

Analysis of AneuRisk65 data: Internal carotid artery shape analysis*

Wen Cheng

*Department of Statistics, University of South Carolina, Le Conte College
Columbia, SC 29208, USA. e-mail: chengwen1985@gmail.com*

Ian L. Dryden†

*School of Mathematical Sciences, University of Nottingham, University Park
Nottingham, NG7 2RD, UK. e-mail: ian.dryden@nottingham.ac.uk*

David B. Hitchcock

*Department of Statistics, University of South Carolina, Le Conte College
Columbia, SC 29208, USA. e-mail: hitchcock@stat.sc.edu*

and

Huiling Le

*School of Mathematical Sciences, University of Nottingham, University Park
Nottingham, NG7 2RD, UK. e-mail: huiling.le@nottingham.ac.uk*

Abstract: The AneuRisk65 data are analysed using methodology from statistical shape analysis. The internal carotid arteries are aligned using translation and rotation in three dimensions, together with shifts of the abscissa coordinate. Spline interpolation and weighted Procrustes methods are used to estimate the mean size-and-shapes in each of the six groups. Differences in torsion and curvature of the group means are highlighted, and permutation and bootstrap tests confirm there is weak evidence for differences in shape between the upper aneurysm group compared to the others. Finally shape variability, analysis of mean radii and classification are explored.

Keywords and phrases: Alignment, bootstrap, curvature, permutation, Procrustes, registration, shape, size-and-shape, torsion.

Received August 2013.

1. Introduction

The AneuRisk65 dataset (Sangalli et al., 2014) consists of various measurements derived from a dataset of 65 arteries, with the centerlines displayed in Figure 1. The six groups are: Upper Aneurysm Left (UL – Black), Lower Aneurysm Left (LL – Red), No Aneurysm Left (Green), Upper Aneurysm Right (UR – Blue),

*Main article [10.1214/14-EJS938E](https://doi.org/10.1214/14-EJS938E).

†To whom correspondence should be addressed.

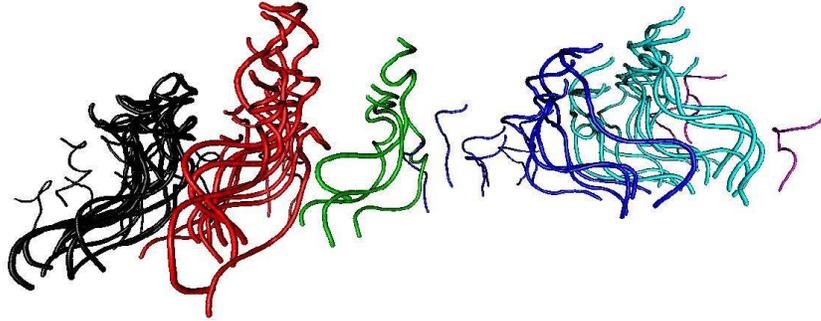


FIG 1. The six groups of arteries. The groups are UL (Black), LL (Red), NL (Green), UR (Blue), LR (Cyan), NR (Pink).

Lower Aneurysm Right (LR – Cyan), No Aneurysm Right (NR – Pink). These data have not been reflected at this stage, but rather these are the raw data. It is of interest to examine the shape differences in the groups of arteries. From Figure 1 we clearly see that the artery centerlines are quite variable in length, and they are quite variable in shape. The Lower groups appear longer than the Upper and No groups. Also, the NL and NR sample sizes are particularly small, and we shall combine the Lower and No groups in some of our analysis, in comparison with the Upper groups (which constitute the more dangerous aneurysms).

2. Procrustes registration

Given that the artery centerlines are different lengths we need to consider carefully how to register the curves. The size-and-shapes of the arteries are invariant under rotation and translation (Kendall, 1989; Le, 1995), and we match the 3D curves by translation, rotation and a shift in the abscissa. The matching procedure is carried out using Procrustes analysis with the `shapes` package in R (Dryden, 2013), although some adaptation is required because the arteries are of different lengths. The method of ordinary Procrustes analysis involves registering two sets of points by translation and rotation (and possibly scale) in order to minimize the sum of square Euclidean distances between the point sets (Dryden and Mardia, 1998, p.84). The extension to matching more than two objects is carried out using generalized Procrustes analysis (Gower, 1975; Goodall, 1991; Dryden and Mardia, 1998), where a mean object is estimated and simultaneously each individual is translated, rotated (and possibly rescaled) to match as closely as possible to the mean. Investigating the mean shape difference between two groups is carried out using multivariate analysis in an appropriate tangent space (Dryden and Mardia, 1998), e.g. using hypothesis tests.

We use 3D x-y-z co-ordinates of the centerline interpolated from the abscissa values $0, 1, \dots, n_t$ using three cubic splines. Initially we use $n_t = 111$, which is the length of the longest artery. For shorter arteries we carry out extrapolation

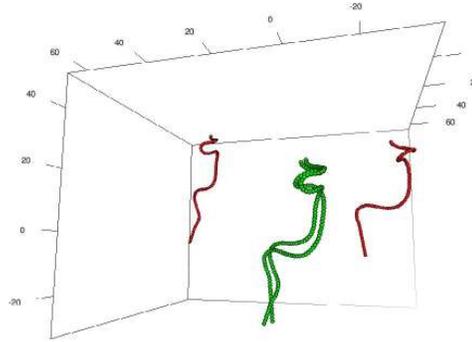


FIG 2. Two extrapolated arteries (red) matched by ordinary Procrustes analysis (green).

using cubic splines, which is then nearly zero weighted in the subsequent analysis. The registration of the arteries is then carried out using weighted Procrustes analysis on the $k = n_t + 1$ points.

Consider two configurations X and μ of $k \geq 3$ points in \mathbb{R}^3 that have been centered, i.e. $1_k^T X = 0 = 1_k^T \mu$. The standard situation of unweighted ordinary Procrustes matching is given by rotating and translating X to μ to minimize the squared Euclidean distance, i.e. minimize

$$D^2(X, \mu) = \|\mu - X\Gamma - 1_p\gamma^T\|^2,$$

over rotations $\Gamma \in SO(3)$ and translations $\gamma \in \mathbb{R}^3$ (Dryden and Mardia, 1998). The solution is $\hat{\gamma} = 0, \hat{\Gamma} = UW^T$ where

$$\mu^T X = W\Lambda U^T, \quad U, W \in SO(3)$$

with Λ a diagonal 3×3 matrix, and the solution is unique if X is non-degenerate with respect to μ (see Kent and Mardia, 2001). Note that $SO(3)$ denotes the set of 3×3 rotation matrices. The size-and-shape space tangent co-ordinates are given by

$$V = X\hat{\Gamma} - \mu,$$

e.g. see Dryden et al. (2007). In Figure 2 we see two arteries (with extrapolated values) matched by ordinary Procrustes analysis.

However, the extrapolated values could lead to problems with ordinary Procrustes matching, in that the extrapolated values may have an undue influence on the registration. Instead the extrapolated part can be given negligible weight in a weighted Procrustes analysis, where we replace $\|X\|$ with $\|X\|_\Sigma = \sqrt{\text{tr}(X^T \Sigma^{-1} X)}$. For example, if Σ is a diagonal matrix then the diagonal elements of Σ^{-1} are the weights.

Given a sample of arteries we match them together and estimate a group mean size-and-shape as follows.

1. Carry out non-weighted generalized Procrustes analysis (with translation and rotation but not scale) on the interpolated and extrapolated curves, and obtain the estimated mean size-and-shape $\hat{\mu}_{template}$.



FIG 3. *Weighted Procrustes analysis: Black/Blue UL/UR, Red/Cyan LL/LR, Green/Pink NL/NR (Reflected data).*

2. Choose the estimated shift in abscissa to minimize the Procrustes distance to $\hat{\mu}_{template}$.
3. Match each individual using weighted ordinary Procrustes to $\hat{\mu}_{template}$, with weight close to zero if the point is extrapolated, and weight 1 if interpolated.
4. Compute the weighted mean $\hat{\mu}$ by averaging the weighted Procrustes registered data over only the specimens which have interpolated but not extrapolated values for a given abscissa.

In order to compare the left and right internal carotid arteries it is useful to consider reflections (e.g. reflect the left arteries), which makes the groups more similar in shape and easier to compare. We carry out the above weighted Procrustes procedure on the reflected data and the results are displayed in Figure 3.

In Figure 4 we see the group means plotted for each of the six groups. The display is such that a small sphere is drawn at the interpolated mean centerline at integer abscissa values, with radius proportional to the number of specimens that contribute to that average point. Clearly at the lower points there are fewer data points in the mean calculation, and hence these parts of the curves will be estimated less reliably.

3. Shape differences

In order to describe the differences in the shapes of 3D curves it is helpful to consider the curvature and torsion of the curve. Given a curve $\{x(t), y(t), z(t)\}$ the torsion function is given by $\tau(t) = \tau_1(t)/\tau_2(t)$, where

$$\begin{aligned}\tau_1 &= x'''(y'z'' - y''z') + y'''(x''z' - x'z'') + z'''(x'y'' - x''y') \\ \tau_2 &= (y'z'' - y''z')^2 + (x''z' - x'z'')^2 + (x'y'' - x''y')^2\end{aligned}$$

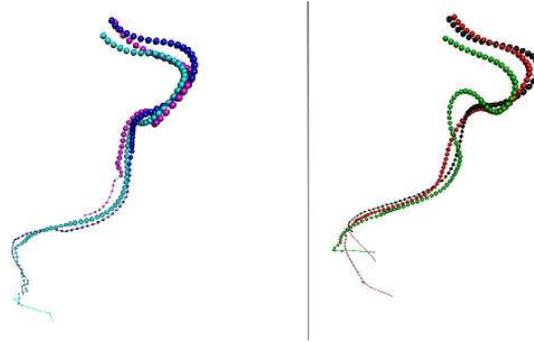


FIG 4. Group means: Black/Blue UL/UR, Red/Cyan LL/LR, Green/Pink NL/NR (Reflected data).

and the curvature function is $\kappa = \sqrt{\tau_2} / \|(x', y', z')\|^{3/2}$. In Figure 5 we see the estimated curvature and torsion for the group means. Note that the amount of smoothing required can be large given that third derivatives need to be estimated, which in turn may bias the estimation.

There are some interesting differences between the groups, and it is also helpful to consider another view of the means in Figure 6 to aid explanation. Some of the main feature in the size-and-shape estimates are:

1. There are three bumps of high curvature, and the highest curvatures are in the NL and NR groups.
2. Groups means UL and UR stay closer to a plane (lower torsion)
3. Group means LL/LR and NL/NR are more twisted, leaving a plane (higher torsion).
4. Group means LL and LR leave a plane on different sides between the bumps 1 and 2 of high curvature (i.e. different signs for torsion), which can be seen in Figure 6.

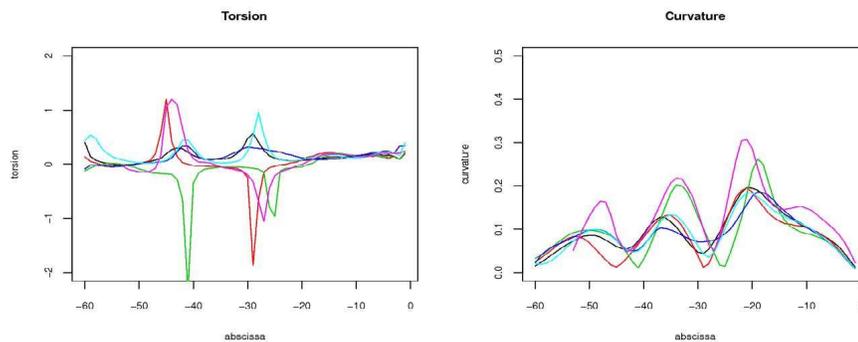


FIG 5. Estimated torsion and curvature of the group means: Black/Blue UL/UR, Red/Cyan LL/LR, Green/Pink NL/NR (Reflected data). The x-axis is the abscissa value.



FIG 6. Mean size-and-shapes: Black/Blue UL/UR, Red/Cyan LL/LR, Green/Pink NL/NR (Reflected data).

5. Groups means NL and NR leave a plane on different sides between the bumps 2 and 3 of high curvature. (i.e. different signs for torsion), which can be seen in Figure 6.

These observations may be helpful in understanding the differences between the artery shapes for the more dangerous Upper aneurysm and less dangerous Lower/No groups. In order to examine whether these observations are statistically significant we consider a hypothesis test for mean shape difference. Given that there are particularly low sample sizes in the No groups, and that we expect them to have some more similarity with the Lower group, we pool the Lower/No groups. We consider just up to length 27 on each artery for this test as all curves are at least length 27. We carry out a permutation and bootstrap test for mean shape difference between the two groups (U vs L/N), based on non-weighted Procrustes matching and include scale invariance. The permutation test involves pooling the data, randomly assigning individuals to group 1 or group 2, and then evaluating the Goodall test statistic (Dryden and Mardia, 1998, p. 162) for each permutation. The p-value of the test is the proportion of random permutations with test statistic values more extreme than that calculated from the data. For the bootstrap test the sample means are first parallel transported to be coincident, and then samples are taken with replacement from each of these translated groups. The p-value is the proportion of bootstrap test statistic values more extreme than the observed statistic. See Amaral et al. (2007) for a more in-depth discussion of permutation and bootstrap tests for shape data. The permutation test gave p-value = 0.079 (1000 permutations) and a bootstrap test gave p-value = 0.068 (1000 resamples) using the `testmeanshapes` command in the `shapes` package in R (Dryden, 2013). Hence, there is weak evidence of a difference in shape between the Upper and Lower/No groups.

We also investigated shape variability, this time on the non-reflected data, and the PC loadings are given graphically in Figure 7. In this analysis we consider up to length 40. The clearest feature here is that PC2 displays the right-

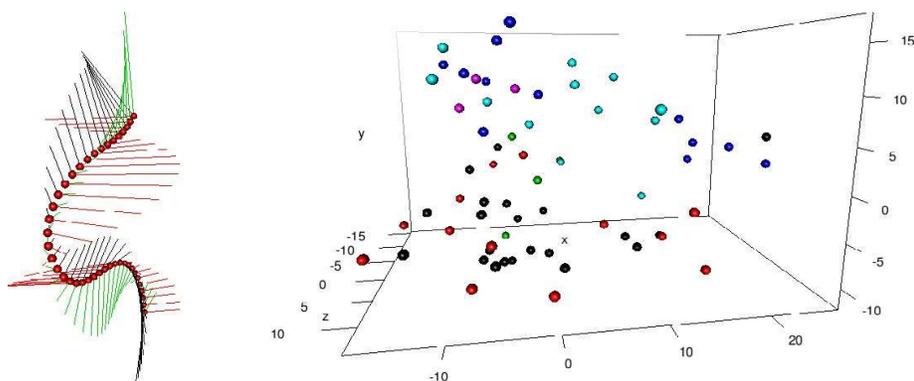


FIG 7. Left: PCA loadings for the non-reflected data: PCA Black PC1, Red PC2, Green PC3. The overall mean is displayed with vectors drawn to the mean +3 s.d. along each PC. Right: PC scores for the individual curves: Black/Blue UL/UR, Red/Cyan LL/LR, Green/Pink NL/NR.

handed/left-handed asymmetry. PC1 and PC3 reflect the size and shape of the first bend in the artery.

The data also contain further measurements as shown in Figure 8, including the maximum inscribed sphere radius along the curve. If we consider an unbalanced one-way ANOVA then for the mean radii along curve the Upper group significantly larger than Lower ($p = 0.027$) and No ($p = 0.074$) groups. The s.d. of radii along curve is not significantly different between the groups. However, for the length of curve the Lower group is longer on average than Upper and No groups ($p < 0.01$). The difference in the lengths, though, is not deemed of

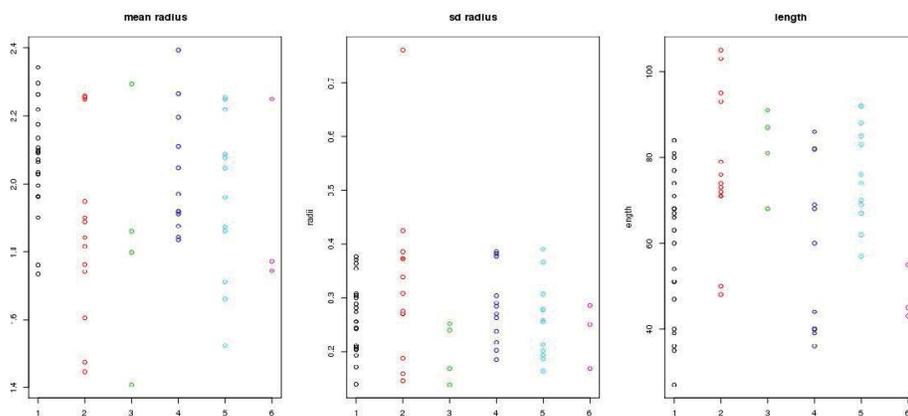


FIG 8. Plots of the mean radius, s.d. of the radius, and length for each of the arteries by group (1: UL, 2: LL, 3: NL, 4: UR, 5: LR, 6: NR).

practical importance, as the position of the image capture equipment is often determined by the position of the aneurysm. We also explored classification of the artery shapes using weighted Procrustes distance to the Upper, Lower and No estimated group means. It is important to make sure that factors that depend on the length of centerline are not included, as this will give a method an unfair advantage given the significant difference in lengths between the groups. Using leave-one-out classification we obtained 67.7% correct classification in comparing Upper versus a combined Lower/No group using only the centerline information.

Finally, it is important in any alignment problem to specify what invariances are required. We have chosen to keep the original geometry of the artery intact and only align by shifting the abscissa, as well as translation and rotation. One could have chosen to match up the bumps and features using non-linear warping, e.g. using the method of Srivastava et al. (2011). Such a procedure should give complementary analysis through analysing the warps jointly with the shapes. However, in this particular application it does seem valuable to be able to explain the shape differences between the groups using torsion and curvature without non-linear warping of the abscissa.

Acknowledgment

The authors are grateful to the MBI for support. The second author acknowledges the support of a Royal Society Wolfson Research Merit Award.

References

- AMARAL, G. J. A., DRYDEN, I. L., and WOOD, A. T. A. (2007). Pivotal bootstrap methods for k -sample problems in directional statistics and shape analysis. *J. Amer. Statist. Assoc.*, 102(478):695–707. [MR2370861](#)
- DRYDEN, I. L. (2013). *shapes: Statistical shape analysis*. R package version 1.1-9.
- DRYDEN, I. L., HIRST, J. D., and MELVILLE, J. L. (2007). Statistical analysis of unlabeled point sets: Comparing molecules in chemoinformatics. *Biometrics*, 63(1):237–251, 315. [MR2345594](#)
- DRYDEN, I. L. and MARDIA, K. V. (1998). *Statistical Shape Analysis*. Wiley, Chichester. [MR1646114](#)
- GOODALL, C. R. (1991). Procrustes methods in the statistical analysis of shape (with discussion). *Journal of the Royal Statistical Society, Series B*, 53:285–339. [MR1108330](#)
- GOWER, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, 40:33–50. [MR0405725](#)
- KENDALL, D. G. (1989). A survey of the statistical theory of shape (with discussion). *Statistical Science*, 4:87–120. [MR1007558](#)
- KENT, J. T. and MARDIA, K. V. (2001). Shape, Procrustes tangent projections and bilateral symmetry. *Biometrika*, 88(2):469–485. [MR1844846](#)

- LE, H.-L. (1995). Mean size-and-shapes and mean shapes: A geometric point of view. *Advances in Applied Probability*, 27:44–55. [MR1315576](#)
- SANGALLI, L. M., SECCHI, P., VANTINI, S., and VENEZIANI, A. (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.
- SRIVASTAVA, A., KLASSEN, E., JOSHI, S. H., and JERMYN, I. H. (2011). Shape analysis of elastic curves in Euclidean spaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(7):1415–1428.