

# RPS (Resistant Procrustes Software): una herramienta novedosa para el Análisis Morfométrico Resistente

Ferraggine, Viviana E.  
Departamento de Computación y Sistemas  
Facultad de Ciencias Exactas, UNCPBA  
Paraje Arroyo Seco, Tandil, Bs.As., Argentina  
[vferra@exa.unicen.edu.ar](mailto:vferra@exa.unicen.edu.ar)

Torcida, Sebastián  
Departamento de Matemática  
Facultad de Ciencias Exactas, UNCPBA  
Paraje Arroyo Seco, Tandil, Bs.As., Argentina

Perez, S. Ivan  
División Antropología,  
Facultad de Ciencias Naturales y Museo, UNLP. CONICET.  
La Plata, Bs.As., Argentina

## Abstract

*En este trabajo se presenta RPS (Resistant Procrustes Software), un paquete integrado para el análisis resistente de la forma de objetos representados por configuraciones de landmarks en 2D y 3D. RPS es una herramienta de código abierto que posee una interfaz de usuario bien diseñada, ofreciendo un ambiente amigable y personalizable. Su performance ha sido testada en configuraciones de hasta 100 landmarks, mostrando tiempos de respuesta aceptables a pesar de la complejidad computacional de los algoritmos resistentes.*

*La principales funcionalidades de RPS son la superposición Procrustes resistente, el cálculo de una distancia y un ordenamiento resistentes y la posibilidad de visualizar en 2D o 3D los resultados anteriores.*

## 1. Introduction

Para estudiar la forma de un objeto la *morfometría geométrica* (MG) utiliza las coordenadas cartesianas en 2D ó en 3D de puntos anatómicos denominados *landmarks*; éstos puntos se caracterizan por su significado funcional o geométrico y deben ser

claramente identificados por el especialista en todos los objetos cuya forma se quiere analizar y/o comparar (homología). Así, la MG representa una forma mediante una *configuración* de landmarks: una matriz  $n \times p$  en la que se almacenan las coordenadas cartesianas de  $n$  landmarks en  $p$  (2 ó 3) dimensiones. El conjunto de todas las configuraciones de landmarks correspondientes a los diferentes objetos en estudio se denomina *dataset*.

Estudios biológicos y antropológicos utilizan habitualmente landmarks para captar información sobre la forma de estructuras fenotípicas y analizar su variación entre individuos, poblaciones y/o especies [14, 10, 1]. Recientemente, el análisis de la variación de la forma se ha convertido en un tema de interés en campos tan diversos como la ecología y los estudios evolutivos [6, 5, 4]; al intentar extraer información sobre la forma de diferentes configuraciones de landmarks el mayor desafío es establecer un marco de análisis adecuado [12, 20, 21].

El primer paso para comparar la forma de dos o más configuraciones de landmarks es superponerlas; para ello se utilizan las llamadas *superposiciones Procrustes* (SP) que determinan y aplican sobre cada configuración las operaciones geométricas de *traslación*, *rotación/reflexión* y *cambio de escala uniforme* a fin de minimizar las diferencias en la *posición*, la *orientación* y el *tamaño*

(respectivamente) entre las configuraciones [1][5]. Para completar el estudio, con las coordenadas que resultan de una SP se suelen realizar diversos análisis estadísticos multivariados.

La superposición Procrustes *por cuadrados mínimos* (GlsP, por sus siglas en inglés) es el método estándar para analizar la variación de forma empleando landmarks [12, 15, 23]. Sin embargo, es ampliamente sabido que GlsP puede arrojar resultados engañosos cuando las diferencias de forma se concentran en menos del 50% de los landmarks. La superposición Procrustes *resistente* (GrP) es tal vez la alternativa más elegante y eficiente [19, 21]; este método utiliza la técnica de las *medianas repetidas* para estimar esos cambios de escala, traslaciones y rotaciones/reflexiones que logran un ajuste óptimo entre las configuraciones. Si las diferencias de forma se localizan en menos del 50% de los landmarks, luego de aplicar una GrP los landmarks que no cambian se superponen perfectamente; las diferencias quedan entonces a la vista a través del desajuste en los landmarks que sí cambiaron. Esta característica del GrP es atractiva e intuitiva geoméricamente, pero el método no es muy utilizado probablemente por la ausencia de una herramienta de software libre y amigable que la implemente.

RPS (*resistant Procrustes software*) está basado en [21], donde se presenta un enfoque integral para realizar un análisis de forma resistente de datasets en 2D o 3D. Específicamente, RPS permite: 1) realizar una GrP; 2) calcular la *matriz de distancias resistente* entre las configuraciones; y 3) obtener un *ordenamiento resistente*, mediante un *escalamiento multidimensional*. Para ofrecer un análisis de forma más completo, RPS también tiene implementadas las versiones por cuadrados mínimos de las tres funcionalidades nombradas.

El rasgo distintivo de RPS es entonces implementar un marco integrado y compacto para realizar un análisis de forma *resistente* con conjuntos de configuraciones de landmarks en 2D y 3D. Además, su diseño modular hace posible integrarlo con otros tipos de análisis morfométricos y cálculos estadísticos multivariados. Así, RPS trae una perspectiva novedosa que puede enriquecer estudios evolutivos y del desarrollo en organismos.

A continuación se ofrece un breve panorama de las funcionalidades de RPS.

## 2. Superposición Procrustes (SP)

Para comparar la forma de varios objetos utilizando landmarks es necesario identificar un sistema de coordenadas común. El propósito de los métodos de superposición es entonces encontrar el marco de referencia adecuado, porque las diferencias de forma resultantes dependen fuertemente del criterio de

superposición seleccionado. Una superposición útil debe remover las diferencias de posición, orientación y escala entre las configuraciones, porque tales diferencias no afectan la forma: éstas son las superposiciones Procrustes. Como se mencionó anteriormente, dos versiones de ellas se implementan en RPS: la novedosa GrP [21] y la conocida GlsP [7, 15].

### 2.1. Superposición Procrustes resistente (GrP)

Intuitivamente, una superposición resistente de dos configuraciones de landmarks se logra cuando los landmarks que no muestran ninguna diferencia se superponen perfectamente. A diferencia de la GlsP, la GrP no resulta de minimizar un criterio matemático explícito; en cambio, se basa en un cálculo de medianas repetidas [19, 21]. Esta técnica posee el máximo *breakdownvalue* [19]: casi el 50%. Esto significa que cada vez que las diferencias de forma se localizan en menos del 50% de los landmarks, se logra una superposición perfecta de aquellos landmarks que no tienen diferencias. Y las diferencias de forma quedan evidenciadas en la falta de ajuste y los grandes residuales de los landmarks (o grupos de landmarks) que sí difieren. Es así que en tales casos el método *resiste* la deformación presente en unos pocos landmarks (aunque se trate de una *gran* deformación) y la superposición sigue la tendencia mayoritaria de no cambio.

Vale la pena mencionar que cuando las diferencias de forma se encuentran en más del 50% de los landmarks, su medición y/o su identificación empieza a ser poco clara. La implementación de GrP sigue los resultados de [21]; el procedimiento ajusta iterativamente cada configuración de landmarks a una configuración *pivote* o *consenso* resistente cuyos landmarks son, respectivamente, *las medianas espaciales* de los landmarks correspondientes; este *consenso resistente* se obtiene mediante un algoritmo conocido [22]. Luego de una GrP las diferencias de forma entre las configuraciones suelen quedar representadas de manera más precisa y clara; esto facilita la interpretación biológica/evolutiva de la variación de la forma en muchas situaciones reales.

### 2.2. Superposición Procrustes por Cuadrados Mínimos (GlsP)

Como se dijo, la GlsP o superposición Procrustes clásica es el criterio de superposición estándar en la morfometría geométrica [1]. Este método minimiza la suma de los cuadrados de las distancias Euclídeas entre las coordenadas cartesianas de los landmarks correspondientes después la superposición;

iterativamente, se ajusta cada configuración a la una configuración *promedio* [7] y así se logra el ajuste óptimo en términos de cuadrados mínimos. Es ampliamente sabido que este método produce resultados engañosos cuando tienen lugar variaciones de forma localizadas, porque el método distribuye tales variaciones entre todos los landmarks.

### 2.3. Distancias Morfométricas

Cuando se realiza una SP se busca medir adecuadamente las diferencias de forma entre dos configuraciones cualesquiera; es necesario entonces calcular una *distancia* que cuantifique esas diferencias y almacenar los resultados en una matriz de distancias. Esta matriz puede posteriormente utilizarse para obtener un *ordenamiento*, o para llevar a cabo un análisis multivariado. Para cada estrategia de superposición, RPS calcula una matriz de distancia coherente.

Después de una GrP, una *distancia resistente* ( $rD$ ) coherente entre cualquier dos configuraciones es la dada por la suma de las distancias Euclídeas no cuadradas entre los landmarks correspondientes [21]; ésta distancia se puede calcular en RPS. Además se puede hallar la distancia por cuadrados de mínimos ( $lsD$ ), luego de una GlsP; la magnitud de la diferencia de forma entre dos configuraciones de landmarks se mide en este caso con la raíz cuadrada de la suma de distancias Euclídeas cuadradas entre landmarks correspondientes.

### 2.4. Ordenamientos

Cuando se realiza una SP es frecuente utilizar una matriz de distancias para obtener un gráfico en 2D o en 3D que facilite la visualización de las diferencias de forma: éste gráfico se denomina *ordenamiento*. RPS implementa el escalamiento multidimensional universal (UMDS) propuesto en [3], que incluye el UMDS por cuadrados mínimos ( $lsUMDS$ ) y el correspondiente UMDS resistente ( $rUMDS$ ).

## 3. RPS (Resistant Procrustes Software)

RPS fue específicamente diseñado para ofrecer al usuario un conjunto de herramientas analíticas y de visualización que permitan realizar análisis resistente de la forma. Todas las funcionalidades de RPS pueden invocarse desde la interfaz de usuario, que está organizada como una barra de herramientas con menús emergentes y consta de dos áreas principales: un área ubicada a la izquierda de la interfaz, en la que se aprecia la lista de elementos de cada proyecto, y el área de

informes y gráficos ubicada a la derecha de la anterior (Fig. 1).

Los elementos que pueden ser encontrados dentro de un proyecto (ver por ej. en la Fig. 1 el proyecto “Test1”) son los dataset a ser analizados, los resultados de los diferentes análisis, las matrices de distancias y los ordenamientos, con los respectivos gráficos en el área central (no se grafican las matrices de distancias).

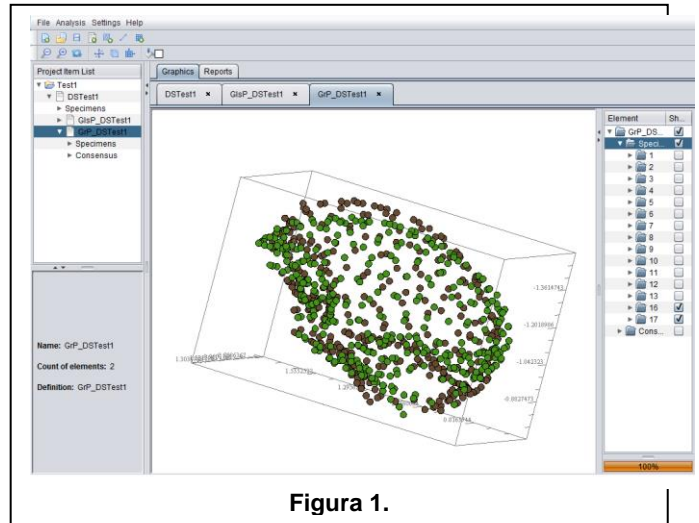


Figura 1.

### 3.1. Área PROJECT ITEM LIST

RPS organiza los conjuntos de datos y los análisis en proyectos, buscando que éstos sean autónomos, y este área ofrece una visión general de los elementos de cada proyecto. Como se muestra en la Figura 2, una estructura de árbol reserva el primer nivel para el nombre del proyecto (por ej., “Test1”). El segundo nivel corresponde al nombre del dataset a analizar (por ej., “DSTest1”). El tercer nivel incluye: la lista de configuraciones de landmarks bajo la etiqueta “Specimens”; la lista de análisis realizados con cada dataset; la configuración consenso “Consensus”, que es opcional, y el dataset que resulta de cada análisis realizado. Los distintos análisis pueden ser distinguidos en RPS por los prefijos automáticos en sus nombres, como se muestra en la Fig. 2:

- GrP\_x** identifica un análisis Procrustes resistente,
- GlsP\_x** identifica un análisis Procrustes por CM,
- rD\_x** identifica al cálculo de una distancia resistente,
- lsD\_x** identifica al cálculo de una distancia por CM,
- 2D\_x** identifica análisis u ordenamiento en 2D,
- 3D\_x** identifica análisis u ordenamiento en 3D,
- rUMDS\_x** identifica un ordenamiento resistente,

*IsUMDS\_x* identifica un ordenamiento por CM.

Asimismo es posible analizar simultáneamente y en un mismo proyecto a diferentes datasets, y compararlos visualmente los resultados.

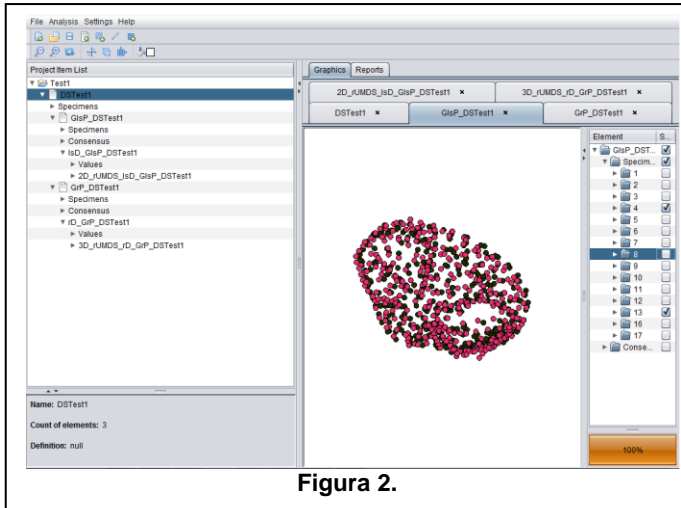


Figura 2.

### 3.2. Área GRAPHICS y REPORTS

Como se aprecia en la Fig. 3, esta área cubre la mayor parte de la ventana de la aplicación. Consta de dos pestañas principales independientes que aparecen en la parte superior: una de gráficos (Graphics) y otra de informes y resultados (Reports). A su vez, la pestaña de gráficos está organizada en sub-pestañas que pueden corresponder a una representación gráfica o a la visualización en 2D o 3D de:

1. un nuevo dataset agregado al proyecto;
2. el resultado de una superposición particular;
3. el resultado de un ordenamiento.

Cada sub-pestaña tiene un marco interactivo donde las características predeterminadas del gráfico pueden ser editadas y personalizadas por el usuario; dichas características incluyen la posibilidad de colocar o quitar los ejes o el borde del gráfico, fijar las escalas, mostrar u ocultar configuraciones o landmarks específicos, etc.

Cuando una sub-pestaña gráfica está activa, la columna ubicada a la derecha del gráfico con las pestañas rotuladas “Element” y “Show” lista las configuraciones exhibidas en dicho gráfico, en el mismo orden que en el dataset original; el usuario puede mostrar u ocultar todas o algunas de las configuraciones o sus landmarks en el gráfico correspondiente. En la parte inferior de esta

columna una barra indica el porcentaje estimado de progreso del análisis que se está ejecutando; esto es útil porque los cálculos resistentes tienen una mayor complejidad computacional que sus análogos por cuadrados mínimos. La pestaña de resultados “Reports” (Fig. 4) ofrece un registro detallado de cada uno de los análisis y cálculos realizados, que se muestran en formato de texto plano; desde la opción de menú “Files” es posible exportar los resultados con el formato de un dataset (para poder usarlo en análisis posteriores) y también en formato texto desde la pestaña (por ej. las matrices de distancias y las coordenadas de los ordenamientos).

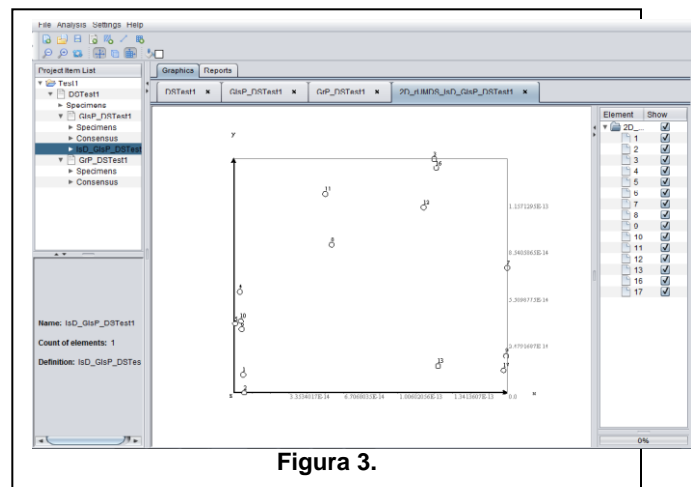


Figura 3.

### 3.3. Opciones de Archivos

El menú pop-up “Files” de la aplicación incluye todos los comandos que realizan acciones en los proyectos, tales como: crear uno nuevo, abrir uno existente y guardar un proyecto; también es posible incorporar un nuevo dataset a un proyecto, exportar gráficos y resultados (ya sea como imágenes o como datasets). Vale la pena mencionar aquí que los software morfométricos existentes utilizan en general tres formatos de archivo diferentes: TPS, NTS y Morphologika [10, 12, 13], con extensiones .tps, .nts y .txt respectivamente. En RPS se brinda la posibilidad de trabajar indistintamente con cualquiera de esos formatos.

Cuando un archivo nuevo con un dataset se carga correctamente, el dataset correspondiente se visualiza automáticamente en el área de gráficos.

### 3.4. Opciones de Análisis

En el menú pop-up “Analysis” se muestran todas las herramientas de análisis que posee RPS para el análisis de la forma. Para efectuar cualquier tipo de análisis es necesario primero seleccionar el dataset; una vez que se ejecuta el algoritmo correspondiente, el resultado aparece como un nuevo ítem en el subnivel correspondiente de la estructura de árbol del proyecto, y el resultado se muestra en el área gráfica.

La matriz de distancias entre todos los pares de configuraciones de un dataset se puede calcular en RPS, como se mencionó anteriormente, por ambos métodos: distancias por cuadrados mínimos (lsD) y distancias resistentes (rD) [21]. Este cálculo también requiere seleccionar el dataset correspondiente; al igual que el resto de los resultados, las distancias calculadas se agregan al subnivel correspondiente del árbol.

Finalmente, una matriz de distancia puede utilizarse para calcular ordenamientos en 2D o en 3D seleccionando el correspondiente comando; el resultado del ordenamiento se muestra luego como un subnivel de dicha matriz de la distancia en la estructura de árbol, y el ordenamiento se visualiza en el área gráfica.

Los resultados y cálculos asociados a los análisis, distancias y ordenamientos se incluyen automáticamente en la pestaña "Reports". (Fig.4)

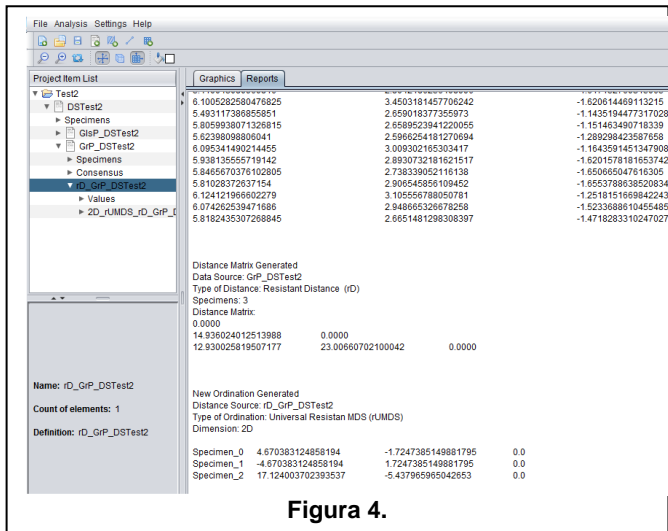


Figura 4.

#### 4. Ejemplo

Para mostrar el funcionamiento de RPS en un caso real, se utilizó un dataset en 2D ( $p=2$ ) que contiene información sobre la forma del cráneo de 160 especies de aves, empleando para ello  $n=12$  landmarks. La Fig. 5 muestra el perfil genérico del cráneo de éstas aves, con la posición de los landmarks utilizados para el análisis de la forma.

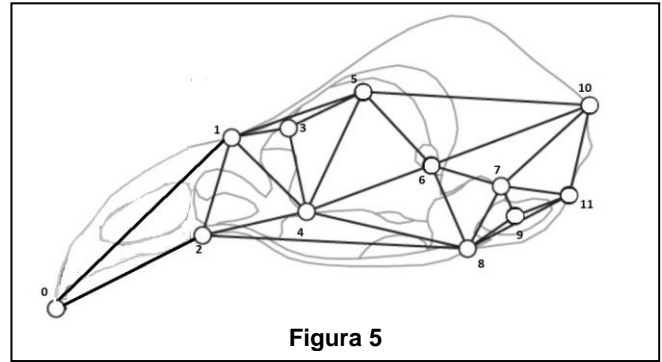


Figura 5

La Fig. 6 muestra la visualización inicial de las 160 especies en el área de gráficos; en el área izquierda de los proyectos se lista el dataset (Birds) y cada una de las especies con sus respectivos landmarks (numerados de 0 a 11), tal como figuran en el archivo.

Luego de realizar ambas superposiciones (GlsP y GrP) se calcularon las correspondientes matrices de distancias y se exploraron los respectivos ordenamientos (CM y resistente), intentando identificar especies en las que los resultados de una y otra superposición fueran sustancialmente distintos. Uno de éstos casos es el de las especies *Eurypyga* y *Balearica pavonina* (Figuras 7 y 8). Puede apreciarse a simple vista en la superposición GrP (Fig. 7) que casi toda la diferencia de forma se concentra en el landmark 0 (la punta del pico); esto puede corroborarse en la Tabla 1, donde se observa que la contribución del landmark 0 a la distancia resistente total entre ambas especies es de aproximadamente el 84%. En cambio, la superposición GlSP (Fig. 8) sugiere que diferencias de forma importante se detectan (además de en el landmark 0) en los landmarks 1, 2, 3, 5 y 12; ésto se corrobora también en las contribuciones de la Tabla 1. Así, el análisis resistente permitiría concluir que ambas especies difieren esencialmente en la punta del pico, mientras que el análisis por cuadrados mínimos indicaría que la diferencia de forma se manifiesta en varios landmarks.

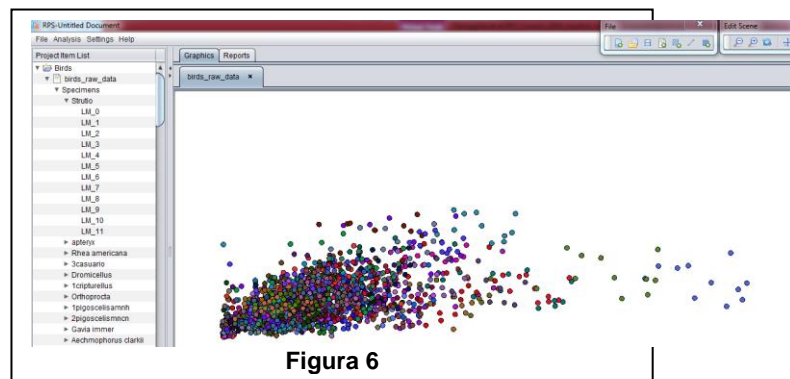
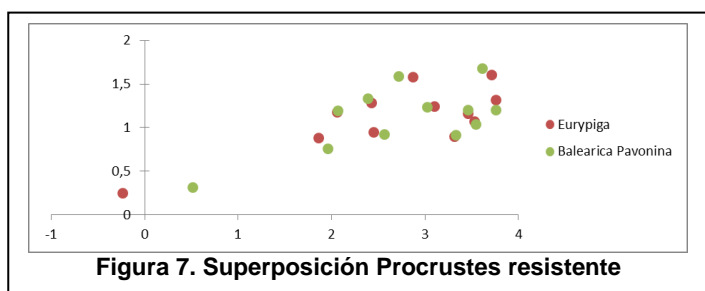


Figura 6

**Tabla 1.**

Landmark	CM	Robusto
0	19,209	84,387
1	10,797	0,086
2	10,791	3,678
3	10,696	0,562
4	0,650	1,853
5	13,679	3,542
6	1,392	0,813
7	2,496	0,175
8	1,762	0,083
9	6,852	0,208
10	5,534	2,463
11	16,142	2,150
	100,000	100,000



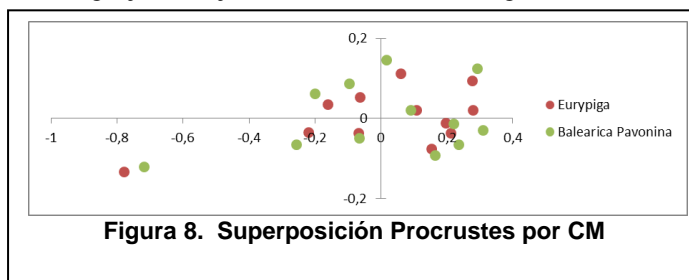
## 5. Comentarios Finales y Trabajos Futuros

En las últimas décadas ha tenido lugar un importante desarrollo de los métodos de morfometría geométrica basados en landmarks; éstos métodos se han convertido en estándares en campos como la biología y la antropología. Sin embargo, durante todo este período la mayoría de los softwares morfométricos (MorphoJ, Geomorph y la serie TPS [8, 2, 13], y otros como los presentados en [11]) han implementado principalmente las técnicas de cuadrados mínimos: la superposición Procrustes clásica y varias herramientas de análisis multivariado de ese tipo; por otro lado, la serie TPS [15] en particular no se trata de un software integrado sino de muchas aplicaciones dispersas, y MorphoJ [8] carece de una interfaz gráfica 3D interactiva que permita visualizar desde diferentes ángulos y con diferentes grados de acercamiento los resultados de las superposiciones. En este contexto, la ausencia de una herramienta integrada y versátil que incluya el enfoque resistente es evidente.

Este artículo presenta entonces la herramienta RPS, que se caracteriza por la implementación de la

superposición resistente inspirada en [21]; además, RPS posee una interfaz de gráficos 2D y 3D interactiva que otros paquetes no ofrecen, e implementa una plataforma integrada y compacta para llevar a cabo un análisis resistente de la forma de objetos representados por configuraciones de landmarks en 2D y 3D. RPS ofrece un entorno amigable y personalizable, y su rendimiento es razonablemente rápido y eficiente a pesar de la complejidad computacional conocida de los algoritmos resistentes. En síntesis, RPS aporta una perspectiva novedosa que podría enriquecer los estudios evolutivos y del desarrollo.

Como líneas de trabajo en lo inmediato, se planea integrar a RPS un motor de bases de datos (BD) que permita compartir de forma más centralizada y ordenada los resultados, guardarlos y combinarlos con otros análisis y ordenamientos básicos que las BD ofrecen. Actualmente se está trabajando en programar las funcionalidades resistentes en el lenguaje estadístico-matemático R, teniendo en cuenta que una gran parte de la comunidad morfométrica utiliza y hace desarrollos en este lenguaje. El objetivo final es también integrar a RPS



con estos paquetes para generalizar y flexibilizar su persistencia, sin afectar la autonomía que la herramienta posee actualmente.

## 6. Agradecimientos

Los autores agradecen al Sr. Leandro Aristide por su colaboración en la generación de las Figuras 1 a 4, y al Dr. Jesús Marugán-Lobón por sus alentadores comentarios.

## 7. Referencias

- [1] Adams, D.C., Rohlf, F.J., & Slice, D.E. (2013). A field comes of age: Geometric morphometrics in the twenty first century. *Hystrix*, 24(1), 7–14.
- [2] Adams, D.C., Otárola-Castillo, E. 2013. geomorph: an R package for the collection and analysis of geometric morphometric shape data. *Methods in Ecology and Evolution*, 4, 393–399
- [3] Agarwal, A., Phillips, J.M., & Venkatasubramanian, S. (2010). Universal multidimensional scaling. In *KDD'10*:

Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1149–1158). New York: ACM.

- [4] Araújo, M.S., Perez, S.I., Magazoni, M.J.C. & Petry, A.C. (2014). Body size and allometric shape variation in the molly *Poecilia vivipara* along a gradient of salinity and predation. *BMC Evolutionary Biology* 14, 251.
- [5] Collyer, M.L., & Adams, D.C. (2013). Phenotypic trajectory analysis: Comparison of shape change patterns in evolution and ecology. *Hystrix*, 24, 75-83
- [6] Drake, A.G., & Klingenberg, C.P. (2008). The pace of morphological change: historical transformation of skull shape in St Bernard dogs. *Proceedings of the Royal Society of London B*, 275, 71–76.
- [7] Gower, J.C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1), 33–51.
- [8] Klingenberg, C.P. (2011). MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11, 353-357.
- [9] Klingenberg, C. P., and J. Marugán-Lobón. (2013) Evolutionary covariation in geometric morphometric data: analyzing integration, modularity and allometry in a phylogenetic context. *Systematic Biology* 62:591–610.
- [10] Mitteroecker, P., & Gunz, P. (2009). Advances in geometric morphometrics. *Evolutionary Biology*, 36, 235–247.
- [11] Morphometrics software web, URL <http://life.bio.sunysb.edu/morph/soft-tps.html> [accedido 22 agosto 2016]
- [12] O'Higgins, P., & Jones, N. (1998). Facial growth in *Cercocebus torquatus*: An application of three dimensional geometric morphometric techniques to the study of morphological variation. *Journal of Anatomy*, 193, 251-272
- [13] The R Project for Statistical Computing. URL <https://www.r-project.org/> [accedido 22 agosto 2016]
- [14] Rohlf, F.J. (1990). Rotational fit (Procrustes) methods. In F.J. Rohlf et al., (Eds.) *Proceedings Michigan morphometrics workshop* (pp. 227–236). Special publication no. 2, Museum of Zoology, Michigan, University of Michigan.
- [15] Rohlf, F.J. (2015). The tps series of software. *Hystrix*, 26(1), 9–12.
- [16] Rohlf, F.J., & Marcus, L.F. (1993). A revolution in morphometrics. *Tree*, 8, 129–132.
- [17] Rohlf, F.J., & Slice, D.E. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, 39(1), 40–59.
- [18] RPS Website, URL <https://sites.google.com/site/resistantprocrustes/> [accessed 02 September 2015]
- [19] Siegel, A.F., & Benson, R.H. (1982). A robust comparison of biological shapes. *Biometrics*, 38, 341–350.
- [20] Slice, D.E. (1996). Three-dimensional generalized resistant fitting and the comparison of least squares and resistant fit residuals. In L. F. Marcus et al., (Eds.), *Advances in morphometrics* (pp.179–199). New York, Plenum Press.
- [21] Torcida, S., Perez, S.I. & Gonzalez, P. N. (2014). An Integrated Approach for Landmark-Based Resistant Shape Analysis in 3D. *Evolutionary Biology*, 41(2), 351-366
- [22] Weiszfeld, E. (1937). Sur le point pour lequel la somme des distances de n points donne ´s est minimum. *Tohoku Mathematical Journal*, 43, 355–386.
- [23] Zelditch, M.L., Swiderski, D.L., Sheets, H D., & Fink, W.L. (2004). *Geometric morphometric for biologists: A primer*. London: Academic Press.