Supplement for "Topo-Geometric Analysis of Variability in Point Clouds using Persistence Landscapes"

James Matuk, Sebastian Kurtek, *Senior Member, IEEE,* Karthik Bharath

APPENDIX A
SIMULATED EXAMPLES WITH ADDITIVE NOISE

We repeat Simulated Example 1 from the main article with additive pointwise noise, i.e., for each point cloud, we independently generate additive noise from a zero-mean bivariate Gaussian distribution with covariance $r(0.1)^2 I_2$, where r is the radius of the circle that underlies the 'noiseless' point cloud. Two examples of point clouds and their corresponding degree $p = 1$, $K = 1$-dimensional landscape functions are shown in panels (a) and (c), and (b) and (d) in Figure 1, respectively. The landscape functions for all 20 point clouds are shown in panel (e). Most of the landscapes generated from the noisy point clouds are similar in shape to those from the 'noiseless' setting. However, sometimes there is an extremely small peak, corresponding to noise, in some of the landscape functions. This is most visually apparent in panel (f), where there is an extremely small peak at $t \approx 0.25$ prior to the major peak in many of the landscape functions. Due to the small magnitude of this noise-induced feature, additive noise appears to have little effect on landscape alignment and mean computation. The mean based on aligned landscapes, visualized in panel (g), is consistent with that of a circle. There is considerable variance reduction in the denoised/transformed persistence diagrams in panel (j), via reparameterizations shown in (i). Importantly, points corresponding to the main feature of the point clouds are collapsed to a single point, while points near the diagonal, corresponding to features created by the additive noise, remain near the diagonal. Such clarity is absent in the noisy persistence diagrams in (h).

Next, we repeat Simulated Example 2 from the main article with additive noise. For point clouds in the blue group (single circle), we independently generate additive noise from a zero-mean bivariate Gaussian distribution with covariance $r(0.1)^2 I_2$, where r is the radius of the circle that underlies the 'noiseless' point cloud. We repeat this procedure for point clouds in the red group, but noise is

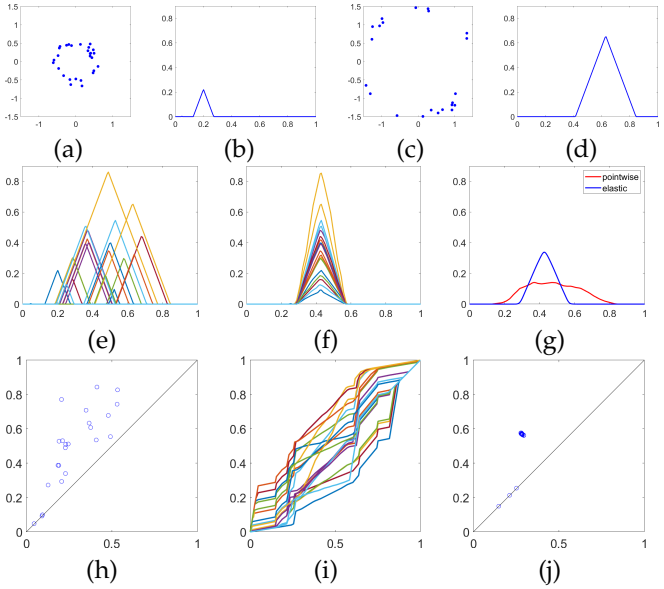

Fig. 1. *Same topology with scale and sampling variabilities as well as additive noise:* (a)&(c): Two examples, from 20, of randomly generated point clouds on circles with randomly chosen radii, random sample sizes and additive noise. (b)&(d): Corresponding persistence landscapes. (e) Persistence landscapes $\{\Lambda_i\}_{i=1}^{20}$ of 20 point clouds. (f) Aligned persistence landscapes $\{\Lambda_i(\gamma_i)\}_{i=1}^{20}$. (g) Mean landscape after (blue) and without (red) alignment. (h) Noisy persistence diagrams $\{(b_{ij}, d_{ij})\}_{i=1}^{20}$ from 20 point clouds. (i) Estimated reparameterizations $\{\gamma_i\}_{i=1}^{20}$. (j) Denoised/transformed persistence diagrams $\{(\gamma_i^{-1}(b_{ij}), \gamma_i^{-1}(d_{ij}))\}_{i=1}^{20}$.

generated such that the covariance depends on the radius of one of the two circles that it belongs to. A single example of a point cloud in the blue and red groups is shown in Figure 2(a)&(c); the corresponding degree $p = 1$, $K = 2$-dimensional landscapes are shown in panels (b) and (d). The shapes of the landscapes in each group are similar to those displayed in the 'noiseless' setting. However, here, we notice some differences in PCA carried out on aligned landscapes. The first direction of variability, viewed in the top row of panels (e)-(g) appears to be associated with scale variability in the point cloud data; the results here are displayed in the same manner as in Figure 6 in the main article. On the other hand, as confirmed in the top row of panel (h), the second direction of variability appears to be associated

- *J. Matuk, Epidemiology Data Center, Univeristy of Pittsburgh*
  E-mail: jam925@pitt.edu
- *S. Kurtek, Department of Statistics, The Ohio State University*
  E-mail: kurtek.1@stat.osu.edu
- *K. Bharath, School of Mathematical Sciences, University of Nottingham*
  E-mail: Karthik.Bharath@nottingham.ac.uk

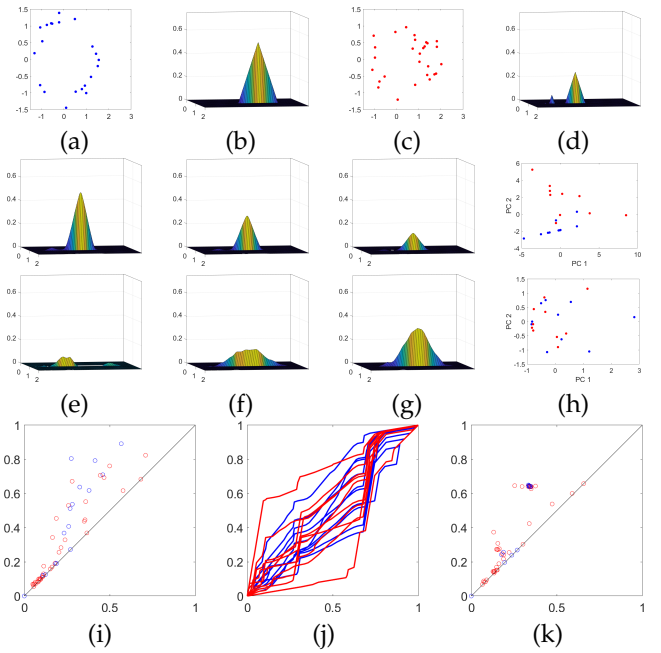

Fig. 2. *Different topology with scale and sampling variabilities as well as additive noise*: (a)&(c) Two examples, from 20, of randomly generated point clouds from topologically different spaces (blue and red, respectively, in all relevant panels). (b)&(d) Corresponding degree $p = 1$, $K = 2$-dimensional persistence landscapes. (e) -1, (f) 0, (g) +1 standard deviation from the mean landscape in the first PC direction, and (h) projection of landscapes onto the first two PC directions: following alignment (top) and without alignment (bottom). (i) Noisy and (k) denoised persistence diagrams. (j) Estimated reparameterizations.

with the homology of the point clouds, i.e., most of the red point clouds, generated from two connected circles, have a positive second PC score, while most of the blue point clouds, generated from a single circle, have a negative second PC score. Since noise can distort topological features, there does appear to be some overlap between the two classes based on the first two PC scores. This observation is in contrast to the 'noiseless' setting, where the two classes are clearly separated based on the first PC score alone. The corresponding PC scores computed based on unaligned landscapes, shown in the bottom of panel (h), provide no such distinction between the two classes. A comparison of the denoised persistence diagrams presented in Figure 2(k) to their noisy counterparts in panel (i) shows the benefits of our approach. While the clustering of features in panel (k) is not as clear as in the 'noiseless' setting, one can still extract useful homological information from the denoised persistence diagrams. On the other hand, this is not possible based on the noisy diagrams in panel (i).

In Figure 3, we illustrate the effects of increasing pointwise noise on alignment of persistence landscapes and denoising of the corresponding persistence diagrams. The point cloud data considered in the left (low noise), middle (medium noise) and right (high noise) columns in this figure were generated by first uniformly sampling 30 points on a circle with radius $r = 1$ and then adding pointwise noise sampled from a zero-mean Gaussian distribution with covariance $(0.1)^2 I_2$, $(0.25)^2 I_2$ and $(0.5)^2 I_2$, respectively. For each of the three examples, we generate 20 point clouds in this manner, and consider their alignment and averaging

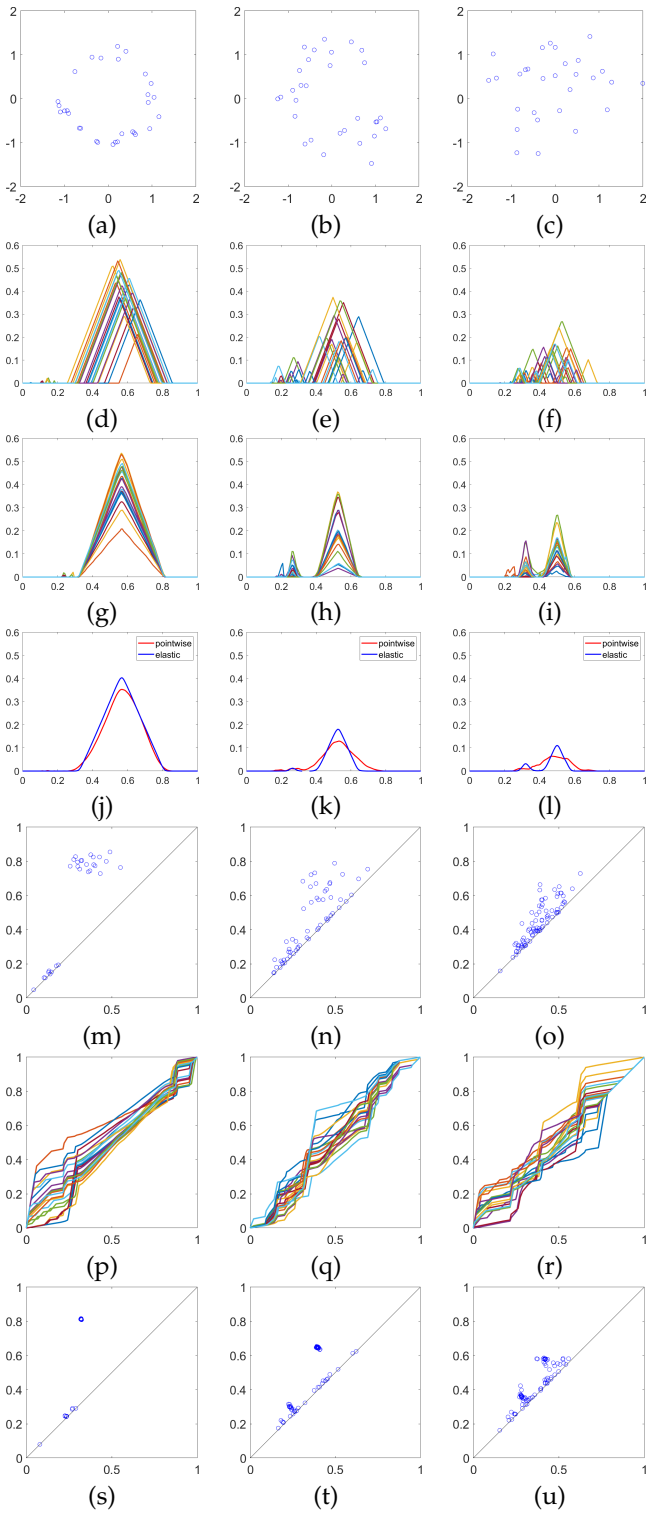

Fig. 3. *Same topology and scale with increasing pointwise noise*: Point clouds in the left, middle and right columns were generated with noise variance of $(0.1)^2$, $(0.25)^2$ and $(0.5)^2$, respectively. (a)-(c) Example of one of the 20 point clouds with additive pointwise noise. (d)-(f) Degree $p = 1$, $K = 1$-dimensional persistence landscapes $\{\Lambda_i\}_{i=1}^{20}$. (g)-(i) Aligned persistence landscapes $\{\Lambda_i(\gamma_i)\}_{i=1}^{20}$. (j)-(l) Mean landscape after (blue) and without (red) alignment. (m)-(o) Noisy persistence diagrams. (p)-(r) Estimated reparameterizations $\{\gamma_i\}_{i=1}^{20}$. (s)-(u) Denoised persistence diagrams $\{(\gamma_i^{-1}(b_{ij}), \gamma_i^{-1}(d_{ij}))\}_{i=1}^{20}$.

using degree $p = 1$, $K = 1$-dimensional persistence land-

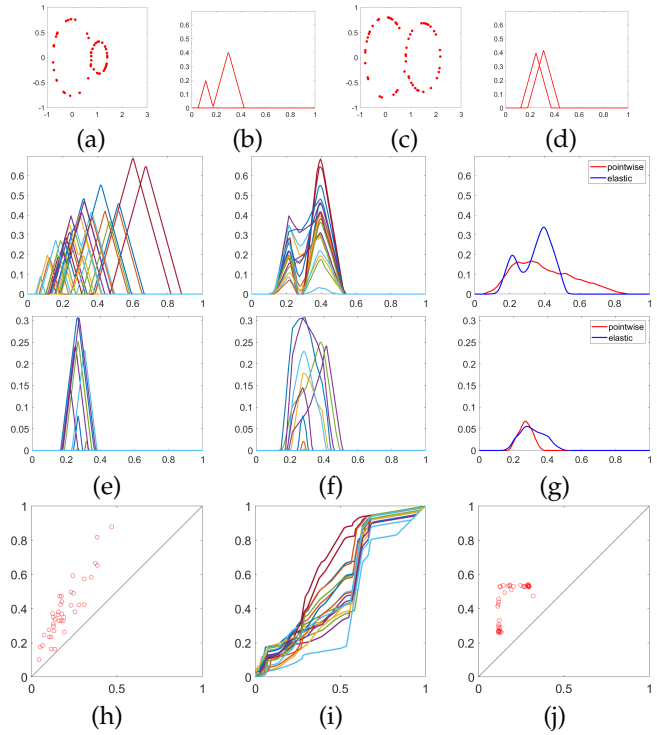

Fig. 4. *Same topology with scale and sampling variabilities*: (a)&(c) Two examples, from 20, of randomly generated point clouds. (b)&(d) Corresponding persistence landscapes. The two rows correspond to the two component functions in each landsacpe: (e) Persistence landscapes $\{\Lambda_i\}_{i=1}^{20}$ of 20 point clouds. (f) Aligned persistence landscapes $\{\Lambda_i(\gamma_i)\}_{i=1}^{20}$. (g) Mean landscape after (blue) and without (red) alignment. (h) Noisy persistence diagrams $\{(b_{ij}, d_{ij})\}_{i=1}^{20}$ from 20 point clouds. (i) Estimated reparameterizations $\{\gamma_i\}_{i=1}^{20}$. (j) Denoised persistence diagrams $\{(\gamma_i^{-1}(b_{ij}), \gamma_i^{-1}(d_{ij}))\}_{i=1}^{20}$.

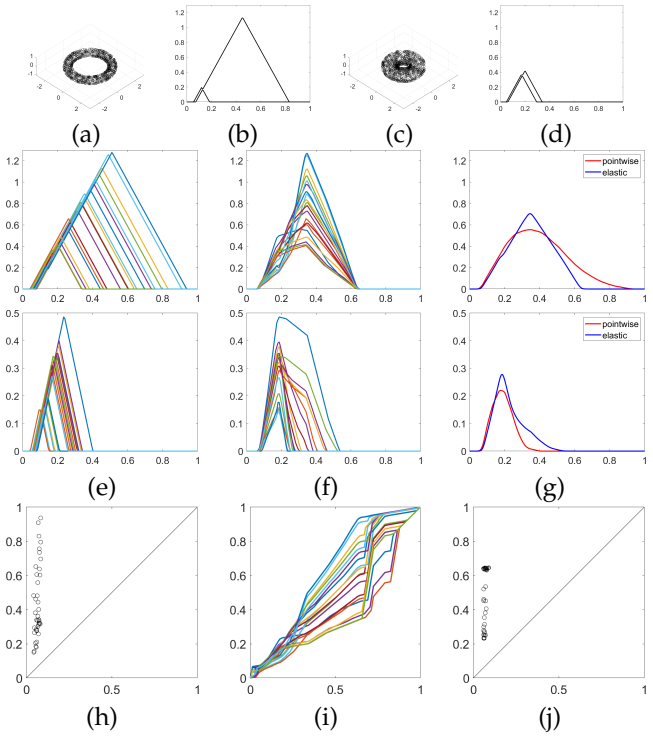

Fig. 5. *Same topology with scale and sampling variabilities:* (a)&(c) Two examples, from 20, of randomly generated point clouds. (b)&(d) Corresponding persistence landscapes. With two rows corresponding to two component landscapes: (e) Persistence landscapes $\{\Lambda_i\}_{i=1}^{20}$ of 20 point clouds. (f) Aligned persistence landscapes $\{\Lambda_i(\gamma_i)\}_{i=1}^{20}$. (g) Mean landscape after (blue) and without (red) alignment. (h) (Rescaled) Noisy persistence diagram $\{(b_{ij}, d_{ij})\}_{i=1}^{20}$ from 20 point clouds. (i) Estimated reparameterizations $\{\gamma_i\}_{i=1}^{20}$. (j) Denoised/transformed persistence diagram $\{(\gamma_i^{-1}(b_{ij}), \gamma_i^{-1}(d_{ij}))\}_{i=1}^{20}$.

scapes. Panels (a)-(c) display one example point cloud for each of the noise settings. The corresponding landscapes for all 20 observations are shown in panels (d)-(f) with corresponding noisy persistence diagrams in panels (m)-(o). While the underlying circle is visibly discernible in panels (a) and (b), the noise overwhelms the structure of the data in panel (c). This is further reflected in the corresponding persistence landscapes, which have one prominent peak in panels (g) and (h), but a relatively smaller prominent peak (and another peak with similar magnitude) in panel (i). Reparameterizations used to align the persistence landscapes in panels (d)-(f), resulting in the aligned landscapes in panels (g)-(i), are shown in (p)-(r). It is evident that alignment is effective in the low and medium noise settings resulting in a mean that is consistent with that of a circle, as seen in panels (j) and (k). Also, as expected, there is considerable variance reduction in the denoised persistence diagrams in panels (s) and (t) as compared to their noisy counterparts shown in panels (m) and (n). As before, points corresponding to the main feature of the point clouds are collapsed to a single point, while points near the diagonal, corresponding to features created by the additive noise, remain near the diagonal. This structure is less clear in the high noise setting. First, the mean shown in (l) contains two clear peaks: a small peak near $t = 0.3$ and a much larger peak near $t = 0.5$. Second, while some variance reduction is observed when comparing the noisy diagrams in (o) to their denoised versions in (u), alignment in this case does not appear to be as effective in distinguishing the underlying structure in the point clouds from noise.

APPENDIX B
ADDITIONAL MEAN ESTIMATION EXAMPLES

In Figure 4, we consider mean estimation based on degree $p = 1, K = 2$-dimensional persistence landscapes computed from 20 point clouds that consist of uniformly sampled points along two circles with different radii. The point clouds in this example are generated in the same way as the data in the red group in Simulated Example 3 in the main article. Panels (a) and (c) show two examples of randomly generated point clouds with their corresponding landscapes in panels (b) and (d). Panels (e)-(g) show all 20 $k = 1$ (top) and $k = 2$ (bottom) landscape component functions, their alignment, and a comparison of the mean before (red) and after (blue) alignment, respectively. The proposed alignment procedure results in a mean landscape that better preserves the major features along both landscape components. On the other hand, the unaligned mean landscape destroys the prominent two peak structure in the first component. In panel (j), the points in the denoised persistence diagrams, corresponding to the two loops in the point clouds, form two separate clusters making the presence of these features

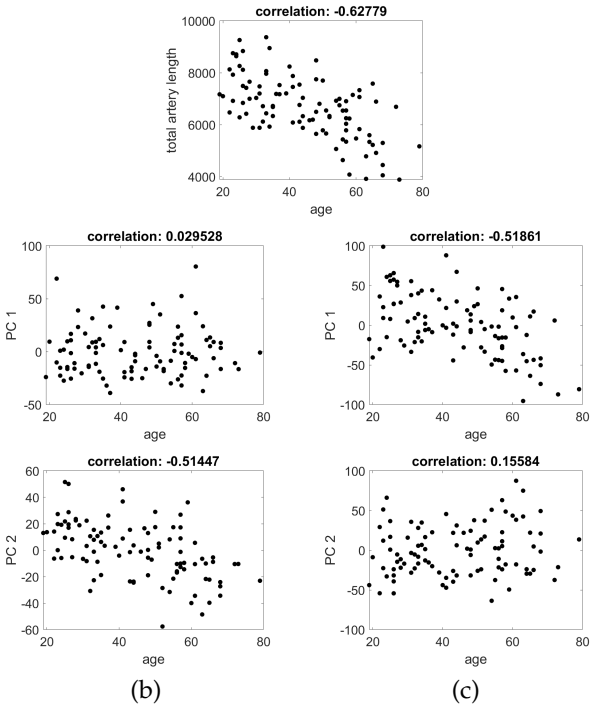

Fig. 6. *(a) Correlation and scatterplot of total artery length versus age. Correlations and scatterplots of PC 1 (top) and PC 2 (bottom) estimated using (b) aligned and (c) unaligned landscapes, computed from original persistence diagrams, versus age.*

in the data much clearer; the noisy diagrams shown in panel (h) do not provide such a distinction.

In Figure 5, we consider mean estimation based on degree $p = 1$, $K = 2$-dimensional persistence landscapes computed from 20 point clouds that consist of 1000 points sampled uniformly on a ringed torus. The major radius of each torus is sampled from a $|N(2, .3^2)|$, while the minor radius is a proportion $\text{Beta}(10, 10)$ of the major radius. Two example point clouds are shown in panels (a) and (c). We preprocess the persistence diagrams used to compute the persistence landscapes by disregarding all points except for the two with longest persistence, as these points correspond to the two loops formed by the torii that underlie the point clouds. The landscapes corresponding to point clouds in (a) and (c) are shown in (b) and (d), respectively. Clearly, the estimated landscapes can vary widely depending on the relationship between the major and minor radii. Panels (e)-(g) show the $k = 1$ (top) and $k = 2$ (bottom) landscape component functions, their alignment, and a comparison of the mean before (red) and after (blue) alignment. The first landscape function is automatically weighted higher during alignment due to the relatively large magnitude of the peak as compared to the second landscape function. In panel (j), the points in the denoised persistence diagrams, using the reparameterizations shown in (i), concentrate to make the presence of the features more clear; the two detected features describe the two loops. In comparison, the noisy diagrams shown in panel (h) do not provide a clear distinction.

APPENDIX C
CORRELATION BETWEEN AGE AND TOPOLOGICAL STRUCTURE OF BRAIN ARTERY TREES BASED ON UNSCALED PERSISTENCE DIAGRAMS

Finally, we repeat the analysis conducted in Section 4.2 in the main article, but using $p = 1$, $K = 100$-dimensional persistence landscapes derived from persistence diagrams that were not rescaled using total artery length. First, in Figure 6(a) we show a scatterplot of total artery length versus age; it is clear that they are significantly correlated with total artery length generally decreasing with age. In Figure 6(b), we show the scatterplots of the first (top) and second (bottom) PC scores, derived from aligned persistence landscapes. Figure 6(c) shows the same, but based on unaligned landscapes. It appears that the first PC in panel (c) captures scale differences rather than topological ones (there is a high correlation with age, which is also highly correlated with total artery length). On the other hand, the first PC in panel (b) does not capture such scale differences as the correlation with age is low.