Technical Report

TR-2007-026

Efficient estimation of 3-dimensional centerlines of inner carotid arteries and their curvature functions by free knot regression splines

by

Laura M. Sangalli, Piercesare Secchi, Simone Vantini, Alessandro Veneziani

Mathematics and Computer Science

EMORY UNIVERSITY

Efficient estimation of 3-dimensional centerlines of inner carotid arteries and their curvature functions by free knot regression splines

L.M. Sangalli,P.Secchi, S.Vantini, A.Veneziani[1]

MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
laura.sangalli@polimi.it, piercesare.secchi@polimi.it,
simone.vantini@polimi.it, alessandro.veneziani@polimi.it

Keywords: functional data analysis, curve fitting, free knot regression splines, local polynomial smoothing.

AMS Subject Classification: 62H12, 65D07.

Abstract

This work stems from the need for accurate estimation of the curvature function of an artery, that emerged within ANEURISK Project, a research program that aims at investigating the role of vascular morphology and hemodynamics on the pathogenesis of cerebral aneurysms. We develop here a regression technique that exploits free knot splines in a novel setting, to estimate 3-dimensional curves, and their derivatives. We thoroughly compare this technique to a classical regression method, local polynomial smoothing, showing that 3-dimensional free knot regression splines yield more accurate and efficient estimates.

1 Introduction

We show how to efficiently estimate the curvature function of a 3-dimensional (3D) curve, starting from its noisy pointwise observation. This work stems from the need for accurate estimation of artery curvature, that emerged within

1. Alessandro Veneziani's present address: Department of Mathematics and Computer Science, Emory University, Atlanta, GA, USA

ANEURISK Project[1]. This project involves researchers of several scientific fields, and aims at investigating the role of vessel morphology, blood fluid-dynamics, and biomechanical properties of the vascular wall, on the pathogenesis of cerebral aneurysms.

Cerebral aneurysms are lesions of cerebral vessels characterized by a bulge of the vessel wall. The origin of this pathology is still unclear. Possible explanations discussed in the medical literature focus on the interaction between biomechanical properties of artery walls and hemodynamic factors, such as wall shear stress and pressure. The hemodynamics is in turn strictly dependent on vascular geometry. This circle of ideas has motivated investigations on the shape of aneurysms and on parent vessel morphologies, and their possible impact on the genesis and development of this pathology. See e.g. Hoi et al. (2004), Hassan et al. (2005), Castro, Putman, and Cebral (2006). ANEURISK Project belongs to this line of research.

This paper, in particular, is devoted to the accurate estimation of vessel curvature. It is well known in fact that vessel curvature strongly affects the local hemodynamics. See for instance Caro, Doorly, and Tarnakasky (1996) and Chandran (1993). More in general, the relevance of the curvature of a pipe, on the inner fluid motion, has been extensively investigated since the pioneristic works by Dean. See e.g. Dean (1927a), Dean (1927b), Smith (1976), Berger, Talbot, and Yao (1983), and Chandran (1993). In particular, an adimensional index called *Dean number* has been introduced, as an extension of Reynolds number, in order to quantify fluid stability in a circular curved pipe. Let R be the radius of the pipe section, and let R_{curv} be the radius of curvature of the pipe centerline; moreover, denote by μ the viscosity of the fluid at hand, by ρ its density, and by U_0 its mean velocity. The Dean number is

$$D = \frac{2\,\rho\,U_0\,R}{\mu}\left(\frac{R}{R_{curv}}\right)^{1/2}$$

Blood dynamics can be really different for different values of D. This explains why accurate estimation of the profile of the curvature radius of artery vessels (or equivalently, the profile of the curvature $curv = 1/R_{curv}$) is of fundamental importance for the analysis of hemodynamics, and its possible consequences on aneurysm pathology.

[1]ANEURISK Project is a joint research program involving *MOX Laboratory for Modeling and Scientific Computing*, Dipartimento di Matematica, Politecnico di Milano, *Laboratory of Biological Structures*, Dipartimento di Ingegneria Strutturale, Politecnico di Milano, *Istituto Mario Negri*, Bergamo, *Ospedale Niguarda Ca' Granda*, Milano, and *Ospedale Maggiore Policlinico*, Milano. The Project is supported by *Fondazione Politecnico di Milano* and *Siemens-Medical Solutions, Italia*.

Figure 1: 3D image of an internal carotid artery with an aneurysm [patient 1].

The starting point of our study is a set of 3D angiographic images taken from patients suspected to be affected by cerebral aneurysms. Our sample is currently based on 65 patients treated at Ca' Granda Hospital, in Milan (Italy). Some of these patients are affected by an aneurysm on the Internal Carotid Artery (ICA), other patients have an aneurysm downstream the ICA, in the so-called Willis circle (see Ustun (2005)); finally, a few patients are healthy. The analyses conducted so far within the ANEURISK Project focus on the ICA, which is clearly recognizable in each of the 65 angiographies. This artery features a great curvature variability not only between different patients, but also within different parts of the vessel for the same patient. Starting from the 3D array of grey-scaled pixels given by the angiography (with lighter pixels showing presence of flowing blood), the artery lumen, i.e. the volume occupied by flowing blood, is identified by a reconstruction algorithm coded in the *Vascular Modeling ToolKit* (VMTK). See Antiga et al. (2003), and Piccinelli et al. (2007). The outcome of this process yields a set of points in a three dimensional space, that are computed as the centers of maximal spheres inscribed in the artery lumen. This set of points identifies the vessel centerline. Moreover, an estimate of the radius along the vessel is provided. Figure 1 shows the draw of the reconstruction of an ICA with its centerline (patient 1). There are two kinds of errors that might occur in this procedure. *Measurement errors*, due to possible (in general unavoidable) errors in the images, related for instance to a movement of the patient during the image acquisition. *Reconstruction errors*, i.e. numerical errors in the algorithm for the 3D reconstruction. Because of these errors, reconstructed centerlines may be quite wiggly and thus need to be smoothed, and so do the estimates of their derivatives, in order to obtain sensible estimates of their curvature functions. We propose estimates of the 3D centerlines and their curvature functions, based on *free knot regression splines*. We show that this technique has many comparative advantages over a classical regression method,

local polynomial smoothing. First of all, free knot regression splines are more efficient than local polynomial smoothing, in the sense that they are able to attain lower approximation errors fitting more parsimonious models. Moreover, they better estimate the salient features of the centerline, as it is shown by comparing the estimates of centerline derivatives, obtained with the two different techniques, and their rough estimates obtained by central differences. Because of this, free knot splines are better able to detect special landmarks in the centerline curvature, that are particularly important for our problem. Finally, the proposed technique leads to data dimension reduction, a fundamental issue for our highly dimensional dataset.

Local polynomial smoothing has been formerly used to obtain pointwise estimates of vessel curvature in Sangalli, Secchi, Vantini, and Veneziani (2007b), where explorative analyses of vessel curvature and radius profiles supported the existence of a strong relationship between vessel geometry and aneurysm location. In particular, Functional Principal Component Analysis of ICA curvature and radius profiles, followed by Quadratic Discriminant Analysis of principal component scores, evidenced significant differences on the distribution of these two geometric features for patients having an aneurysm along the ICA and patients having an aneurysm on the Circle of Willis. The results of this first study have been confirmed in Sangalli, Secchi, and Vantini (2007a), where the same analysis are performed using centerline estimates based on free knot regression splines.

The paper is organized as follows. In Section 2 we present the dataset and the main features of the statistical model we deal with. In Section 3 we recall the essentials of free knot regression splines for the case of functions with one dependent variable, and then describe the extension to our 3D problem. In Section 4 we illustrate the results, and discuss the properties of the estimates obtained. Section 5 is devoted to an extensive comparison between the proposed technique and local polynomial smoothing. Some conclusive considerations are drawn in Section 6.

2 Data and model

For every patient i in our dataset ($i = 1, \ldots, 65$), VMTK reconstruction of ICA centerline is a set of points in $\mathbb{R}^3$, $\{(x_{ij}, y_{ij}, z_{ij}) : j = 1, 2, \ldots, n_i\}$, where x, y and z denote the three space coordinates of each point. Points are ordered moving downward along the ICA, from the point closest to its terminal bifurcation (detected by VMTK) towards the proximal districts, i.e. aorta and heart. The reason for this choice is that the terminal bifurcation of ICA is present in each angiography, even if the portion of ICA captured by the angiography

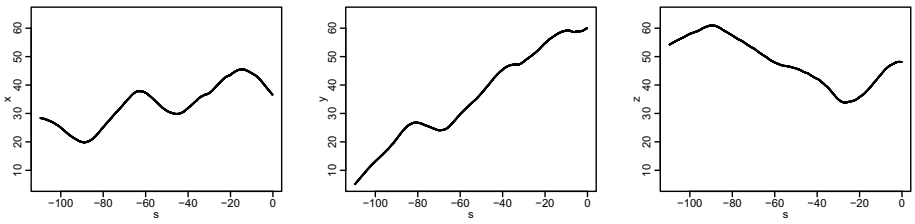

Figure 2: Reconstructed space coordinates of ICA centerline for patient 1, (x_{1j}, y_{1j}, z_{1j}), versus the abscissa parameter s_{1j}, for $j = 1, \ldots, n_1$ ($n_1 = 1350$).

varies from patient to patient (depending on where the angiographic image has been centered). We associate the set of space coordinates with an index set $\{s_{ij} : j = 1, 2, \ldots, n_i\}$, which measures an approximate distance along the ICA. More precisely, s_{i1} is the distance of the point (x_{i1}, y_{i1}, z_{i1}) from the terminal bifurcation of the ICA (as determined by VMTK), and

$$s_{ij} - s_{ij-1} = -\sqrt{(x_{ij} - x_{ij-1})^2 + (y_{ij} - y_{ij-1})^2 + (z_{ij} - z_{ij-1})^2}\,, \quad \text{for } j = 2, \ldots, n_i.$$

The conventional negative sign highlights that we are moving upstream, i.e. in opposite direction with respect to blood flow. Figure 2, for instance, displays the reconstructed space coordinates of ICA centerline for patient 1, (x_{1j}, y_{1j}, z_{1j}), versus the abscissa parameter s_{1j}, for $j = 1, \ldots, n_1$ ($n_1 = 1350$). The number n_i of data points available for each patient ranges from 350 to 1380, and is almost perfectly correlated to the approximate length $|s_{i\,n_i} - s_{i\,1}|$ of the reconstructed centerlines (correlation coefficient=0.999), which in turn varies from 27.219mm to 110.136mm. In other words, the grid density of the 65 reconstructions is the same. The average step of these grids is 0.079mm.

Our task is to estimate, for each patient i, the curvature of the true ICA centerline $\mathbf{c}_i(s) = (x_i(s), y_i(s), z_i(s))$, starting from its noisy pointwise reconstruction $\{(x_{ij}, y_{ij}, z_{ij}) : j = 1, 2, \ldots, n_i\}$. To this aim, independently for each patient (from now on we drop the subscript i), we consider the statistical model

$$\begin{aligned} (x_j, y_j, z_j) = \mathbf{c}(s_j) + \mathbf{e}_j \qquad & j = 1, \ldots, n \\ \mathbf{e}_j = (e_j^{[x]}, e_j^{[y]}, e_j^{[z]}) \qquad & \text{with } \mathbf{e}_j \text{ and } \mathbf{e}_{j'} \text{ independent for } j \neq j', \\ & E(\mathbf{e}_j) = \mathbf{0},\ Var(\mathbf{e}_j) = \sigma^2 \mathbf{I} \end{aligned} \tag{1}$$

and want to estimate the curvature of $\mathbf{c}$. The value of σ^2, in the variance matrix of the error term, is unknown, but assumed to be the same for each patient i, for $i = 1, \ldots, 65$. This assumption is justified by the fact that both the machine used to take the 3D-angiographies and the algorithm used to reconstruct ICA

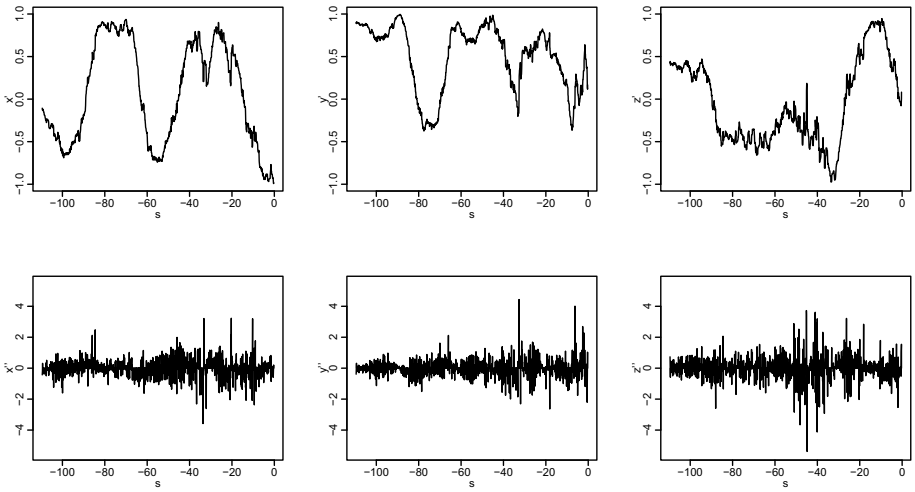

Figure 3: First and second differences (top and bottom, respectively) of reconstructed space coordinates of ICA centerline for patient 1.

centerlines are the same for each patient. Denoting by $\times$ the vector product, the curvature of $\mathbf{c}(s) = (x(s), y(s), z(s))$ is defined by

$$curv_{\mathbf{c}}(s) = \frac{|\mathbf{c}'(s) \times \mathbf{c}''(s)|}{|\mathbf{c}'(s)|^3} =$$

$$\frac{\sqrt{\big(x'(s)y''(s) - x''(s)y'(s)\big)^2 + \big(y'(s)z''(s) - y''(s)z'(s)\big)^2 + \big(z'(s)x''(s) - z''(s)x'(s)\big)^2}}{\big((x'(s))^2 + (y'(s))^2 + (z'(s))^2\big)^{3/2}}$$

where $\mathbf{c}'(s) = (x'(s), y'(s), z'(s))$ and $\mathbf{c}''(s) = (x''(s), y''(s), z''(s))$ are the first and second derivatives of $\mathbf{c}$. Rough pointwise estimates of first derivatives are given by the *first central differences*

$$Dx(s_j) = \frac{x_{j+1} - x_{j-1}}{s_{j+1} - s_{j-1}} \qquad Dy(s_j) = \frac{y_{j+1} - y_{j-1}}{s_{j+1} - s_{j-1}} \qquad Dz(s_j) = \frac{z_{j+1} - z_{j-1}}{s_{j+1} - s_{j-1}}$$

for $j = 2, \ldots, n-1$; the *second central differences*

$$D^2x(s_j) = 2\frac{\frac{x_{j+1}-x_j}{s_{j+1}-s_j} - \frac{x_j-x_{j-1}}{s_j-s_{j-1}}}{s_{j+1} - s_{j-1}} \qquad D^2y(s_j) = 2\frac{\frac{y_{j+1}-y_j}{s_{j+1}-s_j} - \frac{y_j-y_{j-1}}{s_j-s_{j-1}}}{s_{j+1} - s_{j-1}}$$

$$D^2z(s_j) = 2\frac{\frac{z_{j+1}-z_j}{s_{j+1}-s_j} - \frac{z_j-z_{j-1}}{s_j-s_{j-1}}}{s_{j+1} - s_{j-1}}$$

are instead too noisy to be of any use. See for instance Figure 3, which displays first and second central differences of reconstructed space coordinates of ICA centerline for patient 1.

Sangalli et al. (2007b) obtained smooth pointwise estimates of first and second derivatives by first estimating the true centerline $\mathbf{c}$ by means of local polynomial regression. In particular, pointwise estimates of true centerlines were obtained fitting fourth-degree polynomials with a Gaussian kernel. Details are given in Section 5. Estimates of derivatives, and hence of curvatures, were thus obtained by pointwise differentiation of the fitted local polynomials.

Here we propose a different approach, based on free knot regression splines, i.e. regression splines where the number and position of knots are not fixed in advance, but chosen in a way to minimize a penalized sum of squared errors criterion. Since our data are 3D, the idea is to fit simultaneously the three spatial coordinates of the centerline $\big(x(s), y(s), z(s)\big)$, looking for the optimal spline knots along the abscissa parameter s. This technique is described in Section 3.

3 3D free knot regression splines

Consider an interval $[a, b]$, split into subintervals by knots $k_1, \ldots, k_{n_k}$ (with $a < k_1 < \ldots < k_{n_k} < b$). Assume, for the moment, that the number and position of knots are given. An order-m spline over $[a, b]$, with knot vector $\mathbf{k} = (k_1, \ldots, k_{n_k})$, is a piecewise-polynomial of degree $m-1$, with continuous derivatives of order $m-2$ at the knots. The set of all order-m splines over $[a, b]$, with knot vector $\mathbf{k}$, forms a vector space. The dimension of such space is $m + n_k$. This is given by the balance between the number of parameters needed to identify $(n_k + 1)$ polynomials of order m (one polynomial over each of the subintervals), and the continuity constraints, prescribing $m-1$ equations at each of the n_k knots: $m(n_k+1) - n_k(m-1) = m+n_k$. A computationally convenient basis system for this space is the b-spline basis system $\{b^{[\mathbf{k}]}_{r,m}(s) : r = 1, \ldots, m+n_k\}$. Any order-$m$ spline over $[a, b]$ with knot vector $\mathbf{k}$ can thus be expressed as the expansion in bases $f(s) = \sum_{r=1}^{m+n_k} \lambda_r b^{[\mathbf{k}]}_{r,m}(s)$, for some coefficient vector $\lambda = \big(\lambda_1, \ldots, \lambda_{m+n_k}\big)^T$. Derivatives of splines are still splines (of appropriate order), with coefficients directly computed from the coefficients of the original spline, and with the same knot vector. In particular, denoting by $\mathbf{v}$ the $(2m + n_k)$-vector with entries

$$
\begin{aligned}
&v_1 = \ldots = v_m = a,\\
&v_{m+j} = k_j\,, \text{ for } j = 1, \ldots, n_k,\\
&v_{m+n_k+1} = \ldots = v_{2m+n_k} = b,
\end{aligned}
$$

the first and second derivatives of $f(s)$ are given by $f'(s) = \sum_{r=1}^{m+n_k-1} \lambda_r^{[1]} b_{r,m-1}^{[\mathbf{k}]}(s)$ and $f''(s) = \sum_{r=1}^{m+n_k-2} \lambda_r^{[2]} b_{r,m-2}^{[\mathbf{k}]}(s)$, where

$$\lambda_r^{[1]} = \frac{\lambda_{r+1} - \lambda_r}{v_{m+r} - v_{r+1}} \quad \text{for } r = 1, \ldots, m + n_k - 1$$

and

$$\lambda_r^{[2]} = \frac{\lambda_{r+1}^{[1]} - \lambda_r^{[1]}}{v_{m+r} - v_{r+2}} \quad \text{for } r = 1, \ldots, m + n_k - 2.$$

See de Boor (1978) for a systematic introduction to splines.

Now consider the problem of estimating the function $f : \mathbb{R} \to \mathbb{R}$ in the model

$$w_j = f(s_j) + e_j\,, \qquad \text{with } e_j \text{ and } e_{j'} \text{ independent for } j \neq j', \quad E(e_j) = 0, Var(e_j) = \sigma^2,$$

by means of an order-m spline with knot vector $\mathbf{k}$. The regression spline is the spline

$$\hat{f}(s) = \sum_{r=1}^{m+n_k} \hat{\lambda}_r b_{r,m}^{[\mathbf{k}]}(s)$$

where $\hat{\lambda} = \left(\hat{\lambda}_1, \ldots, \hat{\lambda}_{m+n_k}\right)^T$ minimizes the sum of squared errors:

$$SSE(\lambda) = \sum_{j=1}^{n} \left(w_j - \sum_{r=1}^{m+n_k} \lambda_r b_{r,m}^{[\mathbf{k}]}(s_j)\right)^2. \tag{2}$$

In matrix notation, writing $\{B_m^{[\mathbf{k}]}\}_{j,r} = b_{r,m}^{[\mathbf{k}]}(s_j)$, $\mathbf{w} = (w_1, \ldots, w_n)^T$ and $\hat{\mathbf{w}} = (\hat{f}(s_1), \ldots, \hat{f}(s_n))^T$,

$$\begin{aligned} SSE(\lambda) &= (\mathbf{w} - B_m^{[\mathbf{k}]}\lambda)^T(\mathbf{w} - B_m^{[\mathbf{k}]}\lambda) \\ \hat{\lambda} &= \left(B_m^{[\mathbf{k}]^T} B_m^{[\mathbf{k}]}\right)^{-1} B_m^{[\mathbf{k}]^T} \mathbf{w} \\ \hat{\mathbf{w}} &= B_m^{[\mathbf{k}]}\left(B_m^{[\mathbf{k}]^T} B_m^{[\mathbf{k}]}\right)^{-1} B_m^{[\mathbf{k}]^T} \mathbf{w} = H\,\mathbf{w} \end{aligned}$$

where $H = B_m^{[\mathbf{k}]}\left(B_m^{[\mathbf{k}]^T} B_m^{[\mathbf{k}]}\right)^{-1} B_m^{[\mathbf{k}]^T}$ is an orthogonal projection matrix, with $tr(H) = m + n_k$. Thus, $\hat{\mathbf{w}}$ is obtained as orthogonal projection of $\mathbf{w}$ on the space generated by the vectors of basis evaluations $\left(b_{r,m}^{[\mathbf{k}]}(s_1), \ldots, b_{r,m}^{[\mathbf{k}]}(s_n)\right)^T$, for $r = 1, \ldots, m + n_k$.

The order of the spline is usually chosen according to the problem at hand and the scope of the analysis. Order-4 splines (corresponding to piecewise cubic polynomials) are almost always sufficient when one is only interested in a smooth estimate of the function f. When derivatives of f are also of interest, it is usually

convenient to use order $p+2$ or $p+3$ splines, where p is the highest derivative order that has to be estimated.

To improve spline estimation, number and position of knots can be selected by minimizing a suitable error functional, which takes into account the dimension of the model being fitted. A possible model selector, in this case, is provided by the so-called *Stein's unbiased risk estimate* (see Stein (1981))

$$\begin{aligned} pSSE(\lambda, \mathbf{k}) &= \sum_{j=1}^{n} \left(w_j - \sum_{r=1}^{m+n_k} \lambda_r b_{r,m}^{[\mathbf{k}]}(s_j)\right)^2 + C\hat{\sigma}^2(m+n_k) \\ &= (\mathbf{w} - B_m^{[\mathbf{k}]}\lambda)^T(\mathbf{w} - B_m^{[\mathbf{k}]}\lambda) + C\hat{\sigma}^2(m+n_k) \end{aligned} \tag{3}$$

where C is a penalization constant, and $\hat{\sigma}^2$ is an estimate of errors variance. Regression splines where the number and position of knots are not fixed in advance, but chosen by a data-driven criterion such as (3), are called free knot regression splines. Note that, once the optimal knot vector $\hat{\mathbf{k}}$ has been selected, and thus the model has been chosen, the vector of fitted values $\hat{\mathbf{w}}$ is again obtained as orthogonal projection of $\mathbf{w}$ on the space generated by the vectors of basis evaluations, according to $\hat{\mathbf{w}} = H\,\mathbf{w}$, where $H = B_m^{[\hat{\mathbf{k}}]}\left(B_m^{[\hat{\mathbf{k}}]^T} B_m^{[\hat{\mathbf{k}}]}\right)^{-1} B_m^{[\hat{\mathbf{k}}]^T}$. Algorithms for the search of optimal knots have been proposed e.g. by Friedman (1991), Stone, Hansen, Kooperberg, and Truong (1997), Luo and Wahba (1997), and more recently by Zhou and Shen (2001), and Mao and Zhao (2003). In particular, the algorithm developed by Zhou and Shen (2001) strongly improved the previous stepwise forward/backward knot selection procedures, which suffered from knot confounding problems, by also including knot relocation moves. The choice of C is of paramount importance for the accuracy of the estimation. High values of C yield more parsimonious models, i.e. splines with fewer knots, but with higher sum of squared errors, and viceversa. Several researchers suggest particular values of C. For example, Luo and Wahba (1997) suggest $C = 1.2$, Zhou and Shen (2001) recommend $C = 2$, whilst Friedman (1991) uses $C = 3$ for his MARS.

3.1 Extension to the 3D case

Let us now come to our 3D problem, described by model (1). The idea is to estimate the true centerline $\mathbf{c}(s)$ by simultaneously fitting the three coordinate functions, $x(s)$, $y(s)$ and $z(s)$, with three order-m splines having the same knot vector $\hat{\mathbf{k}}$, with knots chosen along the abscissa parameter s:

$$\hat{x}(s) = \sum_{r=1}^{m+n_{\hat{k}}} \hat{\lambda}_r^{[x]} b_{r,m}^{[\hat{\mathbf{k}}]}(s) \qquad \hat{y}(s) = \sum_{r=1}^{m+n_{\hat{k}}} \hat{\lambda}_r^{[y]} b_{r,m}^{[\hat{\mathbf{k}}]}(s) \qquad \hat{z}(s) = \sum_{r=1}^{m+n_{\hat{k}}} \hat{\lambda}_r^{[z]} b_{r,m}^{[\hat{\mathbf{k}}]}(s).$$

As before, denote by $B_m^{[\mathbf{k}]}$ the $(n \times (m+n_k))$-matrix of basis evaluations. Moreover, denote by W the $(n \times 3)$-matrix of observed values, $W = \big[\mathbf{x}\big|\mathbf{y}\big|\mathbf{z}\big]$, where $\mathbf{x} = (x_1, \ldots, x_n)^T$, $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\mathbf{z} = (z_1, \ldots, z_n)^T$, and by $\hat{W}$ the analogous matrix of fitted values $\hat{W} = \big[\hat{\mathbf{x}}\big|\hat{\mathbf{y}}\big|\hat{\mathbf{z}}\big]$, where $\hat{\mathbf{x}} = \big(\hat{x}(s_1), \ldots, \hat{x}(s_n)\big)^T$, $\hat{\mathbf{y}} = \big(\hat{y}(s_1), \ldots, \hat{y}(s_n)\big)^T$, $\hat{\mathbf{z}} = \big(\hat{z}(s_1), \ldots, \hat{z}(s_n)\big)^T$. Finally, denote by Λ the $((m+n_k) \times 3)$-matrix of λ coefficients $\Lambda = \big[\lambda^{[x]}\big|\lambda^{[y]}\big|\lambda^{[z]}\big]$. For a fixed knot vector $\mathbf{k}$, the natural generalization of least squared errors criterion (2), for the 3D model (1), is given by

$$\begin{aligned} 3DSSE(\Lambda) &= tr\big[(W - B_m^{[\mathbf{k}]}\Lambda)(W - B_m^{[\mathbf{k}]}\Lambda)^T\big] \\ &= \sum_{j=1}^{n}\big(x_j - \sum_{r=1}^{m+n_k}\lambda_r^{[x]} b_{r,m}^{[\mathbf{k}]}(s_j)\big)^2 + \sum_{j=1}^{n}\big(y_j - \sum_{r=1}^{m+n_k}\lambda_r^{[y]} b_{r,m}^{[\mathbf{k}]}(s_j)\big)^2 \\ &\quad + \sum_{j=1}^{n}\big(z_j - \sum_{r=1}^{m+n_k}\lambda_r^{[z]} b_{r,m}^{[\mathbf{k}]}(s_j)\big)^2, \end{aligned}$$

which leads to the least squared estimate $\hat{W} = B_m^{[\mathbf{k}]}\hat{\Lambda}$, where

$$\hat{\Lambda} = \big({B_m^{[\mathbf{k}]}}^T B_m^{[\mathbf{k}]}\big)^{-1} {B_m^{[\mathbf{k}]}}^T W \tag{4}$$

Thus, for the search of the optimal $(\hat{\Lambda}, \hat{\mathbf{k}})$, we will use the following model selector:

$$\begin{aligned} &3DpSSE(\Lambda, \mathbf{k}) \\ &= tr\big[(W - B_m^{[\mathbf{k}]}\Lambda)(W - B_m^{[\mathbf{k}]}\Lambda)^T\big] + \mathcal{C}(m+n_k) \\ &= \sum_{j=1}^{n}\big(x_j - \sum_{r=1}^{m+n_k}\lambda_r^{[x]} b_{r,m}^{[\mathbf{k}]}(s_j)\big)^2 + \sum_{j=1}^{n}\big(y_j - \sum_{r=1}^{m+n_k}\lambda_r^{[y]} b_{r,m}^{[\mathbf{k}]}(s_j)\big)^2 \\ &\quad + \sum_{j=1}^{n}\big(z_j - \sum_{r=1}^{m+n_k}\lambda_r^{[z]} b_{r,m}^{[\mathbf{k}]}(s_j)\big)^2 + \mathcal{C}(m+n_k)\,. \end{aligned} \tag{5}$$

Note that the penalization term in (5) does not intend to take into account error variance, as was instead the case in (3). Hence, the penalization constant $\mathcal{C}$ in (5) cannot be compared with the penalization constant C in (3). This choice is related to the specific problem we deal with. Recall that the value of σ in the variance structure of the error term is assumed to be the same for each patient. Since we are here mainly interested in equally smoothing the ICA centerlines of the 65 patients, we will simply focus on choosing the best value of $\mathcal{C}$ for the overall dataset, without bothering with finding a sensible a priori estimate of σ. We discuss the choice of the penalization constant $\mathcal{C}$ in Section 4.1.

It should also be noted that, likewise in the 1D case, the matrix of fitted values $\hat{W}$ is obtained as orthogonal projection of the columns of the matrix of

observed values W on the space generated by the vectors of basis evaluation $\big(b^{[\hat{\mathbf{k}}]}_{r,m}(s_1),\ldots,b^{[\hat{\mathbf{k}}]}_{r,m}(s_n)\big)^T$, for $r = 1\ldots, m+n_{\hat{k}}$,

$$\hat{W} = B^{[\hat{\mathbf{k}}]}_m \big({B^{[\hat{\mathbf{k}}]}_m}^T B^{[\hat{\mathbf{k}}]}_m\big)^{-1} {B^{[\hat{\mathbf{k}}]}_m}^T W = H\, W\,,$$

In particular, the 3D the spline estimator, likewise the 1D spline estimator, is a so called-linear estimator, i.e. there exists a linear operator S, independent of W, such that the fitted values $\hat{W}$ are obtained according to $\hat{W} = S\, W$. For these estimators it is common to take, as Degrees of Freedom (DF) of the model, the trace of the linear operator S, or alternatively, $tr(S^T S)$ or $tr(2S - S^T S)$. See e.g. Buja, Hastie, and Tibshirani (1989) and Hastie and Tibshirani (1990), who first introduced this notion. We will thus consider, as DF of the spline estimator, the trace of the orthogonal projection matrix H, i.e. $m + n_{\hat{k}}$ (note that when the linear operator is an orthogonal projection matrix, as in this case, the three alternative definitions of DF coincide).

Remark 1 If the three coordinates in the error term were assumed to be correlated, and thus $Var(\mathbf{e}_j) = \sigma^2\mathbf{I}$ in (1) was changed in $Var(\mathbf{e}_j) = \Sigma$, for some symmetric positive-definite matrix Σ, then the natural generalization of least squared errors criterion (2) would become

$$3DSSE(\Lambda) = tr\big[(W - B^{[\mathbf{k}]}_m\Lambda)\Sigma^{-1}(W - B^{[\mathbf{k}]}_m\Lambda)^T\big] \tag{6}$$

It is easy to show that

$$\begin{aligned} &argmin_{(\Lambda)}\big\{tr\big[(W - B^{[\mathbf{k}]}_m\Lambda)\Sigma^{-1}(W - B^{[\mathbf{k}]}_m\Lambda)^T\big]\big\} \\ &= argmin_{(\Lambda)}\big\{tr\big[(W - B^{[\mathbf{k}]}_m\Lambda)(W - B^{[\mathbf{k}]}_m\Lambda)^T\big]\big\} \\ &= \big({B^{[\mathbf{k}]}_m}^T B^{[\mathbf{k}]}_m\big)^{-1}{B^{[\mathbf{k}]}_m}^T W \end{aligned}$$

so that the least square estimate $\hat{W} = B^{[\mathbf{k}]}_m\hat{\Lambda}$ does not depend on Σ. Thus, also in this more general case we could use the model selector (5). It turns out that the solution $\hat{\Lambda} = \big({B^{[\mathbf{k}]}_m}^T B^{[\mathbf{k}]}_m\big)^{-1}{B^{[\mathbf{k}]}_m}^T W$, not only minimizes (6), for any symmetric positive-definite matrix Σ, but also minimizes the general variance $\big|(W - B^{[\mathbf{k}]}_m\Lambda)(W - B^{[\mathbf{k}]}_m\Lambda)^T\big|$, where $|\cdot|$ denotes the determinant.

3.2 Algorithm for the search of optimal knots

For the search of optimal knots we modified Zhou and Shen (2001) algorithm, in order to deal with our 3D problem. For computational convenience, we look for optimal knots among the grid values $s_1,\ldots,s_n$. This is not a limitation, since, as noted in Section 2, the grids are very fine. We start with a set of initial knots

which are evenly spaced along the grid. A more accurate selection of initial knots, along the lines of Zhou and Shen (2001), would speed up the algorithm, but is not necessary, as pointed out by the same authors. The algorithm then alternates between the following two steps.

Knot addition. Let $\mathbf{k} = (k_1, \ldots, k_{n_k})$ be the current knot vector, and Λ the corresponding estimate of the coefficient matrix. Set $k_0 = a$ and $k_{n_k+1} = b$. If $\mathbf{k}$ is the initial knot vector, then each of the subintervals $[k_r, k_{r+1}]$, for $r = 0, \ldots, n_k$, is checked for possible addition of one knot. In particular, in each of the subintervals $[k_r, k_{r+1}]$, the knot s_j, with $k_r < s_j < k_{r+1}$, is added to the current knot vector, leading to the new knot vector $\mathbf{k}_j^*$ and to the new estimate of the coefficient matrix $\Lambda_j^* = \left(B_m^{[\mathbf{k}_j^*]^T} B_m^{[\mathbf{k}_j^*]}\right)^{-1} B_m^{[\mathbf{k}_j^*]^T} W$, if

$$3DpSSE(\Lambda_j^*, \mathbf{k}_j^*) < 3DpSSE(\Lambda, \mathbf{k}) \qquad \text{and}$$
$$s_j = argmin_{\{s_l : k_r < s_l < k_{r+1}\}} 3DpSSE(\Lambda_l^*, \mathbf{k}_l^*).$$

If $\mathbf{k}$ is not the initial knot vector, then the subinterval $[k_r, k_{r+1}]$ is checked for possible addition of one knot only if at least one among the neighbor knots, $k_{r-2}, k_{r-1}, \ldots, k_{r+3}$, has been added in the previous iteration.

Knot relocation/deletion. Let $\mathbf{k} = (k_1, \ldots, k_{n_k})$ be the current knot vector. For $r = 1, \ldots, n_k$, if k_r is adjacent to a knot added in the preceding step, then it is checked for possible relocation/deletion. In particular, k_r is removed from the the current knot vector, or moved to the new position s_j, with $k_{r-1} < s_l < k_{r+1}$, if this improves the criterion (5), likewise in the knot addition move. The same procedure is then applied to all knots that have been added in the preceding step.

When, for the first time, the knot addition move does not yield any new knot, all knots are searched for deletion/relocation. The resulting knot vector is set as new initial knot vector, and the algorithm is run again. When, for the second time, the knot addition move does not yield any new knot, all knots are searched for deletion/relocation, and the procedure ends.

The algorithm for the search of optimal knots, and all other data processing, have been coded in R©.

4 Results

The 65 centerlines are estimated by means of 3D free knot splines of order $m = 5$, with penalization constant $\mathcal{C} = 4$. The choice of the penalization constant is discussed in Section 4.1. The order of the splines has been fixed in order to obtain smooth estimates of the first two derivatives. In the future, the torsion

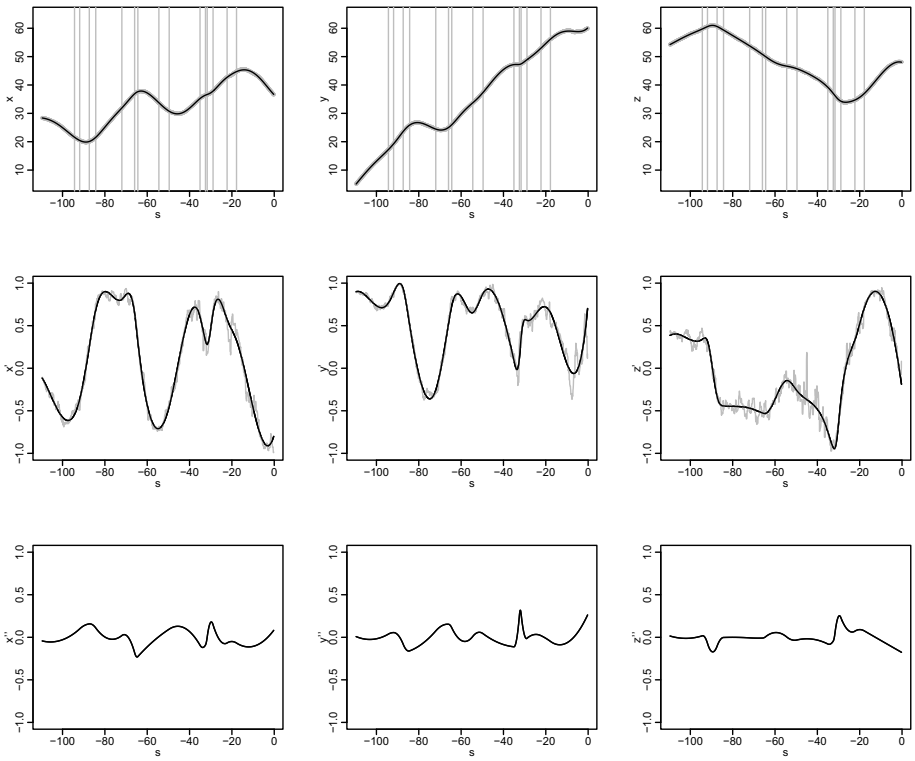

Figure 4: Top: fitted coordinates curves $\hat{x}(s), \hat{y}(s), \hat{z}(s)$, with vertical lines showing the position of the knots along the abscissa parameter s, superimposed to the original $(s_j, x_j), (s_j, y_j), (s_j, z_j)$ in grey [patient 1]. Center: first derivatives of $\hat{x}(s)$, $\hat{y}(s)$ and $\hat{z}(s)$, superimposed to first central differences in grey. Bottom: second derivatives of $\hat{x}(s)$, $\hat{y}(s)$ and $\hat{z}(s)$.

of the centerline may also be object of study. In fact, the hemodynamics is influenced not only by vessel radius and curvature, but also by its torsion, even if in a more complex, and still unclear way. The torsion depends on the first three derivatives. Note that using splines with order $m = 5$, the third derivative is a piecewise continuous line. Thus, order-5 splines can also provide continuous estimates of the torsion.

The top line of Figure 4 shows the estimates, $\hat{x}(s), \hat{y}(s), \hat{z}(s)$, of the three coordinate functions of patient 1 carotid centerline, obtained with order-5 free knot regression splines with $\mathcal{C} = 4$. The vertical lines show the position of the knots along the abscissa parameter s. The estimates are superimposed to the original $(s_j, x_j), (s_j, y_j), (s_j, z_j)$ (grey dots, almost completely hidden by the estimates). Figure 7 is a 3D image of the fitted centerline $\hat{\mathbf{c}}$. Center and bottom lines of

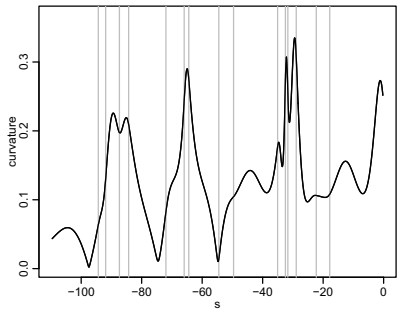

Figure 5: Curvature of $\hat{\mathbf{c}}(s)$ [patient 1]. Points of approximately zero curvature are the siphon delimiters shown in Figure 7.

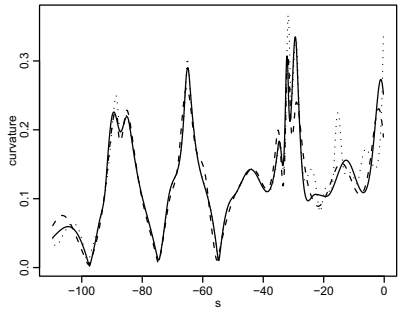

Figure 6: Estimated functions of centerline curvature [patient 1], obtained by free knot splines with $\mathcal{C} = 3, 4, 5$ (dotted, solid and dashed).

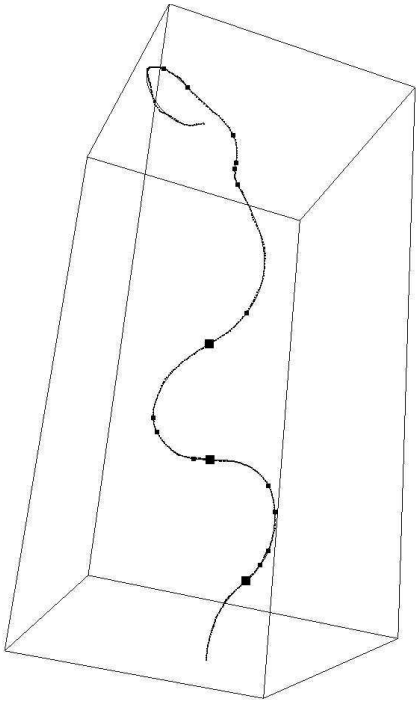

Figure 7: 3D image of fitted centerline (the little bullets show the positions of the spline knots), together with rough data [patient 1]. The big squares are the points of approximately zero curvature in Figures 5, 6.

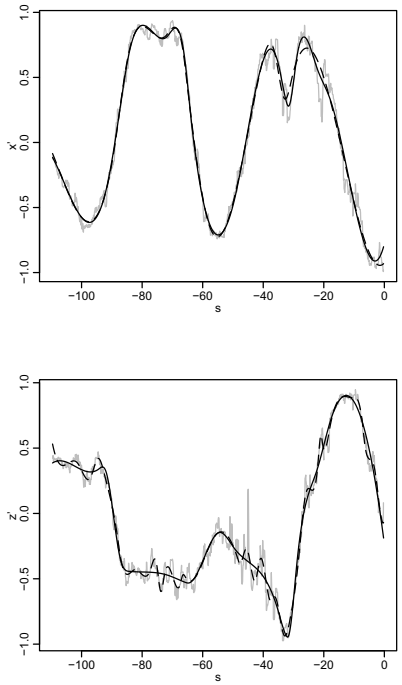

Figure 8: Top: estimates of x' obtained by free knot regression splines with $\mathcal{C} = 4$ (solid) and $\mathcal{C} = 9$ (dashed). Bottom: estimates of z' obtained by free knot regression splines with $\mathcal{C} = 4$ (solid) and $\mathcal{C} = 0.2$ (dashed) [patient 1].

Figure 4 show first and second derivatives of $\hat{x}(s)$, $\hat{y}(s)$ and $\hat{z}(s)$. Spline estimates of first derivatives are superimposed to the rough estimates given by first central differences (grey). Figure 5 shows the corresponding curvature function of $\hat{\mathbf{c}}(s)$.

It is of particular interest to identify the points of approximately zero curvature. In fact, these points can be taken as delimiters of artery bends or siphons, whose identification is important for the morphological analysis of the ICA. Preliminary investigations of the location of the aneurysms in the different bends, and of the position inside each specific bend, yield indeed interesting results in view of the set up of prognostic indexes. See Piccinelli et al. (2007) and Sangalli et al. (2007b). We take as points of approximately zero curvature all points of local minimum for the curvature, with curvature smaller than 1/(68.966mm), where 68.966mm is the mean length of reconstructed centerlines. In words, we consider the centerline to be locally almost linear in the points of minimum curvature within tracts of the centerline that can be locally approximated by a circumference having radius bigger than the average length of the reconstructed centerlines. Figure 7 shows such points in the 3D image of ICA centerline for patient 1.

4.1 Choice of penalization constant

The amount of penalization in the penalized sum of squared errors criterion is fixed taking into account the tradeoff between Average Squared Error (ASE) and DF, which translates the classical bias/variance tradeoff. Indeed, when the penalization increases, the ASE increases while the DF decrease. This is shown in the left part of Figure 9, which displays the boxplots of the distribution of ASE and DF for the fits corresponding to the 65 patients, obtained by free knot regression splines with 5 different penalizations: $\mathcal{C} = 0.2, 3, 4, 5, 9$. Figure 9 also displays the boxplots of the Average Squared Error on first derivatives (ASEder) with respect to the rough estimates given by first central differences:

$$\sum_{j=1}^{n} \big(Dx(s_j) - \hat{x}'(s_j)\big)^2 + \sum_{j=1}^{n} \big(Dy(s_j) - \hat{y}'(s_j)\big)^2 + \sum_{j=1}^{n} \big(Dz(s_j) - \hat{z}'(s_j)\big)^2.$$

We choose $\mathcal{C} = 4$. With this choice of the penalization constant, free knot splines can accurately estimate the salient features of the centerlines, without being too data-adapted. This can be better appreciated looking at spline estimates of first derivatives and comparing them with central differences estimates. Very high values of $\mathcal{C}$ can lead to centerlines estimates that cannot fully get the peaks and through in the first (and thus subsequent) derivatives. See for example the top of Figure 8, which compares the estimates of x' obtained with $\mathcal{C} = 4$

and $\mathcal{C} = 9$, for patient 1 carotid centerline. Very low values of $\mathcal{C}$ may instead yield estimates where also the high frequency variation is fitted. See for example the bottom of Figure 8, which compares the estimates of z' obtained with $\mathcal{C} = 4$ and $\mathcal{C} = 0.2$, always for patient 1. These two opposite mistakes must be avoided to get sensible estimates of curvature profiles. The choice of the particular value $\mathcal{C} = 4$ will be further discussed in next section.

It should be noted that the estimate of the curvature is stable with respect to variations of $\mathcal{C}$ over a reasonable span. Figure 6, for instance, compares the estimates of centerline curvature functions for patient 1 ICA, obtained with $\mathcal{C} = 3, 4, 5$. In particular, the points of (approximately) zero-curvature, whose importance has been previously mentioned, do not change. To show the robustness of the estimation of zero-curvature points, with respect to the choice of the penalization constant in a reasonable span, we compare the results corresponding to the three different penalization $\mathcal{C} = 3, 4, 5$. For 28 out of the 65 patients, no zero-curvature point is found, in the three different estimates corresponding to $\mathcal{C} = 3, 4, 5$. For these patients just a portion of the first siphon is observed. For only 9 out of the 65 patients, the number of zero curvature points is not the same in each of the three estimates: depending on the penalization, 0 points or 1 point are found, so that the three different estimates identify one full siphon, or just a portion of it. For the remaining 28 patients, each of the three estimates find the same number of zero-curvature points, wether 1, or 2 or 3 depending on the patient, for a total of 34 points. Moreover we point out that not only the number of points found by the three different estimates are the same, but also their positions are not statistically different. To show this fact we perform a principal component analysis of the matrix $U = \left[\boldsymbol{u}_1|\boldsymbol{u}_2|\boldsymbol{u}_3\right]$ that has as columns the vectors $\boldsymbol{u}_g = (u_{1g}, \ldots, u_{34g})^T$, for $g = 1, 2, 3$, with the positions of the 34 zero-curvature points, in the three different estimates corresponding to $\mathcal{C} = 3, 4, 5$ (points are ordered according to patient and, when more than 1 point is present per patient, according to the position of the point along the abscissa parameter s). The rows of U thus give the positions of corresponding zero-curvature points in the three different estimates. The first principal component of U is equal to $\frac{1}{\sqrt{3}}(1, 1, 1)$, up to the 4th decimal digit, and the three eigenvalues are respectively 1487.539, 0.187, 0.021. We thus verify that, jointly, the second and third eigenvalue are not significantly different from $(0, 0)$. The p-value of the approximate chi-squared test is 0.20. This supports the assumption that the positions of the 34 zero curvature points are essentially the same in the three different estimates.

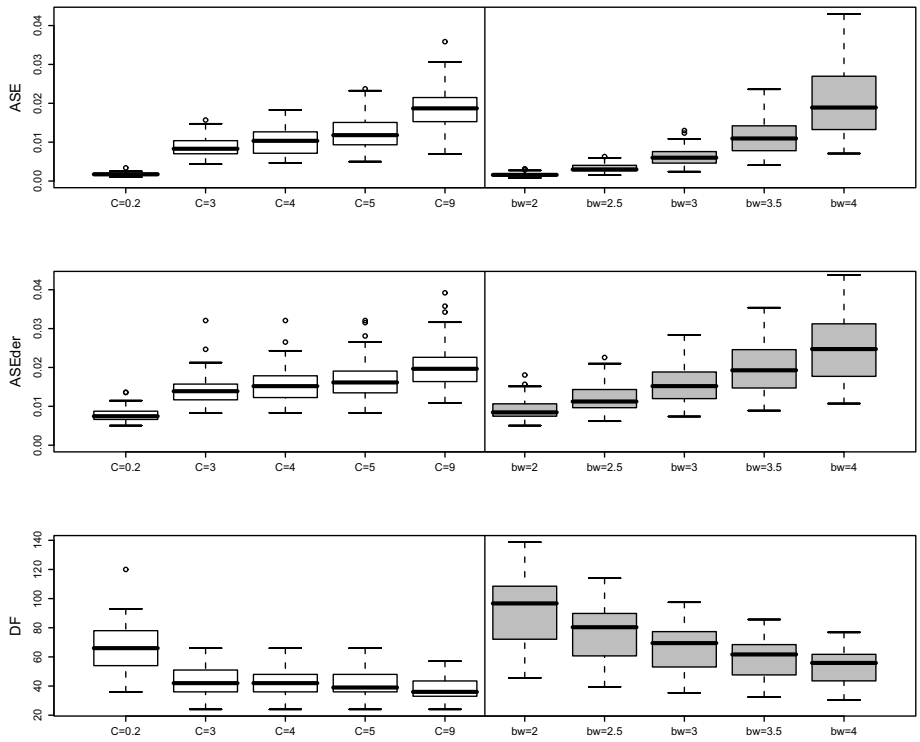

Figure 9: Left: boxplots of the distribution of ASE, ASEder, and DF for the fits corresponding to the 65 patients, obtained by free knot regression splines with penalizations $\mathcal{C} = 0.2, 3, 4, 5, 9$. Right: the same for the fits obtained by local polynomial smoothing with bandwidth $bw = 2, 2.5, 3, 3.5, 4$.

5 Comparison with local polynomial smoothing

In this section we compare free knot spline regression with a more classical model competitor, namely local polynomial smoothing. The latter technique was formerly used to estimate centerlines curvature in Sangalli et al. (2007b), where estimates of centerlines were obtained fitting order-5 polynomials, with a Gaussian kernel and bandwidth equal to 3.

Likewise free knot splines, also local polynomials are a linear estimator, $\hat{W} = LW$. Let m be the order of the polynomials, bw the bandwidth, and $\mathcal{K}(\cdot)$ the kernel. Moreover, denote by $K^{1/2}(s^*)$ the $(n \times n)$-diagonal matrix having j-th entry equal to $\big(\mathcal{K}\big((s^* - s_j)/bw\big)/bw\big)^{1/2}$, and set

$$S(s^*) = \begin{bmatrix} 1 & (s_1 - s^*) & \cdots & (s_1 - s^*)^{m-1} \\ \vdots & & \vdots & \\ 1 & (s_n - s^*) & \cdots & (s_n - s^*)^{m-1} \end{bmatrix}$$

Then the pointwise centerline estimate at s^*, given by local polynomials, is

$$(\hat{x}^*, \hat{y}^*, \hat{z}^*) = \mathbf{e}_{1,m}^T \hat{\Psi}(s^*) \tag{7}$$

where $\mathbf{e}_{r,m}$ represent the r-th column of an $(m \times m)$ identity matrix, and the $(m \times 3)$-matrix $\hat{\Psi}(s^*)$ is given by

$$\begin{aligned}\hat{\Psi}(s^*) &= argmin_{(\Psi)}\big\{tr\big[K^{1/2}(s^*)^T(W - S(s^*)\Psi)(W - S(s^*)\Psi)^T K^{1/2}(s^*)\big]\big\} \\ &= \big[S(s^*)^T K^{1/2}(s^*)K^{1/2}(s^*)S(s^*)\big]^{-1} S(s^*)^T K^{1/2}(s^*)K^{1/2}(s^*)\, W\end{aligned}$$

Using this expression for $\hat{\Psi}(s^*)$, it is possible to rewrite (7) as

$$(\hat{x}^*, \hat{y}^*, \hat{z}^*) = \sum_{j=1}^{n} T\Big(s^*, \frac{s_j - s^*}{bw}\Big)(x_j, y_j, z_j)$$

where

$$T(u, v) = \mathbf{e}_{1,m}^T \big[S(u)^T K^{1/2}(u)K^{1/2}(u)S(u)\big]^{-1} diag\{1, v, \ldots, v^{m-1}\}\mathcal{K}(v)/bw$$

and $diag\{\cdot\}$ denotes a diagonal matrix. It follows that $\hat{W} = L\,W$, where the (i, j)-th entry of the square matrix L is given by

$$L(i, j) = T\Big(s_i, \frac{s_j - s_i}{bw}\Big).$$

To compute the trace of the smoothing matrix L, i.e. the DF of local polynomial estimator, we use the empirical formula provided by Zhang (2003):

$$tr(L) = m + \frac{n}{n-1}\,\mathcal{K}_0\,\frac{|s_n - s_1|}{bw}$$

where

$$\mathcal{K}_0 = \mathcal{K}(0)\,\mathbf{e}_{1,m}^T\, M^{-1}\,\mathbf{e}_{1,m}$$

and the $(m \times m)$-matrix M has (i, j)-th entry given by $M(i, j) = \mu_{i+j-2}$, with $\mu_d = \int t^d \mathcal{K}(t)dt$.

Remark 2 If the three coordinates in the error term in model (1) were assumed to be correlated, as in Remark 1, then it would be natural to look for the value of $\Psi(s^*)$ which minimises

$$tr\big[K^{1/2}(s^*)^T(W - S(s^*)\Psi)\Sigma^{-1}(W - S(s^*)\Psi)^T K^{1/2}(s^*)\big]$$

where Σ is the correlation matrix of the error term. It is easy to see that

$$\begin{aligned}&argmin_{(\Psi)}\big\{tr\big[K^{1/2}(s^*)^T(W - S(s^*)\Psi)\Sigma^{-1}(W - S(s^*)\Psi)^T K^{1/2}(s^*)\big]\big\} \\ &= argmin_{(\Psi)}\big\{tr\big[K^{1/2}(s^*)^T(W - S(s^*)\Psi)(W - S(s^*)\Psi)^T K^{1/2}(s^*)\big]\big\}\end{aligned}$$

so that also the local least square estimate $\mathbf{e}_{1,m}^T \hat{\Psi}(s^*)$ does not depend on Σ.

We now want to argue that, at the cost of an affordable increase in computational cost, estimation by free knot splines has many comparative advantages over kernel smoothing.

Data dimension reduction and computational advantages. Two important gains obtained by using free knot splines are related to their functional nature. Splines naturally yields a dimension reduction of data, which is a fundamental issue for our highly dimensional dataset; moreover, the closed functional form of their estimates can be exploited in subsequent analyses of the fitted curves, avoiding reiterated numerical approximations.

Efficiency. Smoothing by free knot regression splines is more efficient than smoothing by local polynomials, in the sense that the former technique attains lower ASE, and lower ASEder, using less DF. This is shown in Figure 9, which compares the boxplots of the distribution of ASE, ASEder, and DF, for the fits corresponding to the 65 patients, obtained with the two different techniques. Various values of $\mathcal{C}$ and bw are considered, covering a wide range of ASE. Note that there is no value of the bandwidth for which local polynomials can beat free knot splines on both ASE and DF; whereas it is always possible to find a value of $\mathcal{C}$ for which splines estimates do better than local polynomials in both ASE, DF, and also ASEder.

Accuracy. Free knot splines can better estimate the salient features of the curves, expressed by its first derivatives. In fact, even with an higher ASE, splines can attain a lower ASEder. See Figure 9. This means that spline estimates can simultaneously be more accurate and less data adapted. Note that spline estimates with $\mathcal{C} = 4$, the particular value we have chosen, and the local polynomial estimates with $bw = 3$, considered in Sangalli et al. (2007b), have similar ASEder, but splines have higher ASE and less DF. Spline estimates have an higher accuracy thanks to their local adaptivity, which allows them to better detect sharp peaks and troughs in the first and second derivatives. Figure 10, for example, compares estimates of $x'(s)$ for patient 1, obtained by central differences, by free knot regression splines with $\mathcal{C} = 4$ (ASE=0.0115, ASEder=0.0161, and DF=20), and by local polynomial smoothing with $bw = 3$ (ASE=0.0056, ASEder=0.0159, and DF=32.28). Even if for this particular patient, the ASEder is lower for local polynomials than for splines, splines can better get the peaks and troughs in the first derivative, being otherwise smooth, whereas local polynomials cannot fully get these features. Also in subsequent derivatives, spline estimates exhibit more clear-cut peaks and troughs, being otherwise smooth, whereas local polynomial estimates show flatter local features, but are more wiggly over the whole range. Note that the ability of spline of better detecting

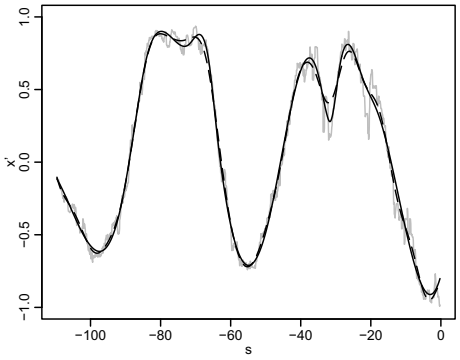

Figure 10: Estimates of x' obtained by free knot splines with $\mathcal{C} = 4$ (solid), and by local polynomial smoothing with $bw = 3$ (dashed), superimposed to first central differences (grey) [patient 1].

salient local features is particular important in our problem, where a big interest lies in curvature landmarks such as peaks and zeros, and their possible influences on hemodynamics. It should be mentioned that this very comparative advantage of free knot splines over other smoothing methods has been also evidenced by Gervini (2006), who carried out comparative simulation studies of free knot splines versus smoothing splines.

Sangalli et al. (2007b) used local polynomials to obtain pointwise estimates of vessel centerlines and their curvature functions, and carried out explorative analyses of vessel curvature and radius profiles. Firstly, centerlines estimates undergo a process of registration, in order to separate their phase variability from their amplitude variability. Then, in order to find the main uncorrelated modes of variability, two separate functional principal component analyses are performed on the registered radius and curvature profiles. Finally, a quadratic discriminant analysis, of principal components scores, identifies the optimal number of principal components that discriminate at best the patients with an aneurysm on the Willis Circle and the remaining patients. This first study supported the existence of a strong relationship between vessel geometry and aneurysm location. This conclusion has been then confirmed in Sangalli et al. (2007a), where the same analyses are performed using centerline estimates based on splines. The consistency of the results is obtained thanks to the fact that the first two principal components of the registered radius and curvature profiles, obtained with splines estimates, do not display significant differences with respect to the old ones, obtained with local polynomials. Differences become instead appreciable in the subsequent principal components of curvature, that were not considered in this study, since discarded by the quadratic discriminant analysis.

6 Discussion

In this paper we have shown how to efficiently estimate a 3D curve and its derivatives by means of free knot regression splines. We have shown that the estimates obtained are more efficient than the estimates based on local polynomials, and that they can better estimate the local salient feature of the curve. Very recently, Gluhovsky and Gluhovsky (2007) proposed a method to choose location-dependent bandwidths in local polynomials. A location-dependent bandwidth may also limit the drawbacks of local polynomials experienced in our 3D problem. An extension to the 3D case of the complex technique proposed by the two authors goes beyond the scope of the present paper. In the same paper, Gluhovsky and Gluhovsky express perplexities about the possibility of obtaining local adaptivity with regression splines methods. We believe our work dispels such doubts.

Acknowledgments

We are grateful to Edoardo Boccardi (*Ospedale Niguarda Ca' Granda*), who provided the 3D-angiographies, and motivated our reaserch by posing fascinating medical questions. Our statistical analysis would not have been possible without the image reconstructions performed by Luca Antiga and Marina Piccinelli (*Istituto Mario Negri*).

References

Antiga, L., Ene-Iordache, B., and Remuzzi, A. (2003), "Computational geometry for patient-specific reconstruction and meshing of blood vessels from MR and CT angiography," *IEEE Trans Med Imaging*, 22(5), 674–684.

Berger, S. A., Talbot, L., and Yao, L. S. (1983), "Flow in curved pipes," *Annual Review of Fluid Mechanics*, 15, 461–512.

Buja, A., Hastie, T., and Tibshirani, R. (1989), "Linear smoothers and additive models," *Ann. Statist.*, 17, 453–555.

Caro, C. G., Doorly, D. J., and Tarnakasky, M. (1996), "Non-planar curvature and branching of arteries and non-planar type-flow," *Proc Roy Soc*, 452, 185–197.

Castro, M. A., Putman, C. M., and Cebral, J. R. (2006), "Computational Fluid Dynamics Modeling of Intracranial Aneurysms: Effect of Parent Artery Segmentation on Intra-Aneurysmal Hemodynamics," *Am J Neuroradiol*, 27, 1703–1709.

Chandran, K. (1993), "Fluid Dynamics in the Human Aorta," *ASME J Biomech Engr*, 115, 611–616.

de Boor, C. (1978), *A practical guide to splines*, vol. 27 of *Applied Mathematical Sciences*, New York: Springer-Verlag.

Dean, W. R. (1927a), "Note on the motion of fluid in a curved pipe," *Phil. Mag.*, 4, 208–223.

— (1927b), "The streamline motion of fluid in a curved pipe," *Phil. Mag.*, 5, 673–695.

Friedman, J. H. (1991), "Multivariate adaptive regression splines," *Ann. Statist.*, 19, 1–141, with discussion and a rejoinder by the author.

Gervini, D. (2006), "Free-knot spline smoothing for functional data," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68, 671–687.

Gluhovsky, I. and Gluhovsky, A. (2007), "Smooth location-dependent bandwith selection for local polynomial regression," *J. Amer. Statist. Assoc.*, 102, 718–725.

Hassan, T., Timofeev, E. V., Saito, T., Shimizu, H., Ezura, M., Matsumoto, Y., Takayama, K., Tominaga, T., and Takahashi, A. (2005), "A proposed parent vessel geometry-based categorization of saccular intracranial aneurysms: computational flow dynamics analysis of the risk factors for lesion rupture," *J. Neurosug.*, 103, 662–680.

Hastie, T. J. and Tibshirani, R. J. (1990), *Generalized additive models*, vol. 43 of *Monographs on Statistics and Applied Probability*, London: Chapman and Hall Ltd.

Hoi, Y., Meng, H., Woodward, S. H., Bendok, B. R., Hanel, R. A., Guterman, L. R., and Hopkins, L. N. (2004), "Effects of arterial geometry on aneurysm growth: three-dimensional computational fluid dynamics study," *J Nerusurg*, 101, 676–681.

Luo, Z. and Wahba, G. (1997), "Hybrid adaptive splines," *J. Amer. Statist. Assoc.*, 92, 107–116.

Mao, W. and Zhao, L. H. (2003), "Free-knot polynomial splines with confidence intervals," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 65, 901–919.

Piccinelli, M., Bacigaluppi, S., Boccardi, E., Ene-Iordache, B., Remuzzi, E., Veneziani, A., and Antiga, L. (2007), "Influence of internal carotid artery geometry on aneurism location and orientation: a computational geometry study," Available at www.maths.emory.edu.

Sangalli, L. M., Secchi, P., and Vantini, S. (2007a), "Functional data analisys for 3D-geometries of the Inner Carotid Artery," in *Book of Short Papers of S.Co. 2007 Conference*, pp. 427–432, available at http://venus.unive.it/sco2007/ocs/index.php.

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2007b), "Functional Data Analisys. A Case Study: Geometrical Features of the Internal Carotid Artery." Tech. rep., MOX, Dipartimento di Matematica, Politecnico di Milano, in preparation.

Smith, F. T. (1976), "Fluid Flow into a Curved Pipe," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 351, 71–87.

Stein, C. M. (1981), "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, 9, 1135–1151.

Stone, C. J., Hansen, M. H., Kooperberg, C., and Truong, Y. K. (1997), "Polynomial splines and their tensor products in extended linear modeling," *Ann. Statist.*, 25, 1371–1470, with discussion and a rejoinder by the authors and Jianhua Z. Huang.

Ustun, C. (2005), "Dr. Thomas Willis' famous eponym: the circle of Willis," *J Hist Neurosci*, 14, 16–21.

Zhou, S. and Shen, X. (2001), "Spatially adaptive regression splines and accurate knot selection schemes," *J. Amer. Statist. Assoc.*, 96, 247–259.