Preface

Ergonomics is the scientific discipline that studies the interactions between human beings and the elements of a system and presents multiple applications in areas such as clothing and footwear design or both working and household environments. In each of these sectors, knowing the anthropometric dimensions of the current target population is fundamental to ensure that products suit as well as possible most of the users who make up the population. Anthropometry refers to the study of the measurements and dimensions of the human body and it is considered a very important branch of Ergonomics because its considerable influence on the ergonomic design of products [12].

Human body measurements have usually been taken using rules, calipers or measuring tapes. These procedures are simple and cheap to carry out. However, they have one major drawback: the obtained body measurements and consequently, the human shape information, is imprecise and inaccurate. Furthermore, they always require interaction with real subjects, which increases the measure time and data collecting. The development of new three-dimensional (3D) scanning techniques has represented a huge step forward in the way of obtaining anthropometric data. This technology allows 3D images of human shape to be captured and at the same time, generates highly detailed and reproducible anthropometric measurements.

The great potential of these new scanning systems for the digitalization of human body has contributed to promoting new anthropometric studies in several countries, such as United Kingdom, Australia, Germany, France or USA, in order to acquire accurate anthropometric data of their current population. In this context, in 2006 the Spanish Ministry of Health commissioned a 3D anthropometric survey of the Spanish female population, following the agreement signed by the Ministry itself with the Spanish associations and companies of manufacturing, distribution, fashion design and knitted sectors. A sample of 10.415 Spanish females from 12 to 70 years old randomly selected from the official Postcode Address File was measured. The two main objectives of this study, which was conducted by the Biomechanics Institute of Valencia, were the following: on the one hand, to characterize the shape and body dimensions of the current Spanish women population to develop a standard sizing system that could be used by all the clothing designers. On the other hand, to promote a healthy image of beauty through the representation of suited mannequins [1]. In order to tackle both objectives, Statistics plays an essential role. Thus, the statistical methodologies presented in this PhD work have been applied to the database obtained from the Spanish anthropometric study.

Clothing sizing systems classify the population into homogeneous groups (size groups) based on some key anthropometric dimensions. All members of the same group are similar in body shape and size, so they can use the same garment. In addition, members of different groups are very different with respect to their body dimensions. An efficient and optimal sizing system aims at accommodating as large a percentage of the population as possible, in the optimum number of size groups that better describes the shape variability of the population. Besides, the garment fit for the accommodated individuals must be as good as possible. A very valuable reference related to sizing systems is the book *Sizing in clothing: Developing effective sizing systems for ready-to-wear clothing*, by Susan Ashdown [2]. Each clothing size is defined from a person whose body measurements are located toward the central value for each of the dimensions considered in the analysis. The central person, which is considered as the size representative (the size prototype), becomes the basic pattern from which the clothing line in the same size is designed.

Clustering is the statistical tool that divides a set of individuals in groups (clusters), in such a way that subjects of the same cluster are more similar to each other than to those in other groups [9]. In addition, clustering defines each group by means of a representative individual. Therefore, it arises in a natural way the idea of using clustering to try to define an efficient sizing system. Specifically, four of the methodologies presented in this PhD thesis aimed at segmenting the population into optimal sizes, use different clustering methods. The first one, called *trimowa*, has been published in *Expert Systems with Applications* [8]. It is based on using an especially defined distance to examine differences between women regarding their body measurements. The second and third ones (called *biclustAnthropom* and *TDDclust*, respectively) will soon be submitted in the same paper [16]. *BiclustAnthropom* adapts to the field of Anthropometry a clustering method addressed in the specific case of gene expression data [11]. Moreover, *TDDclust* uses the concept of statistical depth [10] for grouping according to the most central (deep) observation in each size. As mentioned, current sizing systems are based on using an appropriate set of anthropometric dimensions, so clustering is carried out in the Euclidean space. In the three previous proposals, we have always worked in this way. Instead, in the fourth and last approach, called *kmeansProcrustes*, a clustering procedure is proposed for grouping taking into account the women shape, which is represented by a set of anatomical markers (landmarks). For this purpose, the statistical shape analysis [5] will be fundamental. This contribution has been submitted for publication [18].

A sizing system is intended to cover the so-called "standard" population, discarding those individuals with extreme sizes (both large and small). In mathematical language, these individuals can be considered outliers. An outlier is an observation point that is distant from other observations. In our case, a person with extreme anthopometric measurements would be considered as a statistical outlier. Clothing companies usually design garments for the standard sizes so that their market share is "optimal". Nevertheless, with their foreign expansion, a lot of brands are spreading their collection and they already have a special sizes section. In last years, Internet shopping has been an alternative for consumers with extreme sizes looking for clothes that follow trends. The custom-made fabrication is other possibility with the advantage of making garments according to the customers' preferences. The four aforementioned methodologies (*trimowa*, *biclustAnthropom*, *TDDclust* and *kmeansProcrustes*) have been adapted to only accommodate the "standard" population.

Once a particular garment has been designed, the assessing and analysis of fit is performed using one or more fit models. The fit model represents the body dimensions selected by each company to define the proportional relationships needed to achieve the fit the company has determined. The definition of an efficient sizing system relies heavily on the accuracy and representativeness of the fit models regarding the population to which it is addressed. In this PhD work, a statistical approach is proposed to identify representative fit models. It is based on another clustering method originally developed for grouping gene expression data. This method, called *hipamAnthropom*, has been published in *Decision Support Systems* [17]. From well-defined fit models and prototypes, representative and accurate mannequins of the population can be made.

Unlike clothing design, where representative cases correspond with central individuals, in the design of working and household environments, the variability of human shape is described by extreme individuals, which are those that have the largest or smallest values (or extreme combinations) in the dimensions involved in the study. This is often referred to as the accommodation problem. A very interesting reference in this area is the book entitled *Guidelines for Using Anthropometric Data in Product Design*, published by The Human Factors and Ergonomics Society [3]. The idea behind this way of proceeding is that if a product fits extreme observations, it will also fit the others (less extreme). To that end, in this PhD thesis we propose two methodological contributions based on the statistical archetypal analysis. An archetype in Statistics is an extreme individual that is obtained as a convex combination of other subjects of the sample [4]. The first of these methodologies has been published in *Computers & Industrial Engineering* [6], whereas the second one has been submitted for publication [14].

The outline of this PhD report is as follows:

Chapter 1 reviews the state of the art of Ergonomics and Anthropometry and introduces the anthropometric survey of the Spanish female population.

Chapter 2 presents the *trimowa*, *biclustAnthropom* and *hipamAnthropom* methodologies.

In Chapter 3 the *kmeansProcrustes* proposal is detailed.

The *TDDclust* methodology is explained in Chapter 4.

Chapter 5 presents the two methodologies related to the archetypal analysis.

Since all these contributions have been programmed in the statistical software R [13], Chapter 6 presents the **Anthropometry** R package [15], that gathers together all the algorithms associated with each approach.

In this way, from Chapter 2 to Chapter 6 all the methodologies and results included in this PhD thesis are presented.

At last, Chapter 7 brings together the most important conclusions.

Conclusions

This PhD dissertation has been raised in order to be a rigorous scientific contribution, from the mathematical and statistical point of view, to the disciplines of Ergonomics and Anthropometry. Throughout this report, we have developed different statistical methodologies that may be useful to improve the ergonomic design of products, focusing on the efficient design of clothes and working places.

Current sizing systems used by the apparel industry are not accurately optimized to properly fit the target population. As a consequence, a large part of the population, especially women, does not find clothing that fit well, even after trying several garments. This results in a poor fit, unsold garments and a less competitive business. Furthermore, many people return bought clothes because they are not satisfied with them. Because of this, there are many obsolete stocks. A striking effect of this circumstance in Spain is the proliferation of the so-called *outlet* stores. One of the main problems to develop new patterns and designs is the lack of updated anthropometric data of the current population. Outdated size charts contribute to sizing variance between companies. In this context, The Spanish National Institute of Consumer Affairs (INC, in Spanish) of the Spanish Ministry of Health and Consumer Affairs commissioned in 2006 a 3D anthropometric survey of the Spanish female population, in the frame of the agreement signed with the main Spanish clothing companies. The study was performed by the Biomechanics Insitute of Valencia and the anthropometric information recorded was both 1D and 3D. The main motivation was to characterize the shape and body dimensions of the current Spanish women.

Each of the new statistical approaches presented in this PhD work has been applied to the anthropometric database obtained from this Spanish survey. In this way, one of the main goals of this doctoral work consisted in developing mathematical and statistical techniques and tools for the exploitation of human body databases with a focus on the ergonomic design and functional evaluation of products. This dissertation is part of the activities carried out by the research project related to the Spanish anthropometric survey and can be considered as an example of industry-academia interaction, that help to highlight top-level research and encourage excellence in science.

In Ergonomics and Anthropometry, the body size variability within the target population is characterized through the definition of a limited number of cases. An anthropometric case may be a particular human being or a combination of measurements. Depending on the design problem, there are three types of cases: central, boundary and distributed. Our proposed methods have been developed aimed at identifying central and boundary cases.

The approaches based on clustering algorithms, *trimowa*, *biclustAnthropom*, *hipamAnthropom*, *kmeansProcrustes* and *TDDclust*, allow to define 1D and 3D central cases, which actually are representative statistical models or prototypes (and fit models in the case of *hipamAnthropom*) of the human body of the target population. These prototypes and fit models can be used to make more realistic store mannequins. The five aforementioned methodologies have followed the same scheme. Firstly, the selected data matrix was segmented using a primary control dimension (bust circumference in the case of *trimowa*, *hipamAnthropom*, *kmeansProcrustes* and *TDDclust*, and waist circumference in the case of *biclustAnthropom*). Then, a further segmentation using other secondary control anthropometric variables is carried out. In this way, the first segmentation provides a first easy input to choose the size, while the resulting clusters (subgroups) for each bust (or waist) and other anthropometric measurements optimize sizing. By using a more appropriate statistical strategy, such as clustering, homogeneous subgroups are generated taking into account the anthropometric variability of the secondary dimensions that have a relevant influence on garment fit. Every method has been adapted to only accommodate the “standard” population. In order to choose the primary and control secondary body dimensions, the *European standard to sizing system. Size designation of clothes. Part 3: Measurements and intervals* [7] (EN 13402-3-2004) has been used. This Standard is drawn up by the European Union and it is a set of guidelines for the textile industry. The text, whose compliance is desirable but not obligatory, promotes the implementation of a clothing sizing sytem, adapted to the users, based on the consideration of three body parameters: bust, waist and hip circumference, depending on height.

The comfort and wellbeing feeling and a trendy design are the key elements for consumers to proceed to purchase. A garment can only be comfortable to the wearer if the fit is good. In the case of protective clothing and sportswear, good fit is mandatory to ensure the safety and performance of the user. Clothing fit should be improved with a better garment labelling. Apparel companies should offer consumers a truthful and not confusing information on the garment sizes that they wish to offer for sale, so that people could recognise their size with facility. An understandable labelling system could incorporate pictograms indicating the body measurements associated with the garment, see Fig. 1 for an example based on [2]. This labelling proposal could speed up the purchase process.

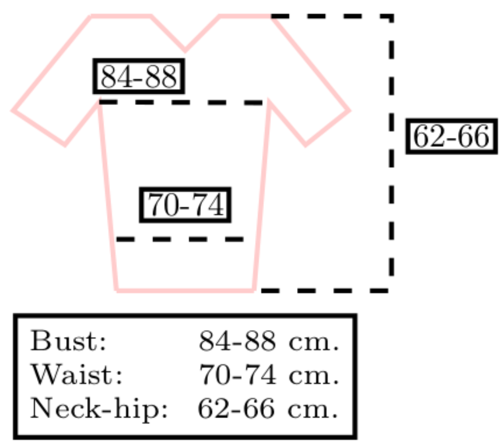

Figure 1: A clothing labelling proposal for upper garments, based on [2].

On the other hand, the approaches based on the statistical archetypal analysis allow to identify boundary cases, that is to say, the individuals who present extreme body measurements. The basic idea is that accommodating boundary cases will accommodate the people who fall within the boundaries (less extreme population). This strategy is valuable in all those problems of human-computer interaction, for example, the design of plane cockpits or truck cabins. When designing workstations or evaluating manual work, it is common to use only a few human models (extreme cases) as virtual test individuals. In this PhD thesis, we have been able to demonstrate that archetypal analysis is a better statistical alternative to determine extreme cases regarding the common used PCA-approach. Unlike PCA, archetypal analysis ensures intended accommodation levels. In addition, the user can decide the number of archetypes to consider or leave the selection by a criterion. In the literature, there is an ongoing discussion about whether the archetypes should be represented by a real observation instead of the standard output, since using the archetypal method, they may or may not be. In some problems, it is crucial that the archetypes are real subjects, observations of the sample, and not fictitious. In this PhD work, a new archetypal concept has been introduced to tackle this problem: the archetypoid. It has been presented an efficient computational algorithm to calculate them and it has been demonstrated some of their advantages regarding classical archetypes. Archetypal and archetypoid analysis could improve industry practice when using human model tools for the design of products and work environments.

All computational algorithms associated with the methods presented in this PhD report have been gathered together into an R package called **Anthropometry**, which is freely available on the Comprehensive R Archive Network at http://cran.r-project.org/package=Anthropometry.

Bibliography

[1] S. Alemany, J. C. González, B. Nácher, C. Soriano, C. Arnáiz, and H. Heras. 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.

[2] S.P. Ashdown. *Sizing in clothing: Developing effective sizing systems for ready-to-wear clothing*. Woodhead Publishing in Textiles, 2007.

[3] HFES 300 Committee. *Guidelines for Using Anthropometric Data in Product Design*. Human Factors and Ergonomics Society, 2004.

[4] Adele Cutler and Leo Breiman. Archetypal Analysis. *Technometrics*, 36(4):338–347, November 1994.

[5] I. E. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, 1998.

[6] I. Epifanio, G. Vinué, and S. Alemany. Archetypal analysis: Contributions for estimating boundary cases in multivariate accommodation problem. *Computers & Industrial Engineering*, 64:757–765, 2013.

[7] European Committee for Standardization. Size designation of clothes. Part 3: Measurements and intervals, 2005.

[8] M.V. Ibáñez, G. Vinué, S. Alemany, A. Simó, I. Epifanio, J. Domingo, and G. Ayala. Apparel sizing using trimmed PAM and OWA operators. *Expert Systems with Applications*, 39(12):10512 – 10520, 2012.

[9] Kaufman, L. and Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley, New York, 1990.

[10] R.Y. Liu, J.M. Parelius, and K. Singh. Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *The Annals of Statistics*, 27(3):783–858, 1999.

[11] S.C. Madeira and A.L. Oliveira. Biclustering Algorithms for Biological Data Analysis: A Survey. *IEEE Transactions on Computational Biology and Bioinformatics*, 1:24–45, 2004.

[12] Stephen Pheasant. *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. Taylor & Francis, Ltd, 2003.

[13] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[14] G. Vinué, I. Epifanio, and S. Alemany. Archetypoids: A new approach to define representative archetypal data. Submitted, 2013.

[15] G. Vinué, I. Epifanio, A. Simó, M.V. Ibáñez, J. Domingo, and G. Ayala. *Anthropometry: An R Package for Analysis of Anthropometric Data*, 2014. R package version 1.0.

[16] G. Vinué and M.V. Ibañez. Data depth and biclustering applied to anthropometric data: Exploring their utility in apparel design. In progress, 2013.

[17] G. Vinué, T. León, S. Alemany, and G. Ayala. Looking for representative fit models for apparel sizing. *Decision Support Systems*, 57:22–33, 2014.

[18] G. Vinué, A. Simó, and S. Alemany. The k-means algorithm for 3D shapes with an application to apparel design. Submitted, 2013.