

Review of the manuscript “Geometric Morphometric Analysis of Projectile Points from the Southwest United States” by Robert J. Bischoff

The manuscript presents a geometric morphometric analysis of projectile points from the Southwest of the United States. The paper’s primary goal is to analyze the unclassified projectile points from Tonto Basin, Arizona. To select the most suitable method for analyzing projectile points from Tonto Basin, the author compared various geometric morphometric approaches on previously classified late prehistoric projectile points from the Southwest of the United States. More precisely, the author compares Elliptical Fourier Analysis (EFA) to landmarks and semilandmarks approaches based on Generalized Procrustes Analysis (GPA). Results showed that the semilandmarks and EFA do not adequately distinguish between projectile points' shapes, especially at identifying notches. However, the landmark approach provided promising results and thus was selected as the primary method to analyze the projectile points from the Tonto Basin sites.

The analysis of projectile points through GM is a relatively novel approach in archaeology. Therefore, one of the main contributions of this paper is the comparison between EFA, semilandmarks and landmarks-based methods in the analysis of projectile points. While the EFA necessitates complete projectile points to extract the shape of the projectile points, the landmark and semilandmark approaches allow the analysis of fragmented or broken projectile points. The methodological approach is innovative, and the study makes a valuable contribution to the knowledge base in the archaeology of the Southwest of the United States (e.g., line 49). The research is replicable. The availability of the R script used to achieve these analyses shows the study's repeatability and the author's best practices.

I recommend the publication of this manuscript; however, the following suggestions need to be considered.

General comments:

- It is recommended to avoid using statements in the first person to maintain an objective, impersonal tone in academic writing. For instance, in **line 119**, talking about the Momocs package, you mention, “... but I have found the tpsDig’s utility to be more superior.” This sentence must be reformulated with data or proof supporting your statement. Also, **line 94**, “In my experience, this can be a frustrating way to spend your time”.
- The structure of the paper needs to be modified. For instance, in **line 154** Geometric Morphometrics section should be preceded by another section entitled methods. These changes will enable the readers to find the section where you discuss the methodology easily. Additionally, the section comparisons (**line 248**) should be named “results” as it presents the results of your analysis. Please follow the specific format used in scientific writing: introduction, materials, methods, results, and discussion (and/or conclusion).
- Be more precise in your statements. For instance, in **line 57**, you talk about “the code” and that it will be available. Please distinct that you’re referring to your R script used for analysis. As you do not mention before this line that you are using R language programming for your study, it is essential to be concise because not all readers are familiar with the terminology. Also, in **line 64**, you say, “...those approaches...” please be precise about which approach you are referring to (GM).
- Distinguish between points when it refers to landmarks and points when it refers to projectile points. For example, in **lines 31-32 and 178**, “points” refer to the landmarks used in the analysis and in **lines 37, 223-225, and 287**, however, “points” refer to projectile points. Please be cautious about the terminology and precise, as not all readers are familiar with GM. This needs to be standardized throughout the manuscript.
- For all PCA graphs, all the axes need to be scaled.

Introduction (line 23)

- **Lines 37 to 41:** lack of references.

- **Line 67:** to be more precise, the statement referring to the code and data availability should be mentioned in the materials and/or methods sections (e.g., line 114), not in the introduction.
- **Lines 69 to 70:** aside from the abstract, it is not previously mentioned that the corner-used landmark analysis is the most successful. Additionally, you are presenting the results of your analysis in the introduction. Be careful not to mix up the study's aims with the study's results; the latter should only be discussed in the results section (in your case, “comparisons”)—same remarks for **line 71**.

Background (line 72)

- This section relates to the materials used; it seems more appropriate to include it within the “data collection” section as you discuss the selected dataset and the reasons for doing so (**lines 73-76**).
- The “background” section needs to be re-evaluated as it starts with the presentation of the sites and materials used in this study (**lines 73 to 84**), then is followed by the lack of documentation (**lines 85 to 89**). It is concluded by the limitations of the method used and the aims of the current study (**lines 90 to 99**). There are a lot of elements presented in this section that need to be reassigned to the appropriate section of the article to avoid repetition.

Data collection (line 100)

- This subsection should be a section on its own (usually referred to as materials, although data collection is also acceptable) that includes “projectile points of the Southwest” and “Tonto Basin Projectile” as subsections.
- **Line 103:** it is mentioned that “Many of Justice’s types could not be included...” based on that statement, and out of the 74 illustrations, how many projectile points were used in this study? The sample needs to be mentioned explicitly.
- **Line 110:** you refer to figure 3 in the text. Figure 3 is a representation of results based on a PCA. The figure needs to be described (briefly) in the text. It is essential to explain to the reader why there should be no hesitation in using illustrations for 2D morphometric analysis based on the results of figure 3.
- **Line 125:** the following sentence needs clarification: “... the authors identify correlates of the types”.
- **Line 126:** please state clearly and explicitly the period you are interested in exploring.

Geometric morphometrics (line 154)

- This subsection should be a section on its own; usually, it is referred to as methods.
- **Line 180:** semilandmarks are not necessarily an alternative approach to landmarks; they can be used in addition to landmarks.
- **Line 183:** what kind or type of error are you referring to here?
- **Line 183:** spelling consistency of semilandmarks (or semi-landmarks).
- **Line 194:** the following sentence needs clarifications: “... using the samples coordinates as landmarks”.
- **Line 199:** precision is needed here – “the principal disadvantage to using landmarks in projectile points analysis”. In addition, reproducibility depends on the objects of study and the accuracy of landmarks’ descriptions.
- **Lines 207 and 208:** automating landmarks registration through templates is possible.
- **Line 234:** to avoid any confusion between the outline used in EFA and the “outline” used here to refer to a landmark configuration, you should add the number of landmarks used (between 30 and 42, as mentioned in line 245). This information should be presented earlier in the paragraph.

Comparisons (line 248)

- This section should be renamed results, as it presents the results of all the analyses.
- **Line 254:** elaborate more on the following statement: “... results were unsatisfactory...”.
- **Tables 2, 3 and 4:** LDA results – what do the values located at the mean/mean cell stand for?

- **Line 264:** this information can also be added in the paragraph between lines 233 and 247 as it is part of the methods used and helps the reader understand why you used different protocols (figures 8 and 9).
- **Figure 10:** what is a morphospace? The term needs to be defined for the reader; not all are familiar with GM terminology. “Morphospace” should be defined in the text before using it in a plot’s title. In addition, the legend needs to be adjusted; it covers part of the plot (top right quadrant).
- **Line 273:** PCA aims “... to reduce the dimensionality and identify patterns in the dataset.”
- **Lines 281-282:** for Pueblo Alto Side Notched results, you can add the value 0.86 in the text so that it is easier for the reader to find the information, just as you did in **line 325** (Ldk in tables 2-4).
- **Line 282:** when you say, “The results are better...”. In what sense? What do you mean by better? As compared to EFA? How did you come to that conclusion?
- **Line 287:** “Figure 11 shows the mean shapes for the selected projectile point types” it is essential to be precise that these are results of the EFA (just as you mention in the title of figure 11, page 18).
- **Lines 301, 302 and 304:** “points” as previously suggested, it is important to distinguish between points as in coordinates or as in projectile.
- **Lines 305-307:** in selecting the number of landmarks to use, you went through an analysis and presented the results between lines 305 and 307. It would be interesting to summarize these results in a table. And when you mention “higher numbers of points...”, how many points? And higher compared to what number?
- **Line 308:** can you elaborate on “similar dimensions of variation”?
- **Figure 12 (page 19):** projectile points in the plot are cropped out (right). Additionally, the legend covers part of the plot.
- **Lines 312 to 316:** the description needs to be more specific. You present a lot of information, some of which seem contradictory. For example, you end your paragraph by saying that the side-notches are still problematic. However, in the previous sentence, you say it is somewhat better for side-notched points. In addition, please mention that you are comparing EFA and semilandmarks approaches here.
- **Line 321:** are you referring to figure 3? As previously commented in the introduction, you need to discuss the preliminary analysis results that helped you determine which GM approach to apply on the Tonto Basin sample.
- **Lines 328:** “... this analysis struggled more so.” – What do you mean by “struggled”? Based on what results/or analysis?
- **Line 329:** be more precise in your statements; “they were previously the worst performing type when using EFA and semilandmarks approaches.”
- **Line 330:** How do you know that these two types are entangled? Based on what analysis?
- **Line 332:** “the final analysis” – which analysis? LDA? Be more precise.
- **Lines 335-336:** what do you mean by consistent? And more consistent and accurate, compared to what?
- **Line 336:** what type of enhancement? What do you propose to enhance results?
- **Line 337:** be more precise in your statement, for example, “... higher accuracy compared to EFA and semilandmarks”.
- **Line 340:** “my purposes”. It would be essential to remind the readers about your study objectives and be explicit.
- **Line 345:** precision – “projectile points proved the most effective when tested on the justice sample”.
- **Line 346:** you mention the PCA analysis here but do not display the plot; why not display the PCA results?
- **Line 350:** you mention that the purpose of the study is not to explore the patterns; however, since you are the first to apply GM and PCA analysis to this collection, it would be pertinent to explore the patterns further.
- **Line 357:** what do you mean by “flatten multidimensional space”? Remember that not all readers are familiar with multivariate statistics and GM analysis.
- **Lines 359-360:** “the results are messier”. What do you mean? Please explain this statement.

- **Lines 363-364:** need of references for the following statement “The side-notched points have a particularly large cluster of typical Hohokam side-notched points.”
- **Line 367:** “GM analysis better captures...” – better compared to what?
- **Figure 13 (page 21):** the figure needs to be referenced and discussed in the text. Describing all the figures/results used in the paper is crucial.
- **Figures 14 and 15:** it is important to add the information related to the sites of origin (information could be mentioned under each projectile point); otherwise, it limits the results’ interpretation.
- The results of the Tonto Basin projectile points analysis can be interpreted further. For example, what were the results of the PCA? Were you able to notice any patterns? Although according to figures 13 and 14, both network graphs based on morphometric distances show some type of clustering, it would have been interesting to see the PCA results. I think you should discuss the results of the PCA further in this section as they seem interesting and innovative for the field. Furthermore, it would be interesting to compare your results to the initial five clusters in the original report (line 142).

Conclusion (line 375)

- **Lines 377 to 379:** need to be clarified.
- **Line 381:** “... the most useful method”. Please mention the GM approach. Using which GM approach? EFA, landmarks or semilandmarks?
- **Lines 376 to 386:** this first paragraph summarizes all the methodological approaches you used (comparing three GM approaches) on the justice sample and a summary of the results. It seems like the paragraph is missing some critical information; for instance, you do not mention the semilandmarks approach. Please be more concise. Also, it does not remind the readers of your objectives and aims.
- **Lines 389-390:** the first part of the sentence mentions that classification was not possible due to sample size, but the second part states that it is possible to classify with these methods. This sentence needs clarification.
- **Line 391:** “... this chapter...”; is this paper part of a book?
- The whole section does not mention any references. Your findings must be contextualized, discussed and compared to other studies.