2005-01-2735

# Multivariate Accommodation Models using Traditional and 3D Anthropometry

**Martin Friess**
Anthrotech Inc



## ABSTRACT

Various statistical approaches have been advocated that aim at creating statistically meaningful and representative models of human variation. While they all have in common the idea to summarize the critical space needed by the user population by a discrete number of cases, substantial differences exist as to how exactly these cases are identified. The choice of statistical procedures also impacts the number of representative cases (i.e. the efficiency of the model) as well as the actual percentage of the accommodated population (accuracy of the model). The purpose of the paper is to test strengths and fallacies of some of the more commonly found approaches using real as well as simulated data. Furthermore, an extension of multivariate accommodation models to 3D coordinate data, which can be used in CAD/CAM environments, are presented.
The results suggest that while overall accommodation percentages tend to improve when the number of variables and representative cases increases, various other factors can be identified that can significantly reduce or even invalidate the model accuracy. Consequently, simplistic approaches based on multiplying variables/cases do not necessarily guarantee pertinent models. Rather, optimization strategies must be sought to reconcile model efficiency and accuracy.

## INTRODUCTION

It has long been recognized that the use of percentile concepts is based on biologically unrealistic assumptions about multivariate distributions and lead to a reduction in potential user accommodations (Moroney & Smith 1972). The key question of how to substitute the simplistic percentile approach has been answered in various ways. While the use of Principal Components Analysis (PCA) has gained widespread popularity, other multivariate methods have been proposed, of which the most important ones will be briefly described in the following section.

PREVIOUS WORK - Robinette and McConville (1981) have proposed regression, which has the advantage of estimating the distribution of multiple variables associated with a specific value, e.g. the 5th percentile stature (see also Flannagan et al. 1998, Manary et al. 1998). A major caveat here is that the predictions are only as good as the correlations underlying the variables and therefore error is introduced notably at the ends of the distribution where residuals are typically higher. This translates into an over- or underestimate of dimensions at the boundary of physical variation, which is often associated with critical safety clearances etc.
Whitestone and Robinette (1996) pursue what they call feature envelopes to ascertain the inclusion of a maximum portion of the population. Feature envelopes are basically extreme values of multiple variables, preferably represented in 3D, that allow for a product (e.g. a helmet) to be built around the outer range of variation of key features. Covariation of these extreme features among and with each other is discarded from this analysis.
Kim et al. (2004) have pursued a hierarchical approach based on cluster analysis. Procrustes Aligned Landmark data were subjected to a classification algorithm not further specified and determined four shape types that were then converted to sizes of facial respirators. This technique appears to have two major drawbacks which can be labeled with subjectivity and lack of statistical power. Cluster analysis, like PCA, is an exploratory tool that involves no hypothesis testing and therefore yields no statistical significance for any of the groups that an investigator might decipher in the resulting hierarchical tree. Furthermore, the number of clusters, if not predefined, is equal to the number of observations at the lowest level of hierarchy and equal to one at the highest, with no statistical criterion to cut anywhere in between.

Bittner et al. (1987) have proposed a routine based on Principal Components Analysis (PCA). In many anthropometric data sets, this allows for an effective reduction of the raw data to 2 or 3 meaningful dimensions, for which the bi- or trivariate equivalent of a percentile (a probability ellipse or ellipsoid) can be calculated. There are a number of factors that can be varied, making the outcome quite dependent on it, so that it might be more appropriate to refer to the family of PCA-based approaches.

These techniques have gained widespread popularity in applied anthropometry/ergonomics (Meindl et al. 1993; Gordon 2002; Friess & Bradtmiller 2003), though no common protocol for their execution exists. Authors use variably the correlation or the covariance matrix, extract any number of components from two up, and identify an equally variable number of representative cases that are thought to entail the range of the key dimensions. The purpose of this paper is to further investigate PCA-based approaches to multivariate accommodation issues.

## MATERIAL AND METHODS

An anthropometric data set containing eight standard head dimensions and 28 landmarks for a total of 585 subjects was used for all analyses.

Two variants of Bittners PCA approach (Bittner et al. 1987) were computed: one on the correlation matrix and one on the covariance matrix. In the first version, the number of extracted components was restricted to two, as was the case in the original study and as is often replicated by others. A 95% probability ellipse was then calculated, and representative subjects were identified at the intersects with the major axes and at the so-called octant points, resulting in 8 estimates or models of extreme and unusual dimensions in the raw variable space. In the alternative approach, the first four components were retained, as many as necessary to account for 95% of total variance in the sample data set. Subjects with extreme values on one PC and zero on the other were identified and used as representatives. In both variants of the PCA approach, computations were done separately for male and female subjects.

Extreme values (smallest and largest) of the identified boundary manikins of the two approaches were used as accommodation criteria for each subject in the data base. A subject was labeled "accommodated" if all of its dimensions were within the range thus defined.

## RESULTS

The "classic" 2-components solution accounts in reality for 58.9% of the total variance. A 3-components solution would raise this value to 72.9%. The smallest and largest value of the estimated raw variables at the four intersects and the four octant points (of the 95% probability ellipse) for this model are listed in Table 1 under PCA1.

The modified approach uses the first four components, accounting for 94.3% of the total variance. The dimensions observed in subjects with extreme PC scores are listed in Table 1 under PCA2. Both models result in a total of 8 boundary manikins and are in this regard of similar efficiency. In terms of validity, however, the two are in sharp contrast. One can easily note that the metric range defined by the PCA2 accommodation model exceeds that of PCA1, with one exception. The maximum interpupillary distance is higher in the PCA1 model than the corresponding value in PCA2.

With respect to the percentage of the sample that is actually accommodated by the respective model, both approaches fail to achieve the targeted 95% mark. The degree to which PCA1 misses the goal, however, is dramatically higher than for PCA2. Just over half of the sample that served to compute the PCA1 model is within the range of the model, while PCA2 achieves 86%. When only one boundary is considered, PCA1 accommodates 74% above the smallest values and 75% below the highest values, while PCA2 achieves 95% and 91% respectively.

**Table 1: Comparison of multivariate accommodation models derived from two PCA approaches. Note the difference in population percentage actually encompassed by each model.**

| | PCA1 | | PCA2 | |
|---|---|---|---|---|
| | Min | Max | Min | Max |
| **% Extremes** | 75% | 74% | 95% | 91% |
| **% Combined** | 56% | | 86% | |
| **Bitragion-Coronale Arc** | 332.1 | 374.5 | 319.0 | 384.0 |
| **Bizygomatic Breadth** | 129.4 | 151.6 | 118.0 | 157.0 |
| **Head Breadth** | 142.2 | 161.2 | 137.0 | 164.0 |
| **Head Circum-ference** | 535.3 | 600.0 | 518.0 | 617.0 |
| **Head Length** | 182.7 | 211.5 | 180.0 | 215.0 |
| **Interpupillary Distance** | 57.8 | 71.5 | 52.0 | 70.0 |
| **Menton-Sellion Length** | 114.3 | 129.6 | 104.0 | 142.0 |
| **Lip Length** | 49.2 | 63.0 | 48.9 | 65.9 |

It is apparent that the logic behind using PCA to define a range of variation in raw variable space is sound, but the procedure serves its purpose poorly or with variable success at best. The cause for this discrepancy lies of course in the way PCA is calculated. Since it “only” identifies directions of greatest correlation or variance in the sample, taking the first few components, no matter how many, will invariably eliminate relatively small correlations/variances, even though these might lead to combinations a raw dimensions that are particularly critical and hard to accommodate. In this sense, accommodating 95% of PC scores is only equivalent to accommodating 95% of the population when all components are included. Consequently, a two- or three-components solution can only yield the desired result in cases with two or three raw variables. Whether dealing with three components provides any advantage over dealing with three raw variables remains to be investigated. In cases of i>3 variables, the use of as many components as necessary to account for 95% of the total variance can be considered a much improved, though still relatively inaccurate solution.

The only accurate solution to the multivariate accommodation problem is a stochastic one: It takes the reverse route by setting a multivariate accommodation goal (e.g. 95%) and determining the required univariate accommodation limits for each raw variable.
In the present case, in order for 8 variables to achieve a combined probability of 0.95, each one must have its own probability of at least 0.993. This way, 95% of the sample is guaranteed to fall within the metric range of this single model.

Another motivation behind the use of PCA is often described as the search for unusual combinations of raw dimensions. In the case of the head, instead of limiting boundary manikins to the smallest and the largest in all dimensions, say from the shortest and narrowest to the longest and broadest, one would like also to represent subjects with a relatively long but narrow head and vice versa. The “classic” approach of Bittner claims to achieve this with the octant points, pc scores at 45 degrees off each axis and intersecting with the probability ellipse.
Figure 1 illustrates the values of each boundary manikin for all raw variables in the form of a polygon graph. Given the difference in scale between the two circumferences and the rest of the raw dimensions, the former are not included in the figure. Each polygon represents the dimensions of one particular boundary manikin, and as can be seen, all manikin yield fairly similar patterns. Unusual combinations of linear measurements would be identifiable as a deviation from the general pattern of the polygons, a graph that slopes upwards from one dimension to the next, where all others slope downwards, in other words, they would “zigzag”.
In the present case, both approaches contain such cases, even though one approach does not use octant points. These unusual combinations are picked up by the PCA2 approach in higher components (4,7 and 8), but the contrast between dimensions is sharper than it is for the octant points. For instance, PC4 represents very broad and very short heads, while the corresponding octant point combines very broad with medium short heads. The octant points identify intermediates rather than extreme combinations, and their contribution to the modeling of statistical boundaries across multiple variables appears therefore less pertinent.

## MULTIVARIATE ACCOMMODATION OF LANDMARK DATA

When landmark data are used, the same statistical approaches can be applied, provided the data are normalized for translation and rotation. In this case, instead of running for instance a PCA on linear dimensions, the normalized xyz coordinates become the input to the computations. Sets of landmark data that represent the geometry of a head or a body can be subjected to statistical analysis after they have been aligned to a common orientation, using for example a least-squares approach (Bookstein 1991). As an alternative to PCA, a meaningful metric for the accommodation rate of aligned landmark data is centroid size, or more precisely the square root of the sum of the squared distances between the center of gravity and all landmarks at original scale of all specimens. This single variable summarizes the overall size variation in a multivariate set of coordinates. As such, it can be subjected to any descriptive statistic like any other variable including the calculation of percentiles. The difference to ordinary percentiles is that centroid size summarizes variation in a multivariate space. When 10 landmarks are used, the 95th percentile centroid size should therefore entail 99.5 percent of all linear landmark distances from the centroid for all landmarks used. This is illustrated below (Figure 2) in a simplified 2D-verison of the head-data. The result is the exact space occupied by 95% of the population.

## CONCLUSION

The present comparison of methods for the computation of boundary models in applied anthropometry suggests that despite its popularity, the common PCA approach fails to fulfill expectations. The results show that in its simplest variant it can lead to enormous portions of the population (nearly 50%) being left out, and therefore lead us to reject its use. An improved version of it requires the use of many components if not all, as well as the covariance matrix (rather than the correlation matrix) for the extraction of components. The contribution of octant points to the determination of multivariate boundaries remains unclear. Still, even this variant did not achieve the level of accommodation it set out to reach, and caution must therefore be used when relying on its outcome. It is recommended that a PCA derived boundary model be systematically tested against the sample from which it was calculated to allow for possible adjustments. Alternatively, Whenever the desired accommodation rate is motivated more by safety issues than by comfort concerns, a more accurate inverse determination is proposed. This approach satisfies also efficiency criteria in that it is easy to perform and simple to apply to real-world design problems.

## REFERENCES

1. Bittner AC; Glenn FA; Harris RM; Iacchevia HP; Wherry RJ. (1987) CADRE: A Family of Manikins for Workstation Design. Asfour SS (ed) Trends in Ergonomics/Human Factors IV. Elsevier
2. Bookstein FL (1991) Morphometric tools for Landmark Data. Cambridge University Press.
3. Flannagan CAC; Manary MA; Schneider LW; Reed MP (1998) An Improved Seating Accommodation Model with Application to Different User Populations. Human Factors in Driving, Vehicle Seating, and Rear Vision. SAE SP 1358.

4. Friess M & Bradtmiller B (2003) 3D Head Models for Protective Helmet Development. Proceedings of the SAE 2003 Digital Human Modeling Symposium, Montreal, CA, June 16-19, 2003.Gordon CC (2002) Multivariate Anthropometric Models For Seated Workstation Design. In PT McCabe (ed.) Contemporary Ergonomics.
5. Gordon CC (2002) Multivariate Anthropometric Models For Seated Workstation Design. In PT McCabe (ed) Contemporary Ergonomics, pp 582-589.Taylor & Francis
6. Kim K; Kim H; Lee J; Lee E; Kim D (2004) Development of a New 3D Test Panel for Half-Mask Respirators by 3D Shape Analysis for Korean Faces. J. Int. Soc. Resp. Protection 21: 125-134.
7. Manary MA; Flannagan CAC; Reed MP; Schneider LW (1998) Development of an Improved Driver Eye Position Model. Human Factors in Driving, Vehicle Seating, and Rear Vision. SAE SP 1358.
8. Meindl RS; Hudson JA; Zehner GF (1993) A Multivariate Anthropometric Method for Crew Station design. Final Report for Period January 1989-January 1993. AFMC, Wright-Patterson Air Force Base, Ohio 45433.
9. Moroney WF & Smith MJ (1972) Empirical Reduction in Potential User Population as the Result of Imposed Multivariate Anthropometric Limits. Naval Aerospace Medical Research Laboratory, Pensacola, FL
10. Robinette KM & McConville JT (1981) An Alternative to Percentile Models. SAE Technical Paper #810217. SAE Inc.
11. Whitestone JJ & Robinette KM (1996) Fitting to Maximize Performance of HMD Systems.


a)

b)

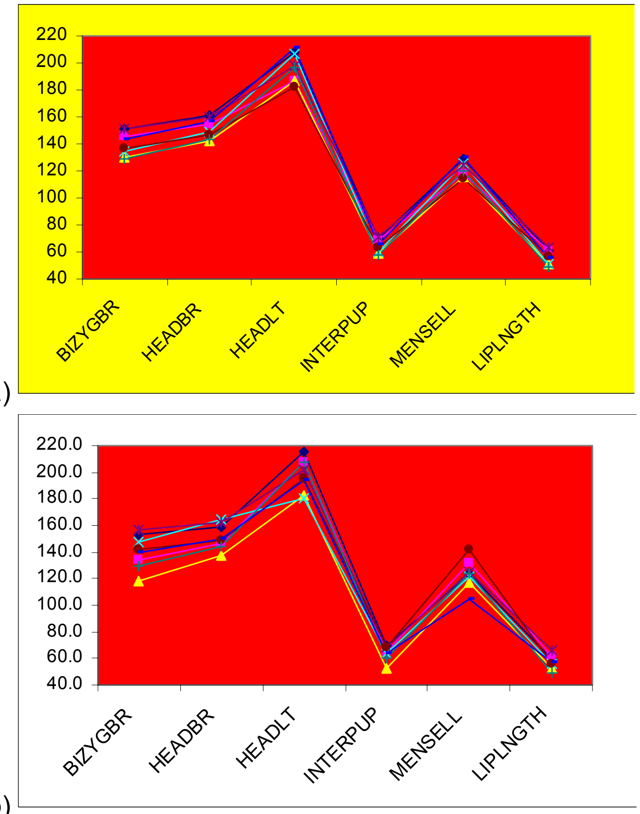


**Figure 1: Comparison of two sets of representative subjects derived by two different PCA approaches. Each polygon links raw dimensions of one of eight representative manikins calculated by the classic Bittner approach (a) and the modified method (b).**

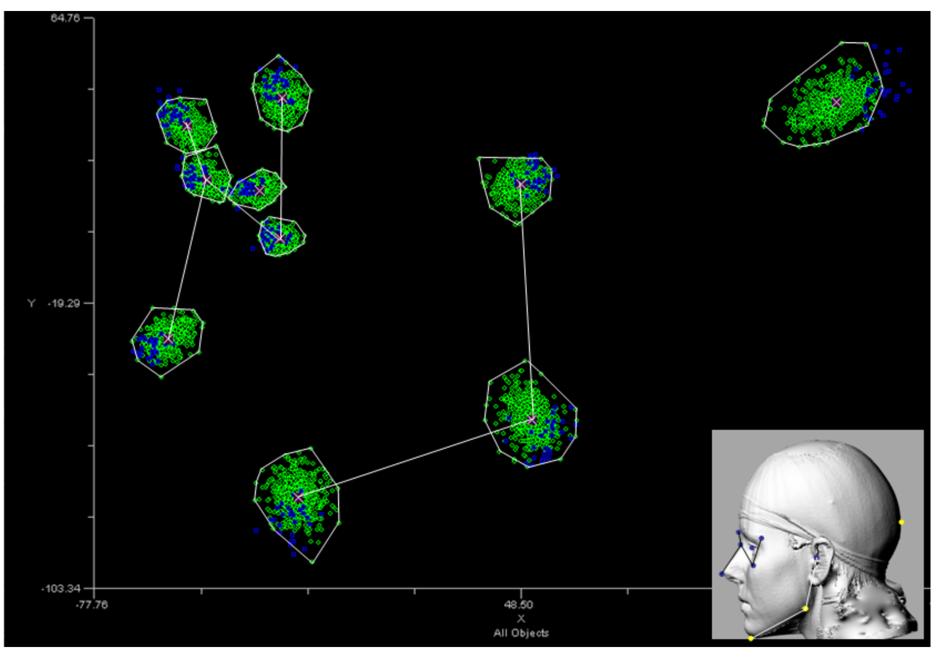


**Figure 2: Space occupied by 95% of the population based on centroid size, exemplified by 2D head landmarks (left lateral view). Green circles: Subjects with 95% centroid size or less. Blue squares: Subjects with more than 95% centroid size. White triangles: Mean head configuration.**