Int. J. Morphol.,
42(2):289-293, 2024.

An R Script to Estimate Which and How Many Landmarks Represent the Most Variable Parts of an Object

Un Script R para Estimar Cuáles y Cuántos Puntos de Referencia Representan las Partes más Variables de un Objeto

Ana Bucchi[1]; Andrés Neumann[2]; Fernanda Quevedo-Díaz[1] & Gabriel M. Fonseca[1]

BUCCHI, A.; NEUMANN, A.; QUEVEDO-DÍAZ, F. & FONSECA, G. M. An R script to estimate which and how many landmarks represent the most variable parts of an object. *Int. J. Morphol., 42(2)*:289-293, 2024.

SUMMARY: The study of the shape variation in geometric morphometrics has an important limitation known as the Pinocchio effect. The Pinocchio effect produces artifactual variances of the landmarks and implies that it is not possible to know the morphological change structure of an object, other than by dividing the landmark sets and then comparing them. This, however, involves making prior assumptions about the pattern of variation of an object. In this study, we provide a code in R to iterate over a complete set of landmarks and test all possible combinations of landmarks until deliver those landmarks associated with the largest to the smallest morphological changes. We tested this on a sample of 28 landmarks in 143 3D models of human skulls. The results indicated that this process can result in a pooled variance of a subset of landmarks that is an order of magnitude larger than that of several other regions of the skull. This method makes it possible to describe the pattern of variation of any 2D or 3D object represented by fixed landmarks, to distinguish the shape features that have more morphological dispersion, and to avoid any aprioristic assumptions about how the morphological changes of an object behave.

KEY WORDS: Geometric morphometrics; Local shape variation; Pinocchio effect; Procrustes superimposition; Variance.

INTRODUCTION

Geometric morphometrics have been one of the predominant approaches in studies of shape variations in recent decades. It consists of the study of the geometry of an object based on Cartesian coordinates of landmarks (Slice, 2007; Toro Ibacache *et al*. 2010) and has been used to study the most varied objects in archaeology and bioarchaeology (among many others, lithics, bones, cut marks, and ceramics (Wilczek *et al*., 2014; Courtenay *et al*., 2019; Bucchi *et al*., 2023). However, this methodology has a limitation in the study of shape variation, known as the Pinocchio effect (Chapman, 1990), which is derived from the use of the most commonly used superimposition method in geometric morphometrics (generalized least squares -GLS- Procrustes superimposition). The least-square criterion distributes the displacement of a few landmarks across all the other landmarks because it penalizes the landmarks with the greatest variation and distributes this variation among the other landmarks (Chapman, 1990; Klingenberg, 2020; Zelditch *et al*., 2004). This results in all landmark coordinate values after Procrustes superimposition having similar dispersion and, consequently, variation cannot be assigned to a particular morphological feature. This produces artifactual results of the variance and certainly represents a limitation in the study of morphological variations in general.

Although other estimates have been proposed to overcome this problem and achieve better calculations of landmark covariation, such as the maximum likelihood criterion (Theobald & Wuttke, 2006) or generalized resistant fit (GRF) (Rohlf & Slice, 1990; Walker, 2000), GLS Procrustes superimposition remain by far the most popular superimposition method.

The Pinocchio effect, coupled with the fact that shape changes cannot be attributed to particular landmarks, but to all tissues or the space included between landmarks (Klingenberg, 2013, 2020), has perhaps hinder the study of the effect of different landmark configurations on the variance of an object, or in which parts of an object the most or least variance is concentrated. However, knowing this would allow us to answer questions such as which morphological features vary more or less, how many and

[1] Centro de Investigación en Odontología Legal y Forense (CIO), Facultad de Odontología, Universidad de La Frontera, Temuco, Chile.
[2] Departamento de Ingeniería Mecánica, Facultad de Ingeniería y Ciencias, Universidad de La Frontera, Temuco, Chile.

Received: 2023-10-31 Accepted: 2023-12-30

which landmarks summarize the most and least variable area of an object, or is this pattern of variation the same between species, populations, sexes, or any other category of interest? In this study, we provided a code in R to address the Pinocchio effect problem and answer these questions. For this, we used the pattern of cranial variation calculated on a sample of 143 3D models of Chilean population described with an initial configuration of 28 landmarks. From these 28 fixed landmarks (complete configuration), we iteratively eliminated those that had associated anatomical regions of low dispersion until we knew the complete structure of the morphological changes. This method provides a sequence of landmarks from the one describing the least to the most variable regions and does not require any a priori assumptions about which anatomical regions have the greatest shape changes. Although a similar proposal to our have been made for Java (van der Linde & Houle, 2009), the absence of reproducible code in R, one of the most widely used programming languages in this area, has made it difficult to know which landmarks should be removed from the analysis.

MATERIAL AND METHOD

The 143 3D models belong to the Subactual collection of Santiago housed at the University of Chile. This collection is composed of people who died in the second half of the 20th century, came from peripheral areas of Santiago de Chile, and belonged to low socioeconomic strata(Meza-Escobar *et al*., 2023). Adult skulls with a good state of preservation and without obvious pathologies or trauma were selected. The 3D models were constructed using photogrammetry as described by Bucchi *et al.* (2023).

A configuration of 28 3D (m) landmarks (k) was digitized in the 143 (n) individuals using the AVIZO software v.9 (Visualization Sciences Group) (Fig. 1). The coordinate values were imported into R (R Core Team, 2021), and all calculations were performed using this software. The original database (prior to the Procrustes superimposition) can be found in Supplementary Material 1. Initially, the pooled variance of the full configuration of landmarks was calculated. The pooled variance provides a combined value

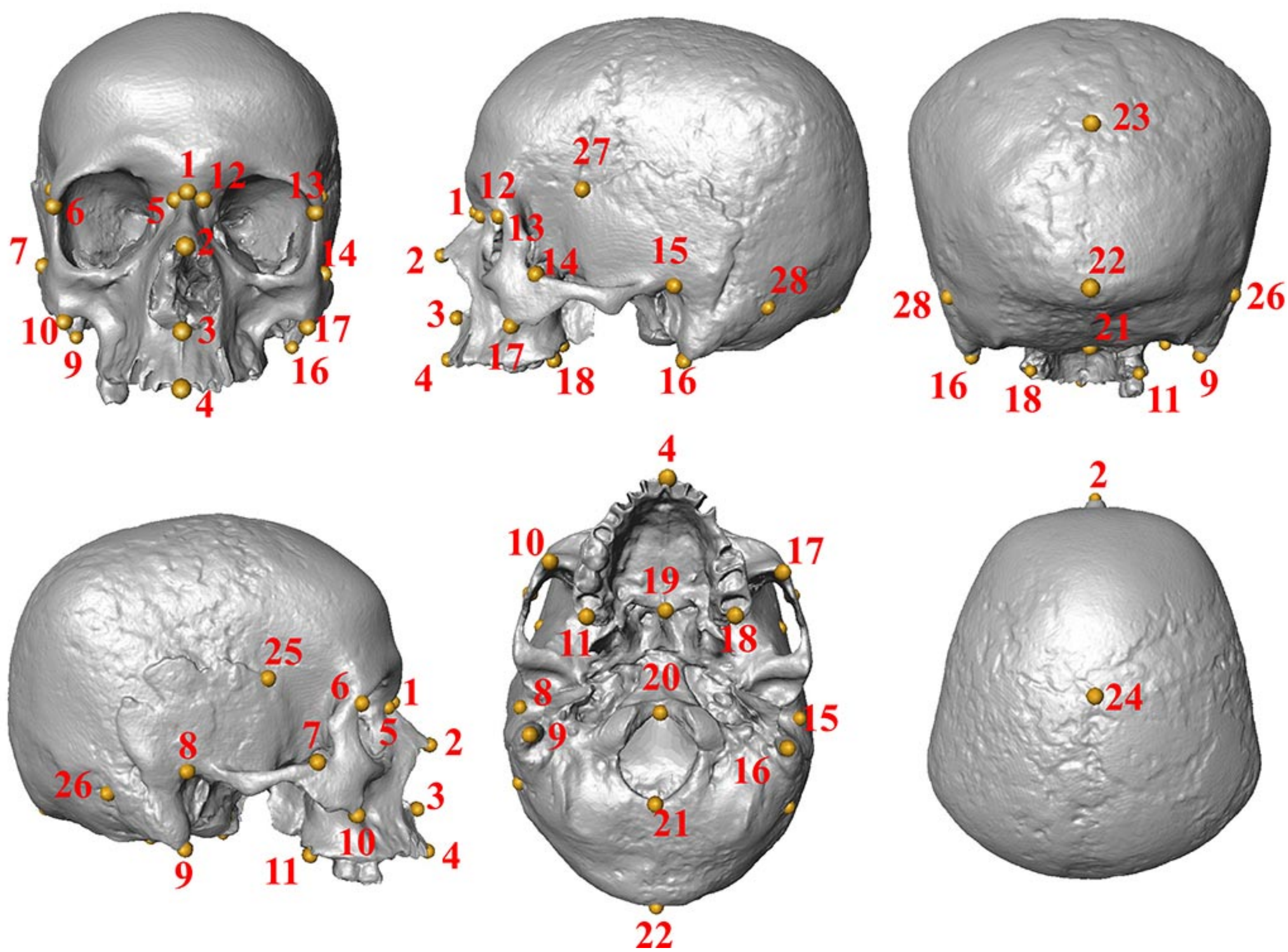

Fig. 1. Landmark configuration used in this study.

of the variance of all anatomical landmarks in a configuration and thus allows comparison of the degree of dispersion of the landmarks as a whole with the pooled variance of other landmark sets.

The R script for this process is provided in Supplementary Material 2.

The iterative process of landmark removal is as follows:

1.From the number of landmarks (k), all possible combinations of k-1 landmarks were extracted.
2.Each subset of landmarks was subject a generalized Procrustes analysis.
3.The pooled variance of each of the k-1 landmark set combinations was calculated.
4. The subset of k-1 landmarks that had the highest composite variance was selected and the landmark from the k-1 configuration that had not been present in that calculation was permanently removed from the analysis.
5. Return to point 1. The landmark removal process continued iteratively until the least number of landmarks possible for a Procrustes superimposition (k=3, iteration 25) was reached.

Therefore, we ended up a sequence of landmarks that were removed with the goal of eliminating regions associated with low dispersion, and the pooled variance associated with each set of landmarks at each iteration.

RESULTS

Table I shows the landmark removal process, and Figure 2 shows the change in the pooled variance at each iteration.

The pooled variance of the original landmark configuration (28 landmarks) was 6.17E-05, whereas the subset of landmarks with the highest pooled variance was an order of magnitude larger and was composed of four landmarks (landmarks 4, 10, 11, and 19) located in the region of the lower side of the maxilla and the lower margin of the right zygomatic (Fig. 1), with a pooled variance of 0.000802. The removal of landmark 10 (located on the inferior margin of the right zygomatic) from this subset of four landmarks caused a drop in the composite variance of the remaining landmarks (4, 11, and 19) (Table I).

Then, in decreasing order of pooled variance, the nasal anatomical region follows (landmarks 2 and 3), followed by the left masticatory region (landmarks 17 and 18), the landmarks of the upper central part of the facial skeleton, and bordering the neurocranium (1, 5, and 12). This is followed by regions of the neurocranium and face, whereas those of lesser dispersion are located, in general, entirely in the neurocranium. Landmarks 22, 23, and 13, which describe a region located between the external occipital protuberance, asterion, and the landmark located at the zygomaticofrontal suture, showed the lowest dispersion, as they were the first to be removed during iteration (Table I).

Table I. Pooled variance associated with each subset of landmarks. In bold are the set of landmarks associated with the highest pooled variance.

Iteration	Subset of landmark associated with the highest variation in each iteration	Landmark removed	Pooled variance
0	1:28	-	6.17E-05
1	1:21, 23:28	22	6.52E-05
2	1:21, 24:28	23	6.92E-05
3	1:12, 14:21, 24:28	13	7.32E-05
4	1:12, 14, 16:21, 24:28	15	7.76E-05
5	1:12, 14, 17:21, 24:28	16	8.23E-05
6	1:12, 14, 17:20, 24:28	21	8.77E-05
7	1:12, 14, 17:20, 24:27	28	9.36E-05
8	1:12, 14, 17:20, 24, 25, 27	26	0.000101
9	1:12, 14, 17:20, 25, 27	24	0.00011
10	1:7, 9:12, 14, 17:20, 25, 27	8	0.000123
11	1:7, 10:12, 14, 17:20, 25, 27	9	0.000135
12	1:7, 10:12, 17:20, 25, 27	14	0.000147
13	1:5, 7, 10:12, 17:20, 25, 27	6	0.000161
14	1:5, 7, 10:12, 17:20, 27	25	0.000174
15	1:5, 10:12, 17:20, 27	7	0.000193
16	1:5, 10:12, 17:19, 27	20	0.000214
17	1:5, 10:12, 17:19	27	0.000248
18	2:5, 10:12, 17:19	1	0.00029
19	2:4, 10:12, 17:19	5	0.000331
20	2:4, 10, 11, 17:19	12	0.000382
21	2:4, 10, 11, 18, 19	17	0.000454
22	2:4, 10, 11, 19	18	0.000549
23	3, 4, 10, 11, 19	2	0.000669
24	**4, 10, 11, 19**	**3**	**0.000802**
25	4, 11, 19	10	0.000718

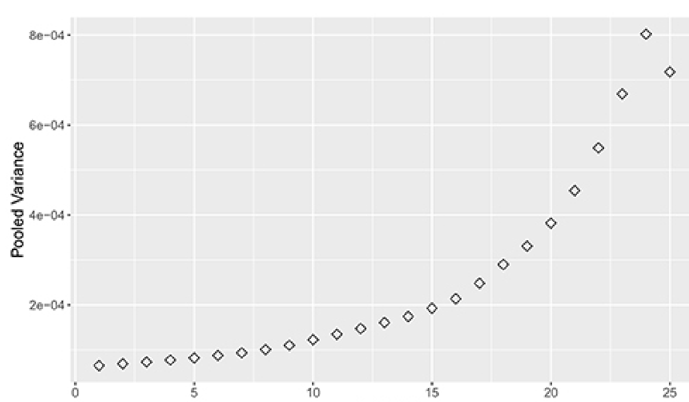

Fig. 2. Change in pooled variance between iterations 1 and 25.

DISCUSSION

In this study, we propose an R script to estimate the variation pattern of an object through geometric morphometrics that addresses the Pinocchio effect constraint. This method allowed us to compare the morphological variation between all anatomical regions included in each of the possible combinations landmarks and to estimate the landmarks associated to the largest and smallest shape changes. It is necessary to mention that this procedure does not eliminate the Pinocchio effect present in all configurations of landmarks and subject to the same Procrustes superimposition, but does propose a way to greatly reduce the influence of this problem by providing a sequence of landmarks describing the effect on the pooled variance of including or removing a new anatomical region in the analysis. To our knowledge, this is the first proposal that provide the code in R to solve this problem efficiently.

Importantly, it also avoids aprioristic assumptions regarding the structure of variation of an object. This is achieved by comparing the pooled variance of all possible combinations of landmarks, which would be impractical if performed individually. In the example of this paper this allowed us to observe that the greatest morphological change among individuals in the sample is in the right masticatory region, followed by the nasal area, and then the left masticatory region, and not in the skull base, which is usually considered a region of important cranial variation in primates, including humans (Bastir *et al*., 2010; Bennett & Goswami, 2012; Neaux *et al*., 2019). We believe that this result makes sense because the study sample came from a collection with widespread tooth loss and significant bone resorption in many individuals (Urzua *et al*., 2009). These results are consistent with Eyquem *et al*. (2019), who, in a sample including the same collection studied here, found that masticatory force influences maxilla shape in a heterogeneous and undetermined manner. This is also consistent with studies indicating that subsistence mode affects craniofacial variation (Von Cramon-Taubadel, 2011; Noback & Harvati, 2015). It is also interesting that the right side of the maxilla and zygomatic have the largest morphological changes and that the pooled variance is larger here than on the left side because it is described in the literature that people tend to chew more with the right side (Nissan *et al*., 2004; Khamnei *et al*., 2019), which indicates that this method of studying variance manages to discriminate relatively small shape changes. We therefore believe that the major variation in cranial shape in this sample is functionally driven.

CONCLUSION

The procedure proposed here makes it possible to describe the variation structure of an object by calculating how many and which landmarks of an initial configuration represent the most and least variable morphological features. Thus, it avoids the limitation of the Pinocchio effect resulting from the use of the Procrustes superimposition in geometric morphometrics. In addition, it describes where the major morphological changes of an object are located, with no previous assumptions about how the variation in parts of an object behaves.

ACKNOWLEDGEMENTS

We would like to thank the Faculty of Social Sciences at the Universidad de Chile and the curator of the Subactual de Santiago collection, Nicole Barreaux, for access to the collection. This work was funded by the Agencia Nacional de Investigación y Desarrollo (ANID) project, Fondecyt Postdoctoral 3210026 (Chile).

BUCCHI, A.; NEUMANN, A.; QUEVEDO-DÍAZ, F. & FONSECA, G. M. Un script R para estimar cuáles y cuántos puntos de referencia representan las partes más variables de un objeto. *Int. J. Morphol., 42(2)*:289-293, 2024.

RESUMEN: El estudio de la variación de la forma en morfometría geométrica tiene una limitación importante conocida como efecto Pinocho. El efecto Pinocho produce variaciones artefactos de los puntos de referencia e implica que no es posible conocer la estructura del cambio morfológico de un objeto, salvo dividiendo los conjuntos de puntos de referencia y luego comparándolos. Sin embargo, esto implica hacer suposiciones previas sobre el patrón de variación de un objeto. En este estudio, proporcionamos un código en R para iterar sobre un conjunto completo de puntos de referencia y probar todas las combinaciones posibles de puntos de referencia hasta entregar aquellos puntos de referencia asociados con los cambios morfológicos más grandes a los más pequeños. Probamos esto en una muestra de 28 puntos de referencia en 143 modelos 3D de cráneos humanos. Los resultados indicaron que este proceso puede dar como resultado una variación combinada de un subconjunto de puntos de referencia que es un orden de magnitud mayor que el de varias otras regiones del cráneo. Este método permite describir el patrón de variación de cualquier objeto 2D o 3D representado por puntos de referencia fijos, distinguir las características de forma que tienen más dispersión morfológica y evitar suposiciones apriorísticas sobre cómo se comportan los cambios morfológicos de un objeto.

PALABRAS CLAVE: Morfometría geométrica; Variación de forma local; Efecto Pinocho; Superposición de Procusto; Diferencia.

REFERENCES

Bastir, M.; Rosas, A.; Stringer, C.; Manuel Cuétara, J.; Kruszynski, R.; Weber, G. W.; Ross, C. F. & Ravosa, M. J. Effects of brain and facial size on basicranial form in human and primate evolution. *J. Hum. Evol., 58(5)*:424-431, 2010.

Bennett, C. V. & Goswami, A. Morphometric analysis of cranial shape in fossil and recent euprimates. *Anat. Res. Int., 2012*:1-7, 2012.

Bucchi, A.; Püschel, T. A. & Lorenzo, C. Modularity of the wrist in extant hominines. *Anthropol. Anz., 80(1)*:1-12, 2023.

Chapman, R. E. Conventional procrustes approaches ralph. *Syst. Rev., 41(3)*:392, 1990.

Courtenay, L. A.; Yravedra, J.; Mate-González, M. Á.; Aramendi, J. & González-Aguilera, D. 3D analysis of cut marks using a new geometric morphometric methodological approach. *Archaeol. Anthropol. Sci., 11(2)*:651-65, 2019.

Eyquem, A. P.; Kuzminsky, S. C.; Aguilera, J.; Astudillo, W. & Toro-Ibacache, V. Normal and altered masticatory load impact on the range of craniofacial shape variation: An analysis of pre-Hispanic and modern populations of the American Southern Cone. *PLoS One, 14(12)*:e0225369, 2019.

Khamnei, S.; Sadat-Ebrahimi, S. R.; Salarilak, S.; Oskoee, S. S.; Houshyar, Y.; Shakouri, S. K.; Salekzamani, Y. & Zamanlu, M. Manifestation of hemispheric laterality in chewing side preference and handedness. *Bioimpacts, 9(3)*:189-93, 2019.

Klingenberg, C. P. How exactly did the nose get that long? A critical rethinking of the Pinocchio effect and how shape changes relate to landmarks. *Evol. Biol., 48(1)*:115-27, 2020.

Klingenberg, C. P. Visualizations in geometric morphometrics: How to read and how to make graphs showing shape changes. *Hystrix, 24(1)*:15-24, 2013.

Meza-Escobar, O.; Galimany, J.; González-Oyarce, R. & Barreaux Höpfl, N. The Colección Osteológica Subactual de Santiago: Origin and current state of a documented skeletal collection from Chile, Latin America. *Forensic Sci. Int., 3(1)*:80-93, 2023.

Neaux, D.; Wroe, S.; Ledogar, J. A.; Heins Ledogar, S. & Sansalone, G. Morphological integration affects the evolution of midline cranial base, lateral basicranium, and face across primates. *Am. J. Phys. Anthropol., 170(1)*:37-47, 2019.

Nissan, J.; Gross, M. D.; Shifman, A.; Tzadok, L. & Assif, D. Chewing side preference as a type of hemispheric laterality. *J. Oral. Rehabil., 31(5)*:412-6, 2004.

Noback, M. L. & Harvati, K. The contribution of subsistence to global human cranial variation. *J. Hum. Evol., 80*:34-50, 2015.

Rohlf, F. J. & Slice, D. Extensions of the procrustes method for the optimal superimposition of landmarks. *Syst. Zool, 39(391)*:40-59, 1990.

Slice, D. E. Geometric Morphometrics. *Annu. Rev. Anthropol., 36*:261-81, 2007.

Theobald, D. L. & Wuttke, D. S. Empirical bayes hierarchical models for regularizing maximum likelihood estimation in the matrix Gaussian procrustes problem. *Proc. Natl. Acad. Sci. U. S. A., 103(49)*:18521-7, 2006.

Toro Ibacache, M. V., Manríquez Soto, G., & Suazo Galdames, I. Geometric morphometry and the biologic shapes study: from the descriptive morphology to the quantitative morphology. *Int. J. Morphol., 28(4)*:977-90, 2010.

Urzua, I.; Huberman, J.; Delgado, I.; Pacheco, A. & Retamal, R. Prevalencia de caries y pérdida de dientes de una población adulta Chilena nacida en el siglo XIX. *Rev. Clin. Periodoncia Implantol. Rehabil. Oral, 2(3)*:175-8, 2009.

van der Linde, K. & Houle, D. Inferring the nature of allometry from geometric data. *Evol. Biol., 36(3)*:311-22, 2009.

Von Cramon-Taubadel, N. Global human mandibular variation reflects differences in agricultural and hunter-gatherer subsistence strategies. *Proc. Natl. Acad. Sci. U. S. A., 108(49)*:19546-51, 2011.

Walker, J. A. Ability of geometric morphometric methods to estimate a known covariance matrix. *Syst. Biol., 49(4)*:686-96, 2000.

Wilczek, J.; Monna, F.; Barral, P.; Burlet, L.; Chateau, C. & Navarro, N. Morphometrics of Second Iron Age ceramics - strengths, weaknesses, and comparison with traditional typology. *J. Archaeol. Sci., 50(1)*:39-50, 2014.

Zelditch, M. L.; Swiderski, D. L.; Sheets, H. D. & Fink, W. L. *Geometric Morphometrics for Biologists: A Primer.* Amsterdam, Elsevier Academic Press, 2004.

Corresponding author:
Ana Bucchi
Universidad de La Frontera
Avenida Francisco Salazar 01145
Temuco
CHILE

E-mail: ana.bucchi@ufrontera.cl