



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Projection of Anthropometric Correlation for Virtual Population Modelling

Rasmussen, John; Waagepetersen, Rasmus Plenge; Rasmussen, Kasper Pihl

Published in:
International Journal of Human Factors Modelling and Simulation

DOI (link to publication from Publisher):
[10.1504/IJHFMS.2018.091353](https://doi.org/10.1504/IJHFMS.2018.091353)

Publication date:
2018

Document Version
Early version, also known as pre-print

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Rasmussen, J., Waagepetersen, R. P., & Rasmussen, K. P. (2018). Projection of Anthropometric Correlation for Virtual Population Modelling. *International Journal of Human Factors Modelling and Simulation*, 6(1), 16-30.
<https://doi.org/10.1504/IJHFMS.2018.091353>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Projection of Anthropometric Correlation for Virtual Population Modelling

John Rasmussen

Department of Mechanical and Manufacturing Engineering,
Aalborg University,
Aalborg, Denmark
E-mail: jr@m-tech.aau.dk

Rasmus Plenge Waagepetersen

Department of Mathematical Sciences,
Aalborg University,
Aalborg, Denmark
E-mail: rw@math.aau.dk

Kasper Pihl Rasmussen

AnyBody Technology A/S,
Niels Jernes Vej 10,
DK-9220 Aalborg, Denmark
E-mail: kpr@anybodytech.com

Abstract: A new statistical method for generation of virtual populations based on anthropometric parameters is developed. The method addresses the problem that most anthropometric information is reported in terms of summary data such as means and standard deviations only, while the underlying raw data, and therefore the correlations between parameters, are not accessible. This problem is solved by projecting correlation from a data set for which raw data are provided. The method is tested and validated by generation of pseudo females from males in the ANSUR anthropometric dataset. Results show that the statistical congruency of the pseudo population with an actual female population is more than 90% for more than 90% of the possible parameter pairs. The method represents a new opportunity to generate virtual populations for specific geographic regions and ethnicities based on summary data only.

Keywords: Anthropometry; Human factors; Statistics; Principal component analysis; PCA; Correlation.

1 Introduction

Anthropometry plays a key role in the fields of ergonomics and industrial product design. Many categories of industrial products derive their value from an ability to interact with humans. Typical examples are vehicles, bicycles, furniture, tools, workplaces, sports equipment and packaging for consumer products. There is also a growing understanding of the importance of compatibility between biomedical products, such as joint replacements and trauma devices, and the bodies into which they are implanted (Palumbo et al. [2012]). For most industrial and biomedical products, specific manufacture for a single user is not feasible, and anthropometric compatibility must be obtained by incorporating adjustment into the product design, such as the height of a bicycle saddle, or by production of a series of sizes that will accommodate the majority of the target population, such as shoes. Accurate and updated anthropometric information for the targeted population is crucially important for such products to be successful.

It is possible to investigate the product's interaction experimentally with a small number of subjects representing the boundary of the targeted population, for instance a 5th percentile female and a 95th percentile male in terms of stature only. Small series of test subjects representing a range is another option. However, neither approach yields a representation of the full anthropometrical variation over the population (Högberg et al. [2015]).

For some time, CAD-integrated digital manikin systems, such as RAMSIS[®], JACK[®] and DELMIA[®], have been available and have offered the opportunity to incorporate anthropometric variation into computer models. More recently, musculoskeletal modelling systems, such as AnyBody (Damsgaard et al. [2006]) and Opensim (Delp et al. [2007]), enable simulation of the detailed biomechanics of the interaction between virtual humans and products at different scales from full body over specific limbs to single joints or muscles. Although technical problems persist, the goal of digital models is to make virtual experiments simple and inexpensive. If successful, the focus shifts from the execution of the experiments to the generation of a virtual population that validly represents the real population of interest.

This defines the challenge of creating virtual populations representative for particular geographic or ethnic groups. Unfortunately, this task is non-trivial. Anthropometric data collection is expensive and time-consuming, and populations change dimensions over time (de Vries et al. [2010]), so data collection must be repeated at regular intervals. It is therefore essential to exploit accessible data as well as possible.

Accessible anthropometric databases vary in collection methods, recorded parameters and presentation of data. In terms of the latter, two categories can be identified.

1. **Raw data** of the measured population, i.e. a table of measured dimensions for each individual.
2. **Summary data**, i.e. means and standard deviations, for the reported dimensions.

The number of accessible anthropometric data sets in the first category, i.e. containing raw data, is quite small. For detailed human anthropometry, the options are in practice limited to the Anthropometric Survey of U.S. Army Personnel (Gordon et al. [1989]), better known as ANSUR, and Civilian American and European Surface Anthropometry Resource, CAESAR (Harrison and Robinette [2002], Robinette and Fleming [2002]). Semi-raw ANSUR data in the form of measured dimensions of individuals have been made accessible on the Internet. CAESAR is based on three-dimensional body scanning from

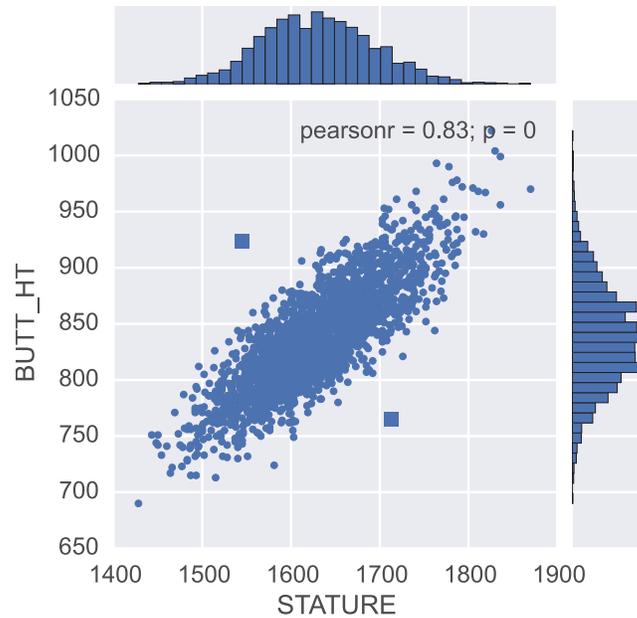


Figure 1 Scatter plot of two dependent parameters from ANSUR, stature and butt height, with their respective univariate distributions indicated on the axes. The circular dots are the actual data points. The square outliers are highly improbable in the population but are probable results of a independent, random variation of the two parameters.

which, in principle, any distinguishable dimension can be derived. It is a commercial product marketed by the Society of Automotive Engineers (SAE). The direct validity of these two databases is naturally limited to the populations and periods in which they were collected.

In contrast to ANSUR and CAESAR, the bulk of accessible databases represent information in terms of summary data only, covering different populations and different subsets of measured dimensions. The National Health and Nutrition Examination Survey (NHANES) (Calafat [2006]) is an American data set covering a large number of subjects but rather few reported dimensions. DINED (Molenbroek [2004]) is an anthropometric database representing the Dutch population and DinBELG (Motmans [2005]) similarly for the Belgian population. Adultdata, Childdata and Older Adultdata (Peebles and Norris [1998]) compile data from many different geographic regions, while the Digital Human Research Center AIST (Kouchi et al. [2000]) manages a data set for the Japanese population. Finally, human factors textbooks such as Pheasant and Hazelgrave [2005] and Chaffin et al. [2006] contain tables of summary data compiled from a variety of sources.

Common to these data sets is that they report means and standard deviations of each variable, implicitly assuming that the measured variables are normally distributed. The mean values provide a coherent description of the dimensions of the average person in the dataset, and the standard deviations describe the variability of the individual parameters over the population, so it would appear that these summary data describe the population well. However, summary data fail to capture correlations between parameters.

Fig. 1 illustrates the strong correlation between two parameters from the ANSUR data set. The central cluster of points represents the measured data, and the two square outliers

represent plausible results of uncorrelated random variations of the two parameters. It is obvious that generation of a pseudo-population by random and uncorrelated variation of the two parameters within their respective normal distributions will result in a circularly shaped pseudo cluster of points that is quite different from the real data and therefore does not represent the population validly.

The goal of this paper is to develop an operational procedure that generates valid, virtual populations matching summary, population-specific target data. This is accomplished by harvesting correlation from a detailed reference data set. The underlying assumption is that the correlation matrix from the reference data set can acceptably represent the inaccessible correlation matrix of the target population. Inter-population differences in anthropometry are known to be significant, but it is possible that different populations exhibit similarities in terms of their correlation matrices. For instance, it is likely that all human populations will share a positive correlation between stature and leg length.

The RASH-method (Pheasant and Hazelgrave [2005]) uses scaling to predict unknown anthropometrical parameters by assuming certain linear relationships between pairs of variables within the same population, e.g. a linear relation between sitting height and stature. Later, more sophisticated, hierarchical regression models based on the ANSUR data set have been developed (Jung et al. [2009], You and Ryu [2005]) to enhance the reliability of prediction of unknown parameters from a few known parameters, e.g. stature and body weight. The resulting methods can generate random dimensions of a virtual person with given input parameters, i.e. predict many parameters from a few. A related problem was addressed by Parkinson and Reed [2010] using principal component analysis to enable the projection of rich data sets like ANSUR and CAESAR onto more updated data with fewer parameters, such as NHANES. Similarly, prediction of missing dimensions from known dimensions was done by linear regression (Brolin et al. [2016]) and by principal component analysis (Brolin [2016]).

The scope of this paper is different. Rather than predicting unknown parameters from known parameters within the same data set, we consider the problem of generating a valid pseudo-population given summary target data and a rich reference data set offering the same parameters as the target. We aim to make the approach general and non-hierarchical, i.e. the approach should not require selection of any particular predictive variables, and it should work for any rich reference and summary target data sets, provided they offer measures of the same dimensions. We validate the method by prediction of known data, i.e. by attempting to generate a virtual female population from a male reference population.

2 Methods

Our approach is divided into two separate steps:

1. Determine a scaling transformation that maps the reference data set into a new data set whose means and standard deviations agree with those of the target data set.
2. Use of the mapped data to generate a virtual population with an arbitrary, user-selected number of individuals by means of principal component analysis.

2.1 Mapping a reference data set

We shall assume that we have a reference population, such as ANSUR, defined by a matrix \mathbf{X} of m anthropometric dimensions of n measured individuals:

$$\mathbf{X} = [x_{ij}], i \in 1..n, j \in 1..m \quad (1)$$

For each column of \mathbf{X} , we can compute empirical means and variances:

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (2)$$

$$s_{x,j}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (3)$$

Similarly, we can compute the entries of the covariance matrix of \mathbf{X} :

$$\mathbf{S}_{x,pq} = \frac{1}{n-1} \sum_{i=1}^n (x_{ip} - \bar{x}_p)(x_{iq} - \bar{x}_q), p, q \in 1..m \quad (4)$$

from which the entries of the correlation matrix is derived:

$$\mathbf{C}_{x,pq} = \frac{\mathbf{S}_{x,pq}}{\sqrt{s_{x,p}^2} \sqrt{s_{x,q}^2}} \quad (5)$$

For the target population, we know the empirical means \bar{y}_j and variances $s_{y,j}^2$ but not individual measurements y_{ij} . However, we assume that to each record x_{ij} in the reference population there corresponds a record y_{ij} in the target population given by the affine relation

$$y_{ij} = a_j x_{ij} + b_j \quad (6)$$

where a_j and b_j are unknown coefficients. We determine these coefficients by matching the known target means and variances with those obtained using (6):

$$\frac{1}{n} \sum_{i=1}^n (a_j x_{ij} + b_j) = \bar{y}_j \quad (7)$$

$$\frac{1}{n-1} \sum_{i=1}^n (a_j x_{ij} + b_j - \bar{y}_j)^2 = s_{y,j}^2 \quad (8)$$

Solving for the coefficients, we obtain

$$\begin{aligned} a_j &= \frac{\sqrt{((\sum x_{ij})^2 - n \sum x_{ij}^2) s_{y,j}^2 (1-n)n}}{(\sum x_{ij})^2 - n \sum x_{ij}^2} \\ &= \sqrt{\frac{s_{y,j}^2}{s_{x,j}^2}} \end{aligned} \quad (9)$$

$$b_j = \bar{y}_j - a_j \bar{x}_j \quad (10)$$

Having thus determined a_j and b_j , we can generate an approximation \mathbf{Y} of the target population using the affine relation (6). Using that each y_{ij} in \mathbf{Y} is given by $y_{ij} = a_j x_{ij} + b_j$ and that $s_{y,j}^2 = a_j^2 s_{x,j}^2$, it is easy to check that the empirical correlation matrix of \mathbf{Y} coincides with that of \mathbf{X} (5), i.e. correlation (but not covariance (4)) remains constant across an affine transformation. In other words, \mathbf{Y} possesses a realistic correlation structure imported from \mathbf{X} .

2.2 *Generation of a virtual population from a rich data set*

Generation of a virtual population, \mathbf{Z} , with an arbitrary number of individuals and validly representing the target population, can be accomplished by random sampling. To do so we transform \mathbf{Y} to a space in which its columns are orthogonal and the covariance matrix reduces to a diagonal (Parkinson and Reed [2010]); in this transformed space, each virtual subject of \mathbf{Z} can be generated by independent sampling. Practically, this is accomplished by Cholesky decomposition or Principal Component Analysis (PCA). The latter is a standard statistical approach, which is available in several numerical libraries, and we shall use one of those implementations to create a mapping from physical to orthogonal space, in which the virtual population will be generated, and an inverse mapping for transferring the sampled data back to physical space. A typical PCA implementation also sorts the columns of the resulting matrix in descending order in terms of variance and provides information about the amount of variance in the data set represented by each column. This carries the additional opportunity of data reduction in the sense that most of the variance can usually be represented by the first few columns in the transformed data. However, data reduction is not the scope of the present work and we use PCA only in the interest of creating a virtual population by independent sampling.

The widespread use of PCA for multiple purposes has brought about extremely efficient and robust numerical implementations of PCA transformation. The computer language Python has been widely adopted by the scientific community in recent years based on an open source strategy. Different collaborative development projects produce high-quality libraries of algorithms for the Python community. Two of these projects are Pandas (McKinney [2012]) for data structures and data analysis tools and scikit-learn (sklearn) (Pedregosa et al. [2011]) for machine learning tasks. PCA is central to the latter, and sklearn consequently provides a robust and efficient set of PCA algorithms. We have used sklearn to develop a compact application for generation of a virtual population upsampled to any number of individuals representative of the ANSUR data set. This will be demonstrated in the results section.

2.3 *Validation*

For validation purposes, we shall use the ANSUR data set (Gordon et al. [1989]) because it is publicly accessible, rich in parameters and contains enough subjects to adequately estimate the correlation structure. From the raw ANSUR data, we remove categorical variables, such as ethnicity and platoon numbers, and subjects for whom one or more parameters are missing. This leaves two data sets for males and females with respectively 1746 and 2205 subjects. The two data sets share the same 131 anthropometric parameters.

We shall investigate the validity of the method by generating a pseudo population and comparing it with empirical data. ANSUR's division into male and female populations

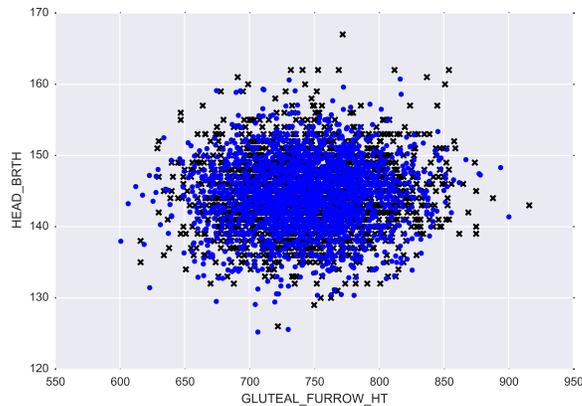


Figure 2 Scatter plot of the parameter pair, gluteal furrow height against head breadth, with the best congruency between the ANSUR female subjects, **Y** (crosses), and the pseudo female subjects, **Z** (dots). Please refer to the ANSUR documentation (Gordon et al. [1989]) for the precise definition of the anthropometric parameters.

represents a convenient opportunity. We compute summary data from the raw female ANSUR data, project the raw male data on them, re-sample the mapped males by means of PCA and compare the result with the original female data.

Prediction of female anthropometry from male subjects is a relatively challenging task, because the anthropometric differences between men and women, even within the same ethnicity, appears to be comparable to or larger than the differences between most ethnic or geographic populations within the same gender.

The authors are not aware of any statistical method to directly quantify the difference between two populations of different sample size in 131-dimensional space. However, two-dimensional combinations, for instance stature against body weight, can be compared both visually and computationally. Provided all parameters are normally distributed, which is assumed for the reference data and ensured by the re-sampling method for the pseudo-data, combinations of any two parameters such as illustrated in Fig. 1 can be investigated for congruency which we here quantify by the relative area of overlap of 95% probability ellipses fitted to the each of the clusters of points (Dueser and Shuggart [1979]).

The 131 parameters can be combined into 8515 different parameter pairs, each of which displays more or less overlap. We shall use results for all combinations as a metric of the validity of the projection method and calculate ellipses and their relative overlaps using the open-source R (R Core Team) package `SIBER` (Jackson et al. [2011]).

3 Results

Figs. 2 and 3 show the parameter pairs with largest respectively smallest relative overlap of ellipses from the original ANSUR female population and the pseudo population generated from the correlation in the male data set.

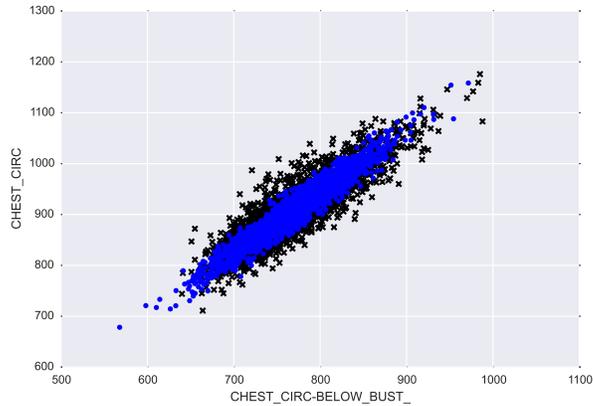


Figure 3 Scatter plot of the parameter pair, chest circumference below bust and chest circumference (i.e. bust circumference), with the worst congruency between the ANSUR female subjects, **Y** (crosses), and the pseudo female subjects, **Z** (dots). Please refer to the ANSUR documentation (Gordon et al. [1989]) for the precise definition of the anthropometric parameters.

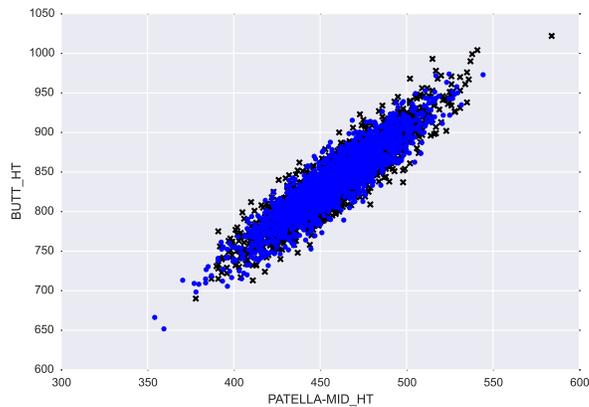


Figure 4 Example of a highly correlated parameter pair, patella mid height and butt height, with a large degree of overlap between the ANSUR female population and the pseudo female population.

The elliptical congruency in Fig. 2 is 99.5% while the value for Fig. 3 is only 51%. The low congruency of the combination of Fig. 3 is an exception; more than 90% of the combinations have congruency better than 90%.

Figs. 2 and 3 may indicate that it is more difficult to obtain a high degree of elliptical overlap for parameter combinations with high correlation, but this is not necessarily the case as illustrated by Fig. 4, which shows the overlap of a highly correlated parameter pair, namely butt height versus patella mid height; its overlap is 99%.

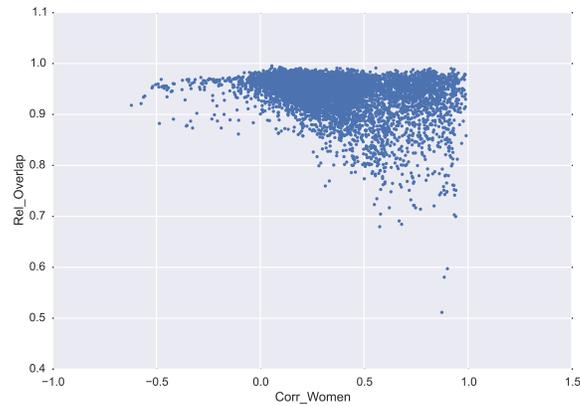


Figure 5 For each pair of variables, the plot shows the relative elliptical overlap between the female and the pseudo female population plotted against the correlation coefficient for the pair in the female population.

An overview of the dependency between the parameter pairs' correlation and the ability of the method to predict them correctly is presented in Fig. 5 and discussed further in Section 4.

4 Discussion

Fig. 3 shows the statistically worst fit out of 8515 combinations with only 51% congruency. Closer examination reveals that the point cluster of the pseudo population is more narrow, i.e. displays a stronger correlation than the original female data. This is not surprising when the nature of the two parameters is considered: They represent the chest circumference at the bust and just below the bust, which is obviously a combination where the gender difference is particularly outspoken.

It is not surprising that prediction of female anthropometry from male anthropometry would be harder for dimensions of large gender differences, and it is important for the assessment of the method to know how often they occur. To this end, Fig. 5 reveals that there are only few parameter combinations that have overlaps less than 70%, and more than 90% of the combinations have relative congruency better than 90%. Although poor relative congruency is exclusively found for highly correlated parameter combinations, most of the highly correlated combinations still have congruency better than 90%.

The results presented above confirm two findings:

- Despite differences in anthropometry, male and female populations appear to share correlation to a large extent and it is not infeasible to apply the correlation of one population to another. This is encouraging because the anthropometrical difference between men and women is probably larger than the differences between most ethnic populations.

- Correlation is unaffected by affine mapping of the data, which allows us to generate pseudo populations by projection of existing reference populations on summary data for a target population.

It is tempting to think that the quality of the pseudo population with respect to the real female population that it simulates can be assessed simply by comparing correlation coefficients between the parameter pairs of the two populations, rather than resorting to elliptical overlap. However, Fig. 2 illustrates why this is insufficient: Pairs of weakly correlated parameters form almost circular point clusters, which have high overlap despite different correlation coefficients. The pair WAIST_NAT_LNTH and WAIST_HT_SIT_NATURAL is a case-in-point. The correlation of this pair is -0.11 for women and 0.25 for pseudo women, but the point clouds are quite similar with a relative area of overlap of 86%.

5 Conclusion

The findings of this paper are encouraging for the opportunity to create valid pseudo populations for different ethnic and geographic populations based on summary data available from the literature or online databases. The applicability of the method is not limited to anthropometry in its traditional form. The method makes no assumptions on the input data other than a physical argument for transfer of correlation from one population to another.

Acknowledgment

This work was partially supported by the Innovation Fund Denmark.

References

- E. Brolin. *Anthropometric diversity and consideration of human capabilities - Methods for virtual product and production development*. PhD thesis, Chalmers University of Technology, 2016. URL <http://publications.lib.chalmers.se/records/fulltext/232940/232940.pdf>.
- E. Brolin, D. Högberg, L. Hanson, and R. Örtengren. Adaptive regression model for prediction of anthropometric data. *International Journal of Human Factors Modelling and Simulation*, In press, 2016.
- A. Calafat. National Health and Nutrition Examination Survey (NHANES) in USA. *Epidemiology*, 17(6):S 76, 2006.
- D. B. Chaffin, G. B. Anderson, and B. J. Martin. *Occupational Biomechanics*. Wiley-Interscience, 2006. ISBN 0-471-72343-6.
- M. Damsgaard, J. Rasmussen, S. T. Christensen, E. Surma, and M. de Zee. Analysis of musculoskeletal systems in the AnyBody Modeling System. *Simulation Modelling Practice and Theory*, 14(8):1100–1111, 2006. doi: 10.1016/j.simpat.2006.09.001.

- C. de Vries, C. J. Garneau, G. Nadadur, and M. B. Parkinson. Considering Secular and Demographic Trends in Designing for Present and Future Populations. In *Proceedings of the ASME 2010 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Montreal, Canada, jan 2010. ASME. ISBN 978-0-7918-4409-0. doi: 10.1115/DETC2010-28879.
- S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, 54(11):1940–1950, 2007. ISSN 00189294. doi: 10.1109/TBME.2007.901024.
- R. D. Dueser and H. H. Shuggart. Niche Pattern in a Forest-Floor Small-Mammal Fauna Published by : Ecological Society of America Stable URL : <http://www.jstor.org/stable/1936473>. *Ecology*, 60(1):108–118, 1979.
- C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, J. T. McConville, I. Tebbetts, and R. A. Walker. Anthropometric Survey of U.S. Army Personnel: Summary Statistics, Interim Report for 1988. Technical report, United States Army Natick Research, Development and Engineering Center, mar 1989.
- C. R. Harrison and K. M. Robinette. CAESAR: Summary Statistics for the Adult Population (Ages 18-65) of the United States of America. Technical report, SAE International, Warrendale, Pennsylvania, jun 2002.
- D. Högberg, E. Brolin, and L. Hanson. Accommodation levels for ellipsoid versus cuboid defined boundary cases. *Procedia Manufacturing* 3, 3(AHFE):3702–3708, 2015. ISSN 23519789. doi: 10.1016/j.promfg.2015.07.799.
- A. L. Jackson, R. Inger, A. C. Parnell, and S. Bearhop. Comparing isotopic niche widths among and within communities: SIBER - Stable Isotope Bayesian Ellipses in R. *Journal of Animal Ecology*, 80(3):595–602, 2011. ISSN 00218790. doi: 10.1111/j.1365-2656.2011.01806.x.
- K. Jung, O. Kwon, and H. You. Development of a digital human model generation method for ergonomic design in virtual environment. *International Journal of Industrial Ergonomics*, 39(5):744–748, 2009. ISSN 01698141. doi: 10.1016/j.ergon.2009.04.001. URL <http://dx.doi.org/10.1016/j.ergon.2009.04.001>.
- M. Kouchi, M. Masaaki, I. Hiromu, and M. Seiji. Anthropometric database for Japanese Population 1997-98. Technical report, Japanese Industrial Standards Center (AIST, MITI), Tokyo, Japan, 2000.
- W. McKinney. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc, Sebastopol, California, 2012. ISBN 978-1-449-31979-3.
- J. F. Molenbroek. DINED anthropometric database, 2004. URL <http://dined.io.tudelft.nl/ergonomics/>.
- R. Motmans. DINBelg 2005, 2005. URL <http://www.dinbelg.be/>.
- B. T. Palumbo, J. Lindsey, and W. Fitz. Patient-specific total knee arthroplasty: A novel technique and implant. *Techniques in Knee Surgery*, 11(4):182–188, 2012. doi: <http://dx.doi.org/10.1097/BTK.0b013e31827db337>.

M. B. Parkinson and M. P. Reed. Creating virtual user populations by analysis of anthropometric data. *International Journal of Industrial Ergonomics*, 40 (1):106–111, 2010. ISSN 01698141. doi: 10.1016/j.ergon.2009.07.003. URL <http://linkinghub.elsevier.com/retrieve/pii/S0169814109000997>.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay. Scikit-learn: Machine Learning in Python. *The Journal of Machine Learning Research*, 12:2825–2830, feb 2011. ISSN 1532-4435. URL <http://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>.

L. Peebles and B. Norris. *Adult data The handbook of adult anthropometric and strength measurements: data for design safety*. Department of Trade and Industry, London, UK, 1998.

S. Pheasant and C. M. Hazelgrave. *Bodyspace: Anthropometry, Ergonomics and the Design of Work*. Taylor and Francis, Boca Raton, Florida, 2005. ISBN 0-415-28520-9.

R Core Team. The R Project for Statistical Computing. URL <https://www.r-project.org>.

K. M. Robinette and S. Fleming. Civilian American AND EUROPEAN SURFACE ANTHROPOMETRY RESOURCE (CAESAR), FINAL REPORT, VOLUME II: DESCRIPTIONS. Technical Report December 1997, United States Airforce Research Laboratory, Dayton, Ohio, 2002.

H. You and T. Ryu. Development of a hierarchical estimation method for anthropometric variables. *International Journal of Industrial Ergonomics*, 35:331–343, 2005. ISSN 01698141. doi: 10.1016/j.ergon.2004.09.007.