RESEARCH ARTICLE

A data-driven classification of 3D foot types by archetypal shapes based on landmarks

Aleix Alcacer[1], Irene Epifanio[1,2]*****, **M. Victoria Ibáñez**[1,2], **Amelia Simó**[1,2], **Alfredo Ballester**[3]

1 Departament de Matemàtiques, Universitat Jaume I, Castelló, Spain, **2** Institut de Matemàtiques i Aplicacions de Castelló, Universitat Jaume I, Castelló, Spain, **3** Institut de Biomecànica de València, València, Spain

* epifanio@uji.es

Abstract

The taxonomy of foot shapes or other parts of the body is important, especially for design purposes. We propose a methodology based on archetypoid analysis (ADA) that overcomes the weaknesses of previous methodologies used to establish typologies. ADA is an objective, data-driven methodology that seeks extreme patterns, the archetypal profiles in the data. ADA also explains the data as percentages of the archetypal patterns, which makes this technique understandable and accessible even for non-experts. Clustering techniques are usually considered for establishing taxonomies, but we will show that finding the purest or most extreme patterns is more appropriate than using the central points returned by clustering techniques. We apply the methodology to an anthropometric database of 775 3D right foot scans representing the Spanish adult female and male population for footwear design. Each foot is described by a 5626 × 3 configuration matrix of landmarks. No multivariate features are used for establishing the taxonomy, but all the information gathered from the 3D scanning is employed. We use ADA for shapes described by landmarks. Women's and men's feet are analyzed separately. We have analyzed 3 archetypal feet for both men and women. These archetypal feet could not have been recovered using multivariate techniques.

1 Introduction

A fundamental issue in the appropriate design of footwear is to know foot shape. In particular, it is important to know the types of foot shapes and how the different feet of users can be explained by this taxonomy, i.e. the foot shape distribution. It is not only important from the shoe manufacturing point of view, since an improper fit prevents shoe purchase, but also because poorly fitting footwear can cause foot pain and deformity [1], especially in women. Therefore, numerous studies have been carried out on foot shapes [2–7].

Identifying foot shapes has a significant impact on design [8–10]. A small group of human models that represents the anthropometric variability of the target population is commonly used in ergonomic design and evaluation. Working with a small group of cases, the test cases,

provides designers with an efficient way to develop and evaluate a product design. Considering the boundary cases or the extreme cases is a common strategy in design [11]. The idea behind considering the boundary cases is that if the design fits for the extreme cases well, then all other less extreme body types in the target population should also be well accommodated.

Knowledge of the types of body part shapes is not only important in the design or apparel industry [12, 13], but also in ergonomics in general [14, 15], and other disciplines such as sport [16–18], medicine [19–21], phylogeny [22], criminalistics [23], etc. Face classification is also important due to its application in forensic anthropology, crime prevention and new human-machine interaction systems and online activities like e-commerce, e-learning, gaming, dating and social media [24, 25]. Furthermore, taxonomy is also very important not only in anthropometry, but also in morphometry in general, such as in animal or plant taxonomy [26, 27] or also in genetics [28].

The method of establishing types of feet, or other parts of the body, is usually based on subjective or visual elements [29]. When objective techniques have been contemplated, these have been very simple [30]. In fact, despite performing 3D scans, that information is then summarized into a series of multivariate measures [5, 7, 31]. These measures are then treated in an ad hoc, heuristic way to couple pre-established types [12], or a cluster analysis is applied to these measures directly or after applying factor analysis or principal component analysis (PCA) to reduce the dimension [2, 14, 24, 32–35].

Our aim is to improve on the previous methodologies used to define taxonomies by removing the subjective steps and making the data speak for themselves. We use archetypoid analysis (ADA) for shapes based on landmarks, which was developed by some of the authors in [36]. ADA is a variant of archetype analysis (AA), which is an unsupervised statistical learning tool. Archetypes lie on the boundary of the convex hull of the data, meaning that they are extreme profiles. ADA returns archetypes in data. On the one hand, this statistical tool allows us to consider all the information contained in the 3D scanners, without the need for extracting variables from them, thus avoiding the step of deciding which variables may or may not be relevant. On the other hand, the tool itself will provide the taxonomy from the data themselves, i.e. it will provide the existing archetypes in the data, while the user only intervenes to specify the number of archetypes to consider. If the user is not sure how many archetypes should be considered, the tool can provide the most reasonable number of archetypes based on the elbow criterion, which will be explained below. Furthermore, the technique returns how the feet are formed as a function of the archetypes by using mixtures of archetypes. In other words, each foot will be represented as a percentage of the archetypal feet; in this way, it can be easily understood by any user who is not expert in this technique. Despite the fact that clustering is the usual technique for defining typologies, we will use a toy example to show that AA or ADA, rather than cluster analysis (CLA), is the most appropriate statistical technique for obtaining a taxonomy. We will use ADA instead of AA in our problem with 3D scans, because we prefer to obtain archetypal feet corresponding to particular individuals in order to describe those archetypal feet by some multivariate measures a posteriori.

The objective of ADA is to represent the cases by means of a convex combination (a mixture) of archetypes that are actual cases, which are referred to as archetypoids. This makes the results returned by ADA easily interpretable, even for non-experts. The difference between AA and ADA is that in AA the archetypes are mixtures of cases, and therefore, they are not necessarily actual cases. In other words, ADA represents the data as mixtures of extreme cases, and not as mixtures of mixtures, as AA does. AA was defined for multivariate data by [37], while ADA was proposed by [13]. ADA has been extended to other kind of data, such as functions [38] or shapes defined by landmarks [36].

AA and ADA applications have been growing at a great rate and they can be found in a diverse range of disciplines, such as biology [39], computer vision [40–45], developmental psychology [46], engineering [11, 13, 47, 48], finance [49], genetics [50], global development [51], machine learning problems [52], market research [53], multi-document summarization [54], neuroscience [55, 56] and sports [57–59].

Archetypal analysis techniques lie somewhere in between two well-known unsupervised statistical techniques: PCA and CLA. Data decomposition techniques aim to find the latent components, and data are expressed as a linear combination of several factors. The constraints on the factors and how they are combined determine the definition of different statistical techniques. In PCA, factors are linear combinations of variables, and therefore their restrictions are minimal. This compromises the interpretability of the factors, but it helps explain the variability of the data. Instead, in CLA, such as k-means algorithm, factors have the greatest restrictions. As factors in k-means are centroids (means of groups of data), they are easily interpretable. However, the modeling flexibility of CLA is reduced due to the binary assignment of data to the clusters. In contrast, AA and ADA enjoy higher modeling flexibility than CLA but without losing the interpretability of their factors. [52] and [13] provide a table summarizing the relationship between several unsupervised multivariate techniques. ADA is also compared with many other unsupervised multivariate techniques in [13].

Percentiles should not be used to find the boundary cases in design since with the exception of $50th$-percentiles, percentile values are not additive [60–62]. Although, different alternatives have been considered, such as the use of CLA [63], the most common approach is based on the use of PCA [61, 64–68]. In this approach, several extreme points are selected from the projection into the first principal components. However, the PCA-approach has several drawbacks [69]. In [61, 67, 68] only the variation in the first two or three components is taken into account, so unconsidered variation may represent cases that are difficult to accommodate, which would be missing. In addition, the number of selected boundary cases with two PCs is eight (fourteen with three PCs), which could be too high in practice. A large numbers of test cases may overwhelm the designer and thus be counterproductive. With ADA we will obtain the extreme cases, since this is precisely the objective of this statistical technique, and we can control the number of extreme cases that the designer wants to consider.

Toy example

In Fig 1 a toy two-dimensional data set is used to illustrate what archetypoids mean and the differences compared with PCA and CLA, as well as to provide some intuition on what these pure and extreme patterns imply in Anthropometry. Two numeric variables are considered from the data set described below: the Foot Length (FL) and Ball Width (BW) of 382 right feet from the adult female Spanish population. We apply k-means and ADA with $k = 3$, i.e. we find 3 clusters and archetypoids, with standardized data. We also apply PCA.

Archetypoids are feet with extreme values, which have clear profiles: archetypoid 1 is characterized by very low FL and BW values, archetypoid 2 has a very high value for BW, but a medium value for FL, while the third archetypoid has a very high FL value together with a medium-high value for BW. Archetypoids are the purest feet. The rest of the feet are expressed as mixtures (collected in alpha coefficients, which is explained in Section 2) of these ideal feet. For example, a foot with values of 244.2 and 86.5 for FL and BW, respectively, is explained by 43% of archetypoid 1 plus 57% of archetypoid 3. From the clustering point of view this foot is assigned to cluster 1, although it is near the border of cluster 2, but clustering does not say anything about the distance of this point with respect to the assigned centroid, or in which direction they are separated. In fact, that foot is quite far from its assigned centroid. This happens
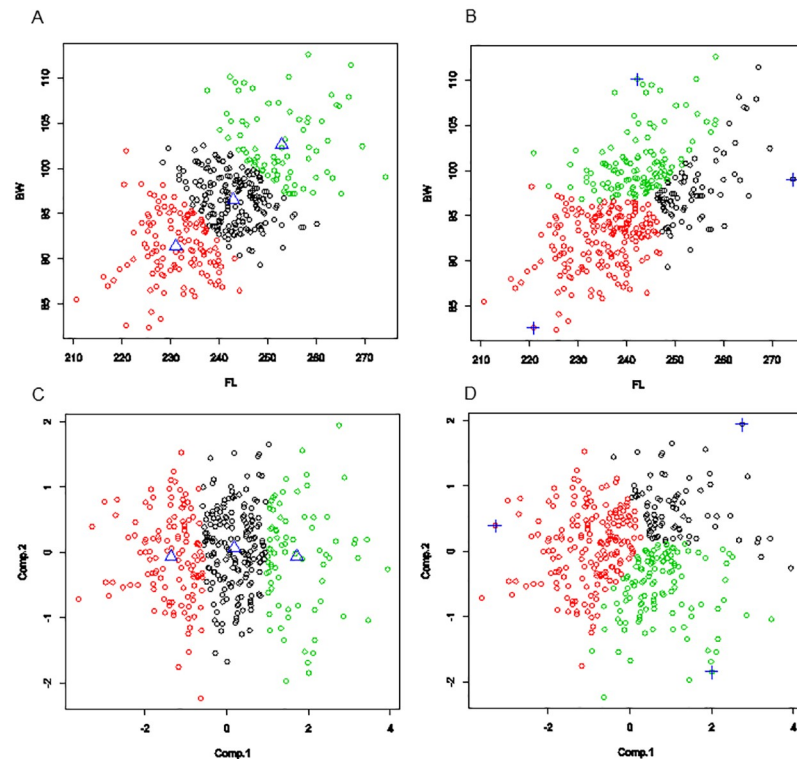
Fig 1. Toy example. (A) Plot of the k-means cluster assignments. The blue triangles represent the centroids of each cluster. (B) ADA assignments by the maximum alpha (see Section 2), i.e. assigned to the archetypoid that best explains the corresponding observation. The blue crosses identify the archetypoids. (C) PC scores with cluster assignments. Projected centroids are represented by blue triangles. (D) PC scores with the ADA assignments. Projected archetypoids are represented by blue crosses.

because the objective of clustering is to assign the data to groups, not to explain the structure of the data more qualitatively.

This is compatible with the natural tendency of humans to represent a group of items by its extreme units [70]. Fig 1B shows the partition of the set generated by assigning the cases to the archetypoid that best explains each observation. However, when we apply k-means to this kind of data set, without differentiated clusters, the centroids are in the middle of the data cloud. Centroid profiles are not as differentiated from each other as archetypoid profiles. This happens because centroids have to cover the set in such a way that the set is partitioned by minimizing the distance with respect to the assigned centroid (see [71] about the connection between set partitioning and clustering). On the one hand, this means that the set partition generated by k-means and ADA would be different (Fig 1A and 1B). On the other hand, centroids are not the purest, and therefore their profiles are not as clear as those of archetypoids. In Fig 2 we show the foot centroids and archetypoids as rectangles. Archetypoids are more intuitively interpretable due to the extremeness of their dimensions: the first archetypoid is a very short and narrow foot (smaller in both dimensions than the smallest centroid); the second archetypoid is very wide, while the second centroid is similar to the mean foot; and the third archetypoid is a very long foot that is longer than the third centroid. All the foot centroids have the same aspect, i.e. the same FL and BW ratio as the mean foot. However, this is not the case with ADA. Archetypoid 1 has the same ratio as the mean foot, but not archetypoids 2 and 3, which are more flattened and elongated, respectively. This can be clearly appreciated in the
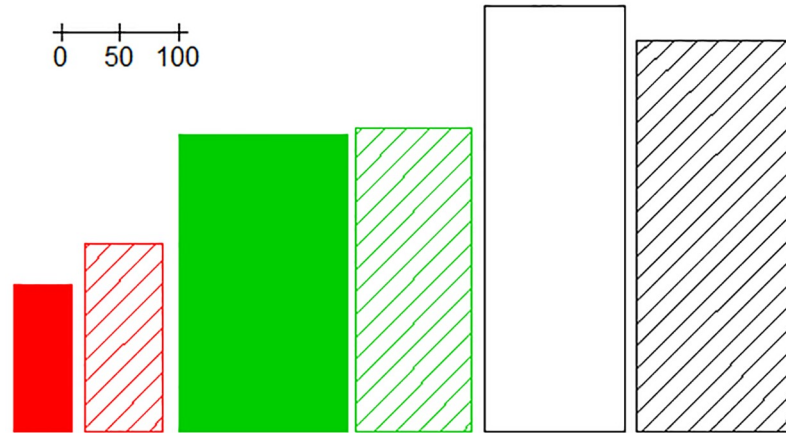
Fig 2. Representative feet of the toy example. The color code of each representant coincides with the color code used in the assignments of Fig 1. The centroids of each cluster are represented by shading lines, while the archetypoids are represented by solid colors. In order to highlight the differences and make them more easily perceptible to the human eye, the percentiles of each representative foot were computed. The rectangles represent the increase or decrease with respect to the median foot measurements, which are: 96 mm (BW) and 241 mm (FL). For example, the percentiles of the first archetypoid are 1 and 2 for each variable, respectively. Therefore, in the plot, it is represented as: 96 · (1 + (0.01-0.5)) = 49 and 241 · (1 + (0.02-0.5)) = 125.

PC projections of Fig 1C and 1D. The first PC is a size component composed of the addition of FL and BW (the loadings are 0.7 and 0.7), while the second PC is a shape component composed of the contraposition of FL and BW (the loadings are 0.7 and -0.7). Note that centroids are all in the zero horizontal line, i.e. centroids do not account for different shapes. However, archetypoids are distributed on the border of the PC score space. Archetypoid 1 is on the zero horizontal line, but with a lower score in PC 1 than the centroids. Archetypoids 2 and 3 have higher scores in PC 1 than the centroids, and additionally they have no zero scores in PC 2, being negative for archetypoid 2 and positive for archetypoid 3. Note also that the feet projected on the first quadrant of the PC space correspond to feet similar to archetypoid 3, those projected on the fourth quadrant correspond to feet similar to archetypoid 2, while the second and third quadrant of the PC space correspond to feet similar to archetypoid 1. The mean foot, located at the origin, coincides with the intersection where the three partitions meet them, i.e. the mean foot is a balanced mixture between the three archetypoids. Finally, note that archetypoids do not coincide with the individuals with the most extreme PC scores (see Fig 1D). Unlike PCA, the objective of ADA is to obtain extreme cases, and individuals with extreme PCA scores do not necessarily return archetypical observations. In fact, archetypes could not be recovered with PCA even if all the components had been considered [11]. Therefore, the appropriate statistical technique for obtaining the extreme cases is ADA.

The outline of the paper is as follows: In Section 2 we introduce our data and review ADA for real-valued multivariate data and for shapes defined by landmarks. In Section 3, our proposal is applied to our women and men data sets from the 3D foot scanner and the results are discussed. Section 4 contains conclusions and some ideas for future work.

2 Materials and methods

Foot database

Our anthropometric database is composed of 775 3D right foot scans representing the Spanish adult male and female population, 393 corresponding to men and 382 to women. The mean,

standard deviation, minimum and maximum age for women (men) were: 40.8 (42.3), 11.3 (10.1), 19 (19) and 68 (67), respectively. The data set was collected from May 3rd 2006 to July 21st 2006 by IBV in the project 'Estudio antropométrico y morfológico 3D de los pies de la población española para su aplicación al diseño de calzado y componentes' (IMPRDA/2005/38) funded by Valencia Region Government (i.e. Instituto de la Mediana y Pequeña Industria Valenciana, IMPIVA) under the programme 'Ayudas a la Promoción del Diseño en la Comunidad Valenciana'. All participants signed an informed consent complying with existing Spanish legislation (Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal, LOPD) granting the use of the data for research purposes. The data were collected by IBV from volunteers recruited in different regions across Spain at shoe shops and workplaces using an INFOOT laser scanner [72]. The scanning process is carried out as can be seen in Fig 3: the user stands upright placing equal weight on each foot, in a specific position and orientation. We obtain a 3D point cloud representing the complete outer surface of the foot, including the sole of the foot. Prior to foot scanning, an expert placed five landmarks at key anatomical locations: tip of the first toe, tip of the second toe, head of the metatarsale tibiale, head of the metatarsale fibulare and pternion (see Fig 4). The landmarks used were non-reflective stickers with a 5 mm diameter provided by the distributor of the 3D foot scanner [72]. The spatial location of theses landmarks was automatically detected and recorded by the software of the 3D scanner. The accuracy of anatomical landmark location in human feet by experts is complex to assess. While [73] reported a median intra-observer error of 2-3 mm, we estimate that our expert had an accuracy of at least 5mm. No personal data was gathered along with the 3D point cloud.

3D foot shapes were registered using the method reported by [74] with a template made up of 5626 vertices, using the five foot landmarks, which enables the automatic computation of 22 key foot measurements (see Fig 5). Put simply, we register the original unorganized point clouds to a common template (template fitting process), which is initialized and guided by the five anatomical landmarks. The template mesh was obtained by uniformly remeshing a watertight mesh representing one foot of the sample. A foot that was randomly selected among those that had an average length and that did not present mild foot conditions such as bunions, hammer toes, claw toes, cavus foot or flat foot. This method provides sufficient template fitting accuracy. The mean, root-mean-square and maximum Hausdorff distance from the scanned point cloud to the registered template are approximately 0.07, 0.1 and 1 mm, respectively, which provides sufficient template fitting accuracy for objects scanned with a resolution of 0.5-1 mm.

The 22 foot measurements are used in product design and in clinical assessment. All 3D registered feet were digitally measured with the algorithms developed by the IBV (Biomechanics Institute of Valencia). Unlike body measurements, foot measurements are not standardized. Only Foot Length, Ball Girth and Ball Width are considered in [75], [76] and [77]. The definitions are those used by the Human Shape Lab of the IBV, which comply with standards and are compatible with the accepted definitions found in the literature [78–82].

However, in contrast to the common procedure in the literature, our working data are not the multivariate measurements, which are a mere summary of the richer information contained in the 3D foot scans. Our data set are the set of landmarks; the foot shape of each individual in our data set was represented by 5626 3D landmarks, i.e. by a 5626×3 configuration matrix. Therefore, we work with 775 configuration matrices.

Other researchers can obtain the data set in the same way. The data set is saved as an R object (.Rdata) [83], in a matrix where each row corresponds with each individual and variables are in columns. The data sets and code in free and open software R [83] for reproducing the results are available at https://figshare.com/articles/adafeet_rar/11553324. Note that the

Fig 3. Infoot® scanner. Scanner used to obtain the foot scans.

availability of the code that implements the methodology allows the methodology to be applied to any data set. In order to demonstrate the procedure in the code we carried out a systematic sample of the landmarks and we retained 5% of the landmarks, since the same results, archety-poids, are obtained using 5626 landmarks and 282 landmarks. In this way, if anybody wants to
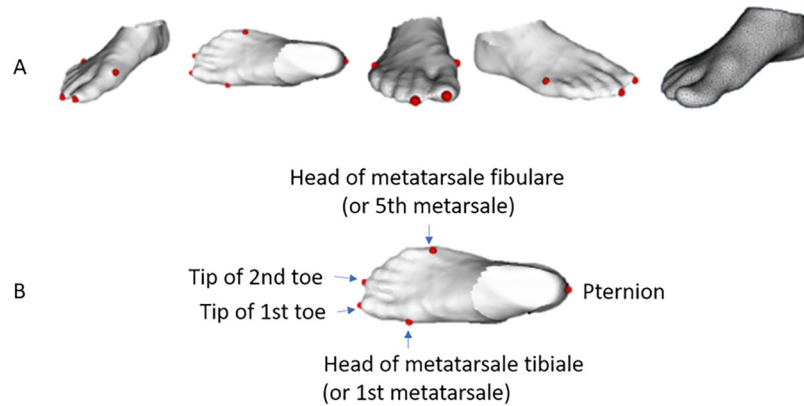
Fig 4. Foot landmarks. (A) Foot landmarks used in the registration of the database and foot template topology (the last image). (B) Names of the five foot landmarks.

reproduce the results, they can obtain the solution faster. Raw data obtained through project IMPRDA/2005/38 are available on request at ibv@ibv.org.

ADA in the shape space

In the multivariate context, let $\{\mathbf{x}_i\}_{i=1}^{n}$ be a set of observations of a variable vector in $\mathbb{R}^k$ taken on n individuals, that is, each observation consists of k measurements $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})$. The archetypoids, $\{\mathbf{z}_j\}_{j=1,\ldots,p}$, are observed data points, so that observations can be approximated by convex combinations of the archetypoids. Then, we will define two matrices of coefficients β and α, such that $\mathbf{x}_i \approx \sum_{j=1}^{p} \alpha_{ij}\mathbf{z}_j$ and $\mathbf{z}_j = \sum_{l=1}^{n} \beta_{jl}\mathbf{x}_l$, with $\beta_{jl} \in \{0, 1\}$, $\forall j, l$. To estimate both matrices of coefficients, the following mixed-integer minimization problem of the residual sum squares (RSS) has to be solved:

$$RSS = \sum_{i=1}^{n}\|\mathbf{x}_i - \sum_{j=1}^{p}\alpha_{ij}\mathbf{z}_j\|^2 = \sum_{i=1}^{n}\|\mathbf{x}_i - \sum_{j=1}^{p}\alpha_{ij}\sum_{l=1}^{n}\beta_{jl}\mathbf{x}_l\|^2, \tag{1}$$

under the constraints

1. $\sum_{j=1}^{p}\alpha_{ij} = 1$ with $\alpha_{ij} \geq 0$ and $i = 1, \ldots, n$ and

2. $\sum_{l=1}^{n}\beta_{jl} = 1$ with $\beta_{jl} \in \{0, 1\}$ and $j = 1, \ldots, p$.

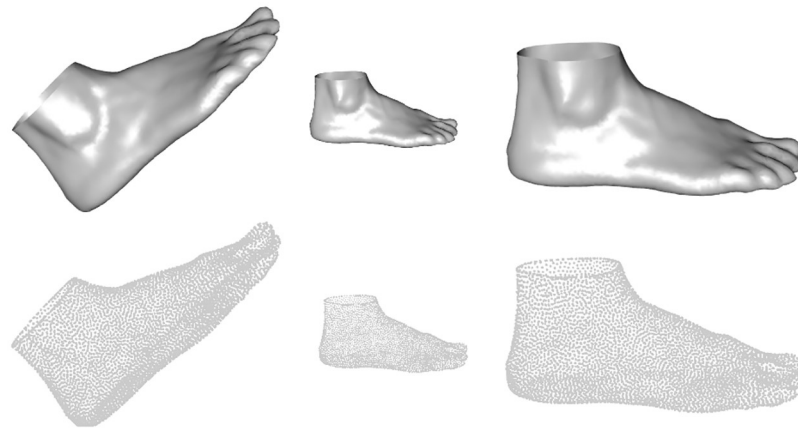Note that $\beta_{jl} = 1$ for one and only one l, otherwise $\beta_{jl} = 0$.
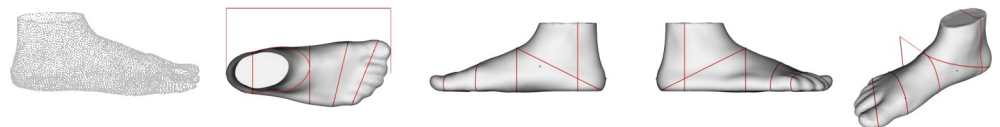
Fig 5. Foot measurements. Examples of digital measurements elicited from a 3D registered foot. Only 8 of the 22 measurements will be used in Section 3, where they will be described in detail. These 8 measurements correspond to the variables that could most influence shoe fitting according to shoe design experts.

Fig 6. Three feet with the same shape. All the objects in this figure correspond to the same shape, i.e. they are equivalent; however, their 3D coordinates are different.

However, as stated above, our data are not multivariate measurements, but a set of landmarks.

Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be $n = 775$ $k \times 3$ configuration matrices, each matrix containing the 3D coordinates of the $k = 5626$ landmarks of each foot. Each matrix could be rearranged to convert it into a vector in $\mathbb{R}^{3k}$ and the above definitions of archetypoids could be used. Nevertheless, these matrices are not representative of the shape of the feet because any translation, rotation or rescaling of them has the same shape. An example can be seen in Fig 6.

Hence, from a theoretical point of view we can define the shape space as:

Definition 1 *The shape space Σ_3^k is the set of equivalence classes $[X]$ of $k \times 3$ configuration matrices $X \in \mathbb{R}^{3 \times k}$ under the action of Euclidean similarity transformations (translation, rotation and scale change).*

In order to obtain a representative element of the shape $[X]$ of a foot, all these transformations have to be removed.

First we remove the location effect. There are different ways to remove location, but we will use the most convenient for mathematical reasons, consisting of multiplying the configuration matrix by the $(k − 1) \times k$ Helmert sub-matrix [84], H, i.e. $X_H = HX$. After removing the location, the representative of a foot is now a $3 \times (k − 1)$ matrix that could be regarded as a vector in the Euclidean space $\mathbb{R}^{3(k−1)}$.

To filter scale we can divide X_H by its Frobenius matrix norm, which is the centroid size, $S(X) = \|X_H\|$:

$$Y = \frac{X_H}{\|X_H\|}. \tag{2}$$

Y is called the pre-shape of the configuration matrix X because all information about location and scale is removed, but rotation information remains.

It is important to note that when scale is removed, the representative of the shape of the foot is still a $(k − 1) \times 3$ matrix, but it cannot be regarded as a vector in a Euclidean space. We are restricted to matrices with the Frobenius norm equal to one and, as a result, they are points in the hypersphere $S^{3(k−1)}$ of $\mathbb{R}^{3(k−1)}$ (a curved subspace). Mathematically, a sphere is a Riemannian manifold.

To choose a single representative of $[X]$ we need to eliminate the rotations and, as a result, our data would be points on the quotient space $S^{3(k-1)}/SO(3)$ where $SO(3)$ is the special orthogonal group of rotation matrices.

Mathematically, this space is a Riemannian submersion of the sphere. The curvature of this space makes the data behave differently than they would do in the Euclidean space; for example, neither the sum nor the multiplication by a scalar is defined i.e. the shape space is not a vectorial space. Fortunately, the theory of Riemannian manifolds tells us that it is possible to work locally in a Riemannian manifold as if we were in a Euclidean space, using the projections of the tangent space at a given point. See Fig 7.

The full Procrustes mean in $S^{3(k-1)}/SO(3)$ of a set of configuration matrices $X_1, \ldots, X_n$ can be defined by

$$[\hat{\mu}] = \arg\inf_{\mu:S(\mu)=1} \sum_{i=1}^{n} d_F^2(X_i, \mu), \qquad (3)$$

where d_F stands for the full Procrustes distance. The mean is estimated by an iterative procedure as described by [85] on pp.90-91. The full Procrustes distance between two configuration matrices X_1 and X_2 is defined by:

$$d_F(X_1, X_2) = \inf_{R \in SO(3), \beta \in \mathbb{R}} \|Y_2 - \beta Y_1 R\|, \qquad (4)$$
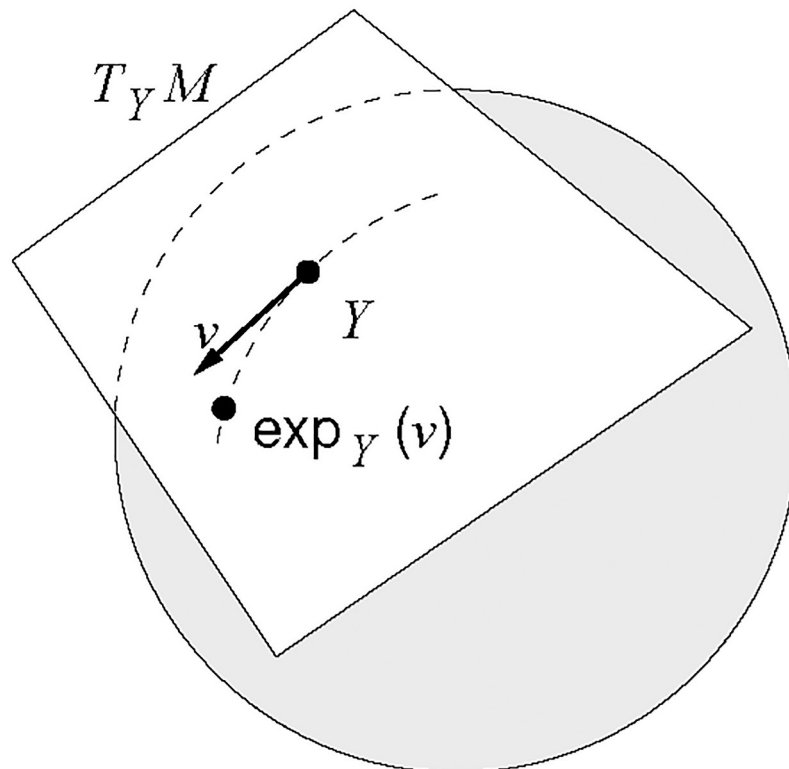
Fig 7. Tangent space at point Y on a sphere. A geometrical view of the tangent plane to a Riemannian manifold M ($S^{3(k-1)}$ in our case) at a point Y, together with the exponential map.

where $SO(3)$ is the orthogonal group of rotations. As explained by [85] on pp. 61-62,

$$d_F(X_1, X_2) = \sqrt{1 - \left(\sum_{i=1}^{m} \lambda_i\right)^2},$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_{m-1} \geq |\lambda_m|$ are the square roots of the eigenvalues of $Y_1^T Y_2 Y_2^T Y_1$, and the smallest value λ_m is the negative square root if and only if $Y_1^T Y_2 < 0$.

So, in view all the above, in [36] we introduced ADA in the tangent space on the mean shape, assuming that our data are sufficiently concentrated around the mean to consider the tangent space a good approximation to shape space. Let us review the main points of this result.

The map that allows us to move from the tangent space to the manifold is called the *exponential map*. And the inverse of the exponential map is called the *logarithmic map*. Their expressions for the shape space are given below.

Let S be the pre-shape of the Procrustes mean μ and $Y_1, \ldots, Y_n$ the preshapes of $X_1, \ldots, X_n$, obtained using Eq 2. To obtain the expression of the projection onto the tangent plane at S of $X_1, \ldots, X_n$, the pre-shape Y_i is rotated to be as close as possible to S.

We write the rotated pre-shape as $Y_i \hat{\Gamma}_i$. The expression of $\hat{\Gamma}_i$ can be found on p. 61 of [85]:

$$\hat{\Gamma}_i = U_i V_i^T,$$

where $U_i, V_i \in SO(3)$ are the left and right matrices of the singular value decomposition of $S^T Y_i$.

Then, the Kent's partial tangent coordinates of Y_i on the tangent space at S, $\mathbf{v}_i$, which will be used in our work, are:

$$\mathbf{v}_i = \log_S(Y_i) \frac{\sin(trace(S^T Y_i \hat{\Gamma}_i))}{trace(S^T Y_i \hat{\Gamma}_i)}, \tag{5}$$

where $\log_S(Y_i)$ is defined by:

$$\log_S(Y_i) = (I_{km-m} - \text{vec}(S)\text{vec}(S)^T)\text{vec}(Y_i \hat{\Gamma}_i) \frac{trace(S^T Y_i \hat{\Gamma}_i)}{\sin(trace(S^T Y_i \hat{\Gamma}_i))}, \tag{6}$$

where I_{km-m} is the $(km - m) \times (km - m)$ identity matrix and *vec* stands for the vectorizing operator. The vectorizing operator of an $l \times m$ matrix A with columns $a_1, a_2, \ldots, a_m$ is defined as: $\text{vec}(A) = (a_1^T, a_2^T, \ldots, a_m^T)^T$.

To project back a point $\mathbf{v}$ in the tangent space to the shape space, the exponential map must be used:

$$Y_v = \text{vec}^{-1}((1 - \mathbf{v}^T\mathbf{v})^{1/2}\text{vec}(S) + \mathbf{v}). \tag{7}$$

Finally, the configuration matrix representing $\mathbf{v}$ would be:

$$X_v = H^T Y_v. \tag{8}$$

Let $\mathbf{v}_1, \ldots, \mathbf{v}_n$ be the tangent coordinates of $\mathbf{X}_1, \ldots, \mathbf{X}_n$. The coordinates in the tangent space $\mathbf{u}_j$ $j = 1, .., p$ of the archetypoids $\mathbf{Z}_j \in \Sigma_3^k$, $j = 1, .., p$ are obtained by minimizing:

$$RSS = \sum_{i=1}^{n} \| \mathbf{v}_i - \sum_{j=1}^{p} \alpha_{ij} \mathbf{u}_j \|^2 = \sum_{i=1}^{n} \| \mathbf{v}_i - \sum_{j=1}^{p} \alpha_{ij} \sum_{l=1}^{n} \beta_{jl} \mathbf{v}_l \|^2, \tag{9}$$

under the constraints

1. $\sum_{j=1}^{p} \alpha_{ij} = 1$ with $\alpha_{ij} \geq 0$ and $i = 1, \ldots, n$ and

2. $\sum_{l=1}^{n} \beta_{jl} = 1$ with $\beta_{jl} \in \{0, 1\}$ and $j = 1, \ldots, p$.

As archetypoids are actual individuals of the sample, the projection of the obtained archetypoids from the tangent space back into the configuration space is immediate.

In summary, we apply multivariate ADA in a tangent space to the shape space.

3 Results and discussion

We have applied ADA separately for men and women, since previous studies have shown gender foot shape differences [5, 7]. Furthermore, footwear designers usually propose different designs for women and men. We have analyzed the whole sample as representative of the population, without removing any possible outlier, since this could be considered part of the population variability. If we were more interested in the archetypal feet of the majority than of the totality, outliers could be identified by computing the Procrustes distances of each foot to the mean, as in [36]. In the same way, if we wanted to accommodate a certain percentage of the population, then only an appropriate part of the sample could be used.

In order to determine the number p of archetypoids for women and men, RSS values have been represented for a series of different p values in Fig 8. Although not very clear, it seems that an elbow is found for $p = 3$, for men and women. In any case, a shoe design expert indicated that this would be a reasonable number for design purposes (a large number of representative cases may overwhelm the designer and thus be counterproductive [11]). Therefore, in the interests of brevity, we examine the results of 3 archetypoids. If the designer decided to choose more archetypoids, our procedure would be the same.

The three archetypoids for women and men are displayed in Fig 9. Archetypoids correspond to actual individuals, so in order to get a concise description of the archetypoids, rather than the whole set of 22 variables, we have computed the percentiles of the most relevant
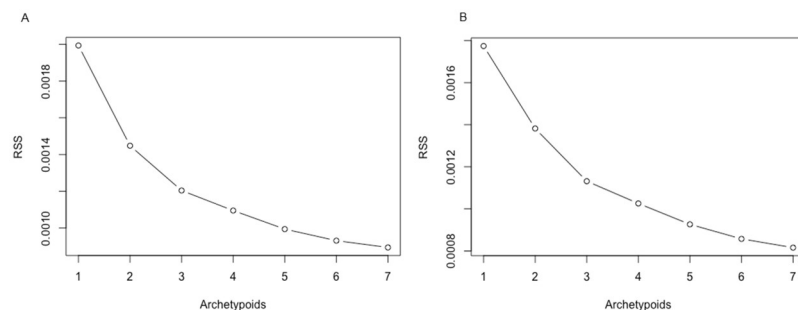
Fig 8. Screeplots for ADA with 3D landmarks. (A) Screeplot for women. (B) Screeplot for men.
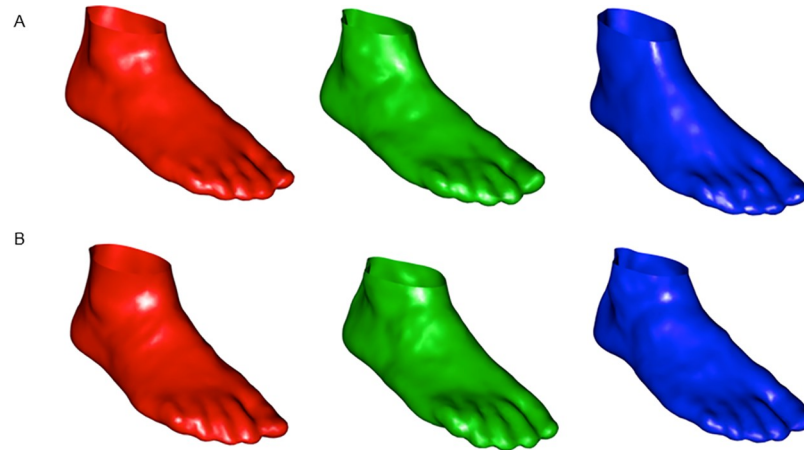
Fig 9. Three archetypoids obtained with 3D landmarks. (A) Archetypoids for women. (B) Archetypoids for men. The first archetypoids are shown in red, the second archetypoids in green, while the third archetypoids are shown in blue.

variables in shoe design. According to shoe design experts, the variables that could most influence shoe fitting are: Foot Length, FL (distance between the rear and foremost point the foot axis); Ball Girth, BG (perimeter of the ball section); Ball Width, BW (maximal distance between the extreme points of the ball section projected onto the ground plane); and Instep Height, IH (maximal height of the instep section, located at 50% of the foot length). But the following variables are also relevant: Toe Height, TH (maximal height of the toe section); Ball Position, BP (distance from the rearmost point of the foot to the intersection of the ball section and the foot axis); Instep Girth, IG (perimeter of the instep section, located at the 50% of foot length); and Instep to Heel Girth, IHG (perimeter of the section that passes through the heel to the instep, located at 50% of the foot length). According to footwear experts, the variable that best describes the size of the foot is FL. As the shape corresponds to the geometrical information that remains once the scale is eliminated, to describe the archetypal foot shapes by variables, we consider the rest of the variables after removing the scale by dividing each of the variables by FL: BG/FL, BP/FL, BW/FL, IG/FL, IH/FL, IHG/FL and TH/FL. Table 1 shows the percentiles of the 3 archetypoids for those variables for women and men, respectively.

According to the percentile profiles (the percentiles of A1W and A1M, A2W and A2M, and A3W and A3M are very much alike), the three archetypoids found for men and women are quite similar. This could indicate that in global terms the three extreme foot shapes for men and women resemble each other. For a larger p values the majority of profiles coincide for men and women but some are different, showing different shapes between genders. Nevertheless, as stated before, we concentrate on the results for $p = 3$ for footwear design in order to create a design that could fit the three archetypal feet.

The percentile profile of the first archetypoid for both women and men is characterized by medium-low percentiles for variables BG/FL, BW/FL, IG/FL, IH/FL, medium-high percentiles for variables IHG/FL and TH/FL, and a low percentile for BP/FL. The percentile profile of the second archetypoid for both women and men is characterized by high percentiles for BG/FL and BW/FL, very low percentiles for BP/FL, IH/FL and IHG/FL and, medium percentiles for IG/FL and TH/FL. Finally, the percentile profile of the third archetypoid for both women and men is characterized by low percentiles for variables BG/FL and IG/FL, a very high percentile for BP/FL, a medium percentile for BW/FL and, very low percentiles for IH/FL, IHG/FL and TH/FL.

Table 1. Percentiles corresponding to the 3 archetypal foot shapes of women and men obtained using 3D landmarks.

Archetypoids	BG/FL	BP/FL	BW/FL	IG/FL	IH/FL	IHG/FL	TH/FL
A1W	24	13	32	36	32	62	58
A2W	74	9	83	42	11	7	45
A3W	26	99	45	28	3	7	6
A1M	22	15	31	38	28	67	58
A2M	75	10	86	44	10	5	46
A3M	25	98	45	29	4	5	3

The first three archetypoids correspond to the sample of women. They are denoted by A1W, A2W and A3W. The last three archetypoids correspond to the sample of men. They are denoted by A1M, A2M and A3M. See details in the text about the meaning of the variables.

In order to view the composition of feet according to the archetypal feet, i.e. to see their distribution, Fig 10 shows the ternary plot for women and men, respectively. The ternary plot represents the alpha values, the sum of which is one, in an equilateral triangle. In both cases, the distributions are quite similar: the majority of feet are a mixture between the three archetypoids, but the second archetypoid has a larger weight than the other archetypoids. There is a small gender difference in the distribution of the purest feet: in women there is a small concentration of feet that are a mixture between archetypoids 2 and 3 (they appear on the side of the triangle that joins archetypoids 2 and 3), but in men this concentration appears on the side of the triangle that joins archetypoids 1 and 2.

Multivariate ADA has been applied to the variables BG/FL, BP/FL, BW/FL, IG/FL, IH/FL, IHG/FL and TH/FL to check if the same results could have been obtained using the variables directly instead of the 3D landmarks. Table 2 shows the percentiles of the 3 multivariate archetypoids for women and men, respectively. The archetypal profiles for men and women coincide again. However, the profiles obtained by multivariate variables and 3D landmarks are somewhat different. The largest differences are found between the profiles of the first archetypoids obtained with multivariate data and 3D landmarks. These differences are found in variables BG/FL, BW/FL, IG/FL, IH/FL, IHG/FL and TH/FL, especially in first four of these variables. The second profiles are similar, with no large differences in variables BG/FL, BW/FL and TH/FL. The third archetypoid profile with 3D landmarks is similar to the third profile obtained with multivariate variables with some not too large differences in variables BG/FL,
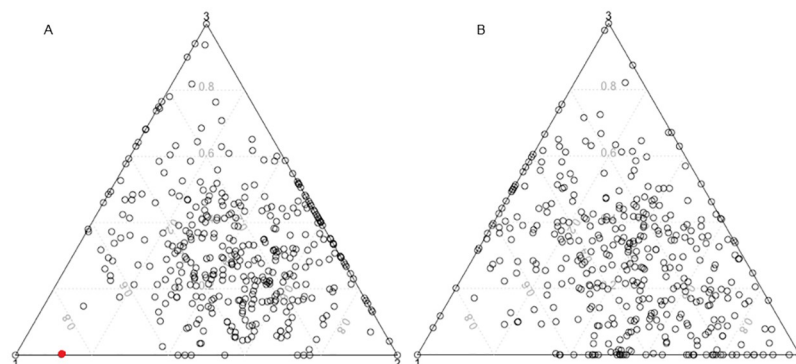
Fig 10. Ternary plots for ADA with 3D landmarks and _p_ = 3. (A) Ternary plot for women. (B) Ternary plot for men. Each point corresponds to a foot, which is described by the alpha values. The corners of the triangle indicate the location of each of the archetypoids. For example, in the first ternary plot the red point represents a foot that is approximated by 88% of archetypoid 1 and 12% of archetypoid 2.

Table 2. Percentiles corresponding to the 3 archetypal foot shapes of women and men obtained using variables.

Archetypoids	BG/FL	BP/FL	BW/FL	IG/FL	IH/FL	IHG/FL	TH/FL
A1W	99	18	99	98	99	100	97
A2W	15	9	18	34	5	1	4
A3W	1	100	4	0	32	1	18
A1M	98	36	95	98	97	96	94
A2M	49	2	59	26	4	16	28
A3M	0	99	0	2	26	13	4

The first three archetypoids correspond to the sample of women. They are denoted by A1W, A2W and A3W. The last three archetypoids correspond to the sample of men. They are denoted by A1M, A2M and A3M. See details in the text about the meaning of the variables.

BW/FL, IG/FL and IH/FL. Therefore, the archetypal profiles obtained using the richer information of 3D landmarks cannot be recovered entirely using multivariate data.

4 Conclusions

We have introduced ADA for the taxonomy of foot shapes defined by 3D landmarks. This procedure avoids the subjective steps of previous methodologies, such as the selection of a set of variables from the 3D foot scans. We have shown that ADA is a more appropriate technique for establishing types of feet (or other parts of the body) than the usual clustering techniques.

We have applied ADA to a sample of foot shapes from the Spanish adult population, and we have analyzed the 3 archetypal feet found using 3D landmarks. We have also shown that these archetypal feet could not be recovered using a multivariate technique. Knowing the archetypal feet can help to design adequate footwear to improve fit and accommodate a great percentage of the population.

As future work, the same methodology could be applied to other databases of other parts of the body or to data sets outside the field of Anthropometry. On the other hand, if landmarks are not the only descriptors of the observations, but other information is available, for example color in biological data sets as described by [26] for ladybird beetles, we can extend the methodology and define ADA in this new space. In that case, the objective function in Eq 1 should be modified to take into account both sets of characteristics. Once the shapes are represented in the tangent space, the information of both vectorial spaces could be (weighted) combined using an adequate interior product to build the corresponding RSS.

If we do not have landmarks to describe the shapes, but instead sets or contour functions, archetypal analysis could also be applied. Preliminary work in two-dimensional sets has been carried out in [86] and [87], respectively, but these ideas could be extended to 3D sets or surfaces.

Author Contributions

Conceptualization: Irene Epifanio, M. Victoria Ibáñez, Amelia Simó.

Data curation: Aleix Alcacer, Alfredo Ballester.

Formal analysis: Aleix Alcacer, Irene Epifanio.

Funding acquisition: Irene Epifanio, M. Victoria Ibáñez.

Investigation: Alfredo Ballester.

Methodology: Irene Epifanio, M. Victoria Ibáñez, Amelia Simó.

Project administration: Irene Epifanio, M. Victoria Ibáñez.

Resources: Alfredo Ballester.

Software: Aleix Alcacer, Irene Epifanio.

Supervision: Irene Epifanio.

Visualization: Irene Epifanio.

Writing – original draft: Irene Epifanio, M. Victoria Ibáñez, Amelia Simó.

Writing – review & editing: Aleix Alcacer, Irene Epifanio, Alfredo Ballester.

References

1. Mickle KJ, Munro BJ, Lord SR, Menz HB, Steele JR. Foot shape of older people: implications for shoe design. Footwear Science. 2010; 2(3):131–139. https://doi.org/10.1080/19424280.2010.487053

2. Krauss I, Grau S, Mauch M, Maiwald C, Horstmann T. Sex-related differences in foot shape. Ergonomics. 2008; 51(11):1693–1709. https://doi.org/10.1080/00140130802376026 PMID: 18941975

3. Delgado-Abellán L, Aguado X, Jiménez-Ormeño E, Mecerreyes L, Alegre LM. Foot morphology in Spanish school children according to sex and age. Ergonomics. 2014; 57(5):787–797. https://doi.org/10.1080/00140139.2014.895055 PMID: 24650291

4. Hong Y, Wang L, Xu DQ, Li JX. Gender differences in foot shape: a study of Chinese young adults. Sports Biomechanics. 2011; 10(02):85–97. https://doi.org/10.1080/14763141.2011.569567 PMID: 21834393

5. Krauss I, Langbein C, Horstmann T, Grau S. Sex-related differences in foot shape of adult Caucasians— a follow-up study focusing on long and short feet. Ergonomics. 2011; 54(3):294–300. https://doi.org/10.1080/00140139.2010.547605 PMID: 21390959

6. Tomassoni D, Traini E, Amenta F. Gender and age related differences in foot morphology. Maturitas. 2014; 79(4):421–427. https://doi.org/10.1016/j.maturitas.2014.07.019 PMID: 25183323

7. Saghazadeh M, Kitano N, Okura T. Gender differences of foot characteristics in older Japanese adults using a 3D foot scanner. Journal of Foot and Ankle Research. 2015; 8(1):29. https://doi.org/10.1186/s13047-015-0087-4 PMID: 26180554

8. Rodrigo AS, Goonetilleke RS, Witana CP. Model based foot shape classification using 2D foot outlines. Computer-Aided Design. 2012; 44(1):48–55. https://doi.org/10.1016/j.cad.2011.01.005

9. Mochimaru M, Kouchi M, Dohi M. Analysis of 3-D human foot forms using the Free Form Deformation method and its application in grading shoe lasts. Ergonomics. 2000; 43(9):1301–1313. https://doi.org/10.1080/001401300421752 PMID: 11014753

10. Wunderlich RE, Cavanagh PR. Gender differences in adult foot shape: implications for shoe design. Medicine and science in sports and exercise. 2001; 33(4):605–611. https://doi.org/10.1097/00005768-200104000-00015 PMID: 11283437

11. Epifanio I, Vinué G, Alemany S. Archetypal analysis: contributions for estimating boundary cases in multivariate accommodation problem. Computers & Industrial Engineering. 2013; 64(3):757–765. https://doi.org/10.1016/j.cie.2012.12.011

12. Alemany S, González JC, Nácher B, Soriano C, Arnáiz C, Heras H. Anthropometric survey of the Spanish female population aimed at the apparel industry. In: Proceedings of the 2010 Intl. Conference on 3D Body scanning Technologies. Lugano, Switzerland; 2010. p. 1–10.

13. Vinué G, Epifanio I, Alemany S. Archetypoids: A new approach to define representative archetypal data. Computational Statistics & Data Analysis. 2015; 87:102–115. https://doi.org/10.1016/j.csda.2015.01.018

14. Jee S, Yun MH. An anthropometric survey of Korean hand and hand shape types. International Journal of Industrial Ergonomics. 2016; 53:10–18. https://doi.org/10.1016/j.ergon.2015.10.004

15. Lin YL, Lee KL. Investigation of anthropometry basis grouping technique for subject classification. Ergonomics. 1999; 42(10):1311–1316. https://doi.org/10.1080/001401399184965

16. Malousaris GG, Bergeles NK, Barzouka KG, Bayios IA, Nassis GP, Koskolou MD. Somatotype, size and body composition of competitive female volleyball players. Journal of Science and Medicine in Sport. 2008; 11(3):337–344. https://doi.org/10.1016/j.jsams.2006.11.008. PMID: 17697797

17. Sterkowicz-Przybycień K, Sterkowicz S, Biskup L, Żarów R, Kryst u, Ozimek M. Somatotype, body composition, and physical fitness in artistic gymnasts depending on age and preferred event. PLOS ONE. 2019; 14(2):1–21.

18. Ryan-Stewart H, Faulkner J, Jobson S. The influence of somatotype on anaerobic performance. PLOS ONE. 2018; 13(5):1–11. https://doi.org/10.1371/journal.pone.0197761

19. Koleva M, Nacheva A, Boev M. Somatotype and disease prevalence in adults. Reviews on environmental health. 2002; 17(1):65–84. https://doi.org/10.1515/reveh.2002.17.1.65 PMID: 12088094

20. Buffa R, Lodde M, Floris G, Zaru C, Putzu PF, Marini E. Somatotype in Alzheimer's disease. Gerontology. 2007; 53(4):200–204. https://doi.org/10.1159/000100486 PMID: 17347566

21. Singh S. Somatotype and disease: a review. Anthropologist. 2007; 3:251–261.

22. Braga J, Zimmer V, Dumoncel J, Samir C, de Beer F, Zanolli C, et al. Efficacy of diffeomorphic surface matching and 3D geometric morphometrics for taxonomic discrimination of Early Pleistocene hominin mandibular molars. Journal of Human Evolution. 2019; 130:21–35. https://doi.org/10.1016/j.jhevol.2019.01.009 PMID: 31010541

23. Ritz-Timme S, Gabriel P, Obertovà Z, Boguslawski M, Mayer F, Drabik A, et al. A new atlas for the evaluation of facial features: advantages, limits, and applicability. International Journal of Legal Medicine. 2011; 125(2):301–306. https://doi.org/10.1007/s00414-010-0446-4 PMID: 20369248

24. Fuentes-Hurtado F, Diego-Mas JA, Naranjo V, Alcañiz M. Automatic classification of human facial features based on their appearance. PLOS ONE. 2019; 14(1):1–20. https://doi.org/10.1371/journal.pone.0211314

25. Sarakon P, Charoenpong T, Charoensiriwath S. Face shape classification from 3D human data by using SVM. In: The 7th 2014 Biomedical Engineering International Conference; 2014. p. 1–5.

26. MacLeod N. The direct analysis of digital images (eigenimage) with a comment on the use of discriminant analysis in morphometrics. In: Proceedings of the Third International Symposium on Biological Shape Analysis. World Scientific, Singapore; 2015. p. 156–182.

27. Viscosi V, Cardini A. Leaf Morphology, Taxonomy and Geometric Morphometrics: A Simplified Protocol for Beginners. PLOS ONE. 2011; 6(10):1–20. https://doi.org/10.1371/journal.pone.0025630

28. Korem Y, Szekely P, Hart Y, Sheftel H, Hausser J, Mayo A, et al. Geometry of the Gene Expression Space of Individual Cells. PLOS Computational Biology. 2015; 11(7):1–27. https://doi.org/10.1371/journal.pcbi.1004224

29. Simmons K, Istook CL, Devarajan P. Female figure identification technique (FFIT) for apparel. Part I: Describing female body shapes. Journal of Textile and Apparel, Technology and Management. 2004; 4:1–16.

30. Vuruskan A, Bulgun E. Identification of female body shapes based on numerical evaluations. International Journal of Clothing Science and Technology. 2011; 23(1):46–60. https://doi.org/10.1108/09556221111096732

31. Lee YC, Wang MJ. Taiwanese adult foot shape classification using 3D scanning data. Ergonomics. 2015; 58(3):513–523. https://doi.org/10.1080/00140139.2014.974683 PMID: 25361465

32. Kim NS, Do WH. Classification of Elderly Women's Foot Type. Journal of the Korean Society of Clothing and Textiles. 2014; 38(3):305–320. https://doi.org/10.5850/JKSCT.2014.38.3.305

33. Loeffler-Wirth H, Vogel M, Kirsten T, Glock F, Poulain T, Körner A, et al. Body typing of children and adolescents using 3D-body scanning. PLOS ONE. 2017; 12(10):1–11. https://doi.org/10.1371/journal.pone.0186881

34. Loffler-Wirth H, Willscher E, Ahnert P, Wirkner K, Engel C, Loeffler M, et al. Novel Anthropometry Based on 3D-Bodyscans Applied to a Large Population Based Cohort. PLOS ONE. 2016; 11(7):1–20. https://doi.org/10.1371/journal.pone.0159887

35. Nikolaidou ME, Boudolos KD. A footprint-based approach for the rational classification of foot types in young schoolchildren. The Foot. 2006; 16(2):82–90. https://doi.org/10.1016/j.foot.2006.02.001

36. Epifanio I, Ibáñez MV, Simó A. Archetypal shapes based on landmarks and extension to handle missing data. Advances in Data Analysis and Classification. 2018; 12(3):705–735. https://doi.org/10.1007/s11634-017-0297-7

37. Cutler A, Breiman L. Archetypal Analysis. Technometrics. 1994; 36(4):338–347. https://doi.org/10.1080/00401706.1994.10485840

38. Epifanio I. Functional archetype and archetypoid analysis. Computational Statistics & Data Analysis. 2016; 104:24–34. https://doi.org/10.1016/j.csda.2016.06.007

39. D'Esposito MR, Palumbo F, Ragozini G. Interval Archetypes: A New Tool for Interval Data Analysis. Statistical Analysis and Data Mining. 2012; 5(4):322–335. https://doi.org/10.1002/sam.11140

40. Chen Y, Mairal J, Harchaoui Z. Fast and Robust Archetypal Analysis for Representation Learning. In: CVPR 2014—IEEE Conference on Computer Vision & Pattern Recognition; 2014.

41. Bauckhage C, Kersting K, Hoppe F, Thurau C. Archetypal Analysis as an Autoencoder. In: Workshop New Challenges in Neural Computation; 2015.

42. Sun W, Yang G, Wu K, Li W, Zhang D. Pure endmember extraction using robust kernel archetypoid analysis for hyperspectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing. 2017; 131:147–159. https://doi.org/10.1016/j.isprsjprs.2017.08.001

43. Sun W, Zhang D, Xu Y, Tian L, Yang G, Li W. A Probabilistic Weighted Archetypal Analysis Method with Earth Mover's Distance for Endmember Extraction from Hyperspectral Imagery. Remote Sensing. 2017; 9(8):841. https://doi.org/10.3390/rs9080841

44. Mair S, Boubekki A, Brefeld U. Frame-based Data Factorizations. In: Proceedings of the 34th International Conference on Machine Learning. vol. 70 of Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR; 2017. p. 2305–2313.

45. Cabero I, Epifanio I. Archetypal analysis: an alternative to clustering for unsupervised texture segmentation. Image Analysis & Stereology. 2019; 38(2):151–160. https://doi.org/10.5566/ias.2052

46. Ragozini G, Palumbo F, D'Esposito MR. Archetypal analysis for data-driven prototype identification. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2017; 10(1):6–20. https://doi.org/10.1002/sam.11325

47. Vinué G. Anthropometry: An R Package for Analysis of Anthropometric Data. Journal of Statistical Software. 2017; 77(6):1–39.

48. Millán-Roures L, Epifanio I, Martínez V. Detection of anomalies in water networks by functional data analysis. Mathematical Problems in Engineering. 2018; 2018(Article ID 5129735):13.

49. Moliner J, Epifanio I. Robust multivariate and functional archetypal analysis with application to financial time series analysis. Physica A: Statistical Mechanics and its Applications. 2019; 519:195–208. https://doi.org/10.1016/j.physa.2018.12.036

50. Thøgersen JC, Mørup M, Damkiær S, Molin S, Jelsbak L. Archetypal analysis of diverse Pseudomonas aeruginosa transcriptomes reveals adaptation in cystic fibrosis airways. BMC Bioinformatics. 2013; 14:279. https://doi.org/10.1186/1471-2105-14-279 PMID: 24059747

51. Epifanio I, Ibáñez MV, Simó A. Archetypal analysis with missing data: see all samples by looking at a few based on extreme profiles. The American Statistician. 2019.

52. Mørup M, Hansen LK. Archetypal analysis for machine learning and data mining. Neurocomputing. 2012; 80:54–63. https://doi.org/10.1016/j.neucom.2011.06.033

53. Porzio GC, Ragozini G, Vistocco D. On the use of archetypes as benchmarks. Applied Stochastic Models in Business and Industry. 2008; 24:419–437. https://doi.org/10.1002/asmb.727

54. Canhasi E, Kononenko I. Weighted archetypal analysis of the multi-element graph for query-focused multi-document summarization. Expert Systems with Applications. 2014; 41(2):535–543. https://doi.org/10.1016/j.eswa.2013.07.079

55. Tsanousa A, Laskaris N, Angelis L. A novel single-trial methodology for studying brain response variability based on archetypal analysis. Expert Systems with Applications. 2015; 42(22):8454–8462. https://doi.org/10.1016/j.eswa.2015.06.058

56. Hinrich JL, Bardenfleth SE, Roge RE, Churchill NW, Madsen KH, Mørup M. Archetypal Analysis for Modeling Multisubject fMRI Data. IEEE Journal on Selected Topics in Signal Processing. 2016; 10 (7):1160–1171. https://doi.org/10.1109/JSTSP.2016.2595103

57. Eugster MJA. Performance Profiles based on Archetypal Athletes. International Journal of Performance Analysis in Sport. 2012; 12(1):166–187. https://doi.org/10.1080/24748668.2012.11868592

58. Vinué G, Epifanio I. Archetypoid Analysis for Sports Analytics. Data Mining and Knowledge Discovery. 2017; 31(6):1643–1677. https://doi.org/10.1007/s10618-017-0514-1

59. Vinué G, Epifanio I. Forecasting basketball players' performance using sparse functional data. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2019; 12(6):534–547. https://doi.org/10.1002/sam.11436

60. Moroney LWF, MSC, USN, Smith MJ. Empirical reduction in potential user population as the result of imposed multivariate anthropometric limits. Naval Aerospace Medical Research Laboratory; 1972.

61. Zehner GF, Meindl RS, Hudson JA. A Multivariate Anthropometric Method For Crew Station Design: Abridged. Kent State University; 1993.

62. Robinette KM, McConville JT. Alternative to Percentile Models. SAE; 1981.

63. Kim K, Kim H, Lee J, Lee E, Kim D. Development of a New 3D Test Panel for Half-Mask Respirators by 3D Shape Analysis for Korean Faces. Journal of the International Society for Respiratory Protection. 2004; 21:125–134.

64. Bittner AC, Glenn FA, Harris RM, Iavecchia HP, Wherry RJ. CADRE: A family of mannikins for workstation design. In: Asfour, S.S. (ed.) Trends in Ergonomics/Human Factors IV. North Holland; 1987. p. 733–740.

65. Gordon CC, Churchill T, Clauser CE, Bradtmiller B, McConville JT, Tebbetts I, et al. 1988 Anthropometric Survey of U.S. Army personnel: Summary statistics interim report. US Army Natick Research, Development and Engineering Center; 1989.

66. Friess M, Bradtmiller B. 3D Head Models for Protective Helmet Development. In: Proceedings of the SAE 2003; 2003.

67. Hudson JA, Zehner GF, Meindl RD. The USAF Multivariate Accommodation Method. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 1998; 42(10):722–726. https://doi.org/10.1177/154193129804201013

68. Robinson JC, Robinette KM, Zehner GF. User's guide to the anthropometric database at the computerized anthropometric research and design (card) laboratory (U). Systems Research Laboratories Inc; 1992.

69. Friess M. Multivariate Accommodation Models using Traditional and 3D Anthropometry. In: SAE Technical Paper; 2005.

70. Davis T, Love BC. Memory for Category Information is Idealized Through Contrast with Competing Options. Psychological Science. 2010; 21(2):234–242. https://doi.org/10.1177/0956797609357712 PMID: 20424052

71. Wu C, Kamar E, Horvitz E. Clustering for set partitioning with a case study in ridesharing. In: IEEE 19th International Conference on Intelligent Transportation Systems (ITSC); 2016. p. 1384–1388.

72. I-Ware Laboratory; 2018. http://www.i-ware.co.jp/.

73. Kouchi M, Mochimaru M. Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry. Applied Ergonomics. 2011; 42(3):518–527. https://doi.org/10.1016/j.apergo.2010.09.011 PMID: 20947062

74. Allen B, Curless B, Popović Z; ACM. The space of human body shapes: reconstruction and parameterization from range scans. ACM Transactions on Graphics (TOG). 2003; 22(3):587–594. https://doi.org/10.1145/882262.882311

75. ISO 8559-1:2017. Size designation of clothes—Part 1: Anthropometric definitions for body measurement; 2017.

76. ASTM D5219-15. Standard Terminology Relating to Body Dimensions for Apparel Sizing; 2015.

77. ISO 7250-1:2008. Basic human body measurements for technological design - Part 1; 2008.

78. Rossi WA, Tennant R. Professional shoe fitting. National Shoe Retailers Association; 2013.

79. Ramiro J, Alcántara E, Forner A, Ferrandis R, García-Belenguer A, Durá J, et al. Guía de recomendaciones para el diseño de calzado. Instituto de Biomecánica de Valencia. 1995; p. 135–151.

80. AIST, Digital Human Research Group; 2018. http://www.dh.aist.go.jp/research/centered/anthropometry/M_foot.html.en.

81. Goonetilleke RS. The science of footwear. CRC Press; 2012.

82. Luximon A. Handbook of footwear design and manufacture. Elsevier; 2013.

83. R Development Core Team. R: A Language and Environment for Statistical Computing; 2019. Available from: http://www.R-project.org.

84. Dryden IL, Mardia KV. Statistical Shape Analysis: With Applications in R. John Wiley & Sons, Chichester; 2016.

85. Dryden IL, Mardia KV. Statistical Shape Analysis. John Wiley & Sons, Chichester; 1998.

86. Alcacer A, Epifanio I, Ibáñez MV, Simó A. Analysis of 2D foot morphology by functional archetypal analysis. In: Proceedings of the XVII Spanish Biometric Conference and VII Ibero-American Biometric Meeting; 2019. p. 24–27.

87. Alcacer A, Epifanio I, Ibáñez MV, Simó A. Archetypal contour function. In: Proceedings of the 12th Scientific Meeting Classification and Data Analysis Group, CLADAG; 2019. p. 26–29.