International Journal of Industrial Ergonomics 59 (2017) 46–53

Contents lists available at ScienceDirect

International Journal of Industrial Ergonomics

journal homepage: www.elsevier.com/locate/ergon

Adaptive regression model for synthesizing anthropometric population data

Erik Brolin [a, b, *], Dan Högberg [a], Lars Hanson [a, b, c], Roland Örtengren [b]

[a] School of Engineering Science, University of Skövde, Skövde, Sweden
[b] Department of Product and Production Development, Chalmers University of Technology, Gothenburg, Sweden
[c] Industrial Development, Scania CV, Södertälje, Sweden

ARTICLE INFO

Article history:
Received 27 November 2015
Received in revised form
30 July 2016
Accepted 6 March 2017
Available online 14 March 2017

Keywords:
Anthropometry
Regression
Multivariate
Conditional
PCA
Variance

ABSTRACT

This paper presents the development of an adaptive linear regression model for synthesizing of missing anthropometric population data based on a flexible set of known predictive data. The method is based on a conditional regression model and includes use of principal component analysis, to reduce effects of multicollinearity between selected predictive measurements, and incorporation of a stochastic component, using the partial correlation coefficients between predicted measurements. In addition, skewness of the distributions of the dependent variables is considered when incorporating the stochastic components. Results from the study show that the proposed regression models for synthesizing population data give valid results with small errors of the compared percentile values. However, higher accuracy was not achieved when the number of measurements used as independent variables was increased compared to using only stature and weight as independent variables. This indicates problems with multicollinearity that principal component regression were not able to overcome. Descriptive statistics such as mean and standard deviation values together with correlation coefficients is sufficient to perform the conditional regression procedure. However, to incorporate a stochastic component when using principal component regression requires raw data on an individual level.

Relevance to industry: When developing products, workplaces or systems, it is of great importance to consider the anthropometric diversity of the intended users. The proposed regression model offers a procedure that gives valid results, maintains the correlation between the measurements that are predicted and is adaptable regarding which, and number of, predictive measurements that are selected.

1. Introduction

In a design process, when developing products, workplaces or systems, it is of great importance to consider the capabilities and diversity of the intended users. Good ergonomics is achieved when the capabilities of humans match the demands made by the product, workplace or system. Restrictions in these interactions can often be connected to body dimensions of the users. Hence, utilizing anthropometric data and methodologies is often a fundamental part of the process of achieving a good fit between capabilities of humans and design of products, workplaces or systems (Dainoff et al., 2004).

Existing anthropometric data can be acquired from a number of sources such as books (Pheasant and Haslegrave, 2006), articles (Hanson et al., 2009), software (Peoplesize, 2008) and web sources (Delft University of Technology, 2012), where data typically is given as descriptive statistics specifying mean and standard deviation values for each measurement. However, when performing anthropometric statistical analyses it is often desirable to have access to raw data, with values for each measurement given on an individual level. Such data exists but may be outdated or only be available for specific populations that differ significantly in body size and demography from the target population of a product or workplace, e.g. the ANSUR data that was measured 1988 and on U.S. military personnel (Gordon et al., 1989). Another issue with existing anthropometric data is that surveys sometimes include only a few subjects or that all necessary measurements are not included, e.g. Hanson et al. (2009) present Swedish data on only 39 male subjects for some measurements and no circumference measurements are included. Even if an increasing number of measurement studies are

* Corresponding author.
E-mail address: erik.brolin@his.se (E. Brolin).

http://dx.doi.org/10.1016/j.ergon.2017.03.008
0169-8141/

carried out using digital laser scanning techniques in order to get faster measuring processes, more data and data that can be reused for subsequent analyses (Robinette et al., 2002; Godil and Ressler, 2009; Hanson et al., 2009), collecting new and comprehensive anthropometric data is an expensive and time-consuming activity. Hence, due to time and cost limits in design processes there is often a need to utilize existing anthropometric data, while trying to modify the data to suit the target population better. Thus there is a need for methods that can predict and synthesize the variance of anthropometric measurements within the target population, based on a blend of a reduced number of measurements that represents the target population and existing comprehensive anthropometric data. Such a synthesizing procedure can be explained by using data from a detailed sample population to generate regression equations used to predict missing anthropometric population data for a target group (Fig. 1).

Existing methods for predicting missing anthropometric data have previously used either proportionality constants (Drillis et al., 1966) or linear regression with stature and/or weight as independent variables. However, these so-called flat regressions models can make estimations with large errors when there are low correlations between the independent and dependent variables (Gannon et al., 1998; You and Ryu, 2005). You and Ryu (2005) present an alternative hierarchical regression model that uses geometric and statistical relationships between body measurements to create specific linear regression equations in a hierarchical structure. Their results show that using a hierarchical regression model gives better estimates of predicted measurements if more measurements are known and used as input. The hierarchical regression model requires data on measurements highest up in the hierarchy, i.e. stature and body weight, to always be included. It is not certain that both stature and weight are included in all anthropometric sources of interest even if it is the case in most situations. Another issue with the hierarchical regression model is that regression equations need to be constructed manually if a new anthropometric source is to be used. Previous research has shown that a conditional linear regression model that makes use of all known variables to predict the values of unknown measurements is more accurate compared to a flat regression model based on stature and weight, and also to the hierarchical regression model (Brolin et al., 2013, 2016). A conditional linear regression model has the additional advantage that any measurement can be used as an independent variable. An issue when any measurements can be selected as independent variables is multicollinearity, which occurs when there are near-constant linear functions of two or more of the independent variables (Jolliffe, 2002). Multicollinearities between independent variables can lead to very large variances of the estimated regression coefficients and unstable estimates of the regression equations. Principal component regression have shown to be useful to overcome this problem as the principal components are uncorrelated with no multicollinearities between them (Brolin et al., 2016).

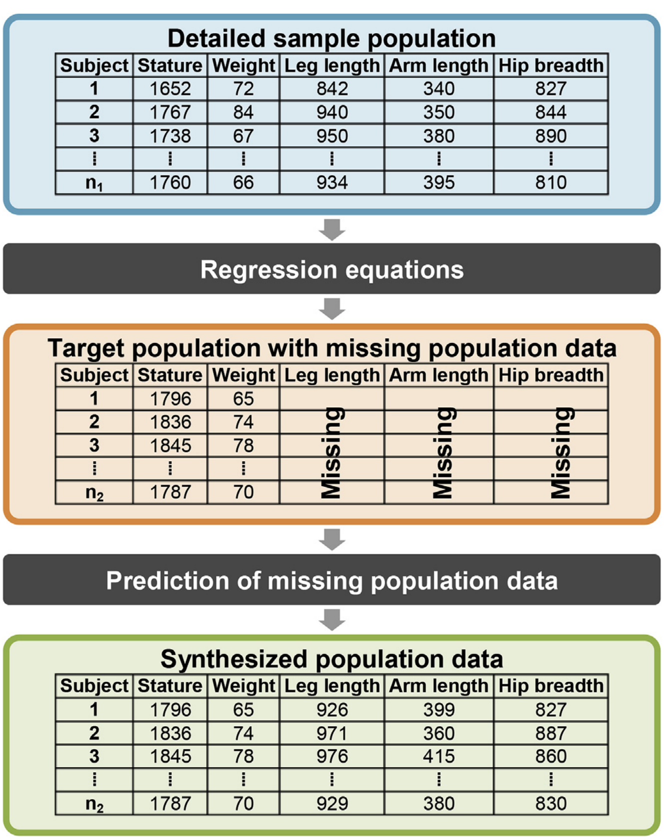

Fig. 1. Synthesizing procedure for missing anthropometric population data.

Even though the conditional regression method makes good prediction of dependent variables, the variance of anthropometric data is not considered. Measurements predicted based on a number of independent measurements will, by using the conditional regression model, always be the same. This is not the case in human populations, e.g. persons of a specific stature will have different weights and proportions (Daniels, 1952). Incorporating a stochastic component to retain residual variance of the anthropometric data increases the accuracy of regression models, especially at percentiles in the tails of the distribution (Nadadur and Parkinson, 2010; Poirson and Parkinson, 2014). The hierarchical regression model presented by You and Ryu (2005) has also been further developed to include a stochastic component. This is achieved by using the corresponding sampling distribution for each regression equation (Jung et al., 2009). Combinations of principal component analysis (PCA) and linear regression to synthesize virtual user populations have been shown to further improve accuracy (Parkinson and Reed, 2010). Incorporation of residual variance has also been shown to give accurate results when predicting preferred design dimensions and behavioural diversity of products (Flannagan et al., 1998; Parkinson and Reed, 2006; Garneau and Parkinson, 2011). When studying reasonably homogeneous populations, most body measurements can be considered normally distributed (Pheasant and Haslegrave, 2006). However, body weight, width and circumference measurements as well as muscular strength often show a positively skewed distribution which would induce errors in the prediction of anthropometric data at percentiles in the tails of the distribution (Pheasant and Haslegrave, 2006).

This paper presents an adaptive regression model which make it possible to synthesize target population data using detailed data from another population, not necessarily closely representative of the target population. The adaptive regression model is based on conditional linear regression and includes use of principal component analysis to consider issues of multicollinearity between selected independent measurements. The developed model also includes incorporation of a stochastic component to retain variance and correlation between measurements. In addition, skewness of the distributions of the dependent variables is considered when incorporating the stochastic components. The developed adaptive regression model and all of its sub-functions are possible to implement in an algorithm, thus making it applicable to any anthropometric data with any number of measurements. The developed model is evaluated by assessing the accuracy of the