Electronic Journal of Statistics
Vol. 8 (2014) 1879–1890
ISSN: 1935-7524
DOI: 10.1214/14-EJS938

AneuRisk65: A dataset of three-dimensional cerebral vascular geometries*

Laura M. Sangalli, Piercesare Secchi† and Simone Vantini

MOX – Department of Mathematics, Politecnico di Milano
Piazza Leonardo da Vinci 32, 20133, Milano, Italy
e-mail: laura.sangalli@polimi.it; piercesare.secchi@polimi.it; simone.vantini@polimi.it

Abstract: We describe the AneuRisk65 data, obtained from image reconstruction of three-dimensional cerebral angiographies. This dataset was collected for the study of the aneurysmal pathology, within the AneuRisk Project. It includes the geometrical reconstructions of one of the main cerebral vessels, the Inner Carotid Artery, described in terms of the vessel centreline and of the vessel radius profile. We briefly illustrate the data derivation and processing, explaining various aspects that are of interest for this applied problem, while also discussing the peculiarities and critical issues concerning the definition of phase and amplitude variabilities for these three-dimensional functional data.

Keywords and phrases: Multidimensional curves, three-dimensional angiographies, AneuRisk65.

Received August 2013.

1. The AneuRisk project and the AneuRisk65 data

The AneuRisk65 data have been collected within the AneuRisk project,[1] a scientific endeavour that aimed at investigating the role of vessel morphology, blood fluid dynamics and biomechanical properties of the vascular wall, on the pathogenesis of cerebral aneurysms. The project has gathered together researchers of different scientific fields, ranging from neurosurgery and neuroradiology to statistics, numerical analysis and bio-engineering.

Detailed descriptions of the project's aims can be found at the AneuRisk webpage http://mox.polimi.it/it/progetti/aneurisk/, where the AneuRisk65 data can be downloaded. These data include the image reconstructions of one of the main cerebral vessels, the Inner Carotid Artery (ICA), described in terms

*Related articles: 10.1214/14-EJS938A, 10.1214/14-EJS938B, 10.1214/14-EJS938C, 10.1214/14-EJS938D, 10.1214/14-EJS938E; rejoinder at 10.1214/14-EJS938REJ.

†Corresponding author.

[1]This project involved the MOX Laboratory for Modeling and Scientific Computing (Dip. di Matematica, Politecnico di Milano), Laboratory of Biological Structure Mechanics (Dip. di Ingegneria Strutturale, Politecnico di Milano), Istituto Mario Negri (Ranica), Ospedale Niguarda Ca' Granda (Milano) and Ospedale Maggiore Policlinico (Milano), and has been supported by the Fondazione Politecnico di Milano and Siemens Medical Solutions Italia.

of the vessel centreline and of the vessel radius profile. In this work we shall briefly describe the data processing, including the accurate estimation of these three-dimensional curves and their derivatives by multidimensional free-knot splines, discuss the issue of phase variation and how this interplays with data classification (see Sangalli et al., 2009a,b, 2010, for details on these matters), highlighting various aspects that are of interest for this applied problem and critical issues that must be considered when analyzing the data. An increasing data warehouse concerning aneurysm pathology can be accessed from the AneuRisk Web Repository http://ecm2.mathcs.emory.edu/aneurisk managed by Emory University and Orobix. These data also include the full three-dimensional reconstructions of the ICA walls, and of the connecting arteries, as well as data concerning hemodynamical quantities, such as wall shear stress and pressure, obtained via computational fluid dynamics in the real subject-specific ICA geometries (see Passerini et al., 2012).

2. Problem and data

Cerebral aneurysms are deformations of cerebral vessels characterized by a bulge of the vessel wall. This is a common pathology in the adult population, usually asymptomatic and not disrupting. Epidemiological statistics (see, e.g., Rinkel et al. (1998)) suggest that between 1% and 6% of adults develop a cerebral aneurysm during their lives. The rupture of a cerebral aneurysm, even if quite uncommon (about one event every 10,000 adults per year), is usually a tragic event, with very high mortality. Unfortunately, rupture-preventing therapies, both endovascular and surgical treatments, are not without risks; this adds to the fact that in clinical practice general indications about rupture risk are still missing. Even the origin of the aneurysmal pathology is still unclear. Possible explanations that have been discussed in the medical literature focus on interactions between the 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. In particular, it has been conjectured that the pathogenesis of these deformations is influenced by the morphological shape of cerebral arteries, through the effect that the morphology has on

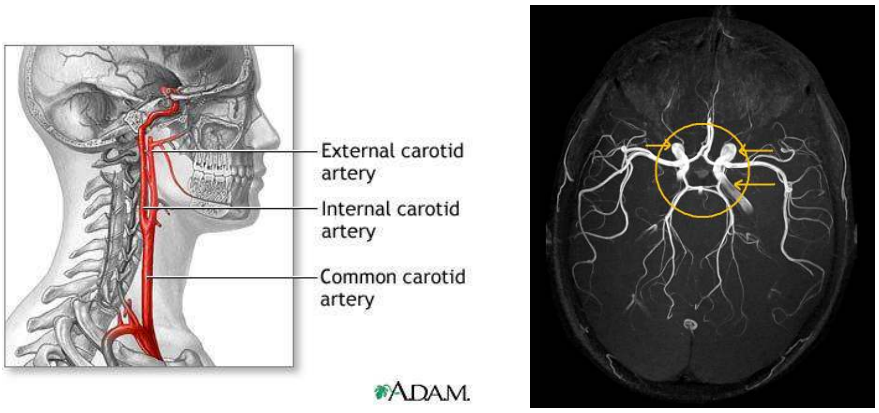

FIG 1. *Left: drawing of an Internal Carotid Artery (http://www.adam.com); the ICA sits for most of its length outside the skull, surrounded by the neck muscle tissues; just before its terminal bifurcation it enters inside the skull, passing through a dural ring (i.e., a hole in the skull bone). Right: the Willis circle, located at the base of the brain, inside the skull, is a net of small arteries and capillaries connecting the main arteries bringing blood to the brain; the terminal parts of the left and right ICAs, clearly visible in the image, are indicated by arrows.*

the hemodynamics. For this reason, the main goal of the AneuRisk project has been the study of relationships between vessel morphology and aneurysm presence and location. The association between vessel geometry and the aneurismal pathology is also explored, e.g., in Cebral et al. (2005); Castro, Putman and Cebral (2006); Ma et al. (2007); Meng et al. (2013)

These lesions may originate along the left or right Internal Carotid Artery, two large arteries bringing blood to the brain, or at or after the terminal bifurcation of the ICA, in the so-called Willis Circle. Each of the two ICAs sits for most of its length outside the skull, along the neck, surrounded by muscle tissues; just before its terminal bifurcation it enters inside the skull, passing through a dural ring (i.e., a hole in the skull bone). See Figures 1 and 2. Arteries downstream of the ICA terminal bifurcation float in the brain humor, inside the skull. For this reason, aneurysms located at or after the ICA terminal bifurcation are more life-threatening; the possible rupture of one such aneurysm is fatal in most cases.

The AneuRisk65 dataset is based on a set of three-dimensional angiographic images taken from 65 subjects, hospitalized at Niguarda Ca' Granda Hospital (Milan), who were suspected of being affected by cerebral aneurysms. Out of these 65 subjects, 33 subjects have an aneurysm at or after the terminal bifurcation of the ICA ("Upper" group), 25 subjects have an aneurysm along the ICA ("Lower" group), and 7 subjects were found without any visible aneurysm during the angiography ("No-aneurysm" group). As commented above, Upper group subjects are those with the most dangerous aneurysms; for this and other clinical reasons, for some statistical analyses it might make sense to join the Lower and No-aneurysm groups in a unique group, to be contrasted to the Upper group. Percentages of females and males and of right and left ICAs do not

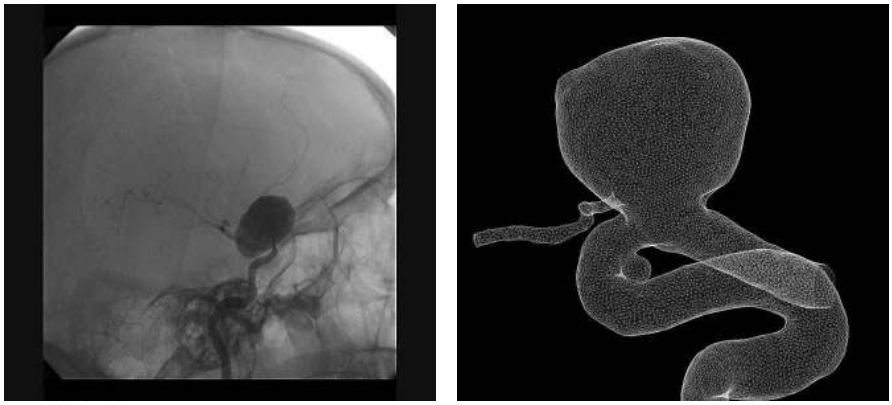

FIG 2. *Left: X-rays image of an aneurysm along an ICA; the artery, with its siphon, is clearly visible in the image. Right: image reconstruction of an ICA with aneurysm (different subject with respect to left panel).*

differ significantly from 50% (the p-values of the test for equal proportions are 0.14 and 0.78, respectively). Age, apart from a superior outlier, appears normally distributed (the p-value of the Shapiro-Wilk test is 0.29), with a sample mean equal to 55.85 years and a sample standard deviation equal to 13.45 years. Gender and age are not included in the dataset because they are supposed to be related to the aneurysmal pathology only through their effect on the morphology of the vessel. Notice that, due to the high radiation quantity implied by the scan, this exam is only performed in case of acute symptoms and if the doctor strongly suspects the presence of an aneurysm. For the same reason, no follow-up is ever performed.

The analyses conducted within the AneuRisk project have focussed on the ICA, which is clearly recognizable in each of the 65 angiographies. Starting from the three-dimensional array of grey-scaled voxels that is generated by the angiography (with lighter voxels showing the presence of flowing blood), the artery lumen (i.e., the volume that is occupied by flowing blood) is identified by a reconstruction algorithm that is coded in the "Vascular Modeling ToolKit" (VMTK). See Piccinelli et al. (2011) and Piccinelli et al. (2009). Figure 3 shows the reconstruction of the ICA of the first subject in the dataset, and also displays the reconstructed centreline of the vessel. The centreline is computed as the set of centers of maximal spheres inscribed in the artery lumen. In particular, for every subject i in the dataset ($i = 1, \ldots, 65$), VMTK reconstruction of the ICA centreline 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 respectively the left/right, up/down and front/back coordinates of each point. It should though be noticed that the x, y, and z coordinates are not absolute, but are relative to the cubic volume analyzed during angiography, that in turn depends on where the angiographic image has been centered; this means that these coordinates are not directly comparable across subjects, since they vary with the location of the scanned volume. Points along the centreline

FIG 3. *Three-dimensional image of an Internal Carotid Artery with an aneurysm [subject 1]; the black line inside the vessel is its centreline.*

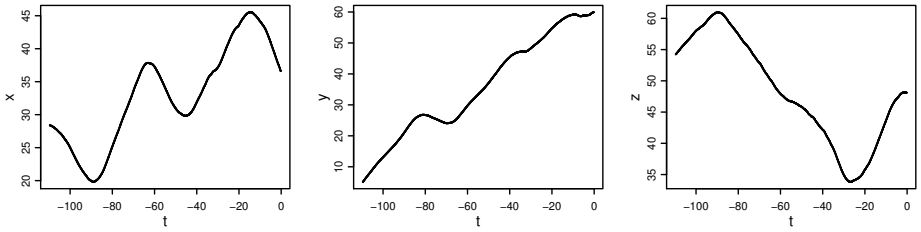

FIG 4. *Reconstructed space coordinates of the ICA centreline for subject 1,* (x_{1j}, y_{1j}, z_{1j})*, versus the abscissa parameter* t_{1j}*, for* $j = 1, \ldots, n_1$ $(n_1 = 1350)$.

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 the ICA is present in each angiography, even if the portion of the ICA captured by the angiography varies from subject to subject (depending on where the angiographic image has been centered). For each subject i, we can associate the set of space coordinates with an index set $\{t_{ij} : j = 1, 2, \ldots, n_i\}$, which measures an approximate distance along the ICA centreline, thus providing an approximate curvilinear abscissa. More precisely, $-t_{i1}$ is the distance of the point (x_{i1}, y_{i1}, z_{i1}) from the terminal bifurcation of the ICA (as determined by VMTK), and, for $j = 2, \ldots, n_i$,

$$t_{ij} - t_{ij-1} = -\sqrt{(x_{ij} - x_{ij-1})^2 + (y_{ij} - y_{ij-1})^2 + (z_{ij} - z_{ij-1})^2}\,.$$

The conventional negative sign highlights that we are moving upstream, i.e., in opposite direction with respect to blood flow. Figure 4, for instance, displays the reconstructed space coordinates of the ICA centreline for subject 1,

(x_{1j}, y_{1j}, z_{1j}), versus the approximate curvilinear abscissa parameter t_{1j}, for $j = 1, \ldots, n_1$ ($n_1 = 1350$). The number n_i of data points available for each subject ranges from 350 to 1380, and is almost perfectly correlated to the approximate length $|t_{i\,n_i} - t_{i\,1}|$ of the reconstructed centreline (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, even if the 65 grids are different. The grids are not evenly spaced; their average step is 0.079mm. The location of the aneurysm along the approximate curvilinear abscissa is provided. Finally, the reconstruction algorithm also provides, for each of the grid points, the radius R_{ij} of the vessel lumen section, computed as the radius of the local maximal inscribed sphere (indicated as Maximal Inscribed Sphere Radius, MISR).

Besides radius, another geometrical quantity that strongly influences the hemodynamics, and hence may in turn play a role on the aneurysm pathogenesis, is the artery curvature, that can be identified by the curvature of the artery centreline. Centreline curvature is not provided by the reconstruction algorithm, but is estimated via the multi-dimensional free knot regression splines techniques summarized in Section 3. See Sangalli et al. (2009a) for details. In general, the three-dimensional shape of the carotid, and in particular the shape of the siphon characterizing the distal part of the ICA, is fundamental in determining the hemodynamics. Of big interest is thus the classification (unsupervised clustering) of ICAs depending on their morphological shape. It is worth mentioning that a classification commonly used in the medical literature, proposed by Krayenbuehl, Huber and Yasargil (1982), discriminates among Γ-shaped, Ω-shaped, and S-shaped ICAs, according to the form of siphon in their distal part, which may resemble the letters Γ, Ω or S, in the presence of zero, one, or two large bends in the ICA siphon, respectively.

3. Data processing

The reconstructed ICA centrelines are of course affected by measurement and reconstruction errors. To compute their curvature it is hence necessary to obtain accurate estimates of the centrelines themselves as well as of their first two derivatives. Figure 5 shows the first and second central differences of reconstructed centrelines, that are rough pointwise estimates of first and second derivatives computed at each grid point as normalized differences of data values at nearby grid points. These rough pointwise estimates appears very noisy. In Sangalli et al. (2009a) we hence proposed a regression technique for the estimation of the true centreline $\mathbf{f}(t) = \big(x(t), y(t), z(t)\big)$, based on free knot regression splines. The proposed technique is shown to provide very good estimates of multidimensional curves and their derivatives. Free-knot splines are regression splines where the number and position of the knots are not fixed in advance, but chosen in a way to minimize a penalized sum of squared errors criterion. Since centreline data are three-dimensional, the idea is to fit simultaneously the three space co-ordinates of the centreline $\big(x(t), y(t), z(t)\big)$, looking for the optimal spline knots along the abscissa parameter t. Estimates of $\mathbf{f}'$ and $\mathbf{f}''$, and

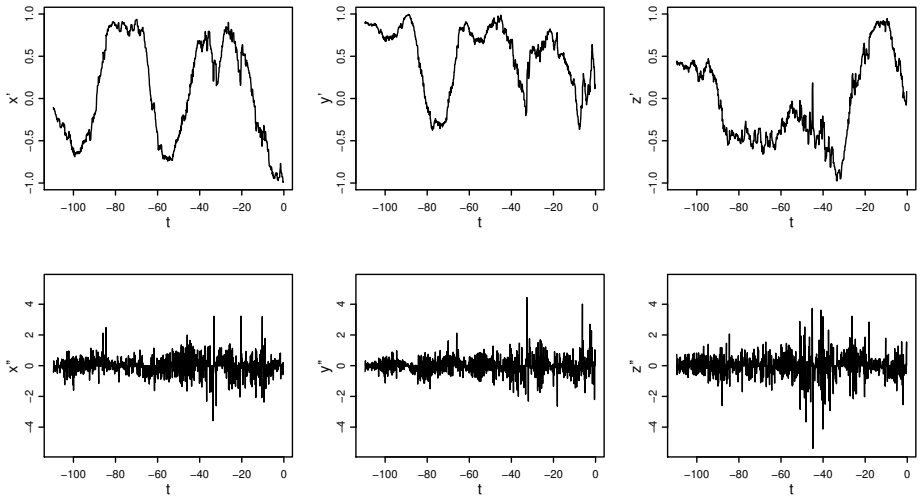

FIG 5. *First and second differences (top and bottom, respectively) of reconstructed space coordinates of the ICA centreline for subject 1.*

hence of $curv_{\mathbf{f}}$, are thus obtained by differentiation of the fitted regression spline. Note that derivatives of splines are still splines (of appropriate order), with the same knot vector and coefficients directly computed from the coefficients of the original spline.

In particular, for each subject i we carry out a separate estimation process. Dropping the subject subscript i, denote by $(t_j, x_j), (t_j, y_j), (t_j, z_j)$, for $j = 1, \ldots, n$, the reconstructed vessel centreline. Let $\{b^{[\mathbf{k}]}_{r,m}(t) : r = 1, \ldots, m + n_k\}$ be the b-spline basis system for the space of splines of order m, over the abscissa interval $[a, b]$, with knot vector $\mathbf{k} = (k_1, \ldots, k_{n_k})$ where $a < k_1 < \cdots < k_{n_k} < b$. Then, the three space coordinates of the vessel centreline for the considered subject are estimated by

$$\hat{x}(t) = \sum_{r=1}^{m+\hat{nk}} \hat{\lambda}^{[x]}_r b^{[\hat{\mathbf{k}}]}_{r,m}(t), \quad \hat{y}(t) = \sum_{r=1}^{m+\hat{nk}} \hat{\lambda}^{[y]}_r b^{[\hat{\mathbf{k}}]}_{r,m}(t), \quad \hat{z}(t) = \sum_{r=1}^{m+\hat{nk}} \hat{\lambda}^{[z]}_r b^{[\hat{\mathbf{k}}]}_{r,m}(t)$$

where the optimal number $\hat{nk}$ of knots, the optimal positions $\hat{\mathbf{k}} = \{\hat{k}_1, \ldots, \hat{k}_{\hat{nk}}\}$ of the knots along the abscissa t, and the optimal vectors of basis expansion coefficients for the three space directions, $\hat{\lambda}^{[x]}_r$, $\hat{\lambda}^{[y]}_r$ and $\hat{\lambda}^{[z]}_r$, are jointly estimated by minimizing the following penalized sum of square error criterion:

$$\sum_{j=1}^{n} \left(x_j - \sum_{r=1}^{m+n_k} \lambda^{[x]}_r b^{[\mathbf{k}]}_{r,m}(t_j) \right)^2 + \sum_{j=1}^{n} \left(y_j - \sum_{r=1}^{m+n_k} \lambda^{[y]}_r b^{[\mathbf{k}]}_{r,m}(t_j) \right)^2 + \sum_{j=1}^{n} \left(z_j - \sum_{r=1}^{m+n_k} \lambda^{[z]}_r b^{[\mathbf{k}]}_{r,m}(t_j) \right)^2 + \mathcal{C}(m + n_k)$$

The optimal knots are searched by a generalization to the multidimensional case of the algorithm described in Zhou and Shen (2001).

The spline order is set to $m = 5$, to obtain smooth estimates of the first two derivatives. A common value of the smoothing parameter C is chosen for the overall dataset of 65 subjects (see Sangalli et al., 2009a, for details on the choice of C). This assumption is justified by the fact that both the machine that was used to take the three-dimensional-angiographies and the reconstruction algorithm are the same for each subject, so that the signal to noise ratio is expected to be the same across subjects.

The top line of Figure 6 shows the estimates, $\hat{x}(t), \hat{y}(t), \hat{z}(t)$, of the three co-ordinate functions of the ICA centreline for subject 1. The vertical lines show the position of the knots along the abscissa parameter t. The estimates are superimposed on the original $(t_j, x_j), (t_j, y_j), (t_j, z_j)$, for $j = 1, \ldots, n$ (grey dots, almost completely hidden by the estimates). Center and bottom lines of the same figure show first and second derivatives of $\hat{x}(t)$, $\hat{y}(t)$ and $\hat{z}(t)$, superimposed

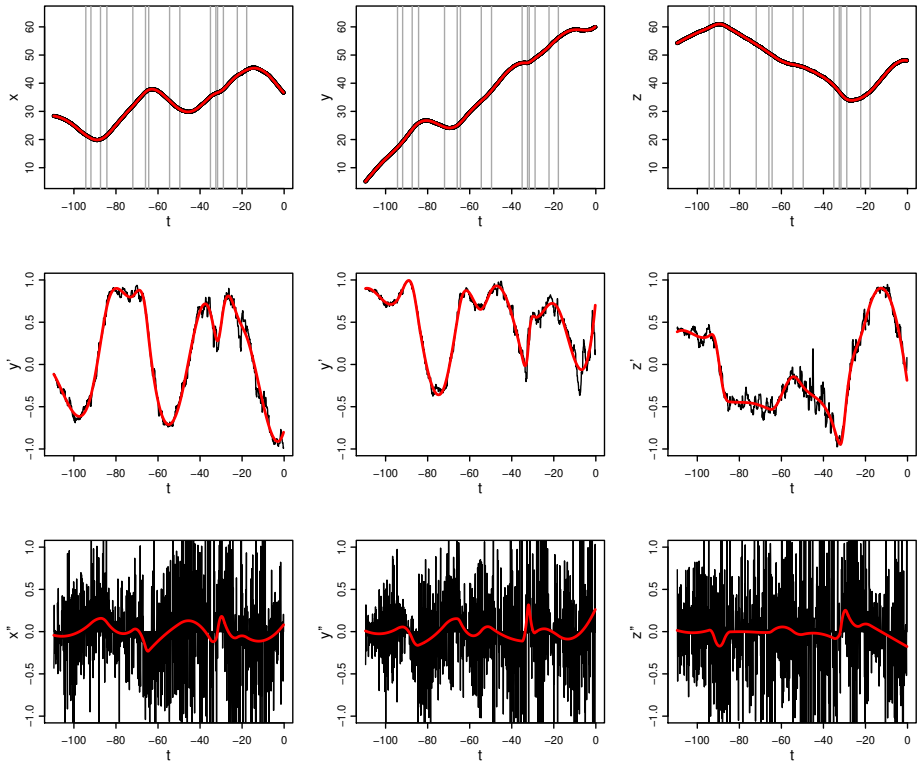

FIG 6. *Top: fitted co-ordinates curves $\hat{x}(t), \hat{y}(t), \hat{z}(t)$, in red, with vertical lines showing the position of the knots along the abscissa parameter t, superimposed on the original data $(t_j, x_j), (t_j, y_j), (t_j, z_j)$, for $j = 1, \ldots, n$ [subject 1]. Center: first derivatives of $\hat{x}(t)$, $\hat{y}(t)$ and $\hat{z}(t)$, in red, superimposed on the first central differences. Bottom: second derivatives of $\hat{x}(t)$, $\hat{y}(t)$ and $\hat{z}(t)$, in red, superimposed on the second central differences.*

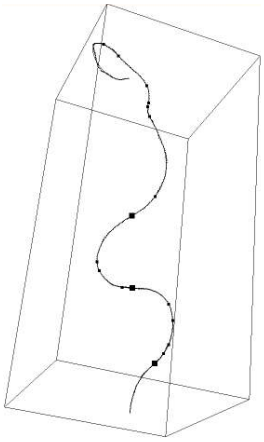

FIG 7. *Three dimensional image of estimated centreline together with raw data, for subject 1; the little bullets show the positions of the spline knots, while the big squares are points of approximately zero curvature.*

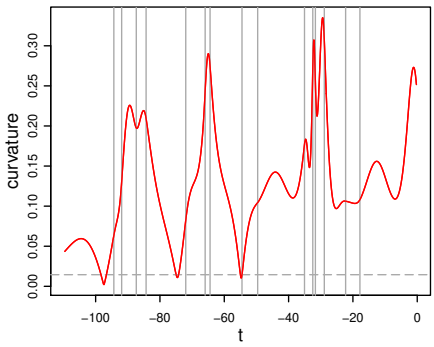

FIG 8. *Centreline curvature for subject 1, computed from first and second derivatives of estimated centreline, displayed in Figure 6. The horizontal dashed line is the approximately zero curvature threshold.*

on the rough estimates that are given respectively by first and second central differences (grey). Thanks to their local adaptivity, free-knot splines estimates efficiently detect the salient features of the curves and are able to fully capture the clear-cut peaks and troughs in the derivatives, but are otherwise smooth. Figure 7 is a three-dimensional image of the fitted centreline, $\hat{\mathbf{f}}(t)$, and Figure 8 shows the corresponding curvature function.

All these processed data (estimates of centrelines and corresponding estimates of centrelines first two derivatives and curvature, for each subject) are included in the AneuRisk65 dataset.

4. A first look at phase variability in the AneuRisk65 data

The first derivatives of the estimated ICA centrelines of the 65 subjects are displayed in Figure 9; in the picture, left carotids have been left-right reflected so that the orientation in the three-dimensional space of all reconstructed ICAs is the same (left-right reflected ICA coordinates are also provided in the AneuRisk65 dataset). As apparent from the figure, the three-dimensional centrelines display a considerable misalignment. This misalignment is the expression of a strong phase variability present among the data, largely due to the different dimensions of the ICA of the various subjects; if not taken properly into account, this misalignment acts as a confounding factor in the data analysis. To enable meaningful comparisons across subjects, it is thus necessary to efficiently decouple the phase and the amplitude variability, the former being mainly due to the differences in the dimensions of the subject's carotids and the latter instead to the differences in their morphological shapes. Sangalli et al. (2009b,

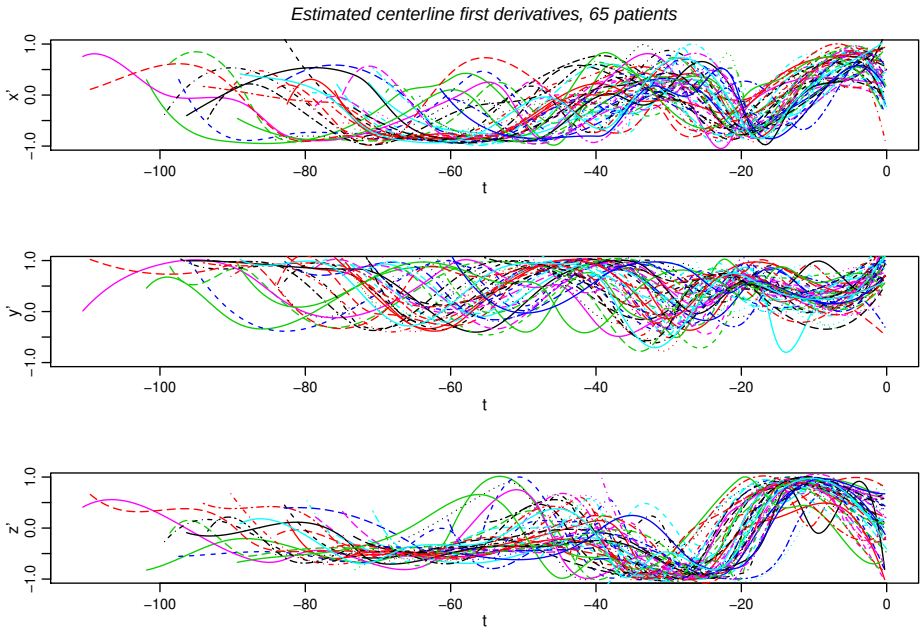

FIG 9. *First derivatives* $\{x'(t), y'(t), z'(t)\}$ *of the estimated ICA centrelines for the 65 subjects.*

2010) describe the solutions to this problem given within the AneuRisk project, considering also the issue of classification of these three-dimensional curves.

It should be stressed again that the portion of ICA captured by the angiography varies from subject to subject, depending on where the angiographic image has been centered. Moreover, even if the terminal bifurcation of the ICA is present in each angiography, the bifurcation point identified by VMTK depends on the bifurcation angle and other geometrical quantities. In other words, neither the starting nor the ending abscissas points of the various curves can be matched. Hence, it is inappropriate to directly use a registration method that forces starting and ending abscissas to be the same across curves.

Acknowledgements

We are grateful to Alessandro Veneziani (P.I. of the AneuRisk Project) and to Tiziano Passerini, Marina Piccinelli and Luca Antiga. The AneuRisk Project has been supported by the Fondazione Politecnico di Milano and Siemens Medical Solutions Italia. All authors are grateful to the MBI Mathematical Biosciences Institute http://mbi.osu.edu/ for support. L. M. Sangalli acknowledges funding by the research program Dote Ricercatore Politecnico di Milano – Regione Lombardia, project "Functional data analysis for life sciences", and by MIUR Ministero dell'Istruzione dell'Università e della Ricerca, *FIRB Futuro in Ricerca* starting grant project "Advanced statistical and numerical methods for

the analysis of high dimensional functional data in life sciences and engineering" http://mox.polimi.it/users/sangalli/firbSNAPLE.html.

References

Castro, M. A., Putman, C. M. and Cebral, J. R. (2006). Patient-specific computational fluid dynamics modeling of anterior communicating artery aneurysms: A study of the sensitivity of intra-aneurysmal flow patterns to flow conditions in the carotid arteries. *AJNR American Journal of Neuroradiology* **27** 2061–2068.

Cebral, J. R., Castro, M. A., Burgess, J. E., Pergolizzi, R. S., Sheridan, M. J. and Putman, C. M. (2005). Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models. *AJNR American Journal of Neuroradiology* **26** 2550–2559.

Krayenbuehl, H., Huber, P. and Yasargil, M. G. (1982). Krayenbuhl/Yasargil Cerebral Angiography. *Thieme Medical Publishers, 2nd ed.*

Ma, B., Lu, J., Harbaugh, R. E. and Raghavan, M. L. (2007). Nonlinear anisotropic stress analysis of anatomically realistic cerebral aneurysms. *Journal of Biomechanical Engineering* **129** 88–96.

Meng, H., Tutino, V. M., Xiang, J. and Siddiqui, A. (2013). High WSS or Low WSS? Complex interactions of hemodynamics with intracranial aneurysm initiation, growth, and rupture: Toward a unifying hypothesis. *AJNR American Journal of Neuroradiology.*

Passerini, T., Sangalli, L. M., Vantini, S., Piccinelli, M., Bacigaluppi, S., Antiga, L., Boccardi, E., Secchi, P. and Veneziani, A. (2012). An integrated CFD-statistical investigation of parent vasculature of cerebral aneurysms. *Cardio. Eng. and Tech.* **3** 26–40.

Piccinelli, M., Veneziani, A., Steinman, D. A., Remuzzi, A. and Antiga, L. (2009). A framework for geometric analysis of 852 vascular structures: applications to cerebral aneurysms. *IEEE Trans. Med. Imaging* **28** 1141–1155.

Piccinelli, M., Bacigaluppi, S., Boccardi, E., Ene-Iordache, B., Remuzzi, A., Veneziani, A. and Antiga, L. (2011). Geometry of the ICA and recurrent patterns in location, orientation and rupture status of lateral aneurysms: An image-based computational study. **68** 1270–1285.

Rinkel, G. J., Djibuti, M., Algra, A. and Van Gijn, J. (1998). Prevalence and risk of rupture of intracranial aneurysms: A systematic review. *Stroke* **29** 251–256.

Sangalli, L. M., Secchi, P., Vantini, S. and Veneziani, A. (2009a). Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines. *Journal of the Royal Statistical Society Ser. C, Applied Statistics* **58** 285–306. MR2750007

Sangalli, L. M., Secchi, P., Vantini, S. and Veneziani, A. (2009b). A case study in exploratory functional data analysis: Geometrical features of the internal carotid artery. *J. Amer. Statist. Assoc.* **104** 37–48. MR2663032

Sangalli, L. M., Secchi, P., Vantini, S. and Vitelli, V. (2010). K-mean alignment for curve clustering. *Computational Statistics and Data Analysis* **54** 1219–1233. MR2600827

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *J. Amer. Statist. Assoc.* **96** 247–259. MR1952735