Analysis of AneuRisk65 data: Elastic shape registration of curves[*]

Qian Xie

Department of Statistics
Florida State University, Tallahassee, FL 32306, USA
e-mail: qxie@stat.fsu.edu

Sebastian Kurtek

Department of Statistics
Ohio State University, Columbus, OH 43210, USA
e-mail: kurtek.1@stat.osu.edu

and

Anuj Srivastava

Department of Statistics
Florida State University, Tallahassee, FL 32306, USA
e-mail: anuj@stat.fsu.edu

Abstract: Shapes of carotid arteries are analyzed here as elastic curves under the framework introduced in [2]. Using a mathematical representation of parameterized curves, termed square-root velocity function (SRVF), in conjunction with an elastic Riemannian metric, the framework provides (1) parameterization-invariant shape metrics for comparing curves, (2) simultaneous registration of coordinate functions across curves, and (3) computation of statistical summaries and models of shapes of given curves. The method is applicable to curves in $\mathbb{R}^n$ for $n \geq 1$. Thus, we study the shapes and alignments of carotid arteries using their 3D coordinates and other geometric properties along the curves, such as radii and curvatures. The results show a significant improvement in curve alignments, leading to a compact phase-amplitude PCA representation and modeling of artery data.

Keywords and phrases: Elastic curve registration, square-root velocity function, parameterization-invariance, Riemannian metric.

Received August 2013.

1. Introduction

In this paper we analyze the AneuRisk65 data introduced in the paper [6, 7]. The goal of this dataset is the study of shapes of cerebral vessels to characterize their physiological attributes, particularly the risk for aneurysms. We mainly treat arteries as curves in $\mathbb{R}^3$ in this paper, although some extensions that include scalar-valued functions along the curves, such as curvatures and radii, are

---

[*]Main article 10.1214/14-EJS938.

also briefly mentioned in the analysis. The shape analysis of curves in $\mathbb{R}^n$ has been of great interest in many parts of science and a variety of tools have been developed. The two main ideas used in statistics literature are: (1) landmark-based shape analysis where one selects matching landmarks across all objects and studies the vectors of landmarks for shape analysis [1], and (2) shape analysis of curves in parameterized forms. We will take the latter approach as it provides a comprehensive joint solution for registration and statistical shape analysis. In this approach, the carotid arteries are studied as parameterized 3D curves of the type $f : I \to \mathbb{R}^3$, where $I = [-1, 0]$.

The problem of joint registration and shape analysis can be studied either as pairwise or groupwise. In a pairwise registration problem between two curves, one solves for which point on one curve matches which point on the other. In case the two curves are parameterized, as is the case for us, the registration is controlled by parameterization. That is, for any $t \in I$, the points $f_1(t)$ and $f_2(t)$ are registered. The change in registration is accomplished by a warping function.

Remark 1. It is important to note the difference in the effects of domain-warping functions for different datasets studied in this special section. In case of real-valued functions, e.g. spike trains [4] or proteomics data, a warping of domain I changes the graph of a function, i.e. a display of the function is actually deformed in the process. In contrast, for curves in $\mathbb{R}^3$, e.g. arteries or juggling trajectories, the shape or the display of the curve remains unchanged under domain warping. The domain warping function is therefore termed a re-parameterization function in this context. This is an important fundamental difference in (real-valued) functional data analysis and shape analysis of curves in higher-dimensional Euclidean spaces.

The next question is: What should be the space of re-parameterization functions? For re-parameterization functions from $\mathbb{R}$ to $\mathbb{R}$, there are several subgroups that can be shortlisted. These include subgroups for translation (one-dimensional), scaling (one-dimensional), and diffeomorphisms that fix an interval. The first two result in linear mappings while the last group provides a nonlinear registration between functions. Assuming that the linear mapping has already been performed, we choose $\mathcal{H}$ to be the set of all positive diffeomorphisms from I to itself. The groupwise registration problem is that given multiple curves $f_1, f_2, \ldots, f_n$, we must find the corresponding re-parameterization functions $h_1, h_2, \ldots, h_n$ such that $\{f_i(h_i(t))\}$ are registered for all $t \in I$. The solution comes from computing a *mean* shape under a proper shape metric and then aligning the individual curves to this mean curve.

For any two curves, the quantity $\|f_1 - f_2\|$, denoting the $\mathbb{L}^2$ distance between them, has often been used to compare shapes of curves. However, this quantity changes with the parameterization of curves even though their shapes are invariant to different parameterizations. So, this is unsuitable for shape analysis in its current form. One may be tempted to fix the issue using the quantity $\inf_{h \in \mathcal{H}} \|f_1 - f_2 \circ h\|$, for removing parameterization variability and for comparing shapes of f_1 and f_2, but there are many problems with that setup. It is

not symmetric and is susceptible to the well-known *pinching problem*. A bigger underlying problem is that this quantity is not a proper metric. (Please refer to the discussion in Srivastava et al. [2, 8] for more details on the limitations of this choice.) An important direction in the literature has been the use of new Riemannian metrics that are invariant to re-parameterizations and, thus, avoid the above-mentioned problem. For curves in $\mathbb{R}^2$ there have been several such invariant metrics, see e.g. [5], etc., but so far there are very few ideas that work for $\mathbb{R}^n$ in general. A solution that is applicable to curves in $\mathbb{R}^n$ is summarized next.

2. Elastic shape analysis framework: A brief introduction

This framework is based on the use of the square-root velocity function (SRVF) of the curve f that is defined as $q : I \to \mathbb{R}^3$ by ([2]) $q(t) = \dot{f}(t)/\sqrt{|\dot{f}(t)|}$. Note that if the original curves are restricted to only the absolutely continuous functions, then the SRVFs are elements of $\mathbb{L}^2(I, \mathbb{R}^3)$, and one can recover the curve (up to a translation) by using the equation $f(t) = \int_0^t q(s)|q(s)|ds$. If a curve f is re-parameterized by any $h \in \mathcal{H}$, then the SRVF of the new curve is given by $(q, h) = (q \circ h)\sqrt{\dot{h}}$. Also, if f is rotated by a rotation matrix $O \in SO(3)$, its SRVF also rotates by the same O. The choice of representing curves by SRVFs is important because the $\mathbb{L}^2$ distance between SRVFs of any two curves is unchanged by simultaneous rotation and re-parameterization of these curves, i.e. $\|q_1 - q_2\| = \|O(q_1, h) - O(q_2, h)\|$ for all $O \in SO(3)$ and $h \in \mathcal{H}$. This isometry property avoids the pinching and other problems mentioned above. It can be shown that the $\mathbb{L}^2$ metric in the SRVF space is an example of an *elastic* Riemannian metric [2] that measures a combination of bending and stretching of curves that deforms one into the other.

The carotid arteries for different patients cannot be measured in a single coordinate system and, therefore, require a global rigid alignment before analysis. In other words, shape analysis of curves is required to be invariant to their rigid motions. One often removes the variability associated with global scalings of curves in shape analysis but if there is interest in incorporating global scales in analysis, they can be easily included in the representation. Two of the variables – rigid translation and global scale – are easily removed as follows. The SRVF representation is invariant of translation by definition. If we want to remove the scale variability, we can restrict f to be a unit length curve such that $\int_I |q(t)|^2 dt = \int_I |\dot{f}(t)|dt = 1$. Thus, the space of SRVFs representing unit-length curves is a unit Hilbert sphere $\mathcal{C}$ endowed with the $\mathbb{L}^2$ metric. It is called the *pre-shape space*. In cases when scale is a variable to be considered in analyses, the underlying space is no longer a hypersphere but the full $\mathbb{L}^2$ space. Given any two SRVFs q_1 and q_2 in $\mathcal{C}$, representing two unit-length curves in $\mathbb{R}^3$, we can compute the shortest path, or a geodesic, between them in $\mathcal{C}$, as $\alpha : I \to \mathcal{C}$ such that $\alpha(\tau) = \frac{1}{\sin(\theta)}(\sin(\theta(1 - \tau))q_1 + \sin(\theta\tau)q_2)$, where θ is the arc-length distance between them given by $\theta = d_c(q_1, q_2) = \cos^{-1}(\langle q_1, q_2 \rangle)$. Here θ provides a quantification of differences in f_1 and f_2 in the pre-shape space while

α provides an optimal deformation between them. The remaining two transformations – rotation and re-parameterization – change SRVFs of curves and are removed using an elaborate algebraic technique as follows. We unify all SRVFs resulting from rotations and re-parameterizations of the same curve, using the notion of equivalence classes (or orbits) under the actions of $SO(3)$ and $\mathcal{H}$. Thus, each shape is uniquely associated with an equivalence class defined as the set: $[q] = \{O(q \circ h)\sqrt{\dot{h}}|(h,O) \in \mathcal{H} \times SO(3)\}$. The shape space of curves in $\mathbb{R}^3$ is then the space of such classes; it can also be viewed as the quotient space of $\mathcal{C}$ modulo $SO(3) \times \mathcal{H}$. Thus, for any two curves f_1 and f_2, represented by their SRVF orbits $[q_1]$ and $[q_2]$, the distance between their shapes is given by:

$$d_s([q_1],[q_2]) = \inf_{(h,O)\in\mathcal{H}\times SO(3)} d_c(q_1, O(q_2 \circ h)\sqrt{\dot{h}}). \qquad (2.1)$$

This equation provides two important ingredients for shape analysis – the minimization over h and O results in optimal registration of points across curves, and the quantity d_s forms a proper metric on the shape space of curves to result in a quantification of shape differences. The geodesic α, this time between q_1 and $O^*(q_2, h^*)$, provides the optimal deformation in the shape space. The optimal rotation O^* is found using the Procrustes method while the optimal re-parameterization h^* is found using the dynamic programming algorithm. In case we have an additional feature function, such as the maximum inscribed sphere radius (MISR), along the curve, we can include it in the analysis by forming curves in $\mathbb{R}^4$, the fourth coordinate being the MISR function. Note that the rotation group does not extend to $SO(4)$ but stays at $SO(3)$ to rotate only the coordinates of the original f. The role of $\mathcal{H}$ also remains the same as earlier.

3. Experimental results

In this section we present some experimental results on the Aneurisk65 dataset obtained using the elastic shape analysis described above. As an example, the original coordinate functions of left carotids are shown in Fig. 1. It is observed that the curvilinear abscissa for individual curves are not equal. However, at the same time, the plot shows that the arteries can have similar shapes but with quite different scales (lengths). Therefore, we standardize the abscissa to range from $[-1, 0]$ with 0 meaning closer to bifurcation. To highlight the need for joint registration of different coordinate functions, we apply the method introduced in [3, 8] for 1D functions separately on each coordinate function and show results in Fig. 2. The original coordinate functions and the registered ones are shown. However, for each single 3D curve, the warping functions used to deform the three axes, x, y and z, are not the same. Which ones should be used to align different curves? The natural solution is to jointly align all three coordinates for 3D curve registration and perform simultaneous analysis.

Pairwise registration and geodesic As the first result, the left column in Fig. 3 shows an example of pairwise registration between a pair of carotid arteries. The curves f_1 and f_2 are drawn with colors along curves representing the val-

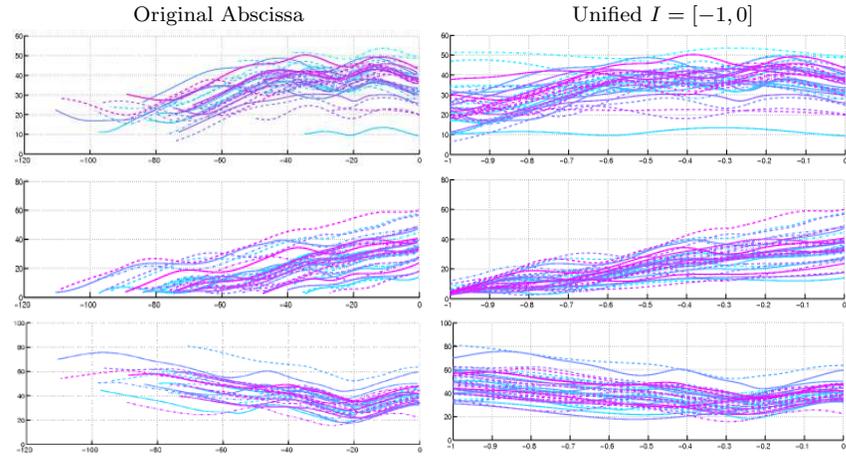Original Abscissa   Unified $I = [-1, 0]$

FIG 1. *Coordinate functions for original curves of left carotids. Each row represents x, y and z coordinate respectively. Left column shows coordinate functions plotted with respect to curvilinear abscissa while the right column presents them plotted with respect to the unified domain interval* $[-1, 0]$.
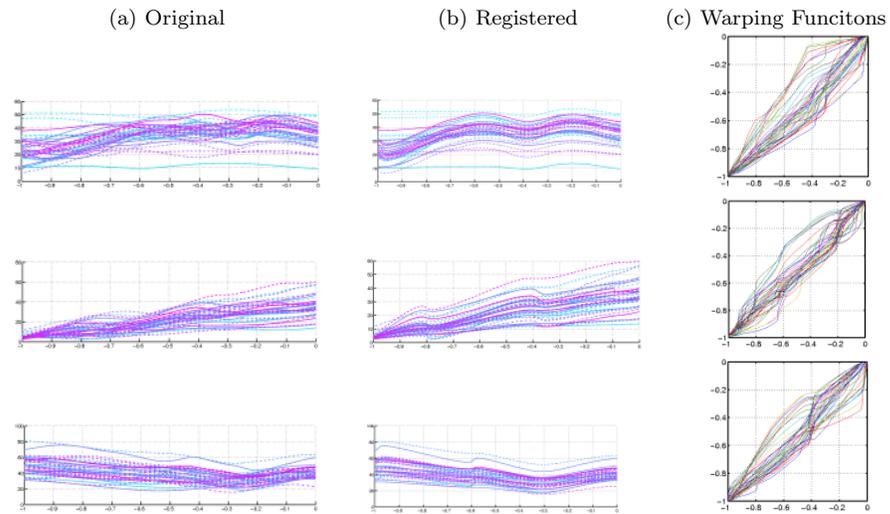
(a) Original   (b) Registered   (c) Warping Funcitons

FIG 2. *Registrations of coordinate functions separately for original curves. Each row represents x, y and z coordinate respectively. Column (a) shows original coordinate functions plotted with respect to the unified domain interval* $[-1, 0]$. *Column (b) shows registered coordinate functions warped separately in each coordinate. Column (c) represents warping functions corresponds to each coordinate.*

ues of MISR. In panel (a) the correspondences between f_1 and f_2 are shown using some dotted lines connecting the two curves. Panel (b) shows the geodesic path, an optimal deformation from f_1 to f_2, in the shape space in equal time steps.

(a) Correspondences (b) Geodesic Path

Registration with/without Feature

(c) Coordinates Only (d) With MISR Feature

X-coordinate

Y-coordinate
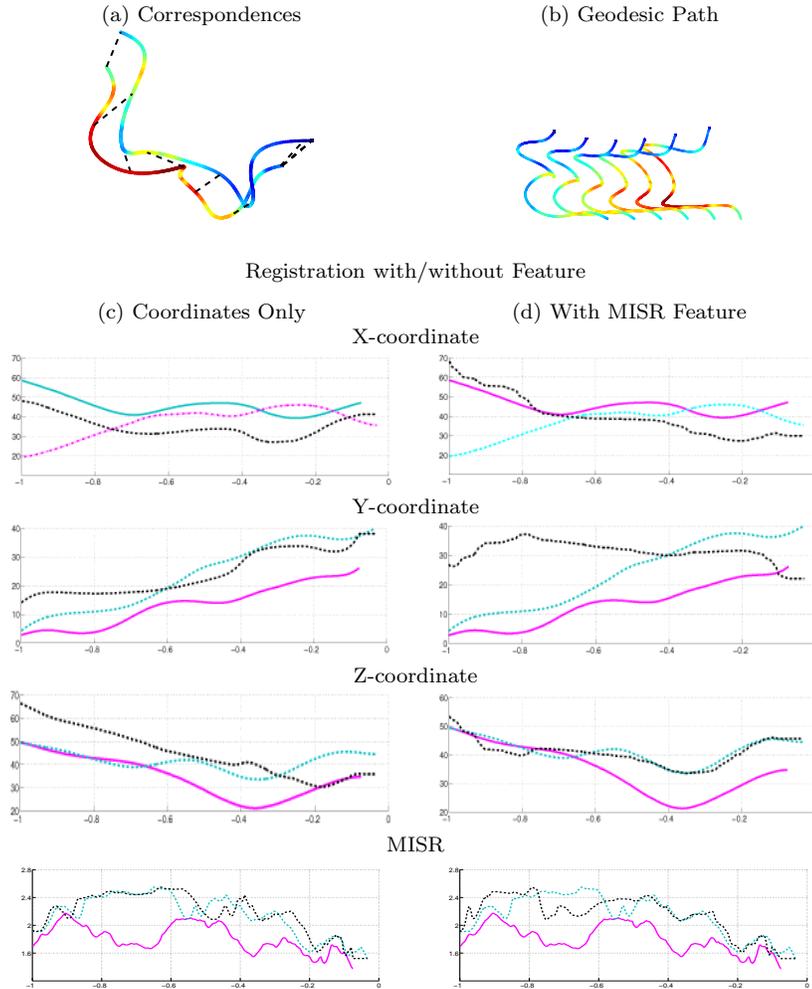

Z-coordinate

MISR

FIG 3. *Pairwise registration results. (a) Correspondences. (b) Geodesic path. (c) Pairwise registration using the coordinate functions only, and (d) Pairwise registration using the coordinate functions and the MISR. Color scheme: f_1 – magenta; f_2 – green; $\tilde{f}_2 = O^*(f_2 \circ h^*)$ – black. In the left case, the coordinate registration improves significantly over the original while in the right the MISR registration improves also.*

Panels (c) and (d) present the x, y, and z coordinate functions in that order using registration based only on these coordinates (left), and these coordinates and the MISR function as the fourth coordinate (right). Note that the end points of the coordinate functions should remain the same after re-parameterization. However, they can change due to 3D rotation. The last row shows the MISR function in the two cases. It can be seen that the quality of MISR registration is better in the right case where it is included in the estimation of h^*.
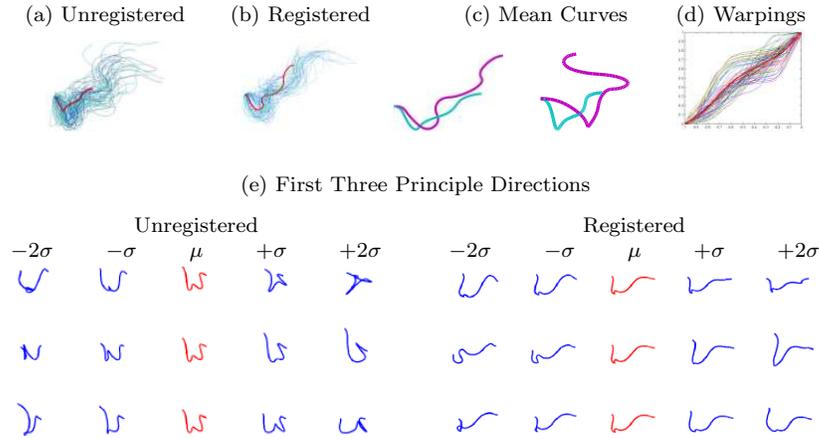
FIG 4. *Statistical modeling of artery curves. (a) Mean curve before registration. (b) Mean curve after registration. (c) Mean curves before and after registration from two views. Unregistered: green; Registered: magenta. (d) Re-parameterization functions with corresponding mean in red. (e) First three principal directions of variation. Left: unregistered. Right: registered.*

Mean shape and groupwise registration Now that we have a proper distance in the shape space, we can compute shape summaries and model shape variability in the given artery data. We use the notion of the Karcher mean to compute a mean and then (in a pairwise manner) align the given curves to this mean using Eqn. 2.1. Then, we perform fPCA of the given curves by first mapping them into the tangent space at the Karcher mean shape, and then mapping the results back to the shape space (see [2]).

In Fig. 4, panels (a) and (b) show the mean curve (red curve) on top of the given artery curves, both without and with registration (optimization in Eqn. 2.1), respectively. In panel (c), two views of the mean curves are displayed for better evaluation. The mean of registered curves is shown in magenta and that of the unregistered ones is in green. It can be seen that the magenta curve captures more local geometric properties due to registration. In panel (d), the warping functions used to register each curve to the mean curve are plotted with the mean of re-parameterization functions (in red). Panel (e) shows the first three principal directions of variation in each row for unregistered and registered curves; red curves represent the mean curves and the blue ones are one and two standard deviation away from the mean along each principal direction.

Furthermore, Fig. 5 shows the registered curves with corresponding mean curves visualized separately as three coordinate functions. Other features along the curve are then warped to display for comparison in Fig. 6. Those include the first derivative of the curves (X1_FKS) along the three coordinates separately, the radius (MISR) and the curvature (Curvature_FKS).

Cluster Analysis Using the shape metric d_s introduced earlier in Eqn. 2.1, we can perform cluster analysis on the given 65 observed curves based on their
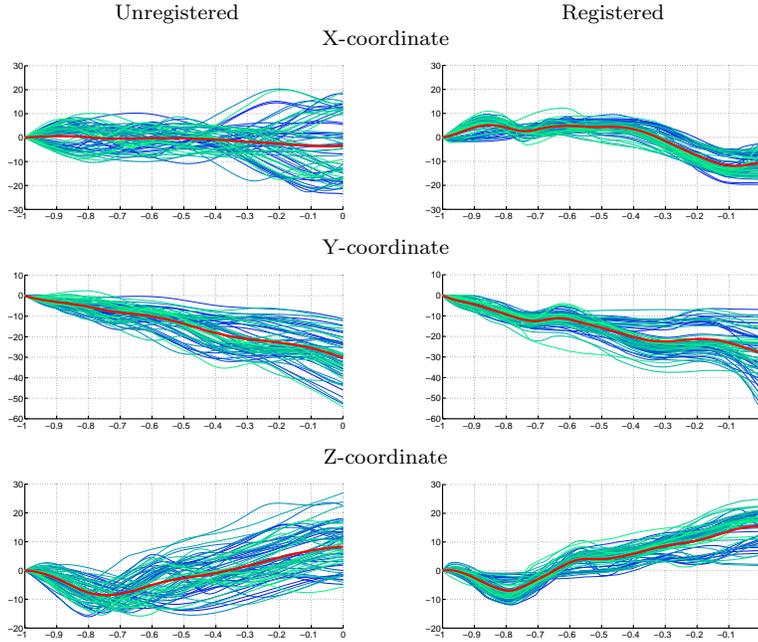
FIG 5. *Mean curves before and after registration. Each coordinate function is shown separately in each row. Mean curves are plotted in red. Left: unregistered. Right: registered.*

shapes. The pairwise distances $d_s([f_i], [f_j])$ are computed for any pair (i, j). Based on the pairwise distances, the agglomerative hierarchical cluster tree using Ward's criterion is applied. The clustering result with three groups is chosen to correspond to the Ω, Γ and S-shaped ICAs. The member curves are then aligned to their respective mean curves within each group. The aligned curves are displayed using the coordinate functions in Fig. 7. The group mean curves capture more detailed shape structure than the overall mean and facilitate more
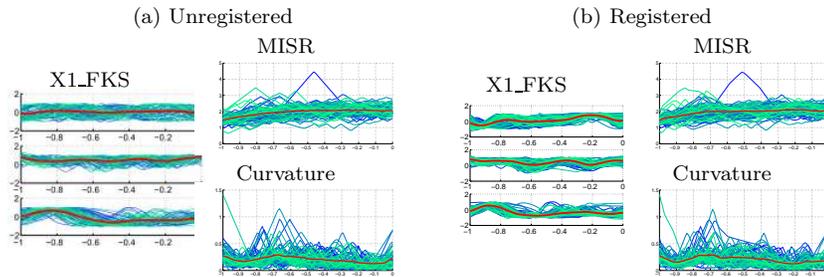

FIG 6. *Computing mean curves. Features are warped with the optimal deformations to register original curves for comparison. X1_FKS functions are shown on each coordinate separately. Left: unregistered. Right: registered.*
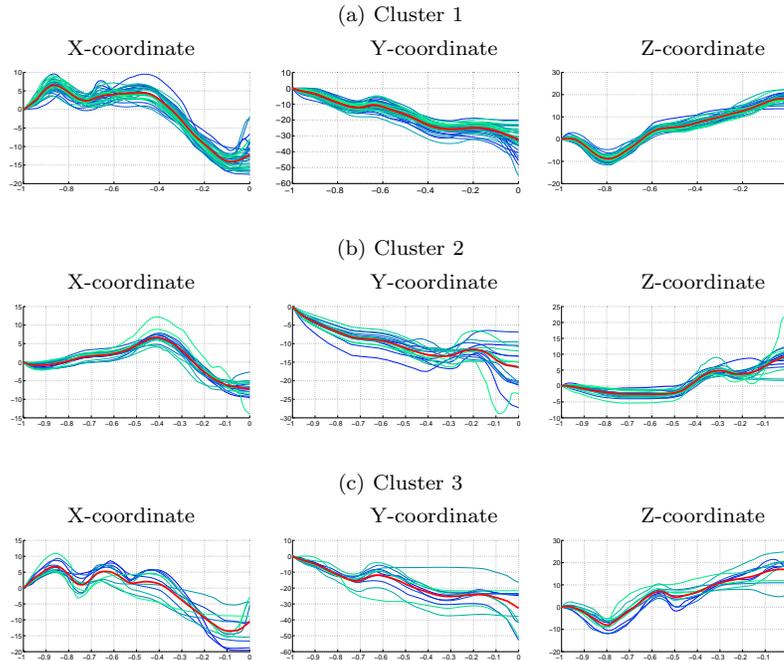
(a) Cluster 1

X-coordinate Y-coordinate Z-coordinate

(b) Cluster 2

X-coordinate Y-coordinate Z-coordinate

(c) Cluster 3

X-coordinate Y-coordinate Z-coordinate

FIG 7. *Mean curves within clusters viewed from three coordinates.*

compact models. It can be observed that cluster 1 and cluster 3 mainly differ in their x-coordinate functions and cluster 2 differs from the other clusters in all coordinates.

4. Conclusion

We have applied elastic shape analysis of 3D and higher dimensional curves to the Anuerisk65 carotid artery data. This framework enables us to align all three coordinate functions simultaneously with one common warping function and perform shape analysis of these curves. This framework also leads to a proper metric on the quotient space that is used in pairwise comparisons, mean computations, fPCA, and clustering of curves according to their shapes.

References

[1] DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis*, John Wiley & Sons. MR1646114

[2] SRIVASTAVA, A., KLASSEN, E., JOSHI, S. H. and JERMYN, I. H. (2011). *Shape analysis of elastic curves in Euclidean spaces*, Pattern Analysis and Machine Intelligence, IEEE Transactions on, **33(7)** 1415–1428.

[3] TUCKER, J. D., WU, W. and SRIVASTAVA, A. (2013). *Generative models for functional data using phase and amplitude separation*, Computational Statistics & Data Analysis, **61** 50–66. MR3063000

[4] WU, W. and SRIVASTAVA, A. (2014). *Analysis of spike train data: Alignment and comparisons using extended Fisher-Rao metric*, Electronic Journal of Statistics, **8** 1776–1785, Special Section on Statistics of Time Warpings and Phase Variations.

[5] MICHOR, W., MUMFORD, D., SHAH, J. and YOUNES, L. (2008). *A metric on shape space with explicit geodesics*, Matematica E Applicazioni, **19(1)** 25–57. MR2383560

[6] SANGALLI, L. M., SECCHI, P., VANTINI, S. and VENEZIANI, A. (2009). *A case study in exploratory functional data analysis: Geometrical features of the internal carotid artery*, Journal of the American Statistical Association, **104(485)** 37–48. MR2663032

[7] SANGALLI, L. M., SECCHI, P. and VANTINI, S. (2014). *AneuRisk65: A dataset of three-dimensional cerebral vascular geometries*, Electronic Journal of Statistics, **8** 1879–1890, Special Section on Statistics of Time Warpings and Phase Variations.

[8] SRIVASTAVA, A., WU, W., KURTEK, S., KLASSEN E. and MARRON, J. S. (2011). *Registration of functional data using Fisher-Rao metric*, arXiv:1103.3817.