

7-10-2019

# The Development and Use of Computational Tools in Forensic Science

Dennis E. Slice

---

# **The Development and Use of Computational Tools in Forensic Science**

Dennis E. Slice<sup>1,2\*</sup>

<sup>1</sup>Department of Scientific Computing, Florida State University, Tallahassee, Florida, USA.

<sup>2</sup>Department of Anthropology, University of Vienna, Vienna, Austria.

\*Correspondence to: Dennis E. Slice, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306-4120 USA. E-mail: dslice@fsu.edu.

Short Title: Computational Tools in Forensic Science

**KEY WORDS: COMPUTATIONAL FORENSICS, MORPHOMETRICS, STATISTICAL MODELS**

## **Abstract**

Modern computational resources make available a rich toolkit of statistical methods that can be applied to forensic questions. This toolkit is built on the foundation of statistical developments dating back to the 19<sup>th</sup> century. To fully and effectively exploit these developments, both the makers and users of software must be keenly aware of the quality, i.e., the accuracy and precision, of the data being modeled or analyzed, and end-users must be sufficiently familiar with the underlying theory to understand the process and results of any analysis or software they use. This is especially important for medico-legal personnel who might be called upon to testify in a court of law and be subject to cross-examination. With respect to the development of computational tools, it is increasingly important that they be made available as open-source code to avoid the pitfalls of commercial software support and the potential dependence of end-users on orphaned software.

What might be called the classical approach to natural history, including general taxonomy, physical anthropology, and forensics, can be quite effective. This involves the proclamations and pronouncements of a seasoned professional who has carefully inspected thousands of individuals and comparative samples. A problem with this approach is that it is neither easily extensible nor scalable nor transferable to practitioners of lesser expertise. Statistical analysis, arising from probability theory and greatly elaborated in the late 19<sup>th</sup> and early-to-mid 20<sup>th</sup> centuries by the likes of Galton, Pearson, Gosset, and Fisher did much to remove the potential subjectivity of the classical process for simple analyses and measurements. The work of Pearson (1928) on the skull of Henry Stewart, Lord Darnley, through the comparison of portraits and contemporary descriptions of the same and of Pearson and Morant (1934) on the authenticity of the skull of Oliver Cromwell represent early, quantitative analyses addressing the very forensic problem of the individual identity of contested human remains. These studies outline and implement much of the morphometric methodology still useful in the toolkit of the forensic anthropologist. Many of the interesting problems in shape analysis, including forensic applications, are, however, multivariate in nature and not readily suited to the inspection or analysis of individual measures or indices. While Galton, Pearson and others were concerned with correlation amongst variables, it required the relatively recent advent of matrix algebra and multivariate analysis to fully elucidate the principles and potential of the joint analysis of multiple measurements. Cayley defined matrix multiplication and inverses in the mid-19<sup>th</sup> century (Cayley 1858), which was exploited and extended by Wishart in the study of covariances (Wishart 1928) and used by Hotelling (1931) and Wilks (1946) to extend the univariate means testing to suites of covarying variables. As with the much of the work published in *Biometrika* contributing to the development of univariate statistics, physical anthropology played a key role as, for instance, Mahalanobis'

work extending Pearson's Coefficient of Racial Likeness to account for covariance structure and leading to the now ubiquitous Mahalanobis distance (1927, 1936). In the realm of modern shape analysis, Mosier (1939) provided a method for comparing configurations of points in any dimension that is the fundamental operation used in the Generalized Procrustes Analysis of shape (Gower 1975, Rohlf and Slice 1990).

Despite the power and desirability of the approaches developed and available in the early 20<sup>th</sup> century, their sophistication and computational requirements limited their general adoption and application to practical problems of interest to physical and forensic anthropologists – Howells' (1973) extensive use of discriminant functions and factor analysis to analyze a vast amount of craniometric data stands as the exception that proves the rule (Oxnard 1974).

This is not to say that early workers were without computational power, but they had names.

Here is a quote from a letter of Karl Pearson to his funding organization, the Worshipful Company of Drapers, reprinted in Egon Pearson's 1936 biographical obituary (Pearson 1936). K. Pearson writes:

“This has been altered by the appointment of Dr. Lee as a computator, and by the occasional payment of additional calculators, e.g., Miss F. Cave of Girton College.”

Today, even ignoring the potential of parallel processing, there is more than enough computing power available on the desktop along with an accessible panoply of data collection modalities for physical and forensic anthropologists around the world to realize much of the potential of multivariate tools. However, there are a number of considerations that should be kept in mind in their development and use.

The first issue to be addressed in the development and use of computational tools is that of the question being addressed. In this area, applied forensic science has an advantage over

many other fields. The important questions in forensics are often external and clearly stated – What sex was the person represented by the skeletal remains? What was their ancestry? How old were they when they died? How long have they been dead? And the more general and possibly more difficult, who were they? With such clear and important questions in hand, one can begin to develop or select tools appropriate for the question being asked and whatever material might be available for analysis.

Given such a decidedly clear question, the Tool Makers (to borrow a phrase commonly used in general morphometric parlance), can begin to address ways in which it can be answered for specific or commonly available data. Here, the importance of the data used to develop a method cannot be overstated, and its quality and relevance should be given extensive consideration. This is especially true for quality, as relevance will reveal itself in the course of method development and testing.

There are a number of issues to consider with respect to data quality, but note that sample size is an issue of quantity, not quality. With sufficient sample size, one is almost guaranteed statistical significance, but that may or may not be useful in the applied setting. Effect size (Kelley 2012) is a more appropriate indicator of model utility, e.g.,  $R^2$  in a regression problem (Yin and Fan 2001), but see the discussion of dimensionality below.

Data quality is an issue of the repeatability of measurements initially within and between observers involved in the method development, but ultimately across end-users. Quantities that cannot be reliably reproduced by different individuals are not particularly helpful for applied use. After repeatability, data accuracy is a primary concern. One must be certain that the data used for method development is correct. It should be checked and rechecked for outliers and simple transcription errors, and here plotting can play a crucial role. To quote the admonition of the late

Dr. Leslie L. Marcus, “Plot the hell out of your data!” Plot the raw data. Plot the principle components. Plot the histograms. Plot. Plot. Plot.

I have, on more than one occasion, been asked by both students, established researchers, and editors to look at striking or unusual results from a morphometric analysis. As I tell my students, I refuse. Instead, I ask for the original data. After checking the data to be sure it has been accurately encoded and has no result-determining outliers, I will attempt to reproduce the results. Only then might I have something to say. Often, though, I have found an issue with the data that can explain the anomalous results. Simply being off by one can randomize relationships within unordered data sets, and a group miscoding can shockingly invert otherwise standard results. Check your data.

While sample size, itself, is not particularly relevant for model development where any useful effect should be strong enough to be evident with modest numbers of individuals, it may be problematic relative to data dimensionality. Data dimensionality refers to the number of variables recorded per individual observation. Simple sets of measurements like the length and diameter of a long bone or the gnathic, cephalic, and other indices present few problems with respect to sample size. With modern data-acquisition techniques and shape-analysis methods data dimensionality can be a substantial consideration. A sample of, say, 1000 individuals should be more than enough to reveal any relationship of morphology to another variable of interest, but consider that if those 1000 individuals have been scanned at even modest resolution, the surfaces may contain an average of 100,000 points each! In three dimensions, this results in 300,000 measured variables. This situation is ripe for overfitting with the ability to completely separate any partitioning of the original data one might specify. Scans of one million three-dimensional points are not uncommon.

This leads to the often necessary issue of dimension reduction, the most common method for which is Principle Components Analysis (PCA) (Jackson 2005). It is common for researchers to reduce dimension by taking only projections of the original data onto the first, first three, or first  $i$  PCs accounting for  $x$  percent of the sample variation. Here it is important to remember that PCA is driven by one thing – directed variance resulting from variable covariance or correlation. Nothing in the machinery of PCA knows about sex or ancestry or age or anything but variance and variable correlation. The first few PCs are sensitive to outliers in the data and even highly variable measurements, either naturally or those just difficult to collect. Plotting the data projections onto the PCs is, in fact, very useful for checking for outliers and highly variable measurements in the data set, but analyzing just the first few or many PCs can completely miss an otherwise strong signal of interest. Again, check your data and its covariance patterns.

Nonetheless, some sort of dimension reduction is often necessary when dealing with complex data sets, and PCA is a familiar and useful one. One needs to be VERY careful in the selection of components to use and aware of the implications of this choice. For a cautionary note and further discussion of PCA see Bookstein (2017, 2018).

Another thing to keep in mind is that models are generally optimal for the data used to build the model. As such, comparing predicted values from a model to the data used to build the model can be especially misleading. Instead, one must carry out some sort of validation procedure comparing predictions to known values for individuals completely removed from model construction. One such approach I find particularly useful is Lachenbruch's method of cross-validation (Lachenbruch and Mickey 1968) where each individual is removed from the dataset, a model built anew, and that individual's trait predicted from the new model. This gives a realistic indication of the performance of a method, and I even include doing this as an option in,



for instance, *3D-ID* (Slice and Ross 2009), so that the user can see for themselves how well the software is expected to perform.

With respect to Tool Makers, I would suggest a few important attributes of any tool expected to be used by the community. First, the implementation should be platform-independent. This is not that hard to achieve. The R-programming environment (R Core Team 2015) provides many tools for data manipulation and can be downloaded to run on any common platform – Windows, OS X, Linux. Java (<https://www.java.com>) is often the most popular programming language at any given time (e.g., <https://www.tiobe.com/tiobe-index/>), has virtual machines available for common operating systems, and development environments allowing for the easy production of simple, intuitive user interfaces. If raw power is required – the previous suggestions are “interpreted” environments and, hence, slower, then C/C++ are good choices with compilers available for any platform.

Secondly, it is critical that software be open-source. That is, the source code should be made freely available for users, but mostly other programmers, to inspect or modify or debug. This is critical, though I have violated this suggestion in the past (but I am trying to do better, e.g., *forAge* (Stoyanova, Algee-Hewitt, and Slice 2015) and *gpsa* (Pomidor and Slice 2015)), because programs distributed as executables, no matter how useful, die with the interest of the programmer or with the programmer, themselves. If source code is available, there is at least a chance that another capable person can take over development and maintenance or find bugs and problems with the original code.

In contrast, the support and maintenance of commercial software is a daunting task, and a lot of money is required to keep on staff a sufficient number of programmers and customer-service representatives. Anything less than a full commercial enterprise behind a package is

mostly doomed to failure. However, the money required to support such an enterprise is likely to be more than a relatively small community can provide or even justify. Open-source software of high quality, even from an individual developer, greatly diminishes the chance of later user dependence on an orphaned program.

The comments thus far have been mostly directed at Tool Makers. The Tool Users also have a responsibility for the appropriate and efficacious use of the tools provided. Here again, data quality is critical. Even if one has but a single bone, the measurements of which are to be submitted to a classification or prediction program, the collection of those measurements must be tested and retested to ensure their greatest precision and accuracy. Method developers can help here by providing raw material through which end-users can compare their data collection results to those assumed by the developer, though this may be practical only for digitally recorded specimens.

Finally, it is absolutely necessary that the end-user has a sufficient understanding of the methods being implemented and the data upon which a method or result is based. This is especially critical for medico-legal professionals who might be called upon to give court testimony and be subject to cross-examination. Any set of appropriately structured numbers submitted to a program will give some sort of appropriately labeled result. Without understanding the underlying processing and models, the end-user really has no idea of their validity. This does not mean that end-users must be able to derive equations or expound on the details of underlying theory, but it does mean they should be able to explain what the program is doing, what relationships within the data are being exploited, and what the results mean, especially as to the uncertainty inherent in them. As a general rule, if you cannot explain what you did to a random person on the street, then you don't understand the method enough to be

using it.

The above provides some basic considerations for the development and use of computational models and software in the forensic setting. These should be quite familiar to most involved in these activities. If they are not, study them carefully and take them to heart.

### **Acknowledgments**

This work was supported by a National Institute of Justice grant (2015-DN-BX-K010) awarded to D. E. Slice and B. F. B. Algee-Hewitt.

*Received 15 September 2018; accepted for publication 9 December 2019.*

## Literature Cited

- Bookstein, F. L. 2017. A method of factor analysis for shape coordinates. *Am. J. Phys. Anthropol.* 164:221–245.
- Bookstein, F. L. 2018. *A Course in Morphometrics for Biologists*. Cambridge: Cambridge University Press.
- Cayley, A. 1858. A memoir on the theory of matrices. *Philos. Trans. R. Soc. Lond.* 148:17–37.
- Gower, J. C. 1975. Generalized procrustes analysis. *Psychometrika* 40:33–51.
- Hotelling, H. 1931. The generalization of Student's ratio. *Ann. Math. Stat.* 2:360–378.
- Howells, W. W. 1973. Cranial variation in man: A study by multivariate analysis of patterns of difference among recent human populations. Papers of the Peabody Museum of Archaeology and Ethnology, Harvard University, v. 67.
- Jackson, J. E. 2005. *A User's Guide to Principal Components*. Hoboken, New Jersey: John Wiley & Sons.
- Kelley, K., and K. J. Preacher. 2012. On effect size. *Psychol. Methods* 17:137–152.
- Lachenbruch, P. A., and M. R. Mickey. 1968. Estimation of error rates in discriminant analysis. *Technometrics* 10:1–11.
- Mahalanobis, P. C. 1927. Analysis of race mixture in Bengal. *J. Proc. Asiat. Soc. Bengal* 23:301–333.
- Mahalanobis, P. C. 1936. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India* 2:49–55.
- Mosier, C. I. 1939. Determining a simple structure when loadings for certain tests are known. *Psychometrika* 4:149–162.
- Oxnard, C. E. 1974. Cranial variation in man: A study by multivariate analysis of patterns of

- difference among recent human populations. *Am. J. Phys. Anthropol.* 41:349–351.
- Pearson, E. S. 1936. Karl Pearson: An appreciation of some aspects of his life and work. *Biometrika* 28:193–257.
- Pearson, K. 1928. The skull and portraits of Henry Stewart, Lord Darnley, and their bearing on the tragedy of Mary, Queen of Scots. *Biometrika* 20B:1–104.
- Pearson, K., and G. M. Morant. 1934. The Wilkinson Head of Oliver Cromwell and its relationship to busts, masks and painted portraits. *Biometrika* 26:1–116.
- Pomidor, B., and D. E. Slice. 2016. *Generalized Procrustes Surface Analysis (GPSA)*. <http://morphlab.sc.fsu.edu/software/gpsa/index.html>.
- R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <https://www.r-project.org>.
- Rohlf, F. J., and D. E. Slice. 1990. Extensions of the Procrustes method for the optimal superimposition of landmarks. *Syst. Zool.* 39:40–59.
- Slice, D. E., and A. Ross. 2010. *Geometric Morphometric Classification of Crania for Forensic Scientists*. <https://www.ncjrs.gov/pdffiles1/nij/grants/231195.pdf>.
- Stoyanova, D., B. F. B. Algee-Hewitt, and D. E. Slice. 2015. *ForAge*. <http://morphlab.sc.fsu.edu/software/forAge/index.html>.
- Wilks, S. S. 1946. Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Ann. Math. Stat.* 17:257–281.
- Wishart, J. 1928. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika* 20A:32–52.
- Yin, P., and X. Fan. 2001. Estimating  $R^2$  shrinkage in multiple regression: A comparison of

different analytical methods. *J. Exp. Educ.* 69:203–224.