Ergonomics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/terg20

Comparison of univariate and multivariate anthropometric accommodation of the northwest Mexico population

Graciela Rodríguez Vega, Ulises Zaldívar Colado, Xiomara Penélope Zaldívar Colado, Dora Aydee Rodríguez Vega & Enrique Javier de la Vega Bustillos

ERGONOMICS
https://doi.org/10.1080/00140139.2021.1892832

Comparison of univariate and multivariate anthropometric accommodation of the northwest Mexico population

Graciela Rodríguez Vega[a,b], Ulises Zaldívar Colado[a], Xiomara Penélope Zaldívar Colado[a], Dora Aydee Rodríguez Vega[c] and Enrique Javier de la Vega Bustillos[d]

[a]Faculty of Informatics, Autonomous University of Sinaloa, Sinaloa, México; [b]Department of Industrial Engineering, University of Sonora, Sonora, México; [c]Department of Mechatronics, Polytechnic University of Sinaloa, Sinaloa, México; [d]Department of Research and Postgraduate Studies, TECNM/Technologic Institute of Hermosillo, Sonora, México

ABSTRCT

Ergonomic workstation design is crucial to prevent work-related musculoskeletal disorders. Many researchers have proposed multivariate analysis for human accommodation. However, no multivariate anthropometric analysis exists for the Mexican population. This study compares common multivariate human accommodation approaches (e.g. principal component and archetypal analyses) and clustering techniques (e.g. *k*-means and Ward's algorithm) with the classical percentile-based univariate accommodation approach, using the Chi-squared goodness-of-fit test and the McNemar's test. The theoretical accommodation percentage obtained by multivariate approaches was higher than those obtained by the percentile univariate approach considering the central 98% data. *k*-means and archetypal analysis obtained similar and the highest accommodation values, followed by Ward's algorithm and principal component analysis. The study findings can be deployed to assess the design of workstations in Mexico, such as electronic components assembly and crew designs, and the effects of different anthropometric measurements in human accommodation.

Practitioner summary: Products and workplaces design are commonly based on the classical univariate approach, using the extreme percentiles. In this study, multivariate approaches were tested on dimensions for sitting workstations, and results showed a bigger accommodation level in comparison to the univariate 1%–99% approaches.

Abbreviations: RHM: representative human model; DHM: digital human model; PCA: principal component analysis; AA: archetypal analysis (AA); PCs: principal components; FA: factor analysis; RSS: residual sum of squares; SSE: sum of squared estimated errors; WA: Ward's algorithm; DBI: Davies–Bouldin index; CHI: Calinski–Harabaz index; SI: silhouette index; SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock–popliteal length; BKL: buttock–knee length; FGR: functional grip reach; AD: anthropometric dimension; E: expected; A: achieved

ARTICLE HISTORY
Received 27 November 2019
Accepted 15 February 2021

KEYWORDS
General ergonomics; industrial ergonomics; anthropometry; standards; workstation design

1. Introduction

Products and workplace design should be user-centred, meeting ergonomic and anthropometric principles (Wichansky 2000). An understanding of human variability, including body shape and size, is appropriate to maximise usability and minimise any negative effects on users (Hanson et al. 2009). Incorrect product and workplace designs can lead to discomfort and musculoskeletal disorder, primarily located in the neck, shoulder, hand, wrist or back (Hanson et al. 2009). In this context, anthropometric data can be helpful, as their objective is the characterisation of the human body using a set of body measurements, such as length, height, width and circumference (Pheasant and Haslegrave 2015).

Despite the importance of anthropometric data, it is difficult to find and maintain accurate and up-to-date databases, even in the most developed countries. Detailed databases are simply not available for some user populations or there may not be enough information about target users. The majority of detailed anthropometric databases are based on military populations, which are not representative of the wider labour population, with few available datasets

CONTACT Graciela Rodríguez Vega graciela.rguez.v@gmail.com; graciela.rodriguez@unison.mx Faculty of Informatics, Autonomous University of Sinaloa, Sinaloa, México

describing civilian populations (Pheasant and Haslegrave 2015). Owing to human variability and the lack of correlation between variables in ergonomic design, individuals unsuited in one dimension may be different from those unsuited in other dimensions (Garneau and Parkinson 2016). Additionally, according to anthropometric principles, all products and spaces should be designed to accommodate the largest possible percentage of the user population (Dianat, Molenbroek, and Castellucci 2018).

Traditionally, percentiles, templates and regression models have been used for workplace accommodation. The use of percentile data is a univariate approach that involves designing boundaries, such as 1 and 5% and 95 and 99%, that represent the extremes of human dimensions (Haslegrave 1986; Bittner 2000; Gordon et al. 1989). In a percentile invariability context, the main disadvantage of this method is that it assumes all dimensions of an individual in the *n*th percentile to be in that percentile (Garneau and Parkinson 2016; Vasu and Mital 2000). Another disadvantage is that, except for the 50%, percentiles are not additive (Robinette and McConville 1981; Zehner, Meindl, and Hudson 1993). Templates consist of a series of numeric data recommended for a specific human dimension (Garneau and Parkinson 2016). By contrast, proportionality constants estimate a body segment based on a ratio to another dimension, such as height (Pheasant and Haslegrave 2015). However, these approaches are often inaccurate for average individuals, and people of the same height may have different segment proportions (Pheasant and Haslegrave 2015; Lin, Wang, and Wang 2004). In regression models, design limits are predicted from multiple regression equations, which use the 5 or 95% values for height and weight predictors (Brolin et al. 2016). Although regression design limits are additive, the predicted values depend on the dimensions correlation: low correlation yields design limits that are reasonably close to the population mean (Gordon, Corner, and Brantley 1997).

Multivariate approaches have been applied as an alternative to univariate and bivariate approaches, on the basis that it is critical to represent how a person fits within a given space at the design stage (Young et al. 2008). This can be done by identifying the representative models of the subjects' anthropometry, usually called 'cases', which represents a set of body dimensions that are to be accommodated in the design (Young et al. 2008; da Silva, Zehner, and Hudson 2020). The representative human model (RHM) and digital human model (DHM) have been used to evaluate the human-workstation interaction, so the iterative process of design evaluation, diagnosis and revision can be more rapidly and economically performed (Jung, Kwon, and You 2009).

In the multidimensional consideration of anthropometric diversity, two general-purpose strategies have been used to improve physical accommodation through workplace design; namely, design based on boundaries or distributed cases (Brolin et al. 2016; Jung, Kwon, and You 2010). The use of boundary methods is based on the same principle of identification of the extreme users in the approach of inclusive design: boundary cases are points located at the edge of a population distribution (Brolin et al. 2016). This approach assumes that the use of manikins – as representative critical test models in the design and evaluation of ergonomic workplaces – can be useful in the accommodation of the less extreme population (Bertilsson, Högberg, and Hanson 2012). Conversely, distributed cases are scattered throughout the distribution (Brolin et al. 2016).

If the objective of the product/environmental design is to accommodate people within a designated percentage of the population, boundary cases are preferred. Contrarily, if the goal is to create a multiple-size design – i.e. adjustable workstations and clothes – the use of distributed cases is recommended (Brolin et al. 2016; Epifanio, Vinué, and Alemany 2013). Additionally, distributed cases can also decrease the risk of missing key areas when using boundary cases (Brolin et al. 2016).

Two of the most common boundary methods used in anthropometry are the principal component analysis (PCA) method, proposed by Bittner et al. (1987), and the archetypal analysis (AA) method (Epifanio, Vinué, and Alemany 2013). Bittner et al. (1987) developed CADRE, a family of manikins for workstation design, and the models were improved in 2000 (Bittner 2000). Zehner, Meindl, and Hudson (1993) simplified Bittner's multivariate approach for US Air Force cockpit design, in which an ellipsoid was adjusted to cover a desired percentage of the total dispersion points obtained using the three principal components (PCs), and the boundary individuals were found on the intersection axis and midpoints of the ellipsoid (Brolin, Högberg, and Hanson 2012). Body size representative cases that could ensure the desired level of accommodation when used appropriately in specifying, designing and testing new aircraft were obtained (Brolin, Högberg, and Hanson 2012; Hudson and Zehner 1998). Gordon, Corner, and Brantley (1997) applied PCA to 12 dimensions of the ANSUR I database to define extreme torso sizes and shapes for the design of integrated body armour and load-bearing systems. Young et al. (2008) constructed 26 representative models for the HSIR database based on the three PCs and obtained better results than the percentile method spread across all measurements. In more recent studies, Guan et al. (2012), applied PCA to human accommodation to obtain the boundary cases for cab design. The percentile values (5th and 95th) were found to be inside the adjusted ellipsoid, concluding that the boundary cases on the surface ellipsoid exhibited a higher accommodation level. Essdai et al. (2018) compared the results obtained using the percentile method and PCA analysis for the 95% confidence interval, and concluded that PCA-based design provided more comfortable accommodation. Biswal and Dahiya (2019) identified 26 boundary individuals from six parameters critical to cockpit design, applying PCA to the IAF aircrew anthropometry survey of 2013 and fitting 96% of the target population.

Comparisons of the results of PCA and factor analysis (FA) in human shape clustering have been made, and differences were found at a significance level of 0.005 (Jianwei et al. 2010). However, this strategy needs to be used carefully because extreme subjects that are difficult to accommodate could be excluded from the designs (Vinué et al. 2014). Brolin et al. (2016) concluded that using the PCA algorithm as a boundary strategy may be problematic when analysing datasets with weak correlation, such as facial dimensions (Hudson and Zehner 1998). Some limitations and recommendations for using this approach were presented by Friess (2005).

Archetypes are another way of identifying boundary human models. Epifanio, Vinué, and Alemany (2013) proposed a methodology that assumes several 'pure' individuals were on the 'edges' of the data, and that all other individuals were considered to be a mixture of theses pure types (Epifanio, Ibáñez, and Simó 2018). This approach has also been used to identify human body shapes (Simo et al. 2020; Abdali et al. 2004).

Although the goal of distributed methods is not to cover the complete distribution of data with a number of well-chosen clusters, they have been tested in anthropometric data analysis to identify different body shapes (Brolin et al. 2016). A review of some clustering techniques in anthropometric data analysis was presented by Abdali et al. (2004). Brolin et al. (2016) analysed the diversity in body size by identifying test cases using three different clustering techniques (*k*-means, hierarchical clustering and Gaussian mixture models) and concluded that implementing PCA – to reduce dimensionality before clustering the anthropometric data – could be useful in identifying clusters when considering the most important part of the variances. Brolin et al. (2016) also suggested that setting the representative case of a cluster as the furthest individual from the population centre could help in identifying the distribution boundaries.

da Silva, Zehner, and Hudson (2020) compared the number of subjects captured by the univariate and multivariate approaches in the accommodation of the Brazilian Air Force pilots' anthropometry. The study concludes that the boundary cases multivariate method for accommodation of the central 90% envelope for the Brazilian Air Force Crew was better than the percentile univariate method. The comparison was made by identifying the individuals that satisfy all the anthropometric limits identified by each method and using Chi-Squared and McNemar's statistical test.

The aim of this study is to compare the RHM of the population of northwest Mexico obtained through different multivariate approaches to the classical univariate-percentile approach. The obtained models will be available for ergonomic design, so that designers can compare their products and workplaces to the data presented and modify them, if necessary. To the best of our knowledge, there have been no similar studies of the Mexican population.

2. Background

Multivariate accommodation was introduced in anthropometric analysis in the 1980s. Its goal was to reduce datasets to a manageable size and achieve better addressing accommodation by removing known noisy variables (Hsiao 2013).

2.1. Boundary approach: PCA

PCA is a linear transformation that converts data into a new dimensional space, such that the new set of variables are linear functions of the original data and are uncorrelated. The aim of the method is to reduce the data dimensionality while preserving as much variability as possible (Jolliffe 2002).

Once PCA is computed, the number of components is selected by considering as much explained variance as possible. Generally, two or three components are chosen to reduce or simplify the case selection as this will represent the total variability by 70% or above (Robinette 2012). PCs can also be chosen based on eigenvalues; according to Jolliffe (2002), Hudson and

Zehner (1998), Zehner, Meindl, and Hudson (1993) and da Silva, Zehner, and Hudson (2020), an eigenvalue is a statistical measure of the explained variance and values over 1.0 often indicate a meaningful PC. When considering two PCs, eight boundary cases can be determined: four at the axis intersections and four in each quadrant midpoint. If three PCs are included, 14 boundary cases can be calculated: six at the axes intersections and eight at the octant midpoints (Zehner, Meindl, and Hudson 1993). The procedure to identify the boundary individuals is described in detail in Zehner, Meindl, and Hudson (1993) and Brolin, Högberg, and Hanson (2012).

2.2. Boundary approach: AA

AA was introduced to anthropometric data analysis by Cutler and Breiman (1994) who classified six head dimensions. According to Epifanio, Vinué, and Alemany (2013), in multivariate analysis, where an *nxm* matrix, *X*, represents a dataset with *n* registers and *m* dimensions, the objective of AA is to find a *kxm* matrix, *Z*, that defines *k* archetypal characteristics. Mathematically, this can be expressed as finding two *nxk* coefficient matrices, α and β, that minimise the residual sum of squares (RSS) (Epifanio, Vinué, and Alemany 2013).

$$\begin{aligned}\mathrm{RSS} &= \sum_{i=1}^{n} |X_{\mathrm{i}} - \sum_{j=1}^{k} \alpha_{ij} Z_{\mathrm{j}}|^2 \\ &= \sum_{i=1}^{n} |X_i - \sum_{j=1}^{k} \alpha_{ij} \sum_{l=1}^{n} \beta_{ij} X_{\mathrm{l}}|^2,\end{aligned} \quad (1)$$

where $\sum_{j=1}^{k} \alpha_{ij} = 1$, $\sum_{i=1}^{n} \beta_{ij} = 1$, and only positive values of α_{ij} and β_{ij} are admitted.

As the number of archetypes should be the input of the analysis, the elbow rule has been widely used, choosing the best *k*-value in grouping individuals. This involves plotting a dissimilarity metric, such as the RSS for different *k*-values. The graphical interpretation implies that when the *k*-value approaches the real number of groups, the sum of squared estimated errors (SSE) declines rapidly; when the *k*-value exceeds the real number of groups, the SSE continues to decline but more slowly (Yuan and Yang 2019).

2.3. Distributed cases methods

Cluster analysis is considered to be an unsupervised learning technique and can be classified as either hierarchical or partitional (Jain 2010). Hierarchical algorithms can be agglomerative (starting clusters consisting of one object), such as Ward's algorithm (WA) or divisive (starting clusters containing the complete dataset, subsequently divided into smaller clusters) (King 2015). Partitional algorithms identify all clusters simultaneously and refine the initial data partitions to obtain a given number of clusters (Jain 2010; King 2015), such as *k*-means, the most widely used partitional method (Saitta, Raphael, and Smith 2008; Saxena et al. 2017).

Clustering results can be compared by analysing the difference between clusters (separation) and the similarity between observations within clusters (compactness); the more compact and separated the clusters are, the better the clustering results are (Jegatha Deborah, Baskaran, and Kannan 2010). Some indexes, such as the Davies–Bouldin index (DBI) (Davies and Bouldin 1979), Calinski–Harabaz index (CHI) (Calinski and Harabasz 1974) and silhouette index (SI), have been proposed for clustering evaluation (Rousseeuw 1987).

2.3.1. Ward's clustering

WA is an agglomerative method based on a minimal sum-of-squares criterion. It forms clusters using the minimal distance between objects and cluster centroids (Equations 2 and 3) (Murtagh and Legendre 2014; Konishi 2014). This algorithm starts with *n* one-element subsets (Ward 1963). In each iteration, two clusters are merged into a new subset, so the evaluation of the objective function for each of the $n(n-1)/2$ possible unions of subsets is required to select the groups to be merged (Ward 1963). Classification analyses have been performed in different areas by implementing WA (Rampado et al. 2019; Tagliabue et al. 2016; Chang et al. 2009).

2.3.2. k-Means clustering

To compute the *k*-means algorithm, the initial mean cluster values are required. These can be randomly selected or based on historical data. These data are then used to compute the first iteration where the algorithm assigns each individual to the closest cluster. Once all individuals are assigned to the best cluster, new cluster means are calculated and a new iteration is computed with the updated cluster means. The objective of the algorithm is to minimise the squared error between the empirical mean of a cluster, μ_k, and the observations, x_i, into cluster, C_k, defined by (Jain 2010):

$$\sum_{x_i \in C_k} |x_i - \mu_k|^2, \quad (2)$$

where *k* is the number of clusters. The function to be optimised is the sum of the squared error over all the

Table 1. Anthropometric dimensions.

Anthropometric dimensions	
Sitting height (SH)	Between a sitting surface and the top of the head
Eye height, sitting (EHS)	Vertical distance from the sitting surface and the ectocanthus landmark on the outer corner of the right eye
Acromial height, sitting (AHS)	Vertical distance from the seat surface to the acromion
Popliteal height (PH)	Vertical distance from a footrest surface to the back of the right knee
Knee height, sitting (KHS)	Vertical distance between a footrest surface and the suprapatellar landmark at the top of the right knee
Buttock–popliteal length (BPL)	Horizontal distance between a buttock plate placed at the most posterior point on either buttock and the back of the right knee juncture
Buttock–knee length (BKL)	Horizontal distance between a buttock plate placed at the most posterior point on either buttock and the anterior point of the right knee
Functional grip reach (FGR)	Horizontal distance between the vertical plane of the back and the centre of a dowel gripped in the right hand of a subject standing erect with the back against a wall and the arm and hand extended forward horizontally

clusters, given by (Jain 2010):

$$\sum_{k=1}^{K}\sum_{x_i \epsilon C_k} |x_i - \mu_k|^2. \quad (3)$$

The main limitations of *k*-means is that the squared error decreases as the number of clusters increases (Jain 2010), and its reliability can be reduced when analysing high-dimensional data, owing to using the Euclidian distance (Saitta, Raphael, and Smith 2008). Several studies have used *k*-means to cluster anthropometric data (Suhardi, Khairina, and Fahma 2017; Niu, Li, and Xu 2009) after applying PCA for data reduction (Jianwei et al. 2010; Lee, Chao, and Wang 2013; Zakaria et al. 2008; Vishnu Vardhana Rao, Kumar, and Brahmam 2013).

3. Materials and methods

In this section, the anthropometric survey conducted for this study is described. Additionally, the statistical data analysis, including the univariate and multivariate approaches, is described.

3.1. Anthropometric survey

To obtain the anthropometric information of the population of northwest Mexico, a survey was conducted between 2010 and 2015. A total of 2603 male participants were randomly selected for the survey. Subjects included graduate and postgraduate students, professors and industrial workers between 18 and 61 years old, who resided in northwestern Mexico at the time of the survey. Eight measurements were taken from the complete anthropometric study (Table 1, Figure 1), according to the Gordon and associates protocol (Gordon et al. 1989); five of them were included in most of the multivariate anthropometric analyses found in literature for the crew design (sitting height, eye height sitting, acromial height sitting, knee height sitting and buttock–knee length) (Zehner, Meindl, and Hudson 1993; Brolin, Högberg, and Hanson 2012; Biswal and Dahiya 2019; Jianwei et al. 2010) and sitting workstations. Buttock–popliteal length and popliteal height sitting were also considered to define the seat length and leg vertical space. Functional grip reach was chosen as the most similar to thumb-tip dimension. All participants wore tight clothes to mitigate measurement errors.

3.2. Measurement process and instruments

Anthropometric measurements were taken while subjects sat with knees flexed at 90°, based on the Gordon and associates protocol (Gordon et al. 1989). A three-segment vertical Martin anthropometer, of 2100 mm in length and 1 mm precision, was used to obtain the required dimensions. This instrument is part of the anthropometric kit Clarita I. The measurements were taken using a chair of variable height and length of 85 cm. Only trained individuals comprised the measuring team. To assure the ability of the measurers and reliability and accuracy of the survey, measurements were obtained twice. If the difference between the measured values was below 2%, the average was registered (Lopez et al. 2019).

3.3. Data analysis

3.3.1. Univariate analysis

To compare univariate and multivariate analysis results, the data were standardised using the normal distribution, and 1, 5, 10, 25, 50, 75, 90, 95, and 99% were calculated on each dimension (Table 2). The mean, standard deviation and maximum and minimum values were also calculated. All variable values were expressed in cm. The nearest neighbours to the 1 and 99% were determined using the Euclidean distance.

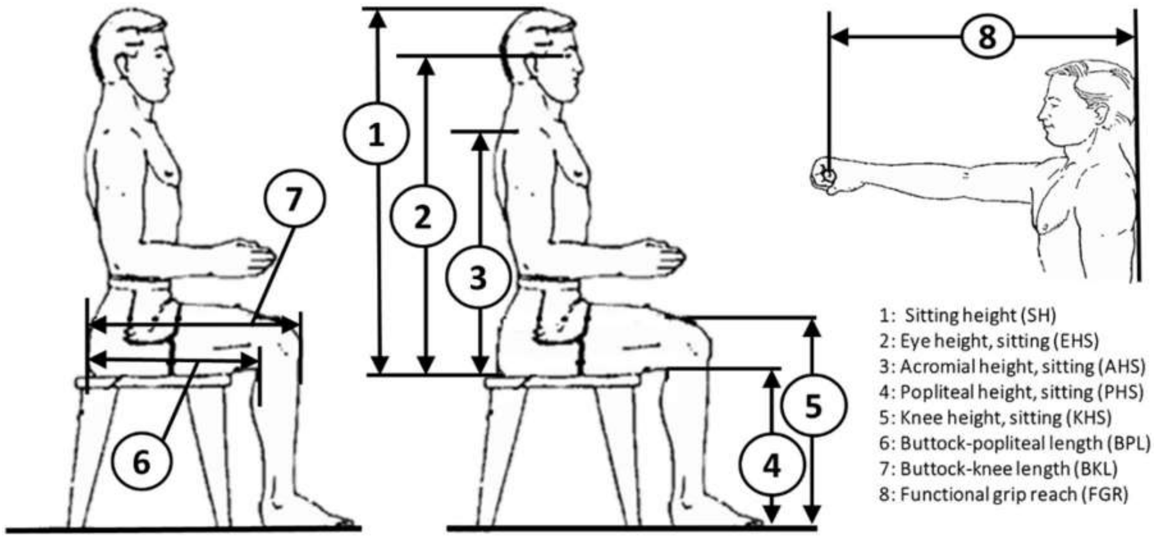

Figure 1. Anthropometric dimensions (Adapted from Gordon et al. 1989).

Table 2. PCA coefficients, eigenvalues and percentage of explained variance.

AD	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
SH	0.42	0.35	−0.07	−0.15	−0.03	−0.37	−0.12	0.72
EHS	0.38	0.43	0.05	−0.13	0.02	−0.22	−0.44	−0.64
AHS	0.34	0.45	0.01	−0.10	−0.08	0.51	0.63	−0.08
PH	0.28	−0.40	−0.50	−0.21	−0.36	−0.40	0.36	−0.22
KHS	0.36	−0.31	−0.46	−0.17	0.33	0.54	−0.37	0.09
BPL	0.29	−0.34	0.56	−0.21	−0.58	0.23	−0.21	0.07
BKL	0.34	−0.33	0.47	−0.10	0.64	−0.24	0.29	−0.07
FGR	0.39	−0.09	−0.02	0.91	−0.09	−0.01	−0.02	−0.01
Eigenvalues	3.46	2.16	1.02	0.53	0.27	0.25	0.20	0.10
% Explained variance	43.31	26.98	12.71	6.68	3.37	3.12	2.55	1.30
Cumulative %	43.31	70.28	82.99	89.67	93.04	96.16	98.70	100.00

AD: anthropometric dimension; SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

Table 3. Basic statistics, $n = 2603$.

AD	Mean	Standard deviation	Minimum	Maximum
SH	89.46	3.26	75.50	101.10
EHS	78.53	3.61	64.00	90.10
AHS	61.02	3.12	50.30	73.60
PH	44.62	3.18	29.40	54.50
KHS	54.01	3.20	32.20	67.50
BPL	47.94	3.23	35.00	67.40
BKL	59.32	3.63	40.00	72.20
FGR	87.47	4.33	66.70	103.20

AD: anthropometric dimension; SH: sitting height; EHS: eye height; sitting; AHS: acromial height; sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

3.3.2. Multivariate accommodation models

The multivariate analysis was performed in two phases: the boundary approach and the distributed approach. In the boundary approach, PCA and AA were implemented. Contrarily, the *k*-means and WA were applied for the distributed approach. All variable values were expressed in cm.

In PCA, the eight anthropometric dimensions were transformed to a new dimensional space, where all components were orthogonal to one another, so that the normality of data did not need to be met. The first three PCs were used to define the body models (Table 3), as the first three components accounted for 82.99% of the total variance – more than the commonly used variability cut-off point of 70% – and their eigenvalues were over 1.0. PC1, which was positive and accounted for 43.31% of the total variation, predicted the overall body size. PC2, accounting for 26.98% of the variation, contrasted the dimensions correlated with the torso (SH, EHS and AHS), lower limbs (PH, KHS, BPL and BKL) and arm reach (FGR). PC3, accounting for 12.71% of the variation, contrasted the calf (PH and KHS) and leg dimensions (BPL and BKL).

Next, the PCA scores were standardised using the normal distribution, and the central 98% sphere – with mean = 0 and standard deviation = 1 – was adjusted to the data, adapting the procedure used by Robinson, Robinette, and Zehner (1992) for the 2D PCA analysis to the 3D PCA analysis. Jolliffe (2002) indicates that using a renormalisation of the PC scores is helpful in the analysis of anthropometric data as it

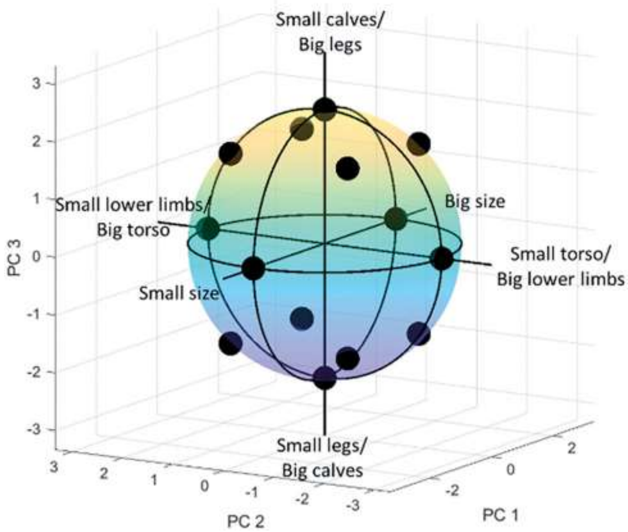

Figure 2. Theoretical boundary cases obtained by PCA.

improves identification of outliers. The boundary cases were represented by the subjects close to the edge of the sphere; hence, a total of 14 boundary cases were identified on the sphere contour. Six of them were at the axes intersections and eight cases were found at the midpoint of each sphere octant (Figure 2).

From the boundary cases on the intersection axes, the real boundary subjects were identified by selecting the nearest neighbour to the case within the sphere, based on the Euclidean distance (Spencer 2013). For octant midpoints, data were segmented into each octant. The Euclidean distance from the midpoint boundary case to every subject in the corresponding octant was computed. The midpoint boundary case's nearest neighbour was chosen as the real midpoint boundary case.

AA was performed in RStudio, using the Anthropometry Package developed by Vinué (2017). The analysis was performed assuming that anthropometric data could be approximated using a normal distribution (Pheasant and Haslegrave 2015). Owing to the number of archetypes (k) being unknown in AA, the analysis was performed for $k = 2, \ldots, 14$ archetypes. The best k-value was determined by applying the elbow rule to the RSS values (points after which the RSS starts decreasing in a linear way). The real boundary cases were defined by the nearest neighbours to the archetypes, based on the Euclidean distance.

The second part of the analysis is related to the distributed approaches: k-means and WA. Clustering analysis was performed on an exploratory basis to find the best subset conformation, considering $k = 2, \ldots, 14$ cases. High values of SI and CHI were preferred, whereas low values of DBI were desirable (Davies and Bouldin 1979; Calinski and Harabasz 1974; Rousseeuw 1987). Once the total population was clustered, the representative individual of each cluster was determined. In order to achieve results based on a boundary approach, as suggested by Brolin et al. (2016), the furthest individual from each population centre was selected.

All the multivariate analyses were performed on the central 98% data, which were obtained from the adjusted sphere using the PCA approach. Then, for both boundary and distributed approaches, the percentile for each anthropometric dimension was calculated based on the complete dataset.

3.3.3. Univariate and multivariate approaches comparison

The comparison of univariate and multivariate approaches was based on the procedure followed by da Silva, Zehner, and Hudson (2020). To determine the differences in the accommodation level obtained by the univariate and multivariate analysis, the 1 and 99% limits were identified for each approach. For the univariate approach, real individuals closest to the 1% and 99% values were defined as the approach limits. For the multivariate approaches, the minimum and maximum anthropometric values obtained from all the cases were selected as the limits for the approach.

In univariate and multivariate approaches, once the limits were identified, each subject in the original database was compared with those limits. A subject was considered accommodated when all their anthropometric dimensions fell within the limits. The resulting accommodation percentage obtained for each approach was compared with the desired level of accommodation (98% = 2551 subjects).

The error (A–E) was computed by subtracting the expected (E) from the achieved (A) accommodated subjects. Positive values indicated that more subjects were accommodated than expected. Two statistical tests were performed: Chi-squared and McNemar's tests. The Chi-squared test ($\alpha = 0.05$) was used for the goodness of fit to determine how well each method estimated the intended accommodation percentage. A p-value of 0.05 was used for testing the null hypothesis (Ho): the model holds for all categories (accommodated vs. not accommodated); Ho is rejected if p-values of the Chi-squared test were smaller than 0.05 (the Chi-squared statistic value was close to the right end). McNemar's test was used to examine if the difference between the achieved accommodation of each method were statistically significant, using matched paired data. If p-value was smaller than $\alpha = 0.05$, the null hypothesis that the

Table 4. Anthropometric dimension (cm) for the basic percentiles ($n = 2603$).

	Theoretical									Real subjects	
AD	1%	5%	10%	25%	50%	75%	90%	95%	99%	1%	99%
SH	82.05	84.20	85.30	87.30	89.50	91.60	93.52	94.80	97.15	80.0	96.1
EHS	70.00	72.60	74.00	76.20	78.30	81.00	83.20	84.50	87.05	69.0	84.2
AHS	54.00	56.00	57.10	59.00	61.00	63.10	64.90	66.04	69.00	54.3	67.3
PH	38.15	39.67	40.58	42.30	44.50	46.80	49.00	50.10	51.70	36.5	50.3
KHS	47.05	49.00	50.20	51.90	54.00	56.10	58.20	59.20	61.25	48.0	62.0
BPL	41.00	43.00	44.00	45.70	47.80	50.00	52.02	53.20	55.65	41.0	52.1
BKL	50.00	53.60	55.00	57.00	59.50	61.70	64.00	65.10	67.45	52.5	66.6
FGR	77.50	80.50	82.00	84.60	87.40	90.20	93.12	94.60	98.15	79.0	94.9

AD: anthropometric dimension; SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach

proportions of subjects captured by each approach were equal, was rejected. In this test, all the multivariate analyses were compared to the percentile approach results.

In summary, the multivariate procedure followed these steps: (1) standardising data; (2) reducing dimensionality; (3) finding boundary cases; (4) finding real boundary/distributed cases and (5) comparing the results of the univariate and multivariate approaches.

4. Results

4.1. Univariate analysis

The basic statistics (mean, standard deviation, minimum and maximum values) shown in Table 3 and Table 4 show the basic percentile values for the complete dataset. The closest individual to the 1 and 99% is also shown in Table 4; those values were defined as the univariate-percentile approach limits. Figure 3 shows the real subjects near to the theoretical 1–99%.

4.2. Multivariate analysis

4.2.1. PCA cases

The closest individual according to the minimum Euclidean distance to the intersection and midpoint cases was defined as the real boundary case. The corresponding percentile values for the cases obtained using PCA are shown in Table 5 and Table 6 present the variable values. Minimum and maximum values were identified and set as the multivariate-boundary-PCA limits. Cases A and B are defined by PC1, big and small individuals respectively, while cases C and B are defined by PC2 contrasting torso and lower limbs. PC3 defines cases D and E, contrasting the leg and calf sizes. Cases G, J, K, and N are closer to the big sizes, while cases H, I, L, and M are closer to the small sizes. Figure 3 shows the 14 PCA real boundary cases.

4.2.2. Archetypal analysis

The RSS for $k = 2, \ldots, 14$ archetypes was obtained and plotted (Figure 4); three points of inflection were defined using the elbow rule: $k = 4$, 6 and 10 archetypes. Tables 7 and 8 show the percentile and variable value for each k-value analysis. Minimum and maximum values were also determined and set as the multivariate-boundary-AA limits. For $k = 4$, case 4 was related to the big size and cases 1 and 3 contrast torso and lower limbs. For $k = 6$, cases 3 and 5, were related to the overall size, while cases 4 and 6 contrast torso and lower limbs. For $k = 10$, cases 2 and 8 referred to the overall size, while cases 2, 5, 9 and 13 were similar to cases D, A, C and J obtained by PCA approach. Figure 3 shows the boundary cases for $k = 4$, 6, 10 and 14.

4.2.3. Distributed models

To select the number of distributed cases, the SI, DBI and CHI values were obtained for $k = 1, \ldots, 14$ cases (Figure 5). In k-means, the best k-values were 3, 6, and 10 according to the SI and CHI calculations, and 7, 6 and 13 were recommended based on the DBI values. For WA, k-values of 3, 4, and 11 were recommended by the SI, 3 and 5 by the CHI calculations, and 11, 12 and 13 were recommended by the DBI calculations. The best or second-best k-value according to each index was chosen to complete the multivariate analysis (k-means: 3, 6 and 10; WA: 3, 5 and 11). Boundary cases obtained for $k = 14$ were also included to compare the values with results obtained by the PCA approach. The following sections describe the results for both algorithms for the recommended k-values.

4.2.3.1. k-Means. Figure 3 shows the boundary cases for $k = 3$, 6, 10 and 14; percentile and variable values are presented in Tables 9 and 10, respectively. Minimum and maximum values were also identified.

For $k = 3$, Case 3 corresponds to small size; while cases 3 and 4, and 2 and 1 for $k = 6$, contrast the overall size and torso and lower limbs, respectively. For $k = 10$, cases 2 and

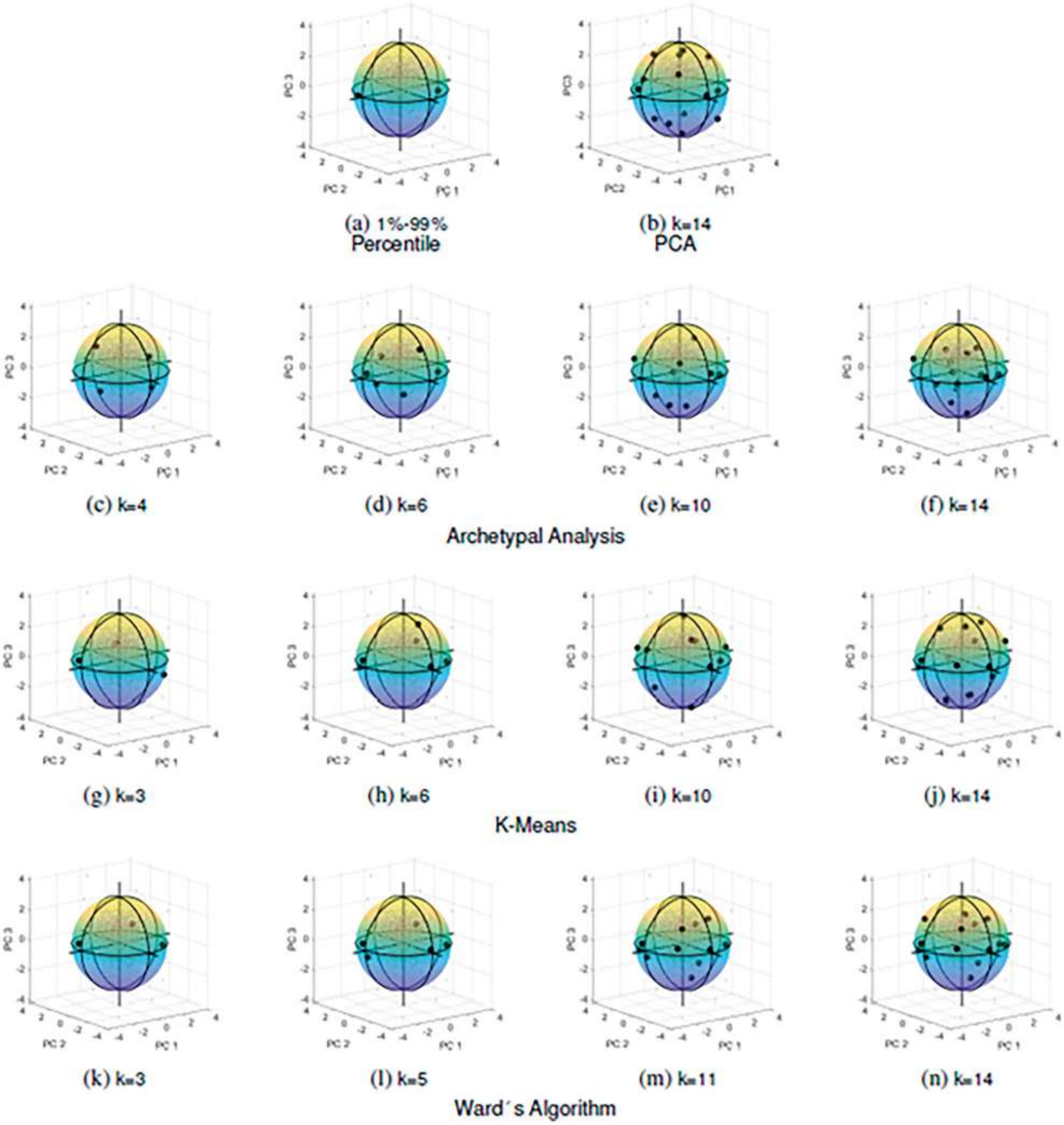

Figure 3. Boundary cases for univariate and multivariate approach.

1, and 8 and 9 contrast the overall size and torso and lower limbs, respectively. For $k = 14$, cases 7 and 11, 6 and 10, and 1, contrast the overall size, the torso and lower limbs, and the calves and legs, respectively.

4.2.3.2. Ward's algorithm. Figure 3 shows the boundary cases for $k = 3$, 5, 11, and 14, and percentile and variable values are presented in Tables 11 and 12, respectively. Minimum and maximum values were also identified.

For $k = 3$, cases 1 and 3 refer to the big and small sizes. For $k = 5$, cases 4 and 5 correspond to big and small sizes, while cases 1 and 2 contrast torso and lower limbs. For $k = 11$, cases 4 and 11, 7, and 9, and 1, contrast the overall size, the torso and lower limbs, and the calves and legs, respectively. For $k = 14$, cases 5 and 6 are related to the big size, while case 4 refers to the small size; cases 12 and 13 contrast torso and lower limbs, and cases 7 and 9 contrast the calves and legs.

Table 5. Percentile values for the PCA cases, $n = 2603$.

	Anthropometric dimension							
Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
A	98	94	98	96	99	90	98	96
B	79	92	97	3	2	3	6	16
C	1	4	6	1	2	5	3	1
D	2	2	2	86	93	96	86	87
E	44	73	50	10	20	95	89	61
F	42	37	75	95	97	11	12	25
G	94	99	100	13	40	84	83	34
H	57	39	77	67	<1	45	33	59
I	2	2	1	41	13	84	77	17
J	79	85	42	48	88	100	99	70
K	98	89	87	93	52	7	23	99
L	46	39	65	26	45	53	<1	70
M	1	2	1	81	45	33	1	27
N	70	24	42	97	100	79	73	100

SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

Table 6. Variable values for PCA cases, $n = 2603$.

	Anthropometric dimension							
Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
A	96.1	84.2	67.3	50.3	62.0	52.1	66.6	94.9
B	92.1	83.6	67.0	38.6	47.5	41.8[a]	53.8	83.1
C	82.0[a]	72.1	56.2	37.6[a]	47.5	42.6	52.6	77.6[a]
D	82.4	71.3	54.6	48.1	58.7	53.5	63.2	92.4
E	89.0	80.7	61.0	40.6	51.3	53.3	63.7	88.7
F	88.8	77.3	63.1	50.0	59.8	44.0	55.0	84.6
G	94.5	87.7[b]	69.1[b]	41.1	53.2	51.1	62.8	85.7
H	90.0	77.5	63.3	46.0	35.0[a]	47.5	57.7	88.5
I	82.5	71.0	53.4[a]	43.9	50.4	51.2	62.0	83.4
J	92.1	82.2	60.4	44.5	57.7	57.6[b]	67.2[b]	89.7
K	96.2[b]	83.0	64.5	49.2	54.2	43.2	56.7	96.9
L	89.1	77.5	62.2	42.6	53.6	48.2	40.0	89.7
M	82.0	70.7[a]	54.0	47.4	53.6	46.5	50.7[a]	84.8
N	91.2	76.0	60.4	50.5[b]	64.4[b]	50.5	61.5	99.5[b]

[a]Minimum values. [b]Maximum values.
All dimensions are in centimetres (cm).
SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

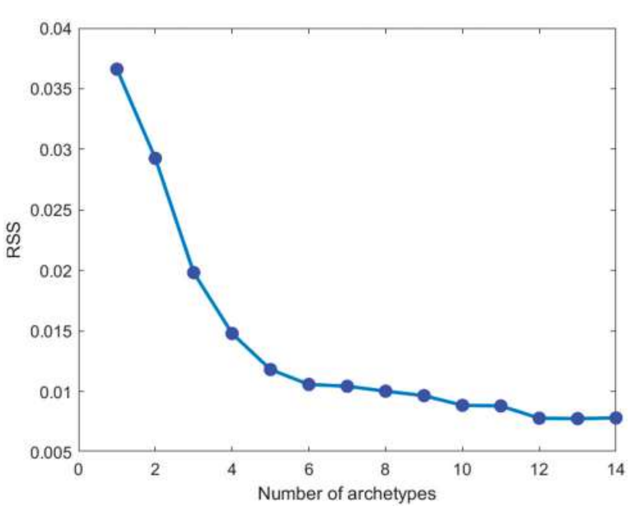

Figure 4. RSS for AA $k = 2, \ldots, 14$.

Table 7. Percentile values for $k = 4$, 6, 10, and 14 archetypes (AA), $n = 2603$.

		Anthropometric dimension							
k	Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
4	1	60	82	60	2	2	34	26	17
	2	5	1	11	27	34	7	6	7
	3	21	36	14	71	89	98	98	84
	4	99	99	95	98	99	69	80	97
6	1	19	48	28	14	14	2	4	65
	2	28	21	19	87	84	30	43	17
	3	5	16	3	7	4	8	7	<1
	4	2	9	4	46	58	96	99	78
	5	98	95	98	97	100	91	98	96
	6	89	87	99	11	14	34	24	13
10	1	38	20	10	99	98	60	54	<1
	2	10	21	21	<1	<1	5	3	2
	3	1	1	1	82	45	32	1	27
	4	93	98	100	22	73	92	95	91
	5	28	29	54	99	<1	100	<1	48
	6	100	100	68	44	83	13	28	38
	7	2	7	3	45	35	53	92	8
	8	93	94	80	99	99	95	96	98
	9	8	7	12	86	90	93	90	100
	10	62	61	78	20	32	<1	5	80
14	1	74	77	54	66	76	8	42	45
	2	1	2	2	87	94	96	86	88
	3	32	59	65	13	35	41	38	63
	4	47	54	60	11	35	3	3	52
	5	93	94	80	99	99	95	96	98
	6	28	29	54	99	<1	100	<1	48
	7	57	48	63	98	85	93	91	22
	8	1	<1	1	62	45	55	36	28
	9	10	21	21	<1	<1	5	3	2
	10	16	14	9	55	11	88	80	91
	11	1	15	80	99	92	2	59	73
	12	93	99	100	8	33	40	46	29
	13	90	85	89	49	72	96	90	88
	14	31	23	20	90	78	23	10	<1

SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height, KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

4.3. Univariate and multivariate accommodation percentage comparison

The resulting accommodation percentage obtained for each approach is shown in Tables 13 and 14. The expected number of subjects to be accommodated (i.e. individuals that meet all the anthropometric dimension limits) for the central 98% accommodation envelope and the error of estimation for each approach are also presented. The univariate-percentile approach provided for 1881 subjects out of the 2603 population, resulting in a loss of 26%, whereas the multivariate approaches accommodated 87–92% of the population. This enclosure difference is due to the ~83% explained variance accepted when choosing the three PCs for the analysis.

The goodness-of-fit test results are shown in Table 15. Large Chi-squared values, close to the upper limit indicate that the model does not fit well, therefore, all null hypotheses are rejected. In other words, all the approaches failed in accommodating the intended range of subjects (p-value <0.05).

McNemar's test was conducted by comparing the proportion of subjects provided by the percentile approach and each of the multivariate analyses. The p-value obtained for the multivariate-distributed Ward analysis for $k=3$, suggested that the proportion of accommodated subjects was similar to that of the univariate percentile, whereas the rest of the tests showed no statistical significance for accommodation ability. Analysing the odds ratio, when compared with the multivariate distributed k-means approach for $k=14$, the odds of not being accommodated using the percentile approach were 85.50 times greater.

Table 8. Variable values for $k=4$, 6, 10 and 14 archetypes (AA), $n=2603$.

		Anthropometric dimension							
k	Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
4	1	90.3	81.7	61.8	38.4[a]	47.5[a]	46.6	57.1	83.4
	2	84.3[a]	70.8[a]	57.3[a]	42.7	52.7	43.3[a]	53.9[a]	81.2[b]
	3	86.9	77.3	57.8	46.3	57.8	54.7[b]	66.3[b]	91.7
	4	96.3[b]	87.0[b]	66.1[b]	51.2[b]	61.8[b]	49.5	62.3	95.5[b]
6	1	86.7	78.4	59.3	41.2	50.7	41.1[a]	53.2[a]	89.1
	2	87.6	75.7	58.4	48.1	57.0	46.3	58.7	83.4
	3	84.2	75.1	55.5[a]	40.0[a]	48.7[a]	43.4	54.1	76.0[a]
	4	82.8[a]	73.8[a]	55.6	44.3	54.6	53.4[b]	67.1[b]	90.7
	5	96.1[b]	84.2[b]	67.3	50.3[b]	62.0[b]	52.1	66.6	94.9[b]
	6	93.4	82.5	67.6[b]	40.8	50.7	46.6	56.8	82.7
10	1	88.5	75.6	57.2	52.5[b]	60.6	48.7	59.7	66.7[a]
	2	85.4	75.7	58.6	36.1[a]	45.3[a]	42.7	52.6	78.5
	3	82.0[a]	70.7[a]	54.0[a]	47.4	53.6	46.5	50.7	84.8
	4	94.1	85.9	69.3[b]	42.2	55.9	52.4	65.3	93.2
	5	87.6	76.6	61.3	51.9	45.5	58.0[b]	45.6[a]	87.2
	6	97.8[b]	89.6[b]	62.4	44.1	56.9	44.4	57.3	86.1
	7	83.0	73.5	55.4	44.2	52.8	48.2	64.4	81.6
	8	94.0	84.0	63.6	51.4	61.4[b]	53.2	65.6[b]	96.4
	9	85.0	73.4	57.4	48.0	57.9	52.5	63.9	100.9[b]
	10	90.4	79.5	63.4	42.0	52.5	37.6[a]	53.4	91.0
14	1	91.5	81.1	61.3	45.9	56.1	43.6	58.6	86.9
	2	82.4	71.3	54.6	48.1	58.7	53.5	63.2	92.4
	3	88.0	79.3	62.2	41.1	52.8	47.2	58.3	88.8
	4	89.2	78.9	61.8	40.8	52.8	42.2	52.7	87.6
	5	94.0	84.0	63.6	51.4	61.4[b]	53.2	65.6[b]	96.4[b]
	6	87.6	76.6	61.3	51.9[b]	45.5	58.0[b]	45.6[a]	87.2
	7	90.0	78.4	62.0	51.3	57.2	52.5	64.0	84.1
	8	81.6[a]	68.9[a]	53.6[a]	45.5	53.6	48.3	58.1	84.9
	9	85.4	75.7	58.6	36.1[a]	45.3[a]	42.7	52.6	78.5
	10	86.4	74.7	57.0	45.0	50.2	51.7	62.3	93.2
	11	82.4	75.0	63.6	51.5	58.2	41.4[a]	60.1	90.0
	12	94.2[b]	86.8[b]	69.2[b]	40.1	52.6	47.1	59.0	85.1
	13	93.5	82.2	64.8	44.5	55.8	53.5	63.8	92.5
	14	87.9	76.0	58.5	48.5	56.3	45.6	54.8	76.4[a]

[a]Minimum values. [b]Maximum values.
All dimensions are in centimetres (cm).
SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

Table 9. Percentile values for $k=3$, 6, 10 and 14 k-means cases, $n=2603$.

		Anthropometric dimension							
k	Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
3	1	79	72	56	100	99	97	93	98
	2	100	99	100	21	41	40	50	86
	3	1	3	6	1	2	5	3	1
6	1	1	2	2	87	94	96	86	88
	2	1	3	6	1	2	5	3	1
	3	96	91	81	98	99	94	98	100
	4	63	38	84	3	2	<1	1	6
	5	2	18	73	22	53	99	98	62
	6	100	98	100	36	75	75	79	98
7	1	80	93	98	3	2	3	6	15
	2	98	95	98	97	100	91	98	96
	3	5	3	8	36	38	6	1	<1
	4	1	1	80	99	92	2	59	73
	5	13	15	16	98	1	100	28	47
	6	59	62	50	95	98	100	100	93
	7	100	98	100	36	75	75	79	98
	8	10	21	21	<1	<1	5	3	2
	9	1	2	2	87	94	96	86	88
	10	87	71	85	13	10	100	46	47
14	1	2	31	18	20	16	96	93	35
	2	58	56	100	54	30	99	96	95
	3	38	20	10	99	98	60	54	<1
	4	<1	<1	<1	50	22	62	41	17
	5	100	98	100	36	75	75	79	98
	6	1	3	6	1	2	5	3	1
	7	99	99	95	98	99	69	80	97
	8	63	38	84	3	2	<1	1	6
	9	20	14	20	68	72	3	2	9
	10	1	2	2	87	94	96	86	88
	11	99	99	100	16	49	10	22	46
	12	65	86	65	92	97	99	100	98
	13	63	91	85	97	93	11	22	86
	14	83	85	72	18	<1	43	41	33

SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

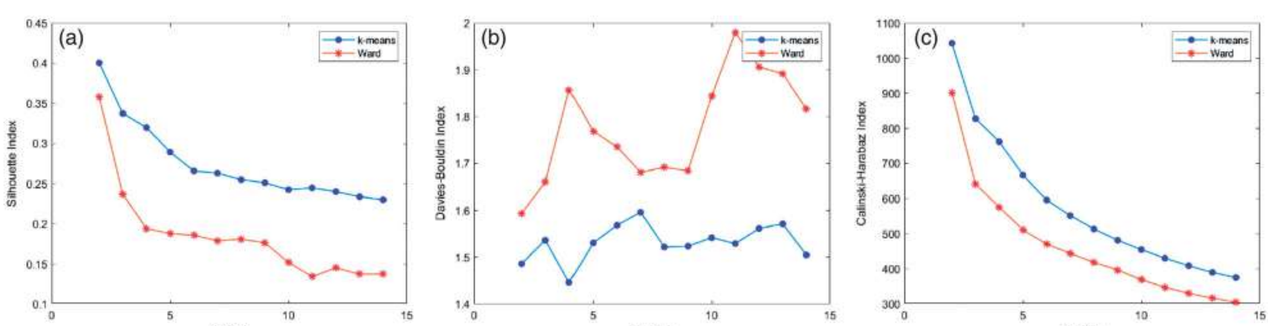

Figure 5. Silhouette, Davies–Bouldin and Calinski–Harabaz plots.

Table 10. Variable values for $k = 3$, 6, 10 and 14 *k*-means cases, $n = 2603$.

		Anthropometric dimension							
k	Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
3	1	92.0	80.6	61.5	54.4[b]	61.3[b]	54.0[b]	64.5[b]	96.5[b]
	2	97.6[b]	87.4[b]	70.9[b]	42.1	53.3	47.1	59.3	92.1
	3	82.0[a]	72.1[a]	56.2[a]	37.6[a]	47.5[a]	42.6[a]	52.6[a]	77.6[a]
6	1	82.4	71.3[a]	54.6[a]	48.1	58.7	53.5	63.2	92.4
	2	82.0[a]	72.1	56.2	37.6[a]	47.5[a]	42.6	52.6	77.6[a]
	3	95.1	83.3	63.7	50.8[b]	61.1[b]	52.8	66.5	101.0[b]
	4	90.5	77.5	64.0	38.7	47.6	39.7[a]	50.4[a]	81.0
	5	83.3	75.3	62.9	42.2	54.2	55.0[b]	66.8[b]	88.7
	6	99.4[b]	85.5[b]	70.7[b]	43.5	56.0	50.1	62.2	96.0
10	1	92.1	83.6	67.0	38.6	47.5	41.8	53.8	83.1
	2	96.1	84.2	67.3	50.3	62.0[b]	52.1	66.6	94.9
	3	84.2	72.0	56.7	43.5	53.0	43.0	51.7[a]	73.5[a]
	4	82.4[a]	70.0[a]	63.6	51.5[b]	58.2	41.4[a]	60.1	90.0
	5	86.0	75.0	58.0	50.8	46.1	58.5[b]	57.3	87.1
	6	90.2	79.6	61.0	49.7	60.4	56.9	68.5[b]	93.6
	7	99.4[b]	85.5[b]	70.7[b]	43.5	56.0	50.1	62.2	96.0[b]
	8	85.4	75.7	58.6	36.1[a]	45.3[a]	42.7	52.6	78.5
	9	82.4	71.3	54.6[a]	48.1	58.7	53.5	63.2	92.4
	10	93.0	80.5	64.2	41.0	50.0	57.6	59.0	87.1
14	1	82.9	76.8	58.2	42.0	50.9	53.3	64.6	85.8
	2	90.1	79.1	69.4	44.9	52.4	55.7[b]	65.6	94.4
	3	88.5	75.6	57.2	52.5[b]	60.6	48.7	59.7	66.7[a]
	4	80.3[a]	67.1[a]	53.0[a]	44.6	51.6	48.9	58.5	83.4
	5	99.4[b]	85.5	70.7[b]	43.5	56.0	50.1	62.2	96.0
	6	82.0	72.1	56.2	37.6[a]	47.5	42.6	52.6	77.6
	7	96.3	87.0[b]	66.1	51.2	61.8[b]	49.5	62.3	95.5
	8	90.5	77.5	64.0	38.7	47.6	39.7[a]	50.4[a]	81.0
	9	86.8	74.7	58.5	46.0	55.8	42.0	51.9	81.8
	10	82.4	71.3	54.6	48.1	58.7	53.5	63.2	92.4
	11	97.0	86.2	69.0	41.5	53.9	43.9	56.6	87.0
	12	90.7	82.3	62.2	49.0	59.7	55.6	70.5[b]	96.3[b]
	13	90.5	83.2	64.1	50.4	58.6	44.1	56.6	92.1
	14	92.5	82.2	62.8	41.7	43.1[a]	47.4	58.5	85.6

[a]Minimum values. [b]Maximum values.
All dimensions are in centimetres (cm).
SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

Table 11. Percentile values for $k = 3$, 5, 11 and 14 Ward's algorithm cases, $n = 2603$.

		Anthropometric dimension							
k	Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
3	1	96	91	81	98	99	94	98	100
	2	100	98	100	36	75	75	79	98
	3	1	3	6	1	2	5	3	1
5	1	44	61	60	6	14	23	<1	24
	2	1	2	2	87	94	96	86	88
	3	100	98	100	36	75	75	79	98
	4	96	91	81	98	99	94	98	100
	5	1	3	6	1	2	5	3	1
11	1	63	91	85	97	93	11	22	86
	2	44	61	60	6	14	23	<1	24
	3	99	92	87	95	89	23	63	95
	4	96	91	81	98	99	94	98	100
	5	1	2	1	41	12	85	78	17
	6	<1	<1	<1	50	22	62	41	17
	7	63	38	84	3	2	<1	1	6
	8	100	98	100	36	75	75	79	98
	9	1	2	2	87	94	96	86	88
	10	86	88	94	98	36	100	1	94
	11	1	3	6	1	2	5	3	1
14	1	99	96	96	21	39	85	82	56
	2	100	98	100	36	75	75	79	98
	3	23	19	54	7	<1	14	15	15
	4	1	3	6	1	2	5	3	1
	5	96	91	81	98	99	94	98	100
	6	98	95	98	97	100	91	98	96
	7	63	91	85	97	93	11	22	86
	8	44	61	60	6	14	23	<1	24
	9	99	92	87	95	89	23	63	95
	10	1	2	1	41	12	85	78	17
	11	<1	<1	<1	50	22	62	41	17
	12	63	38	84	3	2	<1	1	6
	13	1	2	2	87	94	96	86	88
	14	86	88	94	98	36	100	1	94

SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

From the multivariate-boundary-AA results, it can be inferred that the odds of not being accommodated using the percentile approach were 0.48 and 0.4 times lower than the odds of not being accommodated by the $k = 4$ and $k = 6$ AA approaches, respectively.

5. Discussion

Univariate and multivariate approaches were used to obtain the RHM. An advantage of using multivariate over univariate-percentile approaches is that extreme real individuals can be identified, so the 1 and 99 percentiles nearest neighbour were determined. The real human representative models, confirmed that the anthropometric dimension percentile values are not necessarily the same for all the dimensions, as stated by Garneau and Parkinson (2016) and da Silva, Zehner, and Hudson (2020).

Results suggested that the accommodation percentage obtained by the multivariate-boundary-PCA approach enclosed 87% of a central 98% sphere, 15% more than that obtained by the univariate-percentile approach, corroborating the multivariate higher accommodation level obtained by Guan et al. (2012) and Essdai et al. (2018). Similar results of the PCA analysis presented in this study were also obtained by da Silva, Zehner, and Hudson (2020). Their study reveals that the accommodation percentage for multivariate-PCA was 9% less than expected and 12% higher than the univariate-percentile accommodation percentage for the male population, considering two PCs, 85% of explained variance, and a 90% ellipse, using six anthropometric dimensions required for the crew design.

Even though the goodness-of-fit test suggested that all the approaches considered in this study present a significant difference between the desired accommodation percentage and the captured percentage, one method was superior to the others by analysing the Chi-squared values (low Chi-squared values are preferred). Results of the Chi-squared values suggested that $k = 10$ was the most precise AA analysis, accommodating 92% of the population. In the case of the multivariate distributed k-means and WA, the smallest Chi-squares values were obtained with $k = 14$, capturing 92 and 89% of the population, respectively. The AA and k-means showed a small difference in the accommodation percentage. Comparing the results of all the approaches, univariate and multivariate, the most precise approach in estimating the accommodation level was k-means ($k = 14$), followed by AA ($k = 10$), WA ($k = 14$), PCA and finally univariate percentile.

Table 12. Variable values for $k = 3$, 5, 11 and 14 Ward's algorithm cases, $n = 2603$.

		Anthropometric dimension							
k	Case	SH	EHS	AHS	PH	KHS	BPL	BKL	FGR
3	1	95.1	83.3	63.7	50.8[b]	61.1[b]	52.8[b]	66.5[b]	101.0[b]
	2	99.4[b]	85.5[b]	70.7[b]	43.5	56.0	50.1	62.2	96.0
	3	82.0[a]	72.1[a]	56.2[a]	37.6[a]	47.5[a]	42.6[a]	52.6[a]	77.6[a]
5	1	89.0	79.5	61.8	39.7	50.7	45.6	43.7[a]	84.4
	2	82.4	71.3[a]	54.6[a]	48.1	58.7	53.5[b]	63.2	92.4
	3	99.4[b]	85.5[b]	70.7[b]	43.5	56.0	50.1	62.2	96.0
	4	95.1	83.3	63.7	50.8[b]	61.1[b]	52.8	66.5[b]	101.0[b]
	5	82.0[a]	72.1	56.2	37.6[a]	47.5[a]	42.6[a]	52.6	77.6[a]
11	1	90.5	83.2	64.1	50.4	58.6	44.1	56.6	92.1
	2	89.0	79.5	61.8	39.7	50.7	45.6	43.7[a]	84.4
	3	97.5	83.5	64.5	49.8	57.8	45.6	60.5	94.5
	4	95.1	83.3	63.7	50.8[b]	61.1[b]	52.8	66.5[b]	101.0[b]
	5	82.5	71.0	53.4	43.9	50.4	51.2	62.0	83.4
	6	80.3[a]	67.1[a]	53.0[a]	44.6	51.6	48.9	58.5	83.4
	7	90.5	77.5	64.0	38.7	47.6	39.7[a]	50.4	81.0
	8	99.4[b]	85.5[b]	70.7[b]	43.5	56.0	50.1	62.2	96.0
	9	82.4	71.3	54.6	48.1	58.7	53.5	63.2	92.4
	10	92.8	82.6	65.7	50.7	52.9	67.4[b]	50.6	94.0
	11	82.0	72.1	56.2	37.6[a]	47.5[a]	42.6	52.6	77.6[a]
14	1	96.4	84.8	66.2	42.1	53.1	51.2	62.5	88.1
	2	99.4[b]	85.5[b]	70.7[b]	43.5	56.0	50.1	62.2	96.0
	3	87.1	75.5	61.3	39.9	40.6[a]	44.5	55.7	83.0
	4	82.0	72.1	56.2	37.6[a]	47.5	42.6	52.6	77.6[a]
	5	95.1	83.3	63.7	50.8[b]	61.1	52.8	66.5	101.0[b]
	6	96.1	84.2	67.3	50.3	62.0[b]	52.1	66.6[b]	94.9
	7	90.5	83.2	64.1	50.4	58.6	44.1	56.6	92.1
	8	89.0	79.5	61.8	39.7	50.7	45.6	43.7[a]	84.4
	9	97.5	83.5	64.5	49.8	57.8	45.6	60.5	94.5
	10	82.5	71.0	53.4	43.9	50.4	51.2	62.0	83.4
	11	80.3[a]	67.1[a]	53.0[a]	44.6	51.6	48.9	58.5	83.4
	12	90.5	77.5	64.0	38.7	47.6	39.7[a]	50.4	81.0
	13	82.4	71.3	54.6	48.1	58.7	53.5	63.2	92.4
	14	92.8	82.6	65.7	50.7	52.9	67.4[b]	50.6	94.0

[a]Minimum values. [b]Maximum values.
All dimensions are in centimetres (cm).
SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

The McNemar's test suggested that the multivariate approaches were more inclusive than the univariate-percentile approach, when considering a large k-value (close to 14). For a small k-value ($k = 3$), the multivariate-boundary and multivariate-distributed approaches obtained similar accommodation percentages to the univariate approach and lower when using AA. These results agree with da Silva, Zehner, and Hudson (2020), who found that the odds of not being accommodated for 90% of male were higher when using percentile than using the PCA.

This study faced two main limitations. First, this procedure was based on using the maximum and minimum values to test the theoretical multivariate accommodation percentage; this is still a univariate way to perform the analysis. Therefore, the statistical analysis in this study is not able to compare all the data points simultaneously (da Silva, Zehner, and Hudson 2020). A solution to this limitation would be to test cases in real or simulated environments, which is out of scope for this study. Second, only three PCs, accounting for ~83% of the variability, were considered in the multivariate analysis to reduce the complexity of the analysis and the dimensionality of the original dataset; this impacted the multivariate accommodation percentage results.

Further, an advantage of using multivariate over univariate approaches, is that the former can produce models that can be used in more realistic and accurate simulations and evaluation of workplace design. This is done by using a combination of different percentile values excluding only the most extreme human representations.

6. Conclusions and future work

Although distributed-multivariate approaches are mostly implemented to find values across the centre

Table 13. Univariate-percentile and multivariate-boundary PCA theoretical accommodation percentage, $n = 2603$.

		Percentile			PCA		
AD	Expected number	Achieved number	Univariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A–E)	Accommodation, %
SH	2551	2542	−9	97.7	2523	−28	96.9
EHS	2551	2434	−117	93.5	2501	−50	96.1
AHS	2551	2380	−171	91.4	2474	−77	95.0
PH	2551	2279	−272	87.6	2370	−181	91.0
KHS	2551	2232	−319	85.7	2367	−184	90.9
BPL	2551	2035	−516	78.2	2307	−244	88.6
BKL	2551	1973	−578	75.8	2286	−265	87.8
FGR	2551	1881	−670	72.3	2258	−293	86.7

AD: anthropometric dimension; SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

Table 14. Multivariate-boundary-AA and multivariate-distributed k-means and WA theoretical accommodation percentage, $n = 2603$.

	AA												
		$k = 4$			$k = 6$			$k = 10$			$k = 14$		
AD	Expected number	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %
SH	2551	2411	−140	92.6	2499	−52	96.0	2577	26	99.0	2411	−140	92.6
EHS	2551	2400	−151	92.2	2207	−344	84.8	2554	3	98.1	2409	−142	92.5
AHS	2551	2105	−446	80.9	2166	−385	83.2	2522	−29	96.9	2384	−167	91.6
PH	2551	2041	−510	78.4	1934	−617	74.3	2507	−44	96.3	2364	−187	90.8
KHS	2551	2016	−535	77.4	1907	−644	73.3	2483	−68	95.4	2343	−208	90.0
BPL	2551	1847	−704	71.0	1822	−729	70.0	2471	−80	94.9	2304	−247	88.5
BKL	2551	1774	−777	68.2	1764	−787	67.8	2388	−163	91.7	2228	−323	85.6
FGR	2551	1655	−896	63.6	1703	−848	65.4	2387	−164	91.7	2189	−362	84.1
	k-means												
		$k = 3$			$k = 6$			$k = 10$			$k = 14$		
AD	Expected number	Achieved number	Multivariate Error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %
SH	2551	2560	9	98.3	2588	37	99.4	2588	37	99.4	2588	37	99.4
EHS	2551	2470	81	94.9	2525	−26	97.0	2525	−26	97.0	2585	34	99.3
AHS	2551	2367	−184	90.9	2509	−42	96.4	2509	−42	96.4	2565	14	98.5
PH	2551	2349	−202	90.2	2464	−87	94.7	2358	−193	90.6	2546	−5	97.8
KHS	2551	2302	−249	88.4	2439	−112	93.7	2321	−230	89.2	2522	−29	96.9
BPL	2551	2138	−413	82.1	2314	−237	88.9	2259	−292	86.8	2453	−98	94.2
BKL	2551	1993	−558	76.6	2239	−312	86.0	2237	−314	85.9	2442	−109	93.8
FGR	2551	1949	−602	74.9	2196	−355	84.4	2183	−368	83.9	2389	−162	91.8
	WA												
		$k = 3$			$k = 5$			$k = 11$			$k = 14$		
AD	Expected number	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %	Achieved number	Multivariate error (A-E)	Accommodation, %
SH	2551	2571	20	98.8	2571	20	98.8	2588	37	99.4	2588	37	99.4
EHS	2551	2434	−117	93.5	2474	−77	95.0	2525	−26	97.0	2525	−26	97.0
AHS	2551	2332	−219	89.6	2444	−107	93.9	2509	−42	96.4	2509	−42	96.4
PH	2551	2247	−304	86.3	2358	−193	90.6	2421	−130	93.0	2421	−130	93.0
KHS	2551	2214	−337	85.1	2321	−230	89.2	2384	−167	91.6	2413	−138	92.7
BPL	2551	2002	−549	76.9	2147	−404	82.5	2375	−176	91.2	2404	−147	92.4
BKL	2551	1950	−601	74.9	2128	−423	81.8	2325	−226	89.3	2355	−196	90.5
FGR	2551	1932	−619	74.2	2104	−447	80.8	2294	−257	88.1	2324	−227	89.3

AD: anthropometric dimension; SH: sitting height; EHS: eye height, sitting; AHS: acromial height, sitting; PH: popliteal height; KHS: knee height, sitting; BPL: buttock-popliteal length; BKL: buttock-knee length; FGR: functional grip reach.

Table 15. Univariate and multivariate goodness-of-fit and McNemar's test results, $n = 2603$.

	Intended accommodation		Accommodated subjects		Goodness-of-fit test		McNemar's test	
Approach	Frequency	%	Frequency	%	Chi-squared	*p*-Value	Odds ratio[a]	*p*-Value
Univariate								
Percentile	2551	98	1881	72	8808.66	0.0000		
Multivariate boundary PCA								
$k = 14$	2551	98	2258	87	1684.60	0.0000	9.57	0.0000
Multivariate boundary AA								
$k = 4$	2551	98	1655	64	15,753.48	0.0000	0.48	0.0000
$k = 6$	2551	98	1703	65	14,110.81	0.0000	0.40	0.0000
$k = 10$	2551	98	2387	92	527.77	0.0000	24.00	0.0000
$k = 14$	2551	98	2189	84	2571.45	0.0000	5.40	0.0000
Multivariate distributed *k*-means								
$k = 3$	2551	98	1949	75	7111.37	0.0000	1.33	0.0023
$k = 6$	2551	98	2196	84	2472.96	0.0000	9.17	0.0000
$k = 10$	2551	98	2183	84	2657.39	0.0000	12.78	0.0000
$k = 14$	2551	98	2389	92	514.98	0.0000	85.50	0.0000
Multivariate distributed Ward								
$k = 3$	2551	98	1932	74	7518.68	0.0000	1.28	0.0146
$k = 5$	2551	98	2104	81	3920.81	0.0000	3.19	0.0000
$k = 11$	2551	98	2294	88	1296.06	0.0000	52.63	0.0000
$k = 14$	2551	98	2324	89	1011.14	0.0000	64.29	0.0000

[a]Odds ratio for percentile.

of data, *k*-means and WA obtained similar models to boundary approaches when the representative subject of the cluster was identified as the furthest individual from the centre data. The results validated that multivariate approaches obtain higher accommodation levels when performing a theoretical accommodation percentage comparison. Using distributed-multivariate approaches and selecting the subject furthest from the centre data as the real case, a higher number of subjects can be accommodated.

Different model families were produced using different approaches; they can be compared with actual products and adjustable workplace design – such as cab design and automatic adjustable workstations considered in Industry 4.0 – to help fit them to the population of northwest Mexico.

Future studies could include evaluation of the actual design of workstations, based on univariate approaches, using the models obtained herein with real-world or simulated-environment designs and determine how available adjustments influence the ergonomic risk. Additionally, subsequent studies could consider different anthropometric dimensions – i.e. for standing workstations, hand, foot and head models. This study considered basic anthropometry; thus, reproducing it using 3 D anthropometry analysis and range of motion could be of interest. This could be helpful in simulation environments for designing, evaluating and improving products and workstations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the first author fellowship from Programa para el Desarrollo Profesional Docente (PRODEP) as Professor of the University of Sonora.

References

Abdali, O., H. Viktor, E. Paquet, and M. Rioux. 2004. "Exploring Anthropometric Data through Cluster Analysis." SAE Technical Paper No. 2004-01-2187. Warrendale, PA: Society of Automotive Engineers.

Bertilsson, E., D. Högberg, and L. Hanson. 2012. "Using Experimental Design to Define Boundary Manikins." *Work* 41 (Suppl. 1): 4598–4605. doi:10.3233/WOR-2012-0075-4598.

Biswal, P. and Y. S. Dahiya. 2019. "Principal Component Analysis: The Path Ahead for Aircrew-Aircraft Compatibility at the Institute of Aerospace Medicine." *Indian Journal of Aerospace Medicine* 63 (1): 8–15.

Bittner, A. C., F. A. Glenn, R. M. Harris, H. P. Iavecchia, and R. J. Wherry. 1987. "CADRE: A Family of Manikins for Workstation Design." *Trends in Ergonomics/human Factors* IV: 733–7400.

Bittner, A. C. 2000. "A-CADRE: Advanced Family of Manikins for Workstation Design." Paper presented at the Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society: "Ergonomics for the New Millennium", San Diego, CA, USA, July 29–August 4, 774–777.

Brolin, E., D. Högberg, and L. Hanson. 2012. "Description of Boundary Case Methodology for Anthropometric Diversity Consideration." *International Journal of Human Factors Modelling and Simulation* 3 (2): 204–223. doi:10.1504/IJHFMS.2012.051097.

Brolin, E., D. Högberg, L. Hanson, and R. Örtengren. 2016. "Generation and Evaluation of Distributed Cases by Clustering of Diverse Anthropometric Data." *International Journal of Human Factors Modelling and Simulation* 5 (3): 210. doi:10.1504/IJHFMS.2016.079706.

Calinski, T., and J. Harabasz. 1974. "A Dendrite Method for Cluster Analysis." *Communications in Statistics – Theory and Methods* 3 (1): 1–27. doi:10.1080/03610927408827101.

Chang, J. S., C. S. Yoo, S. H. Yi, K. H. Hong, H. S. Oh, J. Y. Hwang, S.-G. Kim, Y. M. Ahn, and Y. S. Kim. 2009. "Differential Pattern of Heart Rate Variability in Patients with Schizophrenia." *Progress in Neuro-Psychopharmacology & Biological Psychiatry* 33 (6): 991–995. doi:10.1016/j.pnpbp.2009.05.004.

Cutler, A., and L. Breiman. 1994. "Archetypal Analysis." *Technometrics* 36 (4): 338–347. doi:10.1080/00401706.1994.10485840.

da Silva, G. V., G. F. Zehner, and J. A. Hudson. 2020. "Comparison of Univariate and Multivariate Anthropometric Design Requirements Methods for Flight Deck Design Application." *Ergonomics* 63 (9): 1133–1149. doi:10.1080/00140139.2020.1765029.

Davies, D. L., and D. W. Bouldin. 1979. "A Cluster Separation Measure." *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-1, no.2*, pp. 224–227. doi:10.1109/TPAMI.1979.4766909.

Dianat, I., J. Molenbroek, and H. I. Castellucci. 2018. "A Review of the Methodology and Applications of Anthropometry in Ergonomics and Product Design." *Ergonomics* 61 (12): 1696–1720. doi:10.1080/00140139.2018.1502817.

Epifanio, I., G. Vinué, and S. Alemany. 2013. "Archetypal Analysis: Contributions for Estimating Boundary Cases in Multivariate Accommodation Problem." *Computers & Industrial Engineering* 64 (3): 757–765. doi:10.1016/j.cie.2012.12.011.

Epifanio, I., M. V. Ibáñez, and A. Simó. 2018. "Archetypal Shapes Based on Landmarks and Extension to Handle Missing Data." *Advances in Data Analysis and Classification* 12 (3): 705–735. doi:10.1007/s11634-017-0297-7.

Essdai, A., V. K. Spasojević Brkić, T. Golubović, A. Brkić, and V. Popović. 2018. "Crane Cabins' Interior Space Multivariate Anthropometric Modeling." *Work* 59 (4): 557–570. doi:10.3233/WOR-182706.

Friess, M. 2005. "Multivariate Accommodation Models Using Traditional and 3D Anthropometry." SAE Technical Paper No. 2005-01-2735. Warrendale, PA: Society of Automotive Engineers.

Garneau, C. J., and M. B. Parkinson. 2016. "A Survey of Anthropometry and Physical Accommodation in Ergonomics Curricula." *Ergonomics* 59 (1): 143–154. doi:10.1080/00140139.2015.1052853.

Gordon, C. C., B. D. Corner, and J. D. Brantley. 1997. "Defining Extreme Sizes and Shapes for Body Armor and Load-Bearing Systems Design: Multivariate Analysis of US Army Torso Dimensions." Technical Report No. TR-97/012. Natick, MA: U.S. Army Natick R&D Center.

Gordon, C. C., T. Churchill, C. E. Clauser, B. Bradtmiller, J. T. McConville, I. Tebbetts, and R. A. Walker. 1989. *Anthropometric Survey of U.S. Army Personnel: Summary Statistics, Interim Report for 1988*. Yellow Springs, OH: Anthropology Research Project Inc.

Guan, J., H. Hsiao, B. Bradtmiller, T.-Y. Kau, M. R. Reed, S. K. Jahns, J. Loczi, H. L. Hardee, and D. P. T. Piamonte. 2012. "U.S. Truck Driver Anthropometric Study and Multivariate Anthropometric Models for Cab Designs." *Human Factors* 54 (5): 849–871. doi:10.1177/0018720812442685.

Hanson, L., L. Sperling, G. Gard, S. Ipsen, and C. Olivares Vergara. 2009. "Swedish Anthropometrics for Product and Workplace Design." *Applied Ergonomics* 40 (4): 797–806. doi:10.1016/j.apergo.2008.08.007.

Haslegrave, C. M. 1986. "Characterizing the Anthropometric Extremes of the Population." *Ergonomics* 29 (2): 281–301. doi:10.1080/00140138608968265.

Hsiao, H. 2013. "Anthropometric Procedures for Protective Equipment Sizing and Design." *Human Factors* 55 (1): 6–35. doi:10.1177/0018720812465640.

Hudson, J. A., and G. F. Zehner. 1998. "The USAF Multivariate Accommodation Method." Paper presented at the Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting, Chicago, IL, USA, October 5–9, 722–726.

Jain, A. K. 2010. "Data Clustering: 50 Years beyond K-Means." *Pattern Recognition Letters* 31 (8): 651–666. doi:10.1016/j.patrec.2009.09.011.

Jegatha Deborah, L., R. Baskaran, and A. Kannan. 2010. "A Survey on Internal Validity Measure for Cluster Validation." *International Journal of Computer Science & Engineering Survey* 1 (2): 85–102. doi:10.5121/ijcses.2010.1207.

Jianwei, N., H. Yiling, L. Muyuan, Z. Xin, R. Linghua, C. Chuzhi, and Z. Baoqin. 2010. "A Comparative Study on Application of Data Mining Technique in Human Shape Clustering: Principal Component Analysis vs. Factor Analysis." Paper presented at the 2010 5th IEEE Conference on Industrial Electronics and Applications, Taichung, Taiwan, June 15–17.

Jolliffe, I. T. 2002. *Principal Component Analysis*. 2nd ed. New York: Springer-Verlag.

Jung, K., O. Kwon, and H. You. 2009. "Development of a Digital Human Model Generation Method for Ergonomic Design in Virtual Environment." *International Journal of Industrial Ergonomics* 39 (5): 744–748. doi:10.1016/j.ergon.2009.04.001.

Jung, K., O. Kwon, and H. You. 2010. "Evaluation of the Multivariate Accommodation Performance of the Grid Method." *Applied Ergonomics* 42 (1): 156–161. doi:10.1016/j.apergo.2010.06.004.

King, R. S. 2015. *Cluster Analysis and Data Mining: An Introduction*. Herndon, VA: Mercury Learning and Information.

Konishi, S. 2014. *Introduction to Multivariate Analysis: Linear and Nonlinear Modeling* (1st ed.). New York: Chapman and Hall/CRC. https://doi.org/10.1201/b17077

Lee, Y.-C., W.-Y. Chao, and M.-J. Wang. 2013. "Developing a New Foot Shape and Size System for Taiwanese Females." In *The 19th International Conference on Industrial Engineering and Engineering Management*, edited by E. Qi, J. Shen, and R. Dou, 959–968. Berlin/Heidelberg: Springer.

Lin, Y.-C., M.-J. J. Wang, and E. M. Wang. 2004. "The Comparisons of Anthropometric Characteristics among Four Peoples in East Asia." *Applied Ergonomics* 35 (2): 173–178. doi:10.1016/j.apergo.2004.01.004.

Lopez, M., E. de la Vega, E. Ramirez, A. Characa, J. M. Velarde, and G. E. Baez. 2019. *Antropometría Para el*

Diseño de Puestos de Trabajo. Primera. Cd. Obregón, México: Editorial ITSON.

Murtagh, F., and P. Legendre. 2014. "Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?" *Journal of Classification* 31 (3): 274–295. doi:10.1007/s00357-014-9161-z.

Niu, J., Z. Li, and S. Xu. 2009. "Comparisons of 3D Shape Clustering with Different Face Area Definitions." In *Digital Human Modeling. ICDHM 2009. Lecture Notes in Computer Science*, edited by V. G. Duffy, Vol. 5620, 55–63. Berlin/Heidelberg: Springer.

Pheasant, S., and C. M. Haslegrave. 2015. *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. Boca Raton, FL: CRC Press.

Rampado, O., L. Gianusso, C. R. Nava, and R. Ropolo. 2019. "Analysis of a CT Patient Dose Database with an Unsupervised Clustering Approach." *Physica Medica* 60: 91–99. doi:10.1016/j.ejmp.2019.03.015.

Robinette, K. M. 2012. "Anthropometry for Product Design." In *Handbook of Human Factor and Ergonomics*, edited by G. Salvendy, 4th ed., 330–381. Hoboken, NJ: Wiley.

Robinette, K. M., and J. T. McConville. 1981. "An Alternative to Percentile Models." SAE Technical Paper Series No. 810217. Warrendale, PA: Society of Automotive Engineers.

Robinson, J. C., K. M. Robinette, and G. F. Zehner. 1992. *User's Guide to Anthropometric Data Base at the Computerized Anthropometric Research and Design (CARD) Laboratory*. 2nd ed., ALTR-1992-0036. Dayton, OH: Crew Systems Directorate, Human Engineering Division.

Rousseeuw, P. J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *The Journal of Computational and Applied Mathematics* 20 (C): 53–65. doi:10.1016/0377-0427(87)90125-7.

Saitta, S., B. Raphael, and I. F. C. Smith. 2008. "A Comprehensive Validity Index for Clustering." *Intelligent Data Analysis* 12 (6): 529–548. doi:10.3233/IDA-2008-12602.

Saxena, A., M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C.-T. Lin. 2017. "A Review of Clustering Techniques and Developments." *Neurocomputing* 267: 664–681. doi:10.1016/j.neucom.2017.06.053.

Simo, A., A. Alcacer, I. E. Id, and M. V. Iba. 2020. "A Data-Driven Classification of 3D Foot Types by Archetypal Shapes Based on Landmarks." *PLOS One* 15 (1): e0228016.

Spencer, N. H. 2013. *Essentials of Multivariate Data Analysis* (1st ed.). New York: Chapman and Hall/CRC. https://doi.org/10.1201/b16344

Suhardi, B., A. Khairina, and F. Fahma. 2017. "Designing Size of Batik Shirt for Male College Student in Surakarta Using Anthropometric Data (Case Study: Male College Student of Industrial Engineering Major in Surakarta)." Paper presented at the 2016 2nd International Conference of Industrial, Mechanical, Electrical, and Chemical Engineering (ICIMECE), Yogyakarta, Indonesia, October 6–7, 192–195.

Tagliabue, M., R. Bagatin, G. Capannelli, C. Cattaneo, A. Bottino, R. Firpo, A. Jezowska, and A. Voena. 2016. "Multivariate Comparison of Reverse Osmosis and Nanofiltration Membranes through Tree Cluster Analysis." *Desalination and Water Treatment* 57 (48–49): 23273–23279. doi:10.1080/19443994.2015.1117824.

Vasu, M., and A. Mital. 2000. "Evaluation of the Validity of Anthropometric Design Assumptions." *International Journal of Industrial Ergonomics* 26 (1): 19–37. doi:10.1016/S0169-8141(99)00060-8.

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

Vinué, G., T. León, S. Alemany, and G. Ayala. 2014. "Looking for Representative Fit Models for Apparel Sizing." *Decision Support Systems* 57 (1): 22–33. doi:10.1016/j.dss.2013.07.007.

Vishnu Vardhana Rao, M., S. Kumar, and G. N. V. Brahmam. 2013. "A Study of the Geographical Clustering of Districts in Uttar Pradesh Using Nutritional Anthropometric Data of Preschool Children." *The Indian Journal of Medical Research* 137 (1): 73–81.

Ward, J. H. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58 (301): 236–244. doi:10.1080/01621459.1963.10500845.

Wichansky, A. M. 2000. "Usability Testing in 2000 and beyond." *Ergonomics* 43 (7): 998–1006. doi:10.1080/001401300409170.

Young, K., S. Margerum, A. Barr, M. A. Ferrer, and S. Rajulu. 2008. "Derivation of Boundary Manikins: A Principal Component Analysis." SAE Technical Paper No. 724. Warrendale, PA: Society of Automotive Engineers.

Yuan, C., and H. Yang. 2019. "Research on K-Value Selection Method of K-Means Clustering Algorithm." *Multidisciplinary Scientific Journal* 2 (2): 226–235. doi:10.3390/j2020016.

Zakaria, N., S. M. Jamil, N. Taib, Y. T. Yong, and Y. B. Wah. 2008. "Using Data Mining Technique to Explore Anthropometric Data towards the Development of Sizing System." Paper presented at the 2008 International Symposium on Information Technology, Vol. 2, 1–7, Kuala Lumpur, Malaysia, August 26–28.

Zehner, G. F., R. S. Meindl, and J. A. Hudson. 1993. "A Multivariate Anthropometric Method for Crew Station Design: Abridged (U)." Technical Report No. AL-TR-1992-0164. Louisville, OH: Crew Systems Directorate, Human Engineering Division, Armstrong Laboratory.