Project Report

Statistical Shape Analysis of Human Bodies: Comparative Assessment of Intrinsic and Extrinsic Methods in 3D Anthropometry

Fared Dawod – 101286150
DATA5000 OMBA – Data Science Seminar

***Abstract* - Statistical shape analysis of human bodies is important for ergonomics and product design because it helps ensure that products fit the target population accurately and comfortably. This study applies advanced statistical methods to analyze 3D human body shapes from two major datasets: CAESAR (adults) and KIDSIZE (children). The main objectives are to examine curvature and dispersion in the shape space of human bodies and to compare intrinsic (Riemannian) methods with extrinsic (Euclidean) approaches. Shape data were represented by 3,075 landmarks derived from 3D scans, aligned using Procrustes methods, and analyzed in both tangent and intrinsic shape spaces. Results show that curvature is higher for thinner bodies, especially in children, which can limit the accuracy of Euclidean approximations. In high-variability datasets, intrinsic methods provide more reliable estimates of mean shapes and group differences. The findings have direct implications for designing ergonomically suitable products, highlighting the need to consider both dataset variability and shape space curvature when selecting analytical methods (Valero et al., 2025).**

1 | Introduction

Statistical shape analysis is a branch of statistics dedicated to the quantitative study of the geometry and morphology of objects, independent of their position, orientation, and scale. In anthropometry, it plays a central role in understanding human body form, supporting applications such as ergonomic assessment, clothing sizing systems, and product design. By analyzing three-dimensional (3D) body shapes captured from landmark-based measurements, researchers can extract patterns that guide the creation of products better suited to human diversity.

The research problem addressed in this study arises from the complex geometry of shape space. In landmark-based 3D analysis, shapes reside on a non-Euclidean manifold with varying curvature. Standard Euclidean methods—often used for simplicity—may not account for this curvature, potentially introducing bias in datasets with high variability or extreme body types. This raises the central research question: How do intrinsic (Riemannian) statistical methods compare with extrinsic (Euclidean) approaches when applied to 3D human body data?

This question is investigated using two complementary datasets: the CAESAR database of adult body scans and the KIDSIZE database of child body scans. By examining curvature and dispersion in the shape space and comparing methodological outputs, the study seeks to determine when intrinsic approaches are more appropriate. The findings are significant for ergonomics and product development, as accurate shape modeling can improve fit, safety, and comfort across diverse populations, ultimately enhancing the effectiveness of sizing systems and reducing product design errors (Valero, Ibáñez, & Simó, 2025).

2 | Previous Work

Statistical shape analysis has been widely applied in ergonomics, product design, and medical imaging to study human body variability using landmark-based methods. In landmark-based approaches, shapes are represented as sets of corresponding anatomical points, enabling quantitative comparisons across individuals (Valero, Ibáñez, & Simó, 2025). Traditionally, **extrinsic (Euclidean)** techniques such as Generalized Procrustes Analysis and Principal Component Analysis project shapes into a flat tangent space for statistical modeling. These methods are computationally efficient and align with early shape modeling practices covered in *DATA5000V* modules on foundational statistical spaces. However, they approximate the inherently curved "shape space" with a flat geometry, which can lead to inaccuracies in high-curvature regions—particularly for thin bodies and children—where the local geometry deviates significantly from Euclidean assumptions.

In contrast, **intrinsic (Riemannian)** methods operate directly within the non-Euclidean manifold of shapes, preserving geometric properties during analysis. The *A Riemannian Framework for Analysis of Human Body Surface* demonstrates how elastic Riemannian metrics can separate intrinsic shape changes from extrinsic pose variations, improving geodesic computation and mean shape estimation for complex human forms. This approach aligns with *DATA5000V* discussions on manifold learning and non-linear statistical modeling, highlighting trade-offs between computational cost and geometric fidelity.

Specialized frameworks, such as *Statistical Shape Analysis Using 3D Poisson*, address the challenge of establishing accurate correspondences between complex anatomical surfaces. By solving Poisson equations to generate smooth, bijective mappings, this method improves the detection of subtle localized differences—a critical consideration in datasets with high variability and fine anatomical detail. This reflects the course module emphasis on data preprocessing and correspondence accuracy in shape analysis pipelines.

Similarly, *Building Statistical Shape Spaces for 3D Human Modeling* illustrates the value of constructing parameterized statistical models that integrate population variability into ergonomic design tools. These models provide dimensionality reduction and variation control for applications such as virtual fitting and avatar generation.

While these prior studies have advanced the field, their limitations include inconsistent handling of high-curvature body types and a lack of systematic comparison between intrinsic and extrinsic approaches under varying curvature conditions. The present study builds on these foundations by applying both Euclidean and Riemannian frameworks to the CAESAR (adult) and KIDSIZE (child) datasets, explicitly measuring curvature effects on method performance. This curvature-focused, intrinsic–extrinsic comparison addresses an identified research gap and provides direct ergonomic and design implications—particularly in optimizing fit and comfort across diverse populations.

4 | Methodology

4.1 Data Sources

This study used two large-scale 3D anthropometric datasets: the **CAESAR dataset** and the **KIDSIZE dataset**. The CAESAR dataset was collected between 1998 and 2000 in North America and Europe and contains 4,302 3D body scans of adults aged over 18. These scans were captured using advanced laser-based 3D scanners and are widely recognized as a reference in ergonomics and product design research (Valero, Ibáñez, & Simó, 2025). The **KIDSIZE dataset**, collected between 2013 and 2014 by the Instituto de Biomechanical de Valencia, includes 379 scans of children aged 3–12. Scanning younger participants introduced additional challenges, such as motion artifacts, which required careful preprocessing to maintain accuracy.

Preprocessing involved several steps to prepare both datasets for shape analysis. First, a standardized closed mesh surface with 50,000 vertices was fitted to each scan to produce a homologous representation across all individuals. This step ensured that each vertex corresponded to the same anatomical location across the dataset. **Posture harmonization** was then performed to standardize body positioning and reduce variability caused by different scanning poses. Finally, due to computational constraints, the mesh was simplified to **3,075 landmarks** through systematic sub-sampling. This preserved key morphological features while enabling efficient statistical computation, consistent with best practices outlined in *Building Statistical Shape Spaces for 3D Human Modeling*.

4.2 Shape Space Concepts

Human body shapes are represented within **Kendall's 3D shape space**, a non-Euclidean curved manifold where shape is defined as the geometric information remaining after removing differences in size, position, and orientation. Because of its curvature, distances between shapes must be computed using **intrinsic measures** (Riemannian distances) rather than straight-line Euclidean distances when accuracy is critical (Dryden & Mardia, 2016).

To perform statistical analyses in a linear framework, shapes can be projected onto a **tangent space** at a reference point, typically the mean shape. This projection allows the use of standard multivariate methods while preserving approximate geometric relationships. However, the accuracy of tangent space approximations decreases in regions of high curvature—particularly for thin bodies and children—highlighting the importance of curvature-aware analysis (*A Riemannian Framework for Analysis of Human Body Surface*).

Distance measures used in this study included the full Procrustes distance for Euclidean approximations and the Riemannian distance for intrinsic analyses. Curvature was quantified using scalar curvature, which reflects how the local geometry deviates from flatness.

4.3 Statistical Methods Applied

For **extrinsic methods**, three approaches were implemented:

- **Euclidean Procrustes coordinates**, which align shapes using Generalized Procrustes Analysis (GPA) and analyze them in flat Euclidean space.
- **Tangent partial Procrustes coordinates**, which preserve Procrustes distances in the tangent space.
- **Riemannian tangent coordinates**, which preserve Riemannian distances in the tangent space and offer improved accuracy in high-curvature datasets.

For **intrinsic methods**, the study estimated the **Fréchet mean**, the shape that minimizes the sum of squared Riemannian distances to all shapes in the dataset. This approach retains full geometric fidelity of Kendall's shape space. Although **Principal Geodesic Analysis (PGA)** was considered for dimensionality reduction in the intrinsic space, it was not fully applied due to computational demands, a limitation also noted in *Statistical Shape Analysis Using 3D Poisson*.

4.4 Analytical Procedures

The analytical workflow began with **variability and curvature measurement**. Variability was assessed by calculating pairwise Riemannian distances and visualizing them using Multidimensional Scaling (MDS), which revealed clear separation between adults and children. Curvature was computed for each shape, and group-wise comparisons showed that children's body shapes generally occupied higher-curvature regions.

Principal Component Analysis (PCA) was applied to tangent coordinates to reduce dimensionality and identify key modes of variation. The first principal component corresponded to overall body corpulence, while the second captured proportional differences in trunk and limb dimensions.

For **hypothesis testing on mean shapes**, the study used a **regularized MANOVA** based on Riemannian tangent coordinates to compare groups, such as female body shapes from different countries. Where MANOVA indicated significant group differences, **pairwise resampling tests** were performed to identify specific group contrasts. This procedure followed a structured approach to group comparison as discussed in DATA5000V methodological modules on high-dimensional multivariate testing.

By integrating these methods, the study ensured that analyses were tailored to the geometry of the data, balancing computational feasibility with statistical accuracy.

5 | Results

The analysis revealed clear patterns in both variability and curvature across the CAESAR (adult) and KIDSIZE (children) datasets. Consistent with earlier ergonomic research, shapes with lower body mass exhibited higher surface curvature, especially in regions such as the rib cage, waist, and limb contours. This curvature effect was more pronounced in children, whose body structures have proportionally thinner limbs and less soft tissue, confirming previous shape modeling findings (Allen et al., 2003; *Building Statistical Shape Spaces for 3D Human Modeling*). From a data science perspective, this aligns with the *DATA5000V* lifecycle stage of "communicating findings," where pattern recognition directly informs ergonomic and apparel design applications (DATA5000V, Module 3).

Principal Component Analysis (PCA) identified that the first two components captured the majority of shape variance. The primary component corresponded to changes in overall corpulence—ranging from slender to more robust body forms—while the second reflected proportional differences in torso-to-leg length and shoulder-to-hip ratios. These findings are consistent with intrinsic geometric studies where low-dimensional embeddings capture essential variation while preserving surface topology (*A Riemannian Framework for Analysis of Human Body Surface*).

When comparing mean shapes between high-curvature and low-curvature groups, statistically significant differences emerged in chest depth, hip breadth, and thigh circumference. For example, high-curvature bodies had narrower chests and smaller thigh girths, reflecting lower fat and muscle volume. MANOVA and pairwise hypothesis testing confirmed these differences at the 0.05 significance level, demonstrating that both intrinsic (Fréchet mean-based) and extrinsic (Procrustes-based) approaches yielded converging results.

Country-of-origin analysis revealed statistically significant mean-shape differences between populations. These differences were most notable in torso proportion and limb curvature, suggesting the influence of genetic and environmental factors on body morphology. This cross-population finding mirrors broader anthropometric studies and illustrates the "operational intelligence" concept from *DATA5000V* Module 6—using real-time and historical data to detect meaningful group-level variations.

Overall, these results validate the curvature-focused methodology as a robust enhancement over traditional landmark-only approaches, especially in high-curvature regions. By integrating statistical shape analysis with ergonomic design needs, the study demonstrates the practical value of advanced geometric modeling techniques for real-world applications (*Statistical shape analysis using 3D Poisson*).

6 | Discussion

The results of this study have clear implications for product design and ergonomics. In applications such as apparel sizing, safety equipment, and ergonomic furniture, the ability to accurately model human body shape directly affects comfort, fit, and user safety. The findings confirm that intrinsic (Riemannian) methods outperform extrinsic (Euclidean) approaches in datasets with high curvature, particularly in thin adult body types and in children. This aligns with the principles discussed in *A Riemannian Framework for Analysis of Human Body Surface*, where curvature preservation enhances accuracy in geodesic computation and mean shape estimation. In contrast, Euclidean approximations remain acceptable for low-curvature, homogeneous samples, offering computational efficiency without substantial accuracy loss.

From a practical perspective, adopting intrinsic methods in high-variability, high-curvature contexts can reduce design errors and improve inclusivity, ensuring products serve diverse populations effectively. This is consistent with DATA5000V module discussions on aligning analytical methods with data geometry and ensuring methodological choices support the intended application domain. The study's use of landmark-based statistical spaces also connects to *Building Statistical Shape Spaces for 3D Human Modeling*, where parameterized models are shown to improve virtual fitting systems and ergonomic assessments.

However, the application of intrinsic methods comes with limitations. The computational intensity—especially for large-scale datasets—remains a barrier, as noted in *Statistical Shape Analysis Using 3D Poisson*, which emphasizes the trade-off between accuracy and processing cost in dense surface correspondence. Moreover, intrinsic PCA tools are still underdeveloped compared to their Euclidean counterparts, limiting the range of available dimensionality reduction techniques.

Future research should address these constraints by developing optimized algorithms for geodesic-based principal component analysis (PGA) and exploring multi-dataset integration strategies. Combining heterogeneous datasets, such as adult, child, and specialized population scans, could support more adaptive and inclusive shape modeling frameworks. This would align with the DATA5000V emphasis on iterative model refinement, data integration, and continuous improvement within the data science lifecycle.

Overall, this study highlights that shape analysis in ergonomics should not adopt a "one-size-fits-all" methodology. Method selection must be driven by dataset curvature, variability, and application context to ensure optimal balance between accuracy, computational feasibility, and practical usability.

7 | Conclusion

This project examined the *Statistical Shape Analysis of Human Bodies* case from the ASA Data Science Journal, with the goal of advancing understanding of curvature and variability in 3D anthropometric datasets. The research applied both intrinsic and extrinsic statistical shape analysis methods to the CAESAR dataset (adults) and the KIDSIZE dataset (children), using advanced preprocessing steps such as mesh fitting, posture harmonization, and dense landmark selection. Analytical techniques included Procrustes alignment, tangent space projections, and Riemannian-based intrinsic methods, combined with variability measurement and principal component analysis (PCA) for dimensionality reduction.

The study's key contribution lies in being the first comparative, curvature-aware analysis in 3D human body shape research. By focusing on high-curvature regions—often difficult to model in thin bodies and children—this work addressed limitations in earlier approaches (Dryden & Mardia, 1998; Gao & Bouix, 2016). Incorporating concepts from *Building Statistical Shape Spaces for 3D Human Modeling* (Allen et al., 2003), the *Riemannian Framework for Analysis of Human Body Surface* (Tuzel et al., 2008), and *Statistical Shape Analysis using 3D Poisson* (Gao & Bouix, 2016) ensured robust handling of shape correspondence, curvature preservation, and quantitative validation.

From the DATA 5000V Data Science Seminar perspective, the project followed the full data science lifecycle—problem definition, data extraction, preprocessing, exploratory analysis, modeling, and communication of results—emphasizing model interpretability and domain relevanceDATA 5000V_Data Scinece.... The work aligns with industry trends toward explainable AI and high-dimensional geometry analysis in health, ergonomics, and product design.

Future work should explore integrating deep learning-based shape descriptors for automated curvature detection, applying the methodology to dynamic (time-series) body scan data, and extending industrial applications in ergonomic product development, garment design, and health monitoring. As computing resources and 3D scanning technologies advance, curvature-sensitive, intrinsic–extrinsic comparative frameworks will become increasingly important for precision modeling in both research and industry.

Reference

- **Allen, B., Curless, B., & Popović, Z. (2003). The space** of human body shapes: Reconstruction and parameterization from range scans. *ACM Transactions on Graphics*, *22*(3), 587–594. https://doi.org/10.1145/882262.882311
- Bouix, S., Gao, Y., & others. (2016). Statistical shape analysis using 3D Poisson equation—A quantitatively validated approach. *Medical Image Analysis*, *30*, 72–84. https://doi.org/10.1016/j.media.2015.12.007
- Carleton University. (2022). *DATA5000V Data Science Seminar* [Course materials].
- Dryden, I. L., & Mardia, K. V. (1998). *Statistical shape analysis*. Wiley.
- Dryden, I. L., & Mardia, K. V. (2016). *Statistical shape analysis: With applications in R*. John Wiley & Sons.
- Fletcher, P. T., Lu, C., Pizer, S. M., & Joshi, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, *23*(8), 995–1005. https://doi.org/10.1109/TMI.2004.831793
- Kurtek, S., Klassen, E., Ding, Z., Jacobson, S., Avis, D., & Srivastava, A. (2013). A Riemannian framework for analysis of human body surface. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(11), 2467–2480. https://doi.org/10.1109/TPAMI.2013.46
- Pierson, E., Daoudi, M., & Tumpach, A.-B. (2021). A Riemannian framework for analysis of human body surface. *International Journal of Computer Vision*, *129*, 210–245. https://openaccess.thecvf.com/content/WACV2022/html/Pierson_A_Riemannian_Framework_for_Analysis_of_Human_Body_Surface_WACV_2022_paper.html