A COPULA-BASED METHOD TO SYNTHESIZE
ANTHROPOMETRIC DATA OF TARGET USER POPULATIONS

A Thesis in
Engineering Design
by
Maria Alessandra Nusiner Bassi

© 2017 Maria Alessandra Nusiner Bassi

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2017
The thesis of Maria Alessandra Nusiner Bassi was reviewed and approved* by the following:

Matthew B. Parkinson  
Associate Professor of Engineering Design and Mechanical Engineering  
Thesis Adviser

Nicholas Meisel  
Assistant Professor of Engineering Design and Mechanical Engineering

Sven G. Bilén  
Professor of Engineering Design, Electrical Engineering, and Aerospace Engineering  
Head of the School of Engineering Design, Technology, and Professional Programs

*Signatures are on file in the Graduate School.
Abstract

The objective of this thesis is to create a method to synthesize the anthropometry of individuals within a target user population through the use of a copula model. This method uses summary statistics from a population’s data in order to synthesize detailed data on body size and shape. The motivation for this method is that, sometimes, designers do not have access to detailed body size and shape (anthropometry) data of target user populations. Instead, these data are available through reports of some key percentiles and/or means and standard deviations (summary statistics). These, while useful to a certain extent, do not contain enough information to carry out the analysis of complex designs. The method developed in this study seeks to overcome the limitations of previous synthesis procedures.

The proposed method is evaluated through a case study that simulates a scenario in which only some key percentiles of a dataset of anthropometry are available. This case study involves the use of two different sets of anthropometric data: U.S. military in 1988 and Japanese in the 1990’s. The results of the evaluation of this method are shown in the form of mean absolute error values in order to compare the distances between the copula synthesized data and the original databases. Statistical analysis in the form of correlation coefficients, Q-Q plots, and a Chi-squared test are also carried out in order to ensure significance of these results.
Contents

List of Figures vi

List of Tables viii

Chapter 1
Introduction 1

1.1 Research Objectives ........................................ 4
1.2 Research Outline ........................................... 5

Chapter 2
Background and Literature Review 6

2.1 Design for Human Variability .......................... 7
2.1.1 Univariate vs. Multivariate Design Practice .......... 7
2.2 Use of Specific Populations in Anthropometric Studies .... 9
2.3 Sources of Detailed Anthropometry ................... 10
2.3.1 U.S. Army Anthropometry Survey (ANSUR) Data ...... 11
2.3.2 Agency of Industrial Science and Technology Industrial Products Research Institute (AIST) Data ............. 12
2.4 Sources of Summary Statistics .......................... 13
2.5 Data Synthesis Methods ................................. 13
2.5.1 Proportionality Constants ............................. 13
2.5.2 Regression Models ..................................... 15
2.5.3 Quantile-Based Methods .............................. 15
2.6 Previous Copula Models ................................. 16
2.7 Background Summary ...................................... 17
List of Figures

2.1 Visualization of univariate analysis for a design variable X when accommodating for a specific measure in a target population as presented by Garneau and Parkinson [1]. .............................. 7

2.2 Visualization of a multivariate (stature and BMI) analysis when accommodating for a specific measure in a target user population as presented by Garneau and Parkinson [1]. .............................. 8

2.3 Anthropometric measures used in this study as illustrated in the Open Lab’s Anthropometric Data Explorer Tool [2] in reference to the ANSUR anthropometric database [3]. .............................. 11

2.4 Japanese anthropometry used in this study as illustrated by the AIST manual[4]. .............................. 12

2.5 Examples of Drillis and Contini’s [5] proportionality constants as illustrated by Fromuth and Parkinson [6]. .............................. 14

3.1 Example of five key percentile values taken from a single body measure on a density distribution. .............................. 20

3.2 Flow diagram of the mapping process of a copula model. .............................. 21

3.3 Flow chart of the proposed copula method to synthesize anthropometric data from summary statistics. This illustration was made specifically for the copula method presented in this study but based off of a similar flow chart seen in Nadadur and Parkinson’s previous research on a quantile-based approach to synthesize anthropometric data [7]. .............................. 23

4.1 Separation method of both halves of a normal distribution to determine the standard deviations (σ₁ and σ₂) used in as the lower (green) and upper (blue) sections of the distributions of each body measure. .............................. 25

4.2 Final step of synthesizing all measure values by combining the dependence model and the information known from the summary statistics using two different standard deviations per measure. .............................. 26
4.3 Overlapping distributions of the single- vs. two-standard deviation methods. .............................................................. 28
4.4 Side by side comparison of the symmetry in probability density functions of a Gaussian (left) vs. non-Gaussian (right) distribution case. .............................................................. 29

5.1 Example of the synthesized head circumference of females (left) and males (right) plotted with its corresponding source data. ........... 38

6.1 Q-Q plots of the female and male Japanese data (specifically, bi-acromial breadth, buttock circumference, head circumference, and forearm-hand length). This comparison was done between the original and the synthesized Japanese anthropometry. ......................... 41
6.2 Q-Q plots of the female and male Japanese data (specifically, head circumference, knee height, sitting height, and waist circumference). This comparison was done between the original and the synthesized Japanese anthropometry. ................................. 42
6.3 Q-Q plots of the female and male ANSUR data (specifically, bi-acromial breadth, buttock circumference, head circumference, and forearm-hand length). This comparison was done between the original and the synthesized ANSUR anthropometry. ......................... 43
6.4 Q-Q plots of the female and male ANSUR data (specifically, head circumference, knee height, sitting height, and waist circumference). This comparison was done between the original and the synthesized ANSUR anthropometry. ......................... 44
List of Tables

3.1 Gathering $k$ percentile values from $l$ different body measures from a target population. ............................................ 21

5.1 Percentile values for the male (M) and female (F) ANSUR [3] anthropometric data. These measures were selected to serve as the percentile values provided in many anthropometric reports and will be input data of the copula-based synthesis method. ............... 32

5.2 Percentile values for the male (M) and female (F) AIST [4] anthropometric data. These measures were selected to serve as the percentile values provided in many anthropometric reports and will be input data of the copula-based synthesis method. .... 33

5.3 Comparison of mean absolute error values of the one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese female synthesized anthropometric values. ......................... 35

5.4 Comparison of mean absolute error values between one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese male synthesized anthropometric values. .......................... 35

5.5 Comparison of the percent error values between one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese female synthesized anthropometric values. ......................... 36

5.6 Comparison of the percent error values between one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese male synthesized anthropometric values. ......................... 36

6.1 Mean absolute error values for male and female ANSUR and Japanese synthesized anthropometry anthropometry using a two-standard deviation method. .............................................. 39

6.2 Percent Error values for male and female ANSUR and Japanese synthesized anthropometry anthropometry using a two-standard deviation method. .............................................. 40
I dedicate this work to my family and friends. I dedicate all of my accomplishments to my loving mother, Iraida Bassi for being my role model and always encouraging me to go after my goals. Special thanks go to my dad and grandmother, Richard Piralla and Gladys Rosalia Bastidas, for always being by my side and supporting me. I also dedicate this thesis to my peers in the Engineering Design program and other friends who I made while at Penn State. I would not have made it through all the stresses without you all.
Chapter 1

Introduction

This thesis presents a method to synthesize anthropometric datasets (i.e., body measures) for target user populations through the use of a copula model. The proposed model takes summary statistics (percentile values and/or means and standard deviations) of anthropometric data from a target population in order to synthesize anthropometry that possess similar characteristics as the target population. Synthesis methods have been proposed to estimate anthropometric measures when needed. In many cases, such methods synthesize anthropometric data that cannot be used in the case of intricate design.

Anthropometric databases consist of the values of body dimensions measured from a sample of individuals of a certain population. Detailed anthropometric data serve as guides of user requirements that relate to body size and shape. Through detailed analysis of anthropometry, designers are able to make decisions on the physical attributes of designs. In some instances, anthropometric data are published in the form of summary statistics through technical reports and other research publications. This creates a problem for designers as they are not able to perform accurate analyses on intricate designs. With detailed databases of anthropometry, designers are able to make well-informed decisions on the physical attributes of a design. These attributes give designers the necessary information to ensure that a high percentage of a target population will be able to use a product/design comfortably. Put differently, designers use detailed anthropometric data in order to ensure that a high percentage of a target population will be able to fit into a specific design.
“One cannot design for an unknown series of men” [8]. This is what McFarland et al. stated when gathering anthropometric data from commercial bus and truck drivers for the design of bus and truck seats. Products that interact directly with the human body must be designed and tailored to a specific population of users. A similar example of this concept can be seen in a study by Mehta et. al [9] of male farmers in India. Here, anthropometric measures were taken from Indian farmers in order to give recommendations on the dimensions of the tools that this population of farmers use. McFarland’s [8] and Mehta’s [9] studies focused on the concept of designing for a targeted population. In these studies, anthropometric measures were taken from chosen populations (U.S. male and female truck and bus drivers and Indian farmers, respectively) in order to design specifically for them. This is done since it is known that populations possess different anthropometric characteristics from each other [10, 11]. Gender, ethnicity, age, and physical activity all play a role in the body size and shape of specific groups of people. For example, research has shown that, over time, height decreases in adults due to spinal shrinkage [12]. Additionally, an individual’s weight is dependent on their gender and varies during their lifetime [13]. Differences in anthropometry based on ethnicity are also seen throughout populations. For example, on average, Blacks in the U.S. present higher amounts of bone and lean mass when compared to other ethnicities [14]. Given that there exist differences in anthropometry based on specific user populations, when a product is designed to fit a given percentage of a target population, designers must use the anthropometry of such population.

Anthropometric data are used in designs that interact with the human body to accommodate them to a given level (e.g., 90%) of a specific population. The accommodation level of a design measures how well a design will fit a population. For example, if a designer aims to accommodate 90% of a target user population in the design of a new car seat, it means that 90% percent of the population will be able to fit in every aspect of this design (i.e., legroom, headroom, and seat width) [15]. An accommodation level of 90% will only be achieved if 90% of the population fits a design on every relevant measure [16]. That is to say, if a person is accommodated in a car seat on legroom and headroom, but not on seat width, then this individual is not accommodated by this design. Hence the need to analyze all the measures that are part of the same design simultaneously in order to be able
to properly accommodate a population.

Copula models have not yet been used to synthesize anthropometric measure values from summarized statistics. Instead, copula models are commonly used to predict trends and similar data behavior. Fields such as finance, applied statistics, insurance services, and economics have benefited from the use of copulas [17]. This study proposes the use of a copula model in the field of anthropometry by estimating the anthropometry of individuals within a population of virtual users.

Copulas are comprised of two parts: a marginal distribution function, which describes the probability distribution, and a copula function, which outline the dependence of the distribution [18]. Copulas are able to model anthropometric data by creating a dependence model which will then be completed with the information taken from the published summary statistics.

A copula approach to synthesize the anthropometry of a given population is described in this study. This method is evaluated through case studies using anthropometric measures from the *Anthropometric Survey of U.S. Army Personnel* (ANSUR) [3] as well as anthropometric measures from the Japanese youth population gathered by the *Agency of Industrial Science and Technology Industrial Products Research Institute* (AIST) [4].

The advantage of using a two-standard deviation method is presented in this new copula model. A justification for such method is also presented by comparing it against a single-standard deviation method. This shows a general improvement in accuracy in the model when using a two-standard deviation method in the new copula approach presented in this study.

A case study was carried out using summary statistics taken from two dissimilar sources of detailed anthropometric data. This enables a synthetic evaluation and validation. This is done by comparing the new copula synthesized data to the original anthropometric data. Both copula synthesized data were compared to the original datasets from which the percentiles were taken. The evaluation demonstrates that this method can work for dissimilar anthropometric databases such as the ones used in this study.
1.1 Research Objectives

The primary motivation for this study is that some anthropometric data for specific populations are published in the form of summary statistics instead of complete, detailed datasets. These data are often published in the form of summary statistics for several reasons; in some cases for convenience when publishing information on the anthropometry of a population, and other times because of legal/financial issues with releasing such data. This presents a problem for designers and researchers since only limited data analysis (e.g. univariate analysis) can be done with summary statistics. This is a problem because univariate analyses are not appropriate for most designs, which are complex in the way that several measurements must be taken into account at the same time. Multivariate analyses should be carried out when the accommodation of a given design depends on the interaction of more than one measurement. Differences between univariate and multivariate analyses are explained in detail in Section 2.1.1.

The proposed study presents a method for synthesizing the anthropometry of individuals of a population. This method is able to create a dataset of detailed anthropometry using \( k \) percentiles from the summary statistics of \( l \) body measures. In addition, the proposed method is able to synthesize anthropometric data if only means and standard deviations are published in the summarized statistics.

This study considers the proposed method through the use of a case study. For convenience, the case study uses a total of \( l = 8 \) body measures from the U.S. military dataset (ANSUR) [3] and the same measures from the Japanese youth population (AIST) [4]. Specifically, biacromial breadth, buttock circumference, chest circumference, forearm-hand length, head circumference, knee height (sitting), and waist circumference (omphalion) are used in this case study. Specific percentiles (\( k = 5 \)) were taken from these same eight measures in both datasets \((l \times k = 40)\) in order to build and validate the proposed method. From this, the copula model will synthesize a total of 2,000 individuals of an anthropometric dataset, each fully defined on the eight selected measures.

There are a total of three research objectives in this current study. These research objectives are as follows:

Objective 1: Review current methods of anthropometric synthesis. A literature
review has been conducted in order to explore what current methods are available to synthesize anthropometric data.

**Objective 2:** Create a new method to synthesize anthropometry for individuals of a particular population. Copulas are investigated as a modeling strategy for the synthesis of anthropometric data.

**Objective 3:** Validate the copula-based method. The proposed method is validated by observing the distances (percent errors) between the synthesized anthropometry and the original data from which the summary statistics were taken.

### 1.2 Research Outline

This Introduction describes the motivation for the work. Chapter 2 presents an extensive literature review and background information on the different uses of anthropometric databases as well as existing synthesis techniques. This second chapter also presents the importance of using target populations in design and the uses of anthropometric data in accommodation. Chapter 3 presents the method for constructing a copula model to synthesize anthropometric data. Chapter 4 shows the method for estimating distributions in the published data and justifies the use of a two-standard deviation method over a single-standard deviation method in Section 4.2. Chapter 5 tests the proposed method through a case study in which summary statistics are taken from two dissimilar anthropometric sources of data. These summary statistics are used in this case study to synthesize databases of anthropometry for a pre-determined number of individuals. Chapter 6 presents the evaluation of the results from the case study in the form mean absolute error, percent error values, correlation coefficients, Q-Q plots, and a Chi-squared test. Similarly, Chapter 6 contains an analysis and interpretation of the results of this study. Finally, Chapter 7 concludes with findings from this study, current limitations, and proposes suggestions for future work.
Chapter 2

Background and Literature Review

The objective of anthropometric design practice is to find accommodation of designs on target populations with the use of detailed anthropometry. A problem exists when the anthropometric data of target user populations are only made available in the form of summary statistics. Without detailed anthropometry, designers cannot perform multivariate analyses on complex designs. For this reason, this study proposes a copula-based method for synthesizing anthropometry using summary statistics. This method employs copulas because of their ability to combine data from different sources and map random variables to a standard uniform distribution [17]. This property of copulas is desired since the proposed model needs to be able to use available data to assign these values to a distribution of choice.

In the past, copulas have been used in a variety of fields other than anthropometric studies. Fields such as finance, statistics, insurance, and economics [19] have benefited from the use of copula models to predict risks and trends. They have not, however, been used to synthesize anthropometric measures of user populations. A study by Cao [20] studied the use of copula models to predict gender and weight from other anthropometric data. Although this study used anthropometric data in order to predict other specific anthropometric measures, it achieved this by using complete datasets. Instead, this study aims to simulate a situation where only summary statistics are made available from the start but the detailed anthropometric data are needed.
2.1 Design for Human Variability

“Accommodation is the condition wherein a user can interact with a device in a preferred or safe way” [21]. Different methods for achieving accommodation have been developed in the past. When designing for human variability, designers seek to “fit” a design to a target population. Methods such as univariate and multivariate analyses seek to determine the physical attributes of a certain design to accommodate populations based on their own differences in anthropometry. The following subsection explain the differences between univariate and multivariate methods.

2.1.1 Univariate vs. Multivariate Design Practice

Univariate analysis is a method used to determine the boundaries within a design to accommodate a particular population. This method uses boundaries (e.g., 5th and 95th percentiles) in the anthropometric measures in order to determine the physical boundaries of such design. Univariate analyses are done one measure at a time [1]. For example, if the design aims to accommodate 90% of a target population in terms of sitting height, hip breadth, and knee height, these three measures will be evaluated separately for accommodation. However, using univariate analysis in...
this case will result in less than 90% accommodation of the design for a target user population since, once all the boundaries are combined, several individuals on the boundaries of the data will be disaccommodated by the design [22, 23]. Figure 2.1 shows a method for visualizing accommodation fit in an univariate case proposed by Garneau and Parkinson [1].

Multivariate analyses are used in order to ensure that accommodation is achieved when more than one measure is affected by the design simultaneously. In contrast with univariate analysis, multivariate analysis considers all the relevant body measures at the same time [24]. It does this to ensure that the target user population will fit the design at the given accommodation level.

Multivariate analyses present a much more complex scenario than univariate analyses. Garneau and Parkinson [1] proposed a method for visualizing accommodation that showed the differences between univariate and multivariate analyses (Figures 2.1 and 2.2). It aids for the visualization of this complex concept and provides designers with a tool to picture the accommodation of spaces of more than one anthropometric measure at a time.
2.2 Use of Specific Populations in Anthropometric Studies

Anthropometric studies provide a better understanding of the human body dimensions and proportions of a target population. The anthropometric data collected in these studies are often used by designers and researchers to design and assess products that will achieve accommodation of a specific user population. Generally, designers will utilize the most suitable set of data depending on the end-user of the product to be designed [10]. For example, if the goal is to design for farmers in the U.S., the anthropometric data used to carry out such designs should represent the farmer population of the U.S. Similarly, if the goal is to design for an Italian soccer team, the anthropometric data used to make a design decision should represent the population of soccer players in Italy. This is attributed to the fact that anthropometric measures are known to be different across different groups of people. For example, a study by Wang et al. looked at the differences in anthropometry between Asian and white Americans [11]. This study concluded that, on average, Asians have a lower body mass index (BMI) than whites. Additionally, this study found that when looking at the U.S. and the Japanese adult populations, on average, the U.S. adult population had higher stature than the Japanese adult population [11]. Due to these anthropometric differences, tasks and products must be tailored to the specific needs of the population for which the designs are made.

Once the target user population is selected, the anthropometric data needs to be obtained. In some cases, such as the case of the 1988 U.S. Army Anthropometry Survey (ANSUR) [3] and the Agency of Industrial Science and Technology Industrial Products Research Institute (AIST) [4] databases, these data are already available to the general public. Although these are cases of detailed anthropometric data, these data cannot be used directly onto designs because of the issues explained in the previous sections of this chapter. However, these data can be and are used in studies of anthropometry such as the present one.

Designers frequently need to perform a multivariate analysis of a design for a population for which detailed anthropometric information has not been published.
Data for some populations exist publicly as summary statistics in journals or technical reports. Unfortunately, because of the complexity of this type of analysis, multivariate analyses cannot be done to make well-informed decisions on what cutoffs are needed in a design to accommodate to a target user population while using summary statistics. Instead, detailed anthropometric is needed in order to perform multivariate analyses. The objective of this thesis is to create a method that will address this issue.

2.3 Sources of Detailed Anthropometry

As mentioned in previous sections, some sources of detailed anthropometry are already available. Examples such as the 1988 U.S. Army Anthropometry Survey (ANSUR) [3] and the Agency of Industrial Science and Technology Industrial Products Research Institute (AIST) [4] databases are a few cases of detailed anthropometry which are already publicly available. These databases are used by researchers and designers to conduct their own studies. Such data are comprised of over 200 different body measures on thousands of individuals.


These data, although extensive and detailed, are in some cases outdated. ANSUR data, for example, was collected over nearly 30 years ago. This means that the data may no longer be accurate in their ability to represent the population from which the sample was originally taken. This phenomena is attributed to secular trends. A secular trend is defined by Merriam-Webster as a change that is “existing or continuing through ages or centuries.” [25]. These changes in anthropometry are observed in populations [26, 27] due to certain factors such as health and affluence [28]. For this reason, anthropometry measures must be taken and studied regularly throughout the years in order to achieve correct accommodation of current user populations.

This thesis uses two dissimilar sources of anthropometry in order to test the
accuracy of the proposed method. This method ensures that the synthesized anthropometry will achieve similar accuracy with different populations. The following subsections describe the two sources of anthropometric data used in the proposed case study.

### 2.3.1 U.S. Army Anthropometry Survey (ANSUR) Data

The ANSUR database used in this study is a compilation of anthropometric data taken from U.S. military personnel in the late 1980’s [3]. These data are from men and women from various ethnic and age groups. Originally, these data were intended to be used by the U.S. military for sizing of clothing and equipment [3].

For the purpose of this study, only eight out of the 298 measures available in this database were considered. The measures included in this study are specifically head circumference, forearm-hand length, bi-acromial breadth, chest circumference, waist circumference (omphalion), buttock circumference and knee height while sitting (all seen in Figure 2.3). These same measures were taken from both men (total of 1774 individuals in the database) and women (2208 individuals in the database) of the population. These eight measures were chosen because it...
Figure 2.4. Japanese anthropometry used in this study as illustrated by the AIST manual[4].

includes measures of *length* and *breadth* which are commonly seen in datasets of anthropometry. Including measures of length and breadth in the proposed model will show its ability to synthesize these types of data.

2.3.2 Agency of Industrial Science and Technology

**Industrial Products Research Institute (AIST) Data**

The AIST database consists of an anthropometric survey measured from a total of around 500 Japanese men and women during the years of 1991 and 1992 [4]. This database includes a total of 204 women and 217 men. The data are separated by age groups as youth (18-29 years old) and elderly (60 years old). The AIST survey includes 266 different body measures that range from chest circumference to hand length. The same eight measurements selected in ANSUR were gathered from this database in order to validate the performance of this method.
2.4 Sources of Summary Statistics

Summary statistics are usually available through journals and other publications. Examples of such sources include a study by Patel and Karmakar [29] in which some statistical information of body measures of Indian farm workers was published. Other sources include summary of anthropometric statistics published by The State Bureau of Technical Supervision in 1989 [30] and the Instituto Nacional de Tecnologia in 1995 [31] which contain anthropometric values at certain percentiles for Chinese and Brazilian civilians, respectively. A similar study by Barroso et al. [32] looked at anthropometry within adult Portuguese workers. These data are all published in the form of summary statistics.

There are many more surveys targeted toward even more specific groups of people. These include Algerian date palm farmers [33], female assembly workers along the Mexico-US border [34], U.S. truck drivers [35], nurses in the Western Cape [36], children workers from 1841 [37], and airline flight attendants [38]. For these populations, only summary statistics are publicly available. Such studies, even though targeted to a specific group of people, do not provide enough information to aid designers make well-informed decisions on design recommendations through multivariate analyses.

2.5 Data Synthesis Methods

In the past, several efforts have been made to synthesize anthropometry of target user populations. These efforts include implementing methods using proportionality constants [5], regression [39], and quantile-based methods [7]. The following subsections explain the research efforts that have been done to synthesize anthropometric data.

2.5.1 Proportionality Constants

Proportionality constants have been commonly used for the synthesis of anthropometry of user populations. This method is based on the belief that an individual’s measure of length of a particular body part is proportional to their stature [40]. Proportion-
Figure 2.5. Examples of Drillis and Contini’s [5] proportionality constants as illustrated by Fromuth and Parkinson [6].

Proportionality constants are calculated by determining either the mean, or 50\textsuperscript{th}, percentile ratio of the length that’s going to be predicted by the stature [6]. Then, by having such ratios, many mean measures can be obtained. One of the earliest efforts to incorporate proportionality constants was by Drillis and Contini [5], as seen in Figure 2.5. They were among the first to publish the mathematical relationships between several body dimensions to stature. The ease of this method has allowed designers to utilize this as common practice to predict certain lengths from known statures. However, this method presents limitations such as the uncertainty of how well some body dimensions can be modeled after a known stature. Another limitation, pointed out by Fromuth and Parkinson [6], is that this method models variability within measures poorly. In other words, this method is not able to model 5\textsuperscript{th} and 95\textsuperscript{th} percentile values in an accurate manner. This could lead to disaccommodation of a big group of a target population. Finally, the last limitation of this method is that it cannot synthesize individuals taking into account that people are not equally proportioned throughout body measures. That is, an individual can present measures on different percentiles (e.g. a 50\textsuperscript{th} percentile male can have a 50\textsuperscript{th} percentile stature but 30\textsuperscript{th} percentile hip breadth).
2.5.2 Regression Models

Another synthesis method used with anthropometric data are regression models. Regression models are used when the anthropometric data of target user populations are already available [41]. Similarly to proportionality constants, regression models are based on the notion that body measures can be estimated from correlations between them. Regression models improve on proportionality constants by taking into account certain available anthropometric measures to create this ratio. Hence being able to predict a body measure more accurately than proportionality constants. Regression models formulate equations that model the dependence of the relevant anthropometry and their predictors [39].

Attempts have been made to make these methods more accurate. One of such attempts consists on modeling the residual variance (variation around the regression line) of the relevant anthropometry [42]. Including residual variance in the regression model increases the accuracy of the predictors at the tails [39]. As determined in a study from 2010, Nadadur and Parkinson [39] compare regression models with and without the use residual variance. This study determined that introducing residual variance in the model can increase the accuracy of the measures at the tail of the distribution (5th and 95th percentiles).

Even though regression models are able to predict data, it only does so when other detailed anthropometric measures of the target user population are available [39]. Regression models cannot synthesize anthropometric data from limited data information such as summary statistics. Therefore, a method to synthesize anthropometric data from summary statistics is still needed.

2.5.3 Quantile-Based Methods

A method for synthesizing anthropometry based on quantiles taken from a target population is presented by Nadadur and Parkinson [7]. This method presents a quantile-based approach to reverse engineer anthropometric data from key percentiles of a target population. This method is built upon the assumption that anthropometric data are non-Gaussian, which makes the data more complex than a normal distribution [7], even though the model still works when using normally distributed data.
This method, although able to synthesize anthropometry from summary statistics, comes with its limitations. The quantile-based approach achieves the anthropometric synthesis by generating the same number of people available in the original population. That is, if the original anthropometric data contains a total of 1,500 individuals, the virtual synthesized population will have exactly 1,500 individuals as well. Thus, an objective for presenting a new anthropometric synthesis method is for it to be able to synthesize anthropometric databases with an arbitrary number of observations.

2.6 Previous Copula Models

In the past, copulas have been used to model and predict different types of data. Because of their ability to model complex data with non-linear dependencies, copulas have been used in financial risk analysis [43, 44], mortality models [45], disease spread [17], climate research [46], and many other fields. In a study by Cao et al. [20], a copula model was used to predict gender and weight from other human body measures. Copula models are yet to be utilized to synthesize anthropometric data from summary statistics of databases of anthropometry.

Copulas provide the means to fit data by combining data from different sources [17] and mapping variables to a standard uniform distribution. Copulas were first defined by Abe Sklar in 1959 to represent an n-dimensional distribution function $F$ in two parts; the marginal distribution function $F_i$ and the copula function $C$, which describes the dependence of the distribution [18].

**Theorem 1** (Sklar’s Theorem). *Let $F \in (F_1, ..., F_n)$ be an n-dimensional distribution function with marginals $F_1, ..., F_n$. Then there exists a copula $C$ such that*

$$F(X_1, ..., X_n) = C(F_1(X_1), ..., F_n(X_n)) \quad (2.1)$$

Similarly, as defined in [47], copulas are defined as a distribution function $C$ with marginals on $[0,1]$ where if $X = (X_1, ..., X_n)$ is a random vector with distribution functions $F_i, X_i \sim F_i, 1 \leq i \leq n$, a function $F$ is defined such that

$$F = C(F_1, ..., F_n) \quad (2.2)$$
Copulas are able to put together datasets by combining data from different sources. What’s most appealing about copulas in anthropometry is that they are able to model dependencies of the variables more widely than other multivariate normal frameworks [48]. This allows the current method to synthesize anthropometry with as many observations (e.g., individuals) as needed.

One key step to building a copula model is the use of a Cholesky factorization to map the correlation of the measures in the model. The first step to build a copula model is to construct the dependency model that will guide the simulation and establish the behavior of the data. The key step towards creating a dependence model for the copula method is to create a covariance matrix via a Cholesky factorization [49]. A Cholesky factorization (in some cases called Cholesky decomposition) is “a positive definite matrix in a unique lower-triangular matrix with positive diagonal entries” [50]. The Cholesky factorization of a positive definite matrix \( \Sigma \) is of the form

\[
\Sigma = CD^{-1}DDD^{-1}C' = LD^2L
\]  

(2.3)

as proven by [50], where \( L = CD^{-1} \) and \( D \) represent the diagonal matrices of \( C \) such that

\[
D = diag(c_{11}, ..., c_{pp})
\]  

(2.4)

### 2.7 Background Summary

When creating designs that interact with the human body, a designer must understand the implications of using anthropometric databases. In order to make well-informed decisions when designing for a level of accommodation, designers must know how to use these data. First, designers must choose to work with the set of anthropometric data that matches the target user population. With the correct set of anthropometric data, analyses can be carried out to determine the cutoffs on the design to achieve a specific level of accommodation.

Different design methods are often used to ensure that some fraction of the target user population will be able to fit in a design or task. These methods can
be generally classified as univariate and multivariate. They identify the boundaries at which a certain percentage of the population will be accommodated. Accommodation methods ensure the proper fit of a product to a target population in order to ensure spatial requirements of users are met.

Since in many cases data of target populations are only made available through reports of summary statistics, designers are in need of a method for synthesizing detailed anthropometry from such sources. Many methods have been proposed in the past to overcome this issue. These efforts include proportionality constants, regression models, and quantile analysis models. These methods have limitations related to their capability to model detailed anthropometry accurately. For this reason, a new method is proposed to use summary statistics in order to accurately synthesize anthropometric values of a target user populations.

Copula models have been used in the past to predict risk behaviors in several fields such as finance, disease control, mortality, and climate research because of their ability to model complex dependencies. Yet no research has been done to study the performance of copula models synthesizing anthropometry. This study proposes a new method for incorporating copulas into the field of synthesis of anthropometric data.

The proposed method is demonstrated in a case study that simulates a scenario in which only key percentiles of two dissimilar user populations are available. In the proposed case study, detailed anthropometry is estimated from summary statistics through the use of a copula model.
Chapter 3

Copula-Based Method for
Synthesizing Anthropometric Data

The copula model portion of this method consists of establishing the dependencies of each distribution. These dependencies are then used in order to estimate anthropometry using summary statistics and information derived from knowledge of the behavior of anthropometric data. The following sections list the steps that are to be taken to create a copula model in order to synthesized anthropometric values.

3.1 Step I: Collecting Summary Statistics

The first step before building the copula-based model is to gather summary statistics of the desired target user population. As mentioned in previous sections, there are a plethora of sources from which to get these data. This specific method can use a total of $k$ percentiles ($P_1$, $P_2$, $P_3$, ..., $P_k$). Each percentile with its respective value ($X_{1st}$, $X_{2nd}$, $X_{3rd}$, ..., $X_{nth}$) as illustrated in Figure 3.1. In the case of having means and standard deviations available in the summary statistics instead of percentiles, this information is used directly after creating the dependence of the copula model. This method explains the case in which only some percentiles are available in the summary statistics since it requires a much more complex computation than the alternate case.

For each body measure chosen (1 through $l$), the same $k$ percentiles should
be selected. For example, if $k$ different percentile values are chosen from $l$ body measures, a total of $k \times l$ data points should be available to start the synthesis method. Table 3.1 shows an example of $k$ percentiles on $l$ body measures.

### 3.2 Step II: Building the Copula Dependence Model

The dependence scheme of the copula model is created via a correlation matrix through a *Cholesky Factorization*. First, a set of marginals $U$ is created (shown on the flow diagram illustrated in Figure 3.2). That is, a set of uniformly distributed numbers between $[0, 1]$ for each measure contained in the data that’s going to be synthesized. In matrix form, this translates to one column for each body measure that is to be synthesized and one row for each individual in the new population. In order to create a correlation matrix $M$ (Figure 3.2), the previous uniform distribution ($U$) is converted to standard normal. This is a specific modification to the original copula model in which the marginal distribution is left uniformly dis-
Figure 3.2. Flow diagram of the mapping process of a copula model.

The normal distribution $K$ serves as mapping for the population that’s going to be synthesized. Finally, the correlation matrix $M$ is created from the distribution $K$ via the Cholesky factorization. As explained in Chapter 2, the Cholesky factorization achieves the creation of a dependence model by creating “a positive definite matrix in a unique lower-triangular matrix with positive diagonal entries” [50].

Once the dependence model has been constructed, another matrix $S$ is created with $m$ number of columns (body measures) and $n$ number of rows (individuals to be synthesized in the new population). Matrix $S$ is in standard normal form. Finally, in order to link the new standard normal distribution stored as $S$ with the dependence model, the correlation matrix $M$ is multiplied with the normal distribution stored in $S$. These results, stored as $H$, will be used later with $Z$-

Table 3.1. Gathering $k$ percentile values from $l$ different body measures from a target population.

<table>
<thead>
<tr>
<th>Anthropometric Measures</th>
<th>Percentiles $P_1$</th>
<th>$P_2$</th>
<th>$P_3$</th>
<th>...</th>
<th>$P_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body Measure 1</td>
<td>$X_{1,1}$</td>
<td>$X_{1,2}$</td>
<td>$X_{1,3}$</td>
<td>...</td>
<td>$X_{1,k}$</td>
</tr>
<tr>
<td>Body Measure 2</td>
<td>$X_{2,1}$</td>
<td>$X_{2,2}$</td>
<td>$X_{2,3}$</td>
<td>...</td>
<td>$X_{2,k}$</td>
</tr>
<tr>
<td>Body Measure 3</td>
<td>$X_{3,1}$</td>
<td>$X_{3,2}$</td>
<td>$X_{3,3}$</td>
<td>...</td>
<td>$X_{3,k}$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Body Measure $l$</td>
<td>$X_{l,1}$</td>
<td>$X_{l,2}$</td>
<td>$X_{l,3}$</td>
<td>...</td>
<td>$X_{l,k}$</td>
</tr>
</tbody>
</table>
scores for the new copula population. This process is illustrated in Figure 3.2.

3.3 Method Outline

The proposed method explained in the sections above is summarized in the following steps and illustrated in Figure 3.3:

1. Select a target population of choice from which summary statistics on its anthropometry can be gathered.

2. Gather information on the data available on the summary statistics reported.

3. Build the copula dependence model.
   
   (a) Create marginal distributions for each body measure to be synthesized.
   
   (b) Create a correlation matrix via Cholesky factorization.

4. Synthesize anthropometry by using the information given in the summary statistics with the copula dependence model.
   
   (a) Gather standard Z-scores from the standard normal distribution for each percentile the synthesis model is to use.
   
   (b) Determine the standard deviations to use for the upper and lower parts of the distribution.
   
   (c) Synthesize anthropometry values for each measure with the information gathered.
Figure 3.3. Flow chart of the proposed copula method to synthesize anthropometric data from summary statistics. This illustration was made specifically for the copula method presented in this study but based off of a similar flow chart seen in Nadadur and Parkinson’s previous research on a quantile-based approach to synthesize anthropometric data [7].
Estimating the Distributions in the Published Data

The third step to the copula-based method to reverse engineer anthropometric data from summary statistics is to estimate the desired anthropometry once the necessary information is gathered. This chapter explains the method for synthesizing anthropometric data from summary statistics of a target user population. Anthropometric data will be synthesized making use of the normal distribution equations derived from the following equation:

$$Z = \frac{X - \mu}{\sigma}$$

(4.1)

where $Z$ refers to the Z-score values taken from the standard normal distribution table, $X$ are the values at $k$ percentiles, $\mu$ refers to the mean of each measure, and $\sigma$ are the standard deviations of each measure of the population. In the case of having only key percentiles available, $\mu$ is substituted by the 50th percentile value given in the summary statistics. In the case having the means and standard deviations of the anthropometric measures available, the given means and standard deviations are used instead in Equation 4.4.

In order to increase the accuracy of the synthesized data, a two-standard deviation method is employed (Figure 4.1). This method uses two different standard deviations per measure in order to account for the assumption that these data are non-Gaussian normal distributions [1]. Assuming that the data is non-Gaussian
Figure 4.1. Separation method of both halves of a normal distribution to determine the standard deviations ($\sigma_1$ and $\sigma_2$) used in as the lower (green) and upper (blue) sections of the distributions of each body measure.

means that the distribution does not have just one standard deviation. In fact, if the distributions had one standard deviation, this would mean that these are standard normal. However, this method would also work for normal distributions.

This two-standard deviation method consists of parting the distributions into two halves (upper and lower), in order to assign one standard deviation per half (Figure 4.1). To choose the best fit for the standard deviations on both halves of each distribution (green and blue as illustrated in Figure 4.1), a best fit method is performed. This best fit procedure involves an optimization procedure using a least square method (Equation 4.2) that gives the standard deviation of each half that will result in the value that’s closest to its respective percentile,

$$\min \sum_{i=1}^{n} (V_{sd} - V_o)^2$$  \hspace{1cm} (4.2)

where $V_{sd}$ is the synthesized measure value to be computed with the new standard deviation, and $V_o$ are the percentile value taken from the summary statistics of the user population. $V_{sd}$ is computed using Equation 4.3 which is taken from the normal Z-score formula, also shown in Equation 4.1.

$$V_{sd} = \mu + (Z \times \sigma)$$  \hspace{1cm} (4.3)
Figure 4.2. Final step of synthesizing all measure values by combining the dependence model and the information known from the summary statistics using two different standard deviations per measure.

Thus, the standard deviation chosen using this method is the one with the percentile measure value that gives the minimum sum of the differences squared between the reported percentile values and the copula-synthesized percentile values ($V_{sd}$). This process to get the standard deviation values ($\sigma$) is computed with an optimization function in MATLAB. This process is computed for the lower and upper halves of the distributions of each measure to gather one standard distribution for each half ($\sigma_1$ and $\sigma_2$).

After the previous data are computed, the measure values are synthesized using the following equation:

$$X = (\text{sim}Z \times \sigma) + \mu \quad (4.4)$$

Where $X$ are the anthropometric measure values to be synthesized, $\text{sim}Z$ are the values from the dependence model created at the end of Step II of this chapter, and $\mu$ are the 50th percentile values of each body measure. The standard deviations ($\sigma$) are used according to the location of the measures. For example, negative $\text{sim}Z$ values are known to be located at the lower (left) side of the distribution. These values correspond to $\sigma_1$. Positive $\text{sim}Z$ values, located in the upper (right) side, are assigned $\sigma_2$. The location of each $\sigma$ value is illustrated in Figure 4.2.
In the hypothetical case of having only means and standard deviations for each measure available in the summary statistics, the method is performed by estimating anthropometric measure values using Equation 4.4. In this particular scenario, the means ($\mu$) and standard deviations ($\sigma$) computed from the two-standard deviation case are used.

Section 4.2 evaluates the overall performance of the two-standard deviations method over a single-standard deviation method (i.e. two standard deviations vs. one standard deviation per measure).

## 4.1 Step IV: Evaluate Results

The evaluation of these results is performed by calculating the mean absolute error (MAE) and percent error (PE) values using a total of 99 percentile pairs (1\textsuperscript{st} through 99\textsuperscript{th} percentiles of the synthesized anthropometric data and their anthropometric source pair). The lower the MAE and PE values are, the closer the synthesized anthropometric data points are to the original anthropometric data. The MAE and PE values are computed using Equations 4.5 and 4.6, respectively.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |X_i - X_o| \quad (4.5)
\]

\[
PE = \frac{|C_i - P_i|}{P_i} \quad (4.6)
\]

In Equation 4.5, $X_i$ represents the percentile measure values of the synthesized data and $X_o$ are the measure values of the original anthropometric data. Equation 4.6 shows $C_i$, which represents the percentile values of the synthesized anthropometric values, $P_i$ are the percentile values from the original anthropometric data. Both metrics are calculated from the 1\textsuperscript{st} through the 99\textsuperscript{th} percentile, which makes $n$ equal to 99 to represent a total of 99 total pairs of anthropometric values. Even though it is assumed that these 99 percentiles won’t be available in summary statistics reports, these were taken from the original dataset for evaluation purposes of this method.
Figure 4.3. Overlapping distributions of the single- vs. two-standard deviation methods.

4.2 Justification of the Two-Standard Deviation Method

A two-standard deviation method is proposed in this copula-based model in order to increase the accuracy of the of the estimated anthropometric measures in comparison to the anthropometric source. To evaluate the performance of the two-standard deviation method, a one-standard deviation method is carried about as well to compare both outcomes. This method can be used in the hypothetical case of having only some key percentiles available from the summary statistics. The standard deviation used in the single-standard deviation method for each measure is taken from the mean of all the standard deviations calculated from each percentile. Such standard deviations are computed using Equation 4.7 which is a derived from the normal distribution Z-score equation (Equation 4.1). This two-standard deviation method could also be used in the alternative case of having means and standard deviations in the summary statistics.

\[ \sigma = \frac{X - \mu}{Z} \]  

(4.7)

In contrast with the process shown previously, these use the same standard deviation (\(\sigma\)) for the lower and upper sections of the distribution. This is different from a two-standard deviation method where there are two standard deviations per measure (\(\sigma_1\) and \(\sigma_2\)) as shown in Figure 4.2. A comparison of both methods is illustrated in Figure 4.3.
In Gaussian (normal) distributions, a single standard deviation describes the variation of the whole distribution [51] (Equation 4.1). Similarly, in Gaussian distributions, the mean ($\mu$) of the distribution is at the very center of the distribution [51], which happens at $\sigma = 0$. These properties allow for symmetry in the distribution (as seen on the left-side distribution of Figure 4.4), which translates into evenly spaced variation. Conversely, in a non-Gaussian distribution, the variation behaves in a more complex way. In a non-Gaussian case, variation is not evenly spaced throughout the whole distribution (right-hand side of Figure 4.4). Therefore, using a single $\sigma$ to predict values in a non-Gaussian distribution will not estimate anthropometric measures accurately. Thus, parting a non-Gaussian distribution in half and treating each side as its own “Gaussian” distribution (process illustrated in Figure 4.1), is more likely to give closer anthropometric measure values when compared to the original source data.

When evaluating these two methods, it is expected that data points from the synthesized anthropometric values using two separate standard deviations will give values that are closer to their original anthropometric counterparts. The performance of these two methods is tested in Chapter 5, Section 5.4 via mean absolute error (MAE), using Equation 4.5, as well as percent errors values (PE), as seen in Equation 4.6. This comparison is done using the 1st through the 99th percentile of the synthesized and original anthropometric data (total of 99 pairs).
Case Study

A case study was carried out in order to test the performance of the proposed copula-based method. The case study simulates a situation in which a person is in need of detailed anthropometry but only has access to some summary statistics of the anthropometry values. This hypothetical case is one which researchers and designers regularly face this issue in their design practice. This case study seeks to prove that the proposed method is accurate enough in solving the scenario presented.

This case study involved gathering summary statistics (i.e., 5\textsuperscript{th}, 25\textsuperscript{th}, 50\textsuperscript{th}, 75\textsuperscript{th}, and 95\textsuperscript{th} percentiles) from full datasets in order to synthesize anthropometric measures with a total of 2,000 individuals. After the anthropometry values for each measure were synthesized, percentiles 1 through 99 were taken to compare them to the original data in terms of mean absolute error (MAE) and percent error (PE) values. The summary statistics were gathered from complete sets of anthropometric data in order to simulate a situation in which only some percentiles from these data are available. Then, the detailed sources of anthropometry were used to assess the equivalence of the real synthesized anthropometry. This case study uses a two-standard deviation method for each anthropometric measure synthesized by the model (process proposed in Chapter 4, Section 4.2, and illustrated in Figure 4.1).

This study was conducted by extracting five percentile values of the same eight body measures from the ANSUR [3] and Japanese [4] female and male anthropometry. These measurements are specifically: biacromial breadth, but-
tock circumference, sitting height, chest circumference, forearm-hand length, head circumference, knee height (sitting), and waist circumference (ombilicus). The same measurements were taken from both anthropometric datasets.

5.1 Preliminary Step: Anthropometric Analysis

Eight different measures from the ANSUR [3] and Japanese [4] anthropometric data were used for the validation of the proposed copula-based method. These populations are dissimilar, making them good candidates to test the robustness of the proposed method. In addition, these eight measures were chosen as part of this case study since they include measurements of length and breadth. Such types of measures are commonly seen in databases of anthropometry. Including measures of breadth and length as part of the validation of this method also helps test its robustness.

5.2 Step I: Collecting summary Statistics

Before computing the copula model, the first step is to identify and collect data on the population of interest. The ANSUR [3] and Japanese [4] detailed anthropometry are incorporated in this copula-based model in order to compare and contrast these results and prove the model’s utility.

In order to simulate a situation where only summary statistics of an anthropometric survey are available through anthropometric reports, summary statistical data were gathered from both datasets. Specifically, the 5th, 25th, 50th, 75th, and 95th (as seen in Tables 5.1 and 5.2) percentiles were calculated from the original data. These specific percentiles were used because these are commonly percentiles that are available in technical and journal articles. The original means and standard deviations were not used as part of this case study. Instead, the mean was substituted by the 50th percentile of each body measure and the standard deviations were computed using a two-standard deviation method.
Table 5.1. Percentile values for the male (M) and female (F) ANSUR [3] anthropometric data. These measures were selected to serve as the percentile values provided in many anthropometric reports and will be input data of the copula-based synthesis method.

<table>
<thead>
<tr>
<th>Anthro. Measures</th>
<th>5th</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>95th</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Biacromial Brth.</td>
<td>367</td>
<td>333</td>
<td>385</td>
<td>351</td>
<td>397</td>
</tr>
<tr>
<td>Buttock Circ.</td>
<td>890</td>
<td>873</td>
<td>939</td>
<td>926</td>
<td>982</td>
</tr>
<tr>
<td>Chest Circ.</td>
<td>886</td>
<td>814</td>
<td>944</td>
<td>863</td>
<td>986</td>
</tr>
<tr>
<td>Forearm-Hand Lnth.</td>
<td>448</td>
<td>406</td>
<td>468</td>
<td>427</td>
<td>483</td>
</tr>
<tr>
<td>Head Circ.</td>
<td>543</td>
<td>523</td>
<td>557</td>
<td>537</td>
<td>567</td>
</tr>
<tr>
<td>Knee Height (Sit.)</td>
<td>514</td>
<td>474</td>
<td>540</td>
<td>498</td>
<td>557</td>
</tr>
<tr>
<td>Sitting Height</td>
<td>855</td>
<td>795</td>
<td>890</td>
<td>827</td>
<td>914</td>
</tr>
<tr>
<td>Waist Circ. (Omph.)</td>
<td>735</td>
<td>675</td>
<td>798</td>
<td>731</td>
<td>856</td>
</tr>
</tbody>
</table>
Table 5.2. Percentile values for the male (M) and female (F) AIST [4] anthropometric data. These measures were selected to serve as the percentile values provided in many anthropometric reports and will be input data of the copula-based synthesis method.

<table>
<thead>
<tr>
<th>Anthro. Measures</th>
<th>Percentiles (mm)</th>
<th>5&lt;sup&gt;th&lt;/sup&gt;</th>
<th>25&lt;sup&gt;th&lt;/sup&gt;</th>
<th>50&lt;sup&gt;th&lt;/sup&gt;</th>
<th>75&lt;sup&gt;th&lt;/sup&gt;</th>
<th>95&lt;sup&gt;th&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M  F</td>
<td>M  F</td>
<td>M  F</td>
<td>M  F</td>
<td>M  F</td>
</tr>
<tr>
<td>Biacromial Brth.</td>
<td></td>
<td>361 327</td>
<td>378 346</td>
<td>393 356</td>
<td>405 365</td>
<td>427 381</td>
</tr>
<tr>
<td>Buttock Circ.</td>
<td></td>
<td>824 831</td>
<td>867 865</td>
<td>899 891</td>
<td>928 928</td>
<td>978 965</td>
</tr>
<tr>
<td>Chest Circ.</td>
<td></td>
<td>793 759</td>
<td>851 800</td>
<td>882 836</td>
<td>920 871</td>
<td>986 941</td>
</tr>
<tr>
<td>Forearm-Hand Lnth.</td>
<td></td>
<td>225 205</td>
<td>239 217</td>
<td>249 225</td>
<td>257 234</td>
<td>271 246</td>
</tr>
<tr>
<td>Head Circ.</td>
<td></td>
<td>545 524</td>
<td>559 535</td>
<td>569 544</td>
<td>579 553</td>
<td>594 565</td>
</tr>
<tr>
<td>Knee Height (Sit.)</td>
<td></td>
<td>469 431</td>
<td>498 455</td>
<td>514 469</td>
<td>530 485</td>
<td>558 506</td>
</tr>
<tr>
<td>Sitting Height</td>
<td></td>
<td>841 772</td>
<td>888 836</td>
<td>888 858</td>
<td>941 880</td>
<td>978 915</td>
</tr>
<tr>
<td>Waist Circ. (Omph.)</td>
<td></td>
<td>333 315</td>
<td>350 329</td>
<td>362 342</td>
<td>375 356</td>
<td>392 377</td>
</tr>
</tbody>
</table>
5.3 Step II: Building the Dependence Model

A dependence scheme of the copula model has been created via *Cholesky Factorization*. First, a set of marginals (uniformly distributed numbers between [0,1]) were synthesized for each measure that was going to be synthesized. In matrix form, this translates to a total of eight columns (eight body measures) and 2,000 rows (2,000 individuals). These marginals were then converted to standard normal. Finally, the *correlation matrix* is created from the marginal normal distribution via the *Cholesky factorization*.

Once the correlation matrix has been constructed, another matrix is created with eight number of columns (body measures) and 2,000 number of rows (individuals to be synthesized). This matrix is in standard normal form. Finally, in order to link the new standard normal distribution with the dependence model, the correlation matrix is multiplied with the normal distribution that was computed last. These results will be used later with Z-scores (Equation 4.4) for the new copula population.

5.4 Comparison of the Performance of Standard Deviation Methods

A comparison of both standard deviation methods is performed to show the higher reliability of the two-standard deviation over the one-standard deviation method. This comparison is done in terms of the MAE and PE values for each anthropometric measure selected in this chapter. Results from the computation of the MAE and PE values (Equations 4.5 and 4.6, respectively) using synthesized anthropometry in reference to this case study are noted in Tables 5.3, 5.4, 5.5, and 5.6. These tables show the performance of the ANSUR and Japanese synthesized anthropometric values against the original source data, separated by males and females, computed using both standard deviation methods.
### Table 5.3. Comparison of mean absolute error values of the one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese female synthesized anthropometric values.

<table>
<thead>
<tr>
<th>Anthropometric Measures</th>
<th>ANSUR</th>
<th>Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One S.D.</td>
<td>Two S.D.</td>
</tr>
<tr>
<td>Biacromial Breadth</td>
<td>0.70192</td>
<td>0.68535</td>
</tr>
<tr>
<td>Buttock Circumference</td>
<td>3.05190</td>
<td>2.05230</td>
</tr>
<tr>
<td>Chest Circumference</td>
<td>6.29010</td>
<td>2.79100</td>
</tr>
<tr>
<td>Forearm-Hand Length</td>
<td>0.91547</td>
<td>0.96342</td>
</tr>
<tr>
<td>Head Circumference</td>
<td>1.20750</td>
<td>0.62759</td>
</tr>
<tr>
<td>Knee Height (Sitting)</td>
<td>1.04730</td>
<td>0.96937</td>
</tr>
<tr>
<td>Sitting Height</td>
<td>1.38770</td>
<td>1.09740</td>
</tr>
<tr>
<td>Waist Circumference (Omphalion)</td>
<td>10.3040</td>
<td>4.91790</td>
</tr>
</tbody>
</table>

### Table 5.4. Comparison of mean absolute error values between one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese male synthesized anthropometric values.

<table>
<thead>
<tr>
<th>Anthropometric Measures</th>
<th>ANSUR</th>
<th>Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One S.D.</td>
<td>Two S.D.</td>
</tr>
<tr>
<td>Biacromial Breadth</td>
<td>0.60732</td>
<td>0.59552</td>
</tr>
<tr>
<td>Buttock Circumference</td>
<td>2.78010</td>
<td>2.51470</td>
</tr>
<tr>
<td>Chest Circumference</td>
<td>5.23670</td>
<td>2.32770</td>
</tr>
<tr>
<td>Forearm-Hand Length</td>
<td>1.33010</td>
<td>0.77798</td>
</tr>
<tr>
<td>Head Circumference</td>
<td>0.73020</td>
<td>0.55995</td>
</tr>
<tr>
<td>Knee Height (Sitting)</td>
<td>1.78440</td>
<td>0.87942</td>
</tr>
<tr>
<td>Sitting Height</td>
<td>1.16250</td>
<td>1.14880</td>
</tr>
<tr>
<td>Waist Circumference (Omphalion)</td>
<td>7.12750</td>
<td>4.24650</td>
</tr>
</tbody>
</table>
Table 5.5. Comparison of the percent error values between one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese female synthesized anthropometric values.

<table>
<thead>
<tr>
<th>Anthropometric Measures</th>
<th>Percent Error Values (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANSUR</td>
<td>One S.D.</td>
<td>Two S.D.</td>
</tr>
<tr>
<td>Biacromial Breadth</td>
<td></td>
<td>0.19528</td>
<td>0.19004</td>
</tr>
<tr>
<td>Buttock Circumference</td>
<td></td>
<td>0.31620</td>
<td>0.21115</td>
</tr>
<tr>
<td>Chest Circumference</td>
<td></td>
<td>0.69639</td>
<td>0.30857</td>
</tr>
<tr>
<td>Forearm-Hand Length</td>
<td></td>
<td>0.20891</td>
<td>0.22038</td>
</tr>
<tr>
<td>Head Circumference</td>
<td></td>
<td>0.21902</td>
<td>0.11474</td>
</tr>
<tr>
<td>Knee Height (Sitting)</td>
<td></td>
<td>0.20041</td>
<td>0.18712</td>
</tr>
<tr>
<td>Sitting Height</td>
<td></td>
<td>0.16445</td>
<td>0.12745</td>
</tr>
<tr>
<td>Waist Circumference (Omphalion)</td>
<td></td>
<td>1.29790</td>
<td>0.62304</td>
</tr>
</tbody>
</table>

Table 5.6. Comparison of the percent error values between one- and two-standard deviation (S.D.) methods for the ANSUR and Japanese male synthesized anthropometric values.

<table>
<thead>
<tr>
<th>Anthropometric Measures</th>
<th>Percent Error Values (%)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANSUR</td>
<td>One S.D.</td>
<td>Two S.D.</td>
</tr>
<tr>
<td>Biacromial Breadth</td>
<td></td>
<td>0.15183</td>
<td>0.14902</td>
</tr>
<tr>
<td>Buttock Circumference</td>
<td></td>
<td>0.31620</td>
<td>0.21115</td>
</tr>
<tr>
<td>Chest Circumference</td>
<td></td>
<td>0.53131</td>
<td>0.23420</td>
</tr>
<tr>
<td>Forearm-Hand Length</td>
<td></td>
<td>0.27497</td>
<td>0.15918</td>
</tr>
<tr>
<td>Head Circumference</td>
<td></td>
<td>0.12884</td>
<td>0.09889</td>
</tr>
<tr>
<td>Knee Height (Sitting)</td>
<td></td>
<td>0.32000</td>
<td>0.15701</td>
</tr>
<tr>
<td>Sitting Height</td>
<td></td>
<td>0.12653</td>
<td>0.12548</td>
</tr>
<tr>
<td>Waist Circumference (Omphalion)</td>
<td></td>
<td>0.84683</td>
<td>0.50733</td>
</tr>
</tbody>
</table>
Most of the measures synthesized by using the two-standard deviation method had a lower MAE and PE values than the measures synthesized using only one standard deviation for each measure. Specific instances, such as the forearm-hand length and the biacromial breadth of the female anthropometry synthesized from ANSUR (Table 5.3) show that the MAE values for the two-standard deviation method are higher (farther apart from the original data) than when using only one standard deviation per measure. These instances were determined to not be of relevance since they’re outnumbered by the instances in which the two-standard deviation method performs more accurately. In addition, the one-standard deviation method synthesized the waist circumference (omphalion) of females (Table 5.3) and males (Table 5.4) from the ANSUR data with almost twice the error (5.3861 mm higher for females and 2.8810 mm higher for males). Conversely, the highest instance in which the two-standard deviation method gives a higher MAE than the one-standard deviation value is of only 0.2127 mm higher (Japan male chest circumference in Table 5.4). Thus, making the two-standard deviation method more reliable than the one standard deviation for each synthesized anthropometric measure.

5.5 Step III: Synthesizing Anthropometric Values

Up to this point, the copula dependence model and necessary data (e.g. summary statistics) are available to carry out the anthropometric data synthesis. Synthesis of anthropometric data is done using Equation 4.4. The synthesis process is explained in detail in Chapter 3 and it’s illustrated in Figures 4.1 and 4.2. Here, the dependencies were modeled using a copula approach and a two-standard deviation method was performed in order to achieve higher accuracy than otherwise.

Examples of the synthesized anthropometry plotted with their original counterparts are shown in Figure 5.1. These figures show female and male synthesized head circumferences for both user populations (ANSUR and Japanese databases) plotted with their respective original counterparts for contrast. The synthesized anthropometry (plotted in a continuous, bright line) is shown to behave closely
Figure 5.1. Example of the synthesized head circumference of females (left) and males (right) plotted with its corresponding source data.

to the original data. Still, proper evaluation and statistical analysis is needed. Evaluation of the mean absolute errors and percent errors of the synthesized data compared to the original data as well as statistical analyses of these results are presented in the next chapter.
Results, Statistical Analysis and Discussion

Evaluation of the synthesized anthropometric data consists in computing how far away the synthesized values are from the original data. Percent error (PE) and mean absolute error (MAE) values (equations 4.6 and 4.5, respectively) were computed for both user populations (ANSUR [3] and Japan [4]) using the two-standard deviation method explained and justified in previous chapters. Results of MAE and PE values of the final synthesized anthropometry are shown in tables 6.1 and 6.2, respectively.

Table 6.1. Mean absolute error values for male and female ANSUR and Japanese synthesized anthropometry anthropometry using a two-standard deviation method.

<table>
<thead>
<tr>
<th>Anthropometric Measures</th>
<th>ANSUR</th>
<th>Japan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Females</td>
<td>Males</td>
</tr>
<tr>
<td>Biacromial Breadth</td>
<td>0.68529</td>
<td>0.59171</td>
</tr>
<tr>
<td>Buttock Circumference</td>
<td>2.04245</td>
<td>2.51155</td>
</tr>
<tr>
<td>Chest Circumference</td>
<td>2.81040</td>
<td>2.33910</td>
</tr>
<tr>
<td>Forearm-Hand Length</td>
<td>0.96361</td>
<td>0.76962</td>
</tr>
<tr>
<td>Head Circumference</td>
<td>0.62808</td>
<td>0.56327</td>
</tr>
<tr>
<td>Knee Height (Sitting)</td>
<td>0.96699</td>
<td>0.87974</td>
</tr>
<tr>
<td>Sitting Height</td>
<td>1.08580</td>
<td>1.14970</td>
</tr>
<tr>
<td>Waist Circumference (Omphalion)</td>
<td>4.91935</td>
<td>4.23295</td>
</tr>
</tbody>
</table>
Table 6.2. Percent Error values for male and female ANSUR and Japanese synthesized anthropometry using a two-standard deviation method.

<table>
<thead>
<tr>
<th>Anthropometric Measures</th>
<th>Percent Errors (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ANSUR</td>
</tr>
<tr>
<td></td>
<td>Females Males Females Males</td>
</tr>
<tr>
<td>Biacromial Breadth</td>
<td>0.19004 0.14902</td>
</tr>
<tr>
<td>Buttock Circumference</td>
<td>0.21115 0.25816</td>
</tr>
<tr>
<td>Chest Circumference</td>
<td>0.30857 0.23420</td>
</tr>
<tr>
<td>Forearm-Hand Length</td>
<td>0.22038 0.15918</td>
</tr>
<tr>
<td>Head Circumference</td>
<td>0.11474 0.09889</td>
</tr>
<tr>
<td>Knee Height (Sitting)</td>
<td>0.18712 0.15701</td>
</tr>
<tr>
<td>Sitting Height</td>
<td>0.12745 0.12548</td>
</tr>
<tr>
<td>Waist Circumference (Omphalion)</td>
<td>0.62304 0.50733</td>
</tr>
</tbody>
</table>

In addition, results from the copula method were statistically analyzed using correlation coefficients, Q-Q plots, and a Pearson’s Chi-Squared test. These tests were done in order to determine the magnitude of the closeness of the data as well as its significance. Correlation coefficients from the eight ANSUR anthropometric pairs varied from 0.99786 to 0.99964. Similarly, correlation coefficients from the eight Japanese anthropometric pairs varied from 0.98036 to 0.99924. These coefficients indicate a high correlation between the synthesized anthropometry and their original pairs.

To visualize the strength of these correlation, Q-Q plots are shown in figures 6.1, 6.2, 6.3, and 6.4. Figures 6.1 and 6.2 show Q-Q plots of the eight measures selected from the Japanese user population. Likewise, figures 6.3 and 6.4 contain Q-Q plots of the same eight measures selected from the ANSUR user population. These plots are in line with the correlation coefficients showing a high correlation of the synthesized and the original anthropometry for both user populations.
Figure 6.1. Q-Q plots of the female and male Japanese data (specifically, biacromial breadth, buttock circumference, head circumference, and forearm-hand length). This comparison was done between the original and the synthesized Japanese anthropometry.
Figure 6.2. Q-Q plots of the female and male Japanese data (specifically, head circumference, knee height, sitting height, and waist circumference). This comparison was done between the original and the synthesized Japanese anthropometry.
Figure 6.3. Q-Q plots of the female and male ANSUR data (specifically, biacromial breadth, buttock circumference, head circumference, and forearm-hand length). This comparison was done between the original and the synthesized ANSUR anthropometry.
Figure 6.4. Q-Q plots of the female and male ANSUR data (specifically, head circumference, knee height, sitting height, and waist circumference). This comparison was done between the original and the synthesized ANSUR anthropometry.
In order to test the statistical significance of these results, a Chi-squared test was carried out. This test is set to determine if the high correlations shown are due to chance or not. P-values were analyzed for each anthropometric pair (16 total pairs) on both user populations. The Chi-squared test showed p-values of 0.000999001 (p < 0.05) which, for this test, would signify that the results shown are not likely due to chance.

Results shown in tables 6.1 and 6.2 present the MAE and PE of the synthesized anthropometry compared to their original anthropometric pairs, respectively. MAE values show the absolute value of the mean of the errors (in mm) and PE values show the percentage that these errors represent to each measure. Table 6.2, showing PE values for each measure, corroborates what was suspected from the MAE values shown in Table 6.1, which is that the errors of the synthesized anthropometry are very small (highest of 0.62%). These results show the accuracy and reliability of the copula model presented in this study. In addition, the statistical analysis done on these data (correlation coefficients, Q-Q plots, and Chi-squared test) show that the results are statistically significant. Conclusions from this model are developed and explained in Chapter 7.
A copula-based anthropometric data synthesis method for target populations was proposed in this study. This method showed to be robust for dissimilar populations, proving that the method works well regardless of the source of the summary statistics. In addition, this method is able to overcome the limitations of previous efforts to synthesize data which have been mentioned in this study. Such limitations range from the number of instances these methods can synthesize to the accuracy of the model itself based on the assumptions these models are based off of. This thesis proposed a method to synthesize anthropometry measures based on the limited information given in technical reports about the anthropometry from target populations that it is able to overcome previous limitations.

The proposed copula method has shown to be able to surpass the limitations of previous anthropometry synthesis efforts. Previous synthesis methods have presented the following limitations: (a) not being able to synthesize anthropometric data with different number of instances from the source data, (b) parting from the assumption that all individuals are similarly proportionate, (c) synthesizing individuals with all measures in the same percentiles, and (d) use of detailed anthropometry measures to synthesize other measures. The copula model presented in this study is able to overcome previous models’ limitations by:

(a) synthesizing anthropometric data at a different number of observations from the source of summary statistics.

(b) Creating a dependence model that is able to synthesize anthropometric data
without parting from the assumption that individuals are similarly proportionate.

(c) It is capable of synthesizing individuals with measures at various percentiles (e.g. an individual with 50th percentile stature, 40th percentile forearm-hand length, and 70th knee sitting height).

(d) Synthesizes anthropometric data without the use of other detailed anthropometry. It achieves the synthesis using only summary statistics from a user population.

Before the method presented in this study, copulas had not yet been used to synthesize anthropometry. By incorporating copulas into the field of anthropometric synthesis, the proposed method was able to overcome the limitations that other synthesis methods have. The use of a copula model allowed for the development of a tool for synthesizing anthropometric data with a desired number of instances by only using the data given in summary statistics. Such method was presented in Chapter 3 and its robustness was tested through a case study in Chapter 5.

The case study proposed in this thesis showed a scenario in which a designer only has access to summary statistics of a set of anthropometric data of a user population. From these summary statistics, the copula-based model was able to synthesize a virtual user population. The results presented from this case study showed the accuracy and statistical significance of this method while overcoming the limitations previously mentioned. Using this method, designers are now able to synthesize anthropometric data from summary statistics of a target user population to use for their own ventures.

7.1 Limitations and Future Work

As with many methods for synthesizing anthropometry, this method comes with its limitations. While developing the copula model, it became evident that most of the issues with accuracy were related to how accurately the standard deviations were determined by the model. Even though it was shown that using a two-standard deviation method will improve the accuracy of the model, there are still slight
percentage errors (less than 0.62% error) which can be improved on. An extension of the current work should include a method to optimize the choice of standard deviations for both halves of each measure distribution. In addition, given the non-Gaussian nature of this data, a method that's able to determine more than one standard deviation per half of each distribution (lower and upper halves) can improve the accuracy of the proposed copula-based method. In other words, a method that uses a different standard deviation depending the distance of each percentile from the mean should be explored.

The effects of using synthesized anthropometric data to perform multivariate analysis are not clear yet. A study to investigate the influence of synthesized anthropometric data is to be performed. This, in order to fully understand the limitations in using these types of data on the analyses of intricate designs.

The use of a copula model to synthesized anthropometric body measures was proven to be successful. This current study contributes to the fields anthropometric studies by giving designers a tool for synthesizing anthropometric measure values for target populations from using summarized statistics. The use of the method described in this work will aid designers on their endeavours when they’re in need of anthropometric data from which they can only get limited information on. After using this tool, designers will be able to perform complex analyses for target populations even if the only information they have available are summary statistics of such populations.
Bibliography


[22] **Moroney, W. F. and M. J. Smith** (1972) “Empirical reduction in potential user population as the result of imposed multivariate anthropometric limits,” *Naval Aerospace Medical Research Laboratory in Pensacola, Florida*.


Vita

Maria Alessandra Nusiner Bassi

Maria Alessandra Nusiner Bassi was born in Caracas, Venezuela, on September 17, 1992. After finishing high school in 2010, she enrolled in Universidad Simón Bolívar, in Caracas, as a Production Engineer. In 2011, she transferred to The Pennsylvania State University in University Park to start her Bachelor of Science degree in Industrial Engineering. She received her degree in Industrial Engineering in 2015 and will receive a Master of Science in Engineering Design degree in May of 2017.