M. GREENACRE & J. BLASIUS (2006) *Multiple Correspondence Analysis and Related Methods.* Boca-Raton, FL: Chapman-Hall. 608 pp. US$99.95. ISBN 1584886285.

The editors have homogenized the style and presentation in a very pleasing way in this multi-authored work. The aim is to present a comprehensive overview of the state of play in this area, for a social or behavioral science reader. Chapter 1 is a good overall introduction, noting that "the data is king" – hence that model follows data rather than the other way around, and that correspondence analysis maps data *profiles* endowed with the χ^2 distance into a Euclidean factor space, thereby rendering visualization possible.

For me, the richness of this book lies in its linkages with other areas. This will help greatly if the reader comes from a different statistical or other background. Essentially every one of its 23 chapters deals to a greater or lesser extent with correspondence analysis as well as some other data analysis techniques. The following were for me the most prominent of these other analysis methods.

Greenacre (chapter 2) links up with canonical correlation analysis, and homogeneity analysis which was developed by the Leiden school under the collective Gifi name. Personally I have never bothered with the distinction between "ordinary", joint and multiple correspondence analysis (any more than the distinction between univariate and multivariate since the former rarely is a practical issue!), but the interested reader will find a well laid out treatment of these terms in this chapter.

In chapter 3 Gower gives (yet another) masterly exposition of the principles of data table decomposition. In chapter 4, de Leeuw develops nonlinear principal components analysis, with also a logistic version, and the use of Voronoï diagrams in output display. Rouanet in chapter 5 gives a lovely overview of geometric data analysis issues, and discusses Pierre Bourdieu's use of correspondence analysis in his social investigations of taste. Nishisato in chapter 6 takes and develops the dual scaling perspective. In chapter 7 Lebart explores validation of results. Greenacre and Pardo in chapter 8 look at how one focuses in analysis on subsets of attributes. Warrens and Heiser in chapter 9 look at Guttman, Rasch, Coombs and related one-dimensional data mappings. Continuing the one-dimensional mapping theme, in chapter 10 van Schuur and Blasius study dichotomous (boolean) variables in the item response approach to dominance and proximity data. Takane and Hwang in chapter 11 regularize correspondence analysis relative to (i) a data model,

and (ii) a stochastic model, where the latter is based, for example, on random permutations, thereby furnishing confidence regions.

Matschinger and Angermeyer in chapter 12 look at "don't know" attributes and generalized canonical analysis. (Not considered, though, is work from *Les Cahiers de l'Analyse des Données*, based on coding approaches, including doubling and cumulating complementary attributes, cf. section 3.2.1 of Murtagh, 2005.) Pagès and Bécue-Bertaut in chapter 13 look at multiple factor analysis. They consider equilibrium between the groups of attributes that enter into the analysis, and their discussion leads to the consideration of separate and indeed "pseudoseparate" analyses. On a related theme, Zárraga and Goitisolo in chapter 14 pursue a similar objective, viz. simultaneous analysis of data tables. They focus on whether or not the marginals of the different data tables should be related or fixed. In chapter 15, Abascal, Lautre, and Landaluce extend this same theme to mixed quantitative and qualitative data.

Classification means clustering (unsupervised classification) or discriminant analysis (supervised classification) and the latter is at issue in chapter 16, by Saporta and Niang. Regression, the continuous analog of discriminant analysis, is also considered in chapter 16. In chapter 17, Bougeard, Hanafi, Noçairi and Qannari study canonical correlation analysis for categorical variables. Caussinus and Ruiz-Gazen in chapter 18 use projection pursuit on categorical variables. Projection pursuit, due to Friedman and Tukey, looks for a low dimensional embedding such that interesting relationships in the data show up well.

The final collection of chapters is introduced by the title "Related Methods", but indeed the entire book could have had this as a subtitle. In chapter 19, Torres-Lacomba deals with conjoint analysis, as used in marketing, applied here to categorical data. Blasius and Thiessen in chapter 20 look at survey analysis, starting from principal components analysis, and then proceeding to a categorical principal components analysis (where scores are derived from the principal component projections), and nonlinear principal components analysis, considered earlier. Kroonenberg and Anderson in chapter 21 explore three-way contingency tables, using additive and multiplicative models. Groenen and Koning in chapter 22 investigate analysis of variance. In chapter 23, Vicente-Villardón, Galindo-Villardón and Blázquez-Zaballos look at logistic (for boolean data) biplots (simultaneous row/column displays).

The software appendix relates to software that is to be made available

on the website www.carme-n.org. There is discussion frequently too of other software programs and environments in the individual chapters, with other web addresses provided. Like my R software at www.correspondances.info, a very similar approach is taken here based on assembling and describing snippets of R code to perform the tasks needed. The great benefit then is availing of R for graphics and display. I also find it handy to pare down the programs used to the minimum if, say, I am analyzing a data table with a few tens of thousand or more rows or columns.

Less emphasis is given in this book than I would have liked to the issue of data encoding, and the many possibilities offered by correspondence analysis for this (see my Murtagh, 2005). When correspondence analysis is used in fields such as information retrieval, linguistics, and the physical sciences, then it is very clear that data analysis starts with the choice and encoding of data. In survey response analysis, handling varied kinds of missing and "don't know" situations, and catering for uncertainty in data, all call for very careful consideration of data preprocessing. Such upstream aspects of data analysis are very ably handled within the correspondence analysis perspective.

As mentioned, the richness of topics will provide readers coming from other backgrounds with lots of entry points into correspondence analysis. As I read through this book, I couldn't help but becoming slightly perplexed. Rather like the dog that Watson referred to, I began to feel like Sherlock Holmes that there was something odd about the lack of barking in the night. It dawned on me: page 4 refers to "Benzécri's seminal work, *Analyse des Données*", and the citation chips in with Benzécri et al., *L'Analyse des Données: L'Analyse des Correspondances*, Dunod, 1973. But, you see, this citation is volume 2, under the aegis of Huyghens. Volume 1 is *La Taxinomie*, under the aegis of Linnaeus. Aha! Clustering is more or less completely missing in this book – there is just a very little discussion in Rouanet's chapter 5 – and with it VACOR and FACOR (mutual description of clusters and factors), and a major part of the Benzécri tradition in data analysis. This is a pity, since it curtails the data analysis, not least for the following reasons.

In the Euclidean, factor space, embedding furnished by correspondence analysis we have a ready-made weighting of variables and observations for subsequent hierarchical (or non-hierarchical) clustering. Then, too, the minimum variance compactness criterion is very germane to the inertia-based parsimony principle applied in correspondence analysis. Over and above clustering, clearly many other data analysis techniques become readily and

easily applicable to the Euclidean factor space embedding. What is more, the χ^2 metric endowed profiles are mapped into (i) a Euclidean factor space, facilitating display, and (ii) an ultrametric, or tree metric, space, thereby further facilitating interpretation. The rich palette of data coding strategies available in correspondence analysis may even *per se* facilitate ultrametric embedding (cf. Murtagh, 2004),and hence data understanding.

The importance of this combined way of approaching data analysis, through the embedding topologies, becomes especially pronounced when handling massive, potentially very high dimensional, data sets. As an integrated data analysis environment or platform, correspondence analysis *au sens large* is up to the future challenges of facing vast masses of complex data. As two examples among many we could mention "India SCAN 1996", part of a large international study of social climate produced by the social and marketing research company DYG Inc.; and the Politbarometer longitudinal political survey of the Zentralarchiv für empirische Sozialforschung in Cologne, Germany. Preliminary aspects of analysis of these data sets were published in Benzécri and Benzécri (1997), and at least the Politbarometer data is touched on in this Greenacre and Blasius volume.

All in all, the Greenacre and Blasius volume does very well what it sets out to do, and is a valuable volume to have close to hand.

Fionn Murtagh
Department of Computer Science, Royal Holloway
University of London, Egham TW20 0EX, England
fmurtagh@acm.org

References

Benzécri, J.P. et coll. (1973). *L'Analyse des Données*, *Tome I: La Taxinomie*, *Tome II: Correspondances*, Paris: Dunod. 2nd edn., 1976.

Benzécri, J.P. and Benzécri, F. (1997). Sources de programmes d'analyse de données en langage Pascal. (III) Élaboration de tableaux divers. (IIIG) Codage d'échelles suivant l'équation personnelle. *Les Cahiers de l'Analyse des Données*, XXII, 375–410.

Murtagh, F. (2004). On ultrametricity, data coding, and computation, *Journal of Classification*, 21, 167–184.

Murtagh, F. (2005). *Correspondence Analysis and Data Coding with R and Java*, Boca-Raton, FL: Chapman-Hall. Software: www.correspondances.info