define Σ^{-1} to be a matrix with some function of the image intensities corresponding to the points $\{(X_i, Y_i, Z_i)\}_{i=1}^n$ along the main diagonal and 0's elsewhere. Then consider the weighted regression version of eq. (2),

$$\begin{aligned} \hat{\beta}_c^x &= \hat{\beta}^x - (W'\Sigma^{-1}W)^{-1} \tilde{W} \{ \tilde{W}'(W'\Sigma^{-1}W)^{-1} \}^{-1} (\tilde{W}\hat{\beta}^x - \tilde{x}), \\ \hat{\beta}_c^y &= \hat{\beta}^y - (W'\Sigma^{-1}W)^{-1} \tilde{W} \{ \tilde{W}'(W'\Sigma^{-1}W)^{-1} \}^{-1} (\tilde{W}\hat{\beta}^y - \tilde{y}), \\ \hat{\beta}_c^z &= \hat{\beta}^z - (W'\Sigma^{-1}W)^{-1} \tilde{W} \{ \tilde{W}'(W'\Sigma^{-1}W)^{-1} \}^{-1} (\tilde{W}\hat{\beta}^z - \tilde{z}), \end{aligned}$$

where now $\hat{\beta}^x$, $\hat{\beta}^y$, and $\hat{\beta}^z$ are the unconstrained weighted fitted coefficients, for example,

$$\hat{\beta}^x = (W'\Sigma^{-1}W)^{-1} W'\Sigma^{-1}X.$$

Note that the fit is invariant to scalings of the image intensities. Although we used the raw image intensities to define Σ^{-1} , other definitions could be used to adapt the fit; for example, using the square of the intensities will put greater emphasis on high-intensity points, whereas using the square root will put less emphasis on these points. Users of the algorithm should be advised that it may be useful to trim the intensities to avoid excessive affects on outlying points. This was not necessary in our application.

The most important modification of the principal curve algorithm lies in choosing an appropriate starting curve. We have found that getting the curve in the correct neighborhood is crucial for fitting complicated structures. Therefore, we use a series of warmup runs, using a few degrees of freedom (small K) to obtain starting values that correctly model the gross features of the data. We start the warmup runs at a line connecting the two specified endpoints. For each value of K , the modified principal curve algorithm then can be run to convergence. After convergence, the estimated curve is used as a starting value for a subsequent warmup run of the algorithm with K increased. A final

run with the desired value of K uses the result of the warmup runs as the starting value.

This method tends to mold the curve to the gross features of the data before moving on to the finer ones; therefore, we refer to this procedure as “molding.” We have found that molding is the single most important aspect in obtaining reasonable fits. Furthermore, the most critical aspect of using these warmup runs is systematically increasing the degrees of freedom. Whether or not the algorithm is run until convergence within each warmup value of K does not seem to affect results, other than to slow down the algorithm; thus only one iteration for each value of K was used to warmup to the final run.

Two algorithms were compared to perform the maximization required in step 2 of the principal curve algorithm. First, a modified BFGS algorithm was used that can accommodate constrained endpoints (Byrd, Lu, Nocedal, and Zhu 1995). Second, a simple grid search was used. Both techniques were effective, although the grid search was quite a bit faster; thus the results presented here are from that technique using 1,000 or 10,000 grid points between 0 and 1. All calculations were performed in the R programming language (R Development Core Team 2006).

With regard to stopping the algorithm, we suggest the relative change in the mean squared error of the estimated function summed across the three dimensions. Specifically, define

$$SMSE = \frac{1}{n} \sum_{i=1}^n [\{X_i - f^x(t_i)\}^2 + \{Y_i - f^y(t_i)\}^2 + \{Z_i - f^z(t_i)\}^2],$$

and thus the required stopping criteria,

$$\frac{|SMSE^{current} - SMSE^{old}|}{SMSE^{old}},$$

to be less than a desired tolerance.

4. IDEALIZED TWO-DIMENSIONAL CURVES

To build intuition, the algorithm was applied to highly idealized 2-D images created using image processing software. Figure 5 displays the raw images. Note that they are of constant intensity; thus the weight matrix, Σ^{-1} , was set to an identity matrix. These images were selected because they have several interesting features, such as sharp turns, and are significantly more difficult to fit than the actual data set. Many of these test images were motivated by Kegl, Krzyzak, Linder, and Zeger (2000). The 512×512 pixel images were created in GIMP (GNU Image Manipulation Program; www.gimp.org) by free-hand drawing with a mouse.



Figure 5. Sample images.

Figure 6 shows the sample curves fitted to the spiral using $K = 20$ to define the degrees of freedom (bottom plots), as well as the starting line (top plots). The fixed endpoints are shown as blue points. In each row, the leftmost plots show the fitted curve. The next plots show the (X, Y, t) data and the fitted curve in 3-D space with the (X, Y) data shown projected onto the horizontal plane. The final two plots on the right show the (t, X) and (t, Y) data and the fitted curve projections. Recall that the smoothing portion of the algorithm fits a smoother to these two projections.

The same plots for the other examples are given in Figure 12. In all of the cases, the fit is obviously good. Because of the sampling and the fast grid-search maximization, the algorithms take only roughly 30 seconds to run (on a Pentium dual-core, 2.16-GHz laptop with 2 GB of RAM). Moreover, the process of molding the fit tends to get the curve in a very close neighborhood of the desired limit, thus requiring very few iterations for the final run.

We also explored the use of constrained interior points. Figure 7 shows the results of the fitted models using constrained interior points for the “3” and spiral images (shown in green). Here $K = 20$ was used to define the final degrees of freedom of the smoothers. For the “3” image, the constrained interior points were nicely incorporated into the smooth curve. In contrast, for the spiral, the constrained interior points interfered with the fit, forcing the curve to make unnatural bends to incorporate them. We find that unless constrained interior points are carefully chosen and implemented, their use can greatly reduce the quality of the resulting fit. This can arise from a conflict between our opinion of where the curve should traverse and that of the principal curve algorithm’s. In addition, constrained interior points can prevent the molding procedure from exploring poorly fitting functions that model only gross features but lead the algorithm to a space of curves that fit the finer ones.

4.1 Comparison With Alternative Methods

In this section we briefly compare the modified principal curve approach with related methods and particularly the principal curve algorithm without the molding procedure, but still constraining endpoints. We note that the original principal curve algorithm without constrained endpoints performs very poorly, because it was designed for more variable data.

We first consider an alternative algorithm for curve fitting discussed by Deng (2007). Here the time points were estimated using an objective function that sought to maximize the curve integral of the fitted curve through the image, subject to a length penalty. A stochastic search algorithm was used that made assumptions on the latent time variables. Specifically, a “constant speed” assumption was made that spaced the values

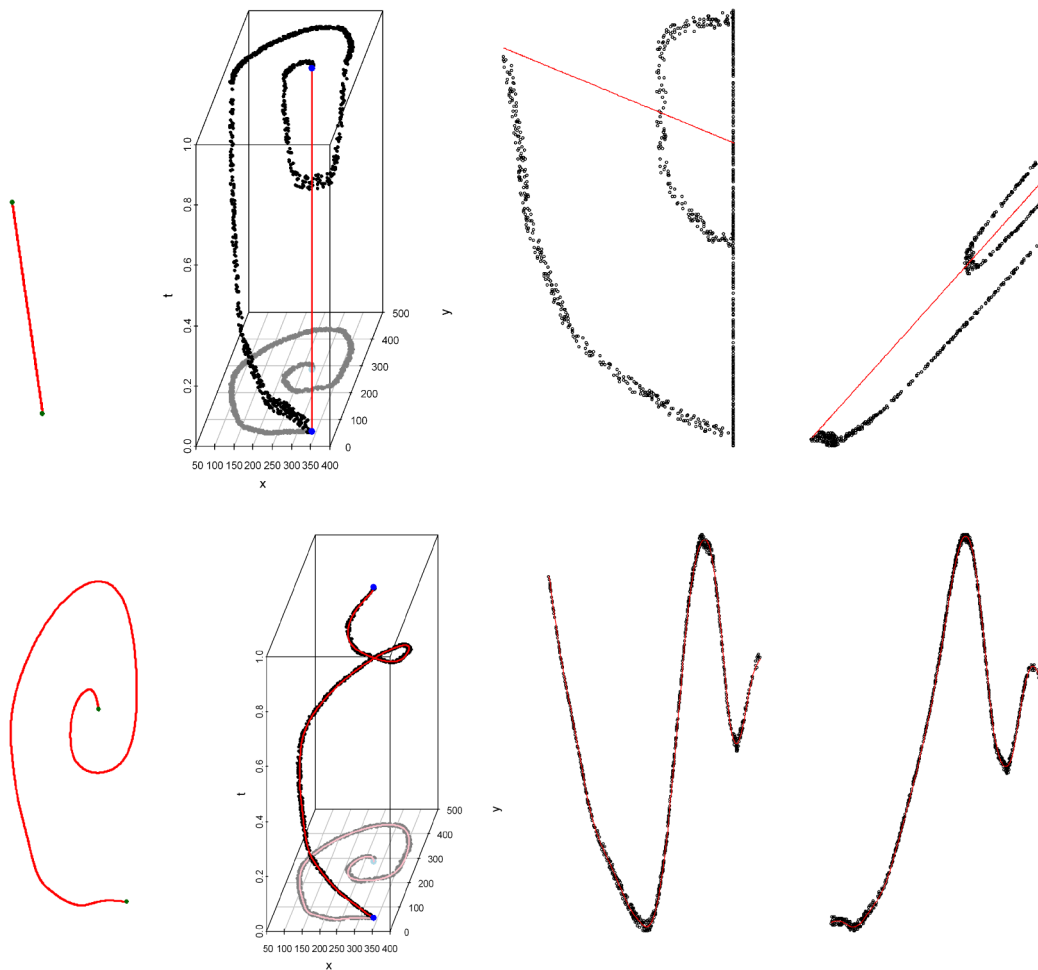


Figure 6. Fitted images for the spiral. The top panel plots the line interpolating the two (blue) specified endpoints is given; the bottom panel plots the fits after being run to 20 degrees of freedom. The first plots on the left show the fitted curve. The second show the data and the fitted curve, both projected onto the (X, Y) plane, as well as the full data with the unobserved variable t . The latter two plots show the (t, X) and (t, Y) projections of the data and the fitted curve.

equally of t_i between 0 and 1. Moreover, it was presumed that the values of t_i were monotonic in the sagittal, axial, or transverse direction. Although the algorithm performed quite well when these assumptions are met, it failed with mild departures from the monotonicity assumption. Figure 8 shows the problem in an idealized 2-D colon-type image created using image

processing software. Specifically around the sigmoid curve, the monotonicity assumption fails, and thus a poor fit is obtained. For figures such as the spiral, where monotonicity fails regardless of the rotation of the image, this method does not apply.

To further emphasize the importance of the molding procedure, Figure 9 shows fitted curves obtained using fixed degrees of freedom starting from a line for $K = 3, 5, 10$, and 20, for the four idealized images. In all cases, the curve is struggling to hit as many points as possible, to minimize the orthogonal projections onto the curve. The curve does not have the ability to fit gross features and so immediately takes as complex a shape as possible. This problem could be overcome by starting the curve using several constrained interior points, as described in Section 3.3. Then perhaps these could be quickly discarded, because a few iterations would place the curve in the correct neighborhood. This would require excessive user input, however, and the molding procedure automatically avoids this problem.

5. APPLICATION TO THE MICROBICIDE IMAGING STUDY

After building intuition with the idealized 2-D data, we applied the methods to the SPECT data. The SPECT data were

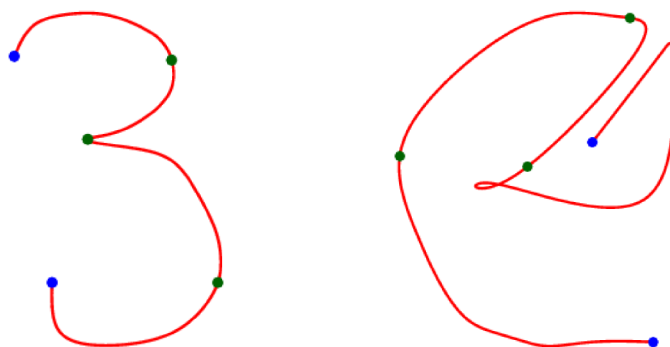


Figure 7. Results of model fits to the “3” and spiral using constrained interior points. The constrained interior points are in green, and the constrained endpoints are shown in blue.

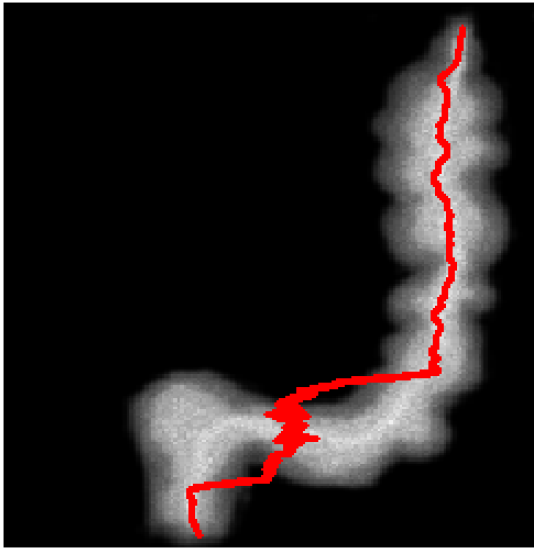


Figure 8. Stochastic search curve fitting algorithm of Deng (2007) used in a idealized 2-D colon image.

processed (reconstructed, filtered, thresholded, and sampled) as outlined in the previous sections. The algorithm was run using $K = 5$ to define the smoother degrees of freedom. The low number was used because the tracer did not extend past the descending colon and followed a very smooth function, with only mild complexity near the sigmoid colon.

A total of 1,000 points were sampled from the image, and a grid search using 10,000 equally spaced points between 0 and 1 was performed in the second step of the principal curve algorithm. A modified BFGS algorithm also was used, although it produced no differences in the results. The constrained endpoints were selected by comparisons with bone structure from the X-ray CT image. No constrained interior points were needed, because the lubricant distribution was fairly continuous.

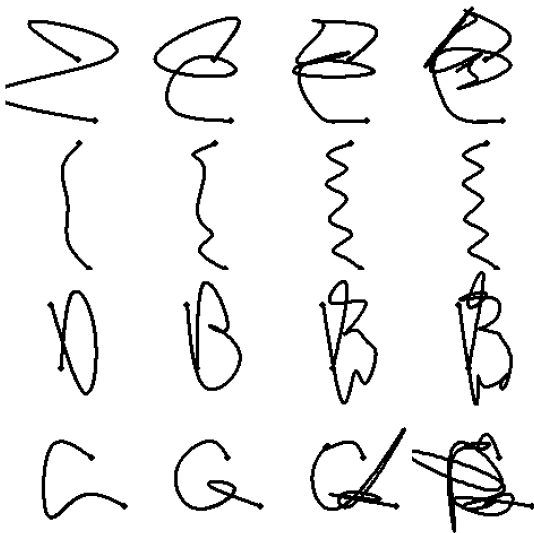


Figure 9. Results of principal curve fits not using the molding procedure. From left to right, the plots use 3, 5, 10, and 20 degrees of freedom. From top to bottom, they are the spiral, sawtooth, “3” and “a” images from Figure 5.

Figure 10 shows the fit in three orientations, two of which display the sampled data, with one showing the intensities and the other displaying the orthogonal projections onto the curve. The shape of the curve was evaluated by physicians, who claimed that it closely followed the knowledge of colon anatomy.

The fitted curve was compared with the sigmoidoscope results. The distance along the curve was calculated as

$$\int_0^T \sqrt{\left(\frac{d}{dt} \hat{f}^x(t)\right)^2 + \left(\frac{d}{dt} \hat{f}^y(t)\right)^2 + \left(\frac{d}{dt} \hat{f}^z(t)\right)^2} dt,$$

where $\hat{f}^x(t)$ denotes the final estimate for $f^x(t)$. Because of the simple form used for the basis, f has easily calculated derivatives. But we did not find a closed form for the resulting integral, which was evaluated numerically. Image intensities were then calculated along the curve using the original, unsampled image in neighborhoods of size 1, 2, 4, and 8 voxels. Here a neighborhood of size c is defined as the box centered at the current voxel with sides $(2c + 1)$. The concentration was estimated as the sum of the image intensities in the neighborhood, divided by the number of voxels.

Figure 11 displays the results, along with the sigmoidoscope data for comparison. Because image processing arbitrarily scales the reconstructed SPECT image, the intensities used in the concentration calculation are proportional only to the actual tracer concentration. Thus, all of the curves were divided by their maximum value, to obtain a unit-free comparison.

Because the SPECT image does not indicate the exact location of the start of the anus, the curves were calibrated so that their maxima matched that of the sigmoidoscope data. There is some degree of error however; inspection of the SPECT image clearly indicates the maximum concentration in the sigmoid colon. In contrast, the maximum for the sigmoidoscope data, 35 cm (1.15 ft), is too far into the colon for the sigmoid. Along with the maximum being too distal, the sigmoidoscope data also appear to be more diffusely distributed than the SPECT data. This is consistent with the hypothesis that repeated application of the scope distributes the tracer/lubricant mixture, thus giving an inaccurate picture of penetrance. Moreover, unlike the sigmoidoscope, the SPECT image can reasonably detect the tracer distribution well into the colon. As such, the image processing tools presented here offer an extremely valuable approach to quantifying the tracer/lubricant distribution without this source of error. Currently we are working with scientific collaborators to obtain accurate measures of the starting distance of the tracer distribution in the SPECT image.

6. DISCUSSION

In this article we have considered the difficult problem of 3-D curve fitting in a novel, scientifically important study. The success of the algorithm will have a direct benefit on the ability to estimate distributional penetrance in SPECT colon imaging. The algorithm’s performance has inspired the possibility of replacing costly and invasive sigmoidoscope collection with statistical image processing tools.

In terms of the future of this scientific application, researchers are now considering dual-isotope studies, in which surrogates for both the lubricant and the HIV-infected semen

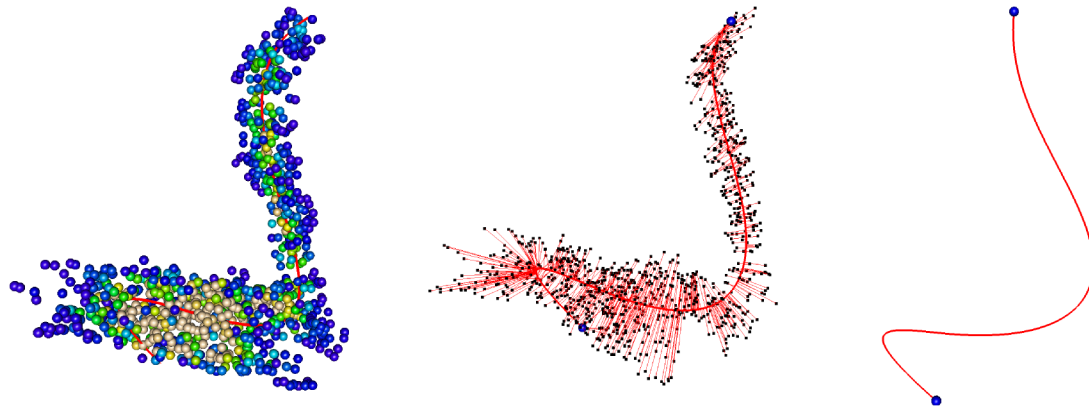


Figure 10. Fitted colon curve in three orientations. The first two plots also show the sampled data points, one with image intensities and another with the end orthogonal projections. The last plot shows the curve and the (blue) constrained endpoints.

are simultaneously imaged after being tagged with tracers at different energy levels. If successful, this experimental technique will offer accurate experimental validation of antimicrobial lubricants.

We also note that the algorithm has further potential scientific applications in unrelated areas. Specifically, it's currently being evaluated for its viability in estimating white matter tracts from brain MRI images.

Our algorithm has proven to be a nearly ideal candidate for obtaining a centerline through the distribution of the SPECT tracer. Techniques for constraining endpoints and interior points were given. Furthermore, in the challenging 2-D test images, the single most important modification clearly was the molding procedure, in which lower smoothing degrees of freedom were used to capture gross features of the data before moving on to the finer ones.

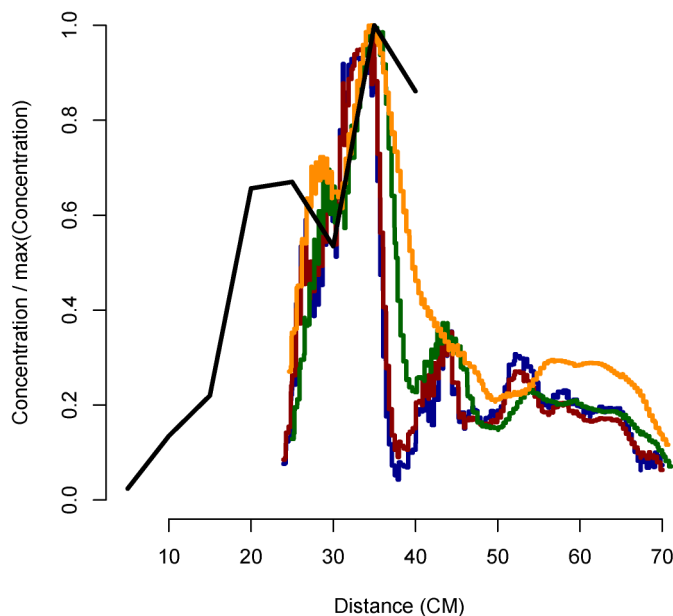


Figure 11. Concentration estimates by distance (in centimeters) from the curve beginning near the anus using various neighborhood sizes around the curve (—, 1; —, 2; —, 4; —, 8). The sigmoidoscope data are shown in black. All curves are normalized relative to their maximum value.

Although this research has provided scientific collaborators with a set of easily implementable tools for processing their images and obtaining concentration/distance curves, the larger, more general problem of arbitrary dimensional curve fitting leaves much room for further methodological development. This collection of problems, although considered in the image processing literature, has received insufficient attention in the statistics literature.

As for potential future directions, note that the algorithm is not completely automated. For example, starting and ending points of the curve must be given, and a final degrees of freedom for the smoothers must be determined. For the latter point, given that the algorithm runs very quickly, we recommend trying several values of K and using visual inspection and previous anatomic knowledge to determine a final value. In particular, as K grows too large, the curve will make visually unnatural bends to minimize the orthogonal projections of the data onto the curve. For future research, we believe that penalized spline approaches for fitting may automate the procedure more completely. In addition, it is possible that separate degrees of freedom, and separate bases, could be used for f^x , f^y , and f^z . Finally, given the variation in complexity of some of 2-D shapes, smoothers that can accommodate variable degrees of freedom across t (adaptive smoothers) may be able to fit more complex structures.

Finally, although the developed algorithm proved to be very successful in this application, other methods for allocating the time points could be considered in other settings. For example, Deng (2007) considered a stochastic search algorithm under a user-specified objective function. In addition, as a referee has pointed out, the latent time variable t could be treated in a Bayesian manner; that is, a prior on the time variable could be set, along with priors on all of the smoothing parameters. An added benefit of such an approach is the built-in ability to quantify uncertainty in the curve fits.

[Received May 2007. Revised December 2007.]

REFERENCES

- Banfield, J., and Raftery, A. (1992), "Ice Floe Identification in Satellite Images Using Mathematical Morphology and Clustering About Principal Curves," *Journal of the American Statistical Association*, 87, 7–16.
- Bitter, I., Kaufman, A., and Sato, M. (2001), "Penalized-Distance Volumetric Skeleton Algorithm," *IEEE Transactions on Visualization and Computer Graphics*, 7, 195–206.

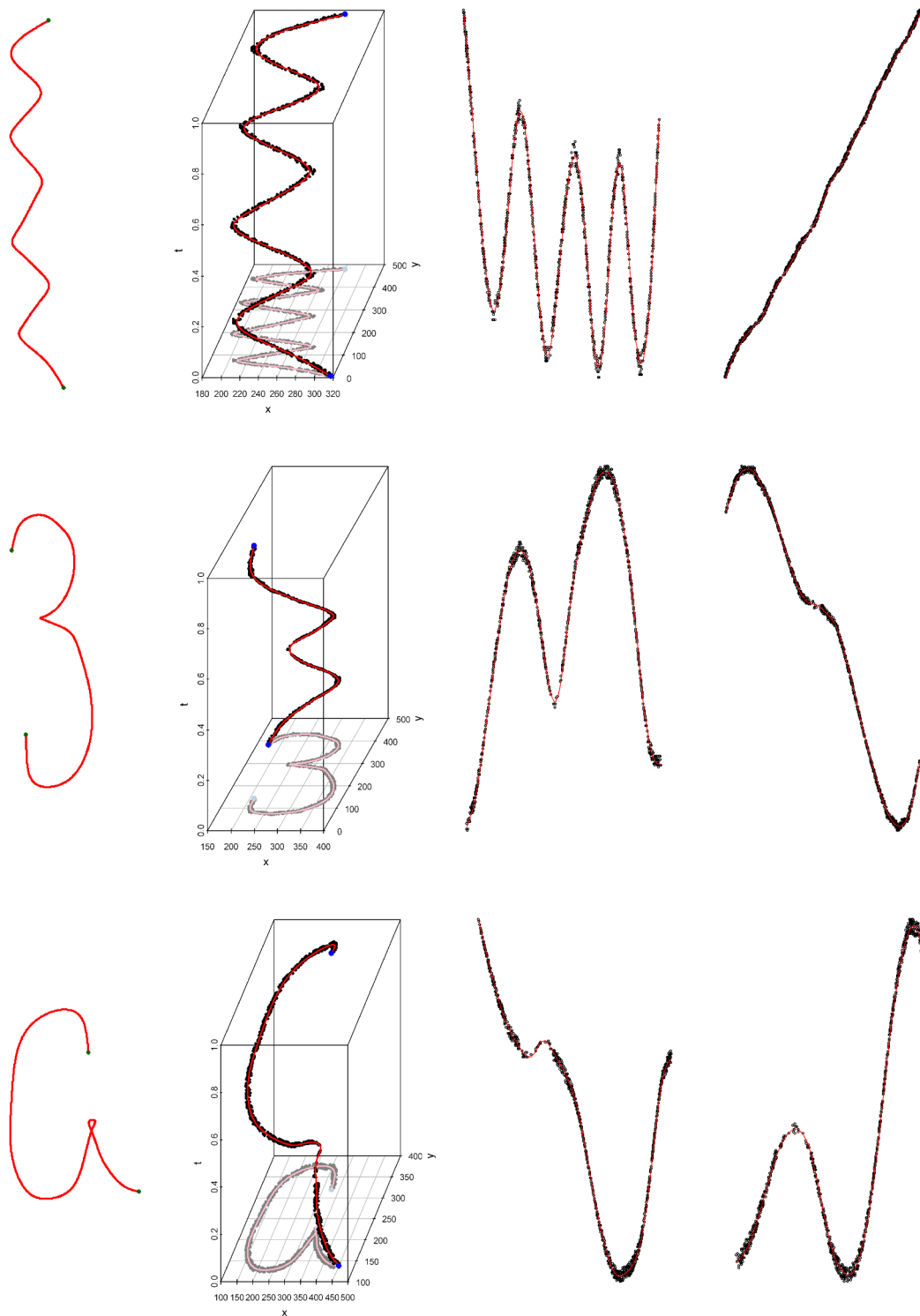


Figure 12. Fitted values for three example data sets at 20 degrees of freedom.

Bouix, S., Siddiqi, K., and Tannenbaum, A. (2003), "Flux-Driven Fly Throughs," in *Proceedings of the 2003 Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, Madison, WI: IEEE Computer Society, pp. 449–454.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995), "A Limited Memory Algorithm for Bound Constrained Optimization," *SIAM Journal on Scientific Computing*, 16, 1190–1208.

Chaudhuri, P., Khandekar, R., Sethi, D., and Kalra, P. (2004), "An Efficient Central Path Algorithm for Virtual Navigation," *Proceedings of the Computer Graphics International (CGI'04)*, Washington, DC: IEEE Computer Society, pp. 188–195.

Chiou, R., Kaufman, A., Liang, Z., Hong, L., and Achiotou, M. (1998), "Interactive Path Planning for Virtual Endoscopy," in *Proceedings of the IEEE Nuclear Science and Medical Imaging Conference*, Toronto, CA: IEEE, pp. 2069–2076.

Cohen, L., and Kimmel, R. (1997), "Global Minimum for Active Contour Models: A Minimal Path Approach," *Journal of Computer Vision*, 24, 57–78.

Deng, L. (2007), "Spline-Based Curve Fitting With Applications to Kinetic Imaging," unpublished master's thesis, Johns Hopkins University, Dept. of Biostatistics.

Deschamps, T., and Cohen, L. (2001), "Fast Extraction of Minimal Paths in 3D Images and Applications to Virtual Endoscopy," *Medical Image Analysis*, 5, 281–299.

- Dijkstra, E. (1959), "A Note on Two Problems in Connection With Graphs," *Numerische Mathematik*, 1, 269–271.
- Ge, Y., Stels, D., Wang, J., and Vining, D. (1999), "Computing Centerline of a Colon: A Robust and Efficient Method Based on 3D Skeletons," *Journal of Computer-Assisted Tomography*, 23, 786–794.
- Hastie, T., and Stuetzle, W. (1989), "Principal Curves," *Journal of the American Statistical Association*, 84, 502–516.
- Hong, L., Muraki, S., Kaufman, A., Bartz, D., and He, T. (1997), "Virtual Voyage: Interactive Navigation in the Human Colon," in *Proceedings of SIGGRAPH'97*, New York, NY: ACM Press, pp. 27–34.
- Hudson, H. M., and Larkin, R. S. (1994), "Accelerated Image Reconstruction Using Ordered Subsets Projection Data," *IEEE Transactions on Medical Imaging*, 13, 601–609.
- Kegl, B., Krzyzak, A., Linder, T., and Zeger, K. (2000), "Learning and Design of Principal Curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 281–297.
- McFarland, E., Wang, G., Brink, J., Balfe, D., Heiken, J., and Vannier, M. (1997), "Spiral Computed Tomographic Colonography: Determination of the Central Axis and Digital Unraveling of the Colon," *Academic Radiology*, 4, 367–373.
- R Development Core Team (2006), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing.
- Samara, Y., Fiebach, M., Dachman, A., Doi, K., and Hoffmann, K. (1998), "Automated Centerline Tracking of the Human Colon," in *Proceedings of the SPIE Conference on Image Processing*, Bellingham, WA: SPIE, pp. 740–746.
- Searle, S. (1971), *Linear Models*, New York: Wiley.
- Telea, A., and Vilanova, A. (2003), "A Robust Level-Set Algorithm for Centerline Extraction," in *Proceedings of the Symposium on Data Visualization 2003*, Aire-la-Ville, Switzerland: Eurographics Association, pp. 185–194.
- Thorpe, J. (1979), *Elementary Topics in Differential Geometry*, New York: Springer-Verlag.
- Wan, M., Dachille, F., and Kaufman, A. (2001), "Distance-Field-Based Skeletons for Virtual Navigation," in *IEEE Visualization 2001*, Washington, DC: IEEE Computer Society, pp. 239–245.