

Analysis of AneuRisk65 data: Classification and curve registration*

Ana-Maria Staicu

Department of Statistics

North Carolina State University Raleigh NC 27695 USA

e-mail: ana-maria.staicu@ncsu.edu

and

Xiaosun Lu

Department of Statistics and Operations Research

University of North Carolina at Chapel Hill NC 27599 USA

e-mail: xiaosun@live.unc.edu

Abstract: This paper concerns the relationship between the geometry of the Inner Carotid Artery, as described by its centerline curvature and its radius, and the location of the aneurysm for the AneuRisk65 data. Fisher Rao curve registration is used to align the curvature of the artery, and this alignment is then used to register both the curvature and the radius profiles. Based on this alignment, interesting results are found regarding the discrepancy between the arteries of patients with aneurysms at or after the terminal bifurcation (upper group) and the arteries of subjects with aneurysms before bifurcation, or without aneurysms (lower-no group).

Keywords and phrases: AneuRisk65 data, principal component analysis, curve registration, classification, functional data analysis.

Received August 2013.

This is a discussion paper for analyzing the AneuRisk65 data found at <http://mox.polimi.it/it/progetti/aneurisk/>. As Sangalli et al. (2014) details, these data include the image reconstruction of the Inner Carotid Artery (ICA), described in terms of the vessel's centerline and its radius profile for many subjects. We focus on the curvature profiles, defined pointwise as a function of the first and the second derivatives of the vessel's centerline, and on the radius profiles, computed pointwise as the maximal inscribed sphere radius (MISR) in the vessel; see Sangalli et al. (2009). The goal is to explore the relationship between the geometry of the ICA – as depicted by the centerline curvature and its radius profiles – and the location of the aneurysm. Specifically, we concentrate on the discrepancy between arteries of patients with aneurysms at or after the terminal bifurcation (upper group) and the arteries of patients with aneurysms before bifurcation of the ICA or without aneurysms (lower-no group).

*Main article [10.1214/14-EJS938](https://doi.org/10.1214/14-EJS938).

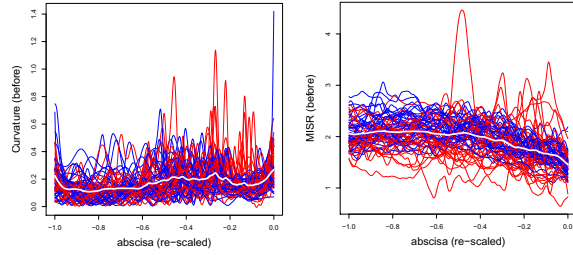


FIG 1. Re-scaled functions of curvature (left) and MISR (right). Color represents the location of the aneurysm (blue for upper group, red for lower-no group). The white curves indicate mean functions.

1. Data objects

Object Oriented Data Analysis, introduced by Wang and Marron (2007), provides useful terminology for the study of the geometry of the artery, where *data objects* are understood as the atoms of the statistical analysis.

An intuitive choice of data objects is the pair of curvature and MISR profiles, estimated via multidimensional free-knot splines (see Sangalli et al. (2009) for more information). These functions are defined on different domains, since the ICA centerlines are of different lengths, and thus we mapped the domain of each function linearly to $[-1, 0]$, as a preliminary step of the analysis. Then a B-spline interpolation is used to define these functions on a common fine grid of points (with 100 basis functions and the smoothing parameter chosen via restricted maximum likelihood, REML). The resulting functions of curvature and MISR are shown in Figure 1. Further analyses are based on these rescaled functions.

The curvature and the MISR functions contain both the phase variation and the amplitude variation. For the curvature functions, these two types of variation are separated using the Fisher Rao curve registration method proposed by Srivastava et al. (2011); they are captured by the resulting domain warping functions and the aligned functions, respectively, shown in the first two panels of Figure 2. This domain warping approach provides two types of data objects:

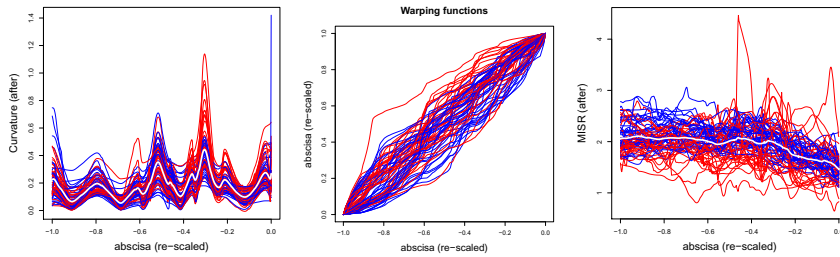


FIG 2. Left: Aligned curvature functions from Fisher Rao curve registration. Middle: Domain warping functions for aligning the curvature functions in Figure 1 (left). Right: Aligned MISR functions obtained by applying these domain warping (middle) to the MISR functions in Figure 1 (right). Color code is the same as that in Figure 1.

warping functions for studying the phase variation, and aligned functions for studying the amplitude variation; see Lu (2013) and Lu and Marron (2013) for more discussion. Let $h_{C,i}$ be the warping functions for registration of the curvature profiles and $\tilde{f}_{C,i}$ be the aligned curvatures. The MISR functions are then aligned using the same domain warping functions, $h_{C,i}$, for registering the curvature functions; denote by $\tilde{f}_{MISR,i}$ the aligned MISR curves, shown in Figure 2 (right). The following data objects are analyzed later in this paper:

- (1) Aligned curvature functions, $\tilde{f}_{C,i}$
- (2) Aligned MISR functions, $\tilde{f}_{MISR,i}$
- (3) Domain warping functions for aligning curvature/MISR functions, $h_{C,i}$.

2. Analysis of arteries geometry using amplitude variation

First we wish to gain insight into how the arteries vary, while accounting for the location of the aneurysm; we consider the amplitude variation of the arteries solely. The aligned MISR and curvature profiles are depicted in the leftmost and rightmost plots of Figure 2 respectively, where the color reflects the group membership: blue for the lower-no group and red for the upper group. For each sample in part, the variation of the curves is assessed by assuming a model based on group mean adjusted functional principal component analysis (FPCA); see Jiang and Wang (2010). Specifically, if c_i is a group indicator variable, with $c_i = 'L'$ for the lower-no group and $c_i = 'U'$ for the upper group, and $f_i(\cdot)$ is a generic random curve observed on $I = [-1, 0]$ with group membership given by c_i , then we assume that $\tilde{f}_i(t) = \mu_{c_i}(t) + \epsilon_i(t)$. Here $\mu_{c_i}(\cdot)$ is a smooth group mean function, and $\epsilon_i(\cdot)$ is a residual process, possibly perturbed by white noise, which is independent of the group membership. To address the measurement error aspect both the aligned curvature and MISR profiles are smoothed by using B-spline basis; the model components are then estimated by using the sample-based estimators.

Figure 3 displays the estimated mean profiles of the aligned curvatures and MISR of the ICA for subjects in the lower-no group and upper group. On aver-

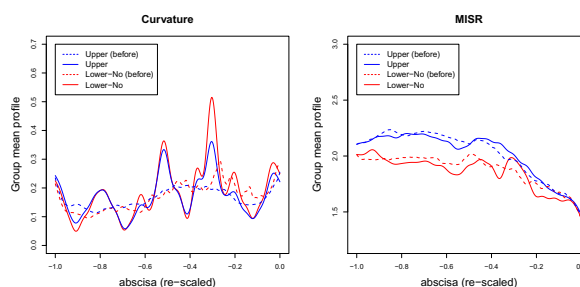


FIG 3. Group mean profiles of the registered curvatures (left panel) and the registered MISR (right panel). Overlaid are the corresponding group means before the registration procedure, indicated by the dashed lines. Color code is the same as that in Figure 1.

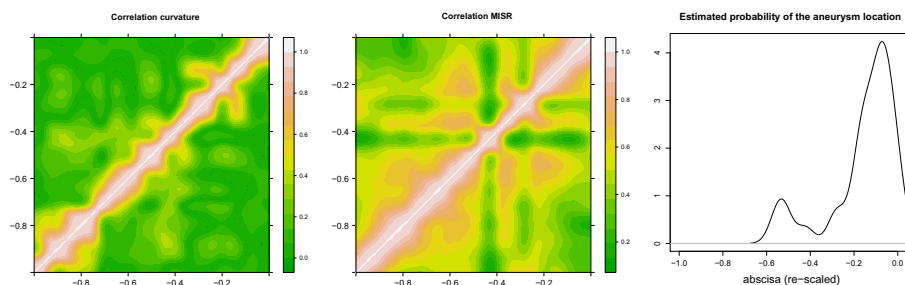


FIG 4. Correlation of the pooled covariance of the registered curvatures (leftmost panel) and the registered MISR (middle). Right: density estimate of the aneurysm location along ICA.

age, the curvature of the ICA shows two main peaks (at around -0.5 and -0.3), which have similar magnitudes in the upper group, but increasing magnitudes for the lower-no group. The peaks correspond to the two siphon centers of the ICA and indicate that the ICA for the upper group is less curved, especially in the proximity of the first siphon center. Moreover the MISR graph shows that the ICA tapers off as it gets closer to the bifurcation for both groups and that the width of the ICA in the upper group is on average larger than its counterpart in the lower-no group.

Next we turn to the random deviation from the group mean trajectory, for both the curvatures and MISR, as depicted by the smooth signal captured by the generic term $\epsilon_i(t)$. Instead of presenting the eigen-analysis of this deviation, we focus on their correlation calculated from the pooled covariance, which provides interesting information. Figure 4 shows the estimated correlation in absolute value of the curvatures (left plot) and MISR (center) profiles. For the curvature profiles, the correlation decays to zero relatively fast. On the other hand the correlation of the MISR profiles is much stronger and furthermore seems to indicate two separate clusters, where the separation is closely related to the location of aneurysm. To see this, the right plot of Figure 4 shows the Gaussian kernel estimate of the probability density of the location of the aneurysm, indicating that most aneurysms are clustered into two groups.

Interestingly these results corroborate the previous findings published in Sangalli et al. (2009), which are obtained using the last 3cm of the ICA; in contrast our analysis is based on the full available data. The two approaches rely on different registration methods, used in the first step of the analysis.

3. Classification of aneurysms using joint analysis of amplitude and phase variation

Now, we consider the arteries, or in fact their shape, as predictors, and we wish to classify them according to the location of the aneurysms. In this attempt we focus on the variability of the arteries, which is assessed through the amplitude and phase variation jointly for the curvature curves and only the amplitude variation for the MISR curves. To reduce the large dimensionality of the predictors,

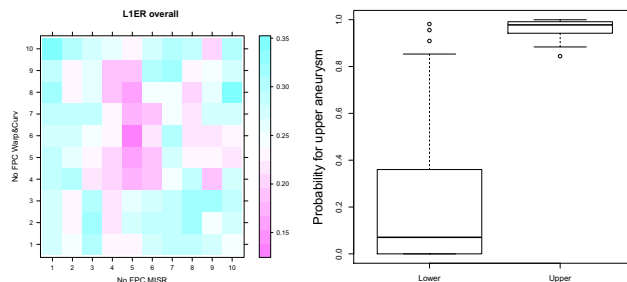


FIG 5. Left: Overall $L1ER$ for various truncation values for the number of scores for MISR, K_{MISR} and curvatures, K_C . Right: Predicted probabilities of a subject with upper aneurysm using the optimal truncation values $K_{MISR} = 5$, $K_C = 6$.

FPCA is employed (Ramsay and Silverman (2005), Crainiceanu et al. (2009), etc). We briefly describe the procedure for the MISR curves.

Let $\tilde{f}_{MISR,i}$ be the aligned MISR curves, and denote by $\hat{\mu}_{MISR}(\cdot)$ and by $\hat{\Sigma}_{MISR}(\cdot, \cdot)$ the estimated mean and covariance function respectively. Furthermore, denote by $\hat{\phi}_{MISR,k}(\cdot)$ and $\hat{\lambda}_{MISR,k}$ the k th eigenfunctions and the k th eigenvalue respectively of $\hat{\Sigma}_{MISR}(\cdot, \cdot)$. The MISR profile can be summarized by the set of scores $(\hat{\xi}_{MISR,i1}, \hat{\xi}_{MISR,i2}, \dots)$ where the scalar generic score $\hat{\xi}_{MISR,ik}$ gives the variation about the k th eigenfunction of the de-trended data and is estimated as

$$\hat{\xi}_{MISR,ik} = \langle \tilde{f}_{MISR,i} - \hat{\mu}_{MISR}, \hat{\phi}_{MISR,k} \rangle; \quad (1)$$

here the inner product is defined as $\langle f_1, f_2 \rangle := \int f_1 f_2$.

The bivariate FPCA of the curvature curves uses a modified inner product to account for different variability between the registered curvature functions, and those of the warping functions (Ramsay and Silverman (2005)). Specifically, if $(h_{C,i}, \tilde{f}_{C,i})^T$ is the two-dimensional vector comprising the domain warping functions and the aligned curvatures, then the inner product for two pairs $(h_{C,i}, \tilde{f}_{C,i})^T$ and $(h_{C,j}, \tilde{f}_{C,j})^T$ is defined as $\int h_{C,i} h_{C,j} + \kappa \int \tilde{f}_{C,i} \tilde{f}_{C,j}$. Here κ is some appropriately chosen constant; we selected κ as the ratio between the overall variability of the warping functions and that of the aligned curvatures. Let $(\hat{\zeta}_{C,i1}, \hat{\zeta}_{C,i2}, \dots)$ be the set of estimated scores summarizing the phase and amplitude variation of the curvatures. The R package **refund** (Crainiceanu et al. (2011)) is used to fit both the bivariate FPCA and the univariate FPCA. Like Sangalli et al. (2009), we too find different distributions of the scores corresponding to the subjects in the two groups (lower-no/upper).

Thus the two sets of scores, $\xi_{MISR,ik}$ and $\zeta_{C,i\ell}$ summarize the vessel geometry of the arteries; we use them to study the relationship with the location of the aneurysm. Regularized discriminant analysis (Friedman (1989)) is used for this purpose. The truncated number of scores for both the radius, K_{MISR} and curvature profiles, K_C is determined according to the leave-one-out error rate (L1ER) criterion. In particular, we find that the choice $K_{MISR} = 5$ and

$K_C = 6$ optimizes this criterion, with the optimal value equal to 13.8%. As Figure 5 illustrates, these truncation estimates yield an excellent classification performance for the subjects with upper aneurysms and 25% misclassification rate for subjects in the lower-no group. These findings improve the prediction results reported in Sangalli et al. (2009).

Overall our analysis, using a two-step procedure with a different registration method (at the first step) and analysis based on the entire observed data (at the second step) provides results that replicate the findings of Sangalli et al. (2009).

Acknowledgements

Staicu was partly supported by NSF grant number DMS 1007466. The authors would like to acknowledge the support of the Mathematical Biosciences Institute for hosting the workshop which led to this work.

References

- CRAINICEANU, C., REISS, P. (Coordinating authors), GOLDSMITH, J., GREVEN, S., HUANG, L., and SCHEIPL, F. (Contributors) (2011). *refund: Regression with functional data*. R package version 0.1-5.
- CRAINICEANU, C., STAIKU, A.-M., and DI, C. (2009). Generalized multilevel functional regression. *J Am Stat Assoc*, 104(488):1550–1561. [MR2750578](#)
- FRIEDMAN, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, 84(405):165–175. [MR0999675](#)
- JIANG, C. R. and WANG, J. L. (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, 38(2):1194–1226. [MR2604710](#)
- LU, X. (2013). Object oriented data analysis of cell images and analysis of elastic functions. *Ph.D. Dissertation*, University of North Carolina, Chapel Hill, NC, USA. [MR3153472](#)
- LU, X. and MARRON, J. S. (2013). Principal nested spheres for time warped functional data analysis. [arXiv:1304.6789](#).
- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*. Springer. [MR2168993](#)
- SANGALLI, L. M., SECCHI, P., and VANTINI, S. (2014). Analysis of Aneurisk65 data: k -mean alignment. *Electronic Journal of Statistics*, 8:1891–1904, Special Section on Statistics of Time Warpings and Phase Variations.
- 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](#)
- SRIVASTAVA, A., WU, W., KURTEK, S., KLASSEN, E., and MARRON, J. S. (2011). Statistical analysis and modeling of elastic functions. [arXiv:1103.3817](#).
- WANG, H. and MARRON, J. S. (2007). Object oriented data analysis: Sets of trees. *The Annals of Statistics*, 35(5):1849–1873. [MR2363955](#)