Annotated Bibliography of Multivariate Statistical Methods in Astronomy

F. Murtagh and A. Heck

Version: 1986

Abstract

Application studies involving the use of multivariate statistical methods in astronomy are referenced, along with many annotations as to the methods employed and the significance of the work. Additionally, general works of reference are listed. In all more than 150 references are listed, and an index of authors is included.

1 Introduction

When faced with large quantities of data, the use of statistical data analysis and pattern recognition algorithms can offer considerable time-savings, together with ensuring consistency and "objectivity" of treatment. Being multivariate (multidimensional), they allow the simultaneous treatment of many variables.

There have been many types of multivariate statistics algorithms, but among the most commonly used are algorithms for *Cluster Analysis*, *Discriminant Analysis*, *Principal Components (or Factor) Analysis*, and *Regression Analysis*.

Given a set of objects, each characterised on the same set of variables, clustering methods will produce groups of the objects. The objects in the resulting groups will either be closer to one another than to non-group members, or satisfy some other homogeneity or compactness criterion. "Closeness" is most often defined by the Euclidean distance, but other metrics may well merit consideration. The question of "standardization" or "normalization" (centring the objects in the multidimensional space and rescaling them to have unit variance) may also have to be addressed before carrying out the clustering.

Discriminant methods allow assigment of objects to already existing groups. Such methods may use locally-defined metrics, and thus be sensitive to different parts of the parameter space; or they may be based on Bayesian probability. In Discriminant Analysis, the first step will be to choose a training set; then, in a second step, new items are assigned to the most appropriate class of items. Discriminant Analysis has been refered to as "supervised classification" (because of the need to define the training set, - perhaps by a visual study of a relatively small number of objects), while Cluster Analysis has been termed "unsupervised classification".

Principal Components Analysis is used for dimensionality reduction The best linear combinations of the axes in the initial parameter space are sought (the criterion of fit used is a least squares one). It can be used to study what the most relevant variables are for the objects or items studied.

Regression, or curve fitting generally, are problem areas which are widely known in the physical sciences.

This bibliography is motivated by increasingly wide interest in the use of multivariate statistical methods in astronomy. The researcher has, however, a basic difficulty in going to one of the available on–line bibliographic databases and, for example, doing a search for all work involving "clusters"! For this reason, it is helpful to have available a select bibliography, both of work carried out in astronomy, and also of the more important works outside astronomy.

In the following, it is attempted to be reasonably comprehensive; the principal objective is that a selection of the literature available on particular topics be listed, and in the case of the general bibliographies, important works — mainly books — be given. In some cases where it was felt useful, references are repeated in different sections; in general, however, it may be noted that books often have material of relevance for topics other than those under which they are listed. Computer packages are sometimes listed: often the relevant documentation and examples provide a quick and painless way to get information on particular techniques.

Finally, a warm acknowledgement is extended to the many colleagues who, at one time or another, said: "Oh, there is an article which might be of interest in a recent issue of ...".

2 Cluster Analysis: Astronomy

Principal Components Analysis has often been used for determining a classification, and these references are not included here.

The problems covered in the following include: star-galaxy separation, using digitized image data; spectral classification, — the prediction of spectral type from photometry; taxonomy construction (for asteroids, stars, and stellar light curves); galaxies; gamma and X–ray astronomy; a clustering approach not widely used elsewhere is employed for studies relating to the moon, to asteroids and to cosmic sources; and work relating to interferogram analysis is represented.

1. J.D. Barrow, S.P. Bhavsar and D.H. Sonoda, "Minimal spanning trees, filaments and galaxy clustering", *Monthly Notices of the Royal Astronomical Society*, **216**, 17-35, 1985.

   (This article follows the seminal approach of Zahn — see reference among the general clustering works — in using the MST for finding visually evident groupings.)

2. R. Bianchi, A. Coradini and M. Fulchignoni, "The statistical approach to the study of planetary surfaces", *The Moon and the Planets*, **22**, 293-304, 1980.

   (This article contains a general discussion which compares the so-called G-mode clustering method to other multivariate statistical methods. Other references by Coradini, Carusi, and others, also use this method.)

3. R. Bianchi, J.C. Butler, A. Coradini and A.I. Gavrishin, "A classification of lunar rock and glass samples using the G-mode central method", *The Moon and the Planets*, **22**, 305-322, 1980.

4. A. Bijaoui, "Méthodes mathématiques pour la classification stellaire", in *Classification Stellaire, Compte Rendu de l'Ecole de Goutelas*, ed. D. Ballereau, Observatoire de Meudon, Meudon, 1979, pp. 1-54.

   (This presents a survey of clustering methods.)

5. R. Buccheri, P. Coffaro, G. Colomba, V. Di Gesù, S. Salemi, "Search of significant features in a direct non-parametric pattern recognition method. Application to the classification of multiwire spark chamber pictures", in (eds.) C. de Jager and Neiuwenhuijzen, *Image Processing Techniques in Astronomy*, D. Reidel, Dordrecht, pp. 397-402, 1975.

   (A technique is developed for classifying γ-ray data.)

6. S.A. Butchins, "Automatic image classification", *Astronomy and Astrophysics*, **109**, 360-365, 1982.

   (A method for determining Gaussian clusters, due to Wolf, is used for star/galaxy separation in photometry.)

7. A. Coradini, M. Fulchignoni and A.I. Gavrishin, "Classification of lunar rocks and glasses by a new statistical technique", *The Moon*, **16**, 175-190, 1976.

   (The above, along with the references of Bianchi and others, make use of a novel clustering technique termed the G-mode method. The above contains a short mathematical description of the technique proposed.)

8. A. Carusi and E. Massaro, "Statistics and mapping of asteroid concentrations in the proper elements' space", *Astronomy and Astrophysics Supplement Series*, **34**, 81-90, 1978.

   (This article also uses the so-called G-mode method, employed by Bianchi, Coradini, and others.)

9. C.R. Cowley and R. Henry, "Numerical taxonomy of Ap and Am stars", *The Astrophysical Journal*, **233**, 633-643, 1979.

   (40 stars are used, characterised on the strength with which particular atomic spectra — the second spectra of yttrium, the lanthanides, and the iron group — are represented in the spectrum. Stars with very similar spectra end up correctly grouped; and anomolous objects are

detected. Clustering using lanthanides, compared to clustering using iron group data, gives different results for A_p stars. This is not the case for A_m stars, which thus appear to be less heterogeneous. The need for physical explanations are thus suggested.)

10. C.R. Cowley, "Cluster analysis of rare earths in stellar spectra", in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 1983, pp. 153-156.

    (About twice the number of stars, as used in the previous reference, are used here. A greater role is seen for chemical explanations of stellar abundances and/or spectroscopic patterns over nuclear hypotheses.)

11. J.K. Davies, N. Eaton, S.F. Green, R.S. McCheyne and A.J. Meadows, "The classification of asteroids", *Vistas in Astronomy*, **26**, 243-251, 1982.

    (Phyiscal properties of 82 asteroids are used. The dendrogram obtained is compared with other classification schemes based on spectral characteristics or colour–colour diagrams. The clustering approach used is justified also in being able to pinpoint objects of particular interest for further observation; and in allowing new forms of data — e.g. broadband infrared photometry — to be quickly incorporated into the overall approach of classification–construction.)

12. G.A. De Biase, V. di Gesù and B. Sacco, "Detection of diffuse clusters in noise background", *Pattern Recognition Letters* **4**, 39-44, 1986.

13. P.A. Devijver, "Cluster analysis by mixture identification", in V. Di Gesù, L. Scarsi, P. Crane, J.H. Friedman and S. Levialdi (eds.), *Data Analysis in Astronomy*, Plenum Press, New York, 1984, pp. 29-44.

    (A very useful review article, with many references. A perspective similar to perspectives adopted by many discriminant analysis methods is used.)

14. V. Di Gesù and B. Sacco, "Some statistical properties of the minimum spanning forest", *Pattern Recognition*, **16**, 525-531, 1983.

    (In this and the following works, the minimal spanning tree or fuzzy set theory — which, is clear from the article titles — are applied to point pattern distinguishing problems involving gamma and X-ray data. For a rejoinder to the foregoing reference, see R.C. Dubes and R.L. Hoffman, "Remarks on some statistical properties of the minimum spanning forest", *Pattern Recognition*, **19**, 49-53, 1986. A reply to this article is forthcoming, from the authors of the original paper.)

15. V. Di Gesù, B. Sacco and G. Tobia, "A clustering method applied to the analysis of sky maps in gamma–ray astronomy", *Memorie della Società Astronomica Italiana*, 517-528, 1980.

16. V. Di Gesù and M.C. Maccarone, "A method to classify celestial shapes based on the possibility theory", in G. Sedmak (ed.), ASTRONET 1983 (Convegno Nazionale Astronet, Brescia, Published under the auspices of the Italian Astronomical Society), 355-363, 1983.

17. V. Di Gesù and M.C. Maccarone, "Method to classify spread shapes based on possibility theory", Proceedings of the 7th International Conference on Pattern Recognition, Vol. 2, IEEE Computer Society, 1984, pp. 869-871.

18. V. Di Gesù and M.C. Maccarone, "Features selection and possibility theory", *Pattern Recognition*, *19*, 63-72, 1986.

19. J.V. Feitzinger and E. Braunsfurth, "The spatial distribution of young objects in the Large Magellanic Cloud — a problem of pattern recognition", in eds. S. van den Bergh and K.S. de Boer, *Structure and Evolution of the Magellanic Clouds*, IAU, 93-94, 1984.

    (In an extended abstract, the use of linkages between objects is described.)

20. I.E. Frank, B.A. Bates and D.E. Brownlee, "Multivariate statistics to analyze extraterrestial particles from the ocean floor", in V. Di Gesù, L. Scarsi, P. Crane, J.H. Friedman and S. Levialdi (eds.), *Data Analysis in Astronomy*, Plenum Press, New York, 1984.

21. A. Fresneau, "Clustering properties of stars outside the galactic disc", in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 1983, pp. 17-20.

(Techniques from the spatial processes area of statistics are used to assess clustering tendencies of stars.)

22. A. Heck, A. Albert, D. Defays and G. Mersch, "Detection of errors in spectral classification by cluster analysis", *Astronomy and Astrophysics*, **61**, 563-566, 1977.

23. A. Heck, D. Egret, Ph. Nobelis and J.C. Turlot, "Statistical confirmation of the UV spectral classification system based on IUE low–dispersion stellar spectra", *Astrophysics and Space Science*, **120**, 223-237, 1986.

    (Among other results, it is found that UV standard stars are located in the neighbourhood of the centres of gravity of groups found, thereby helping to verify the algorithm implemented. A number of other papers, by the same authors, analysing IUE spectra are referenced in this paper. Apart from the use of a large range of clustering methods, these papers also introduce a novel weighting procedure — termed the "variable procrustean bed" — which adjusts for the symmetry/asymmetry of the spectrum. Therefore, a useful study of certain approaches to the coding of data is to be found in these papers.)

24. J.P. Huchra and M.J. Geller, "Groups of galaxies. I. Nearby groups", *The Astrophysical Journal*, **257**, 423-437, 1982.

    (The single linkage hierarchical method, or the minimal spanning tree, have been rediscovered many times — see, for instance, Graham and Hell, 1985, referenced in the general clustering section. In this study, a close variant is used for detecting groups of galaxies using three variables, — two positional variables and redshift.)

25. J.F. Jarvis and J.A. Tyson, "FOCAS: faint object classification and analysis system", *The Astronomical Journal*, **86**, 476-495, 1981.

    (An iterative minimal distance partitioning method is employed in the FOCAS system to arrive at star/galaxy/other classes.)

26. G. Jasniewicz, "The Böhm-Vitense gap in the Geneva photometric system", *Astronomy and Astrophysics*, *141*, 116-126, 1984.

    (The minimal spanning tree is used on colour-colour diagrams.)

27. A. Kruszewski, "Object searching and analyzing commands", in *MIDAS — Munich Image Data Analysis System*, European Southern Observatory Operating Manual No. 1, Chapter 11, 1985.

    (The *Inventory* routine in MIDAS has a non–hierarchical iterative optimization algorithm. It can immediately work on up to 20 parameters, determined for each object in a scanned image.)

28. M.J. Kurtz, "Classification methods: an introductory survey", in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 1983.

    (Kurtz lists a large number of parameters — and functions of these parameters — which have been used to differentiate stars from galaxies.)

29. J. Materne, "The structure of nearby clusters of galaxies. Hierarchical clustering and an application to the Leo region", *Astronomy and Astrophysics*, **63**, 401-409, 1978.

    (Ward's minimum variance hierarchic method is used, following discussion of the properties of other hierarchic methods.)

30. M.O. Mennessier, "A cluster analysis of visual and near–infrared light curves of long period variable stars", in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 1983, pp. 81-84.

    (Light curves — the variation of luminosity with time in a wavelength range — are analysed. Standardization is applied, and then three hierarchical methods. "Stable clusters" are sought from among all of these. The study is continued in the following.)

31. M.O. Mennessier, "A classification of miras from their visual and near-infrared light curves: an attempt to correlate them with their evolution", *Astronomy and Astrophysics*, **144**, 463-470, 1985.

32. MIDAS (Munich Image Data Analysis System), European Southern Observatory, Garching-bei-München (Version 4.1, January 1986). Chapter 13: Multivariate Statistical Methods (F. Murtagh).

    (This premier astronomical data reduction package contains a large number of multivariate algorithms.)

33. M. Moles, A. del Olmo and J. Perea, "Taxonomical analysis of superclusters. I. The Hercules and Perseus superclusters", *Monthly Notices of the Royal Astronomical Society*, **213**, 365-380, 1985.

    (A non–hierarchical descending method, used previously by Paturel, is employed.)

34. F. Murtagh, "Clustering techniques and their applications", *Data Analysis and Astronomy* (Proceedings of International Workshop on Data Analysis and Astronomy, Erice, Italy, April 1986) Plenum Press, New York (1986, forthcoming).

35. F. Murtagh and A. Lauberts, "A curve matching problem in astronomy", (forthcoming), 1986.

    (A dissimilarity is defined between galaxy luminosity profiles, in order to arrive at a spiral–elliptical grouping.)

36. G. Paturel, "Etude de la région de l'amas Virgo par taxonomie", *Astronomy and Astrophysics*, **71**, 106-114, 1979.

    (A descending non–hierarchical method is used.)

37. D.J. Tholen, "Asteroid taxonomy from cluster analysis of photometry", PhD Thesis, University of Arizona, 1984.

    (Between 400 and 600 asteroids using good–quality multi–colour photometric data are analysed.)

38. F. Giovannelli, A. Coradini, J.P. Lasota and M.L. Polimene, "Classification of cosmic sources: a statistical approach", *Astronomy and Astrophysics*, **95**, 138-142, 1981.

39. B. Pirenne, D. Ponz and H. Dekker, "Automatic analysis of interferograms", *The Messenger*, No. 42, 2-3, 1985.

    (The minimal spanning tree is used to distinguish fringes; there is little description of the MST approach in the above article, but further articles are in preparation and the software — and accompanying documentation — are available in the European Southern Observatory's MIDAS image processing system.)

40. A. Zandonella, " Object classification: some methods of interest in astronomical image analysis", in *Image Processing in Astronomy*, eds. G. Sedmak, N. Capaccioli and R.J. Allen, Osservatorio Astronomico di Trieste, Trieste, 304-318, 1979.

(This presents a survey of clustering methods.)

3 Cluster Analysis: General

41. M.R. Anderberg, *Cluster Analysis for Applications*, Academic Press, New York, 1973.

(A little dated, but still very much referenced; good especially for similarities and dissimilarities.)

42. J.P. Benzécri et coll., *L'Analyse des Données. I. La Taxinomie*, Dunod, Paris, 1979 (3rd ed.).

(Very influential in the French speaking world; extensive treatment, and impressive formalism.)

43. R.K. Blashfield and M.S. Aldenderfer, "The literature on cluster analysis", *Multivariate Behavioral Research*, **13**, 271-295, 1978.

44. H.H. Bock, *Automatische Klassifikation*, Vandenhoek und Rupprecht, Göttingen, 1974.

(Encyclopaedic.)

45. CLUSTAN, Clustan Ltd., 16 Kingsburgh Road, Edinburgh EH12 6DZ, Scotland.

(One of the few exclusively clustering packages available.)

46. B. Everitt, *Cluster Analysis*, Heinemann Educational Books, London, 1980 (2nd ed.).

(A very readable, introductory text.)

47. A.D. Gordon, *Classification*, Chapman and Hall, London, 1981.

(Another recommendable introductory text.)

48. R.L. Graham and P. Hell, "On the history of the minimum spanning tree problem", *Annals of the History of Computing*, **7**, 43-57, 1985.

(An interesting historical study.)

49. J.A. Hartigan, *Clustering Algorithms*, Wiley, New York, 1975.

(Often referenced, this book could still be said to be innovative in its treatment of clustering problems; it contains a wealth of sample data sets.)

50. M. Jambu and M.O. Lebeaux, *Cluster Analysis and Data Analysis*, North-Holland, Amsterdam, 1983.

(Some of the algorithms discussed have been overtaken by, for instance, the "nearest neighbour chain" or "reciprocal nearest neighbour" algorithms. These latter are described in the reference of Murtagh, below.)

51. L. Lebart, A. Morineau and K.M. Warwick, *Multivariate Descriptive Statistical Analysis*, Wiley, New York, 1984.

(A useful book, centred on Multiple Correspondence Analysis, but also including clustering, Principal Components Analysis, and other methods.)

52. R.C.T. Lee, "Clustering analysis and its applications", in J.T. Tou (ed.) *Advances in Information Systems Science, Vol. 8*, Plenum Press, New York, 1981, pp. 169-292.

(Practically book–length, this is especially useful for the links between clustering and problems in computing and in Operations Research.)

53. F. Murtagh, *Multidimensional Clustering Algorithms*, COMPSTAT Lectures Volume 4, Physica-Verlag, Wien, 1985.

(Algorithmic details of a range of widely–used clustering methods.)

54. F.J. Rohlf, "Generalization of the gap test for the detection of multivariate outliers", *Biometrics*, **31**, 93-101, 1975.

(One application of the minimal spanning tree.)

55. G. Salton and M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, New York, 1983.

(A central reference in the information retrieval area.)

56. P.H.A. Sneath and R.R. Sokal, *Numerical Taxonomy*, Freeman, San Francisco, 1973.

(Very influential for biological applications, it also has some impressive collections of graph representations of clustering results.)

57. H. Späth, *Cluster Dissection and Analysis: Theory, Fortran Programs, Examples*, Ellis Horwood, Chichester, 1985.

(Recommendable reference for non–hierarchic, partitioning methods.)

58. A. Tucker, *Applied Combinatorics*, Wiley, New York, 1980.

(For background reading on graph theory and combinatorics.)

59. D. Wishart, "Mode analysis: a generalization of nearest neighbour which reduces chaining effects", in ed. A.J. Cole, *Numerical Taxonomy*, Academic Press, New York, 282-311, 1969.

(Discusses various variance–based clustering criteria which, interestingly, are justified by the difficulties experienced by more mainstream algorithms in clustering data of the type found in the H–R diagram.)

60. C.T. Zahn, "Graph-theoretical methods for detecting and describing Gestalt clusters", *IEEE Transactions on Computers*, **C-20**, 68-86, 1971.

(Central reference for the use of the minimal spanning tree for processing point patterns.)

4 Discriminant Analysis: Astronomy

61. H.-M. Adorf, "Classification of low-resolution stellar spectra via template matching - a simulation study", *Data Analysis and Astronomy*, (Proceedings of International Workshop on Data Analysis and Astronomy, Erice, Italy, April 1986) Plenum Press, New York (1986, forthcoming).

62. E. Antonello and G. Raffaelli, "An application of discriminant analysis to variable and nonvariable stars", *Publications of the Astronomical Society of the Pacific*, **95**, 82-85, 1983.

(Multiple Discriminant Analysis is used.)

63. A. Heck, "An application of multivariate statistical analysis to a photometric catalogue", *Astronomy and Astrophysics*, **47**, 129-135, 1976.

(Multiple Discriminant Analysis and a stepwise procedure are applied.)

64. M.J. Kurtz, "Progress in automation techniques for MK classification", in ed. R.F. Garrison, *The MK Process and Stellar Classification*, David Dunlop Observatory, University of Toronto, 1984, pp. 136-152.

(Essentially a k-NN approach is used for assigning spectra to known stellar spectra classes.)

65. J.F. Jarvis and J.A. Tyson, "FOCAS - Faint object classification and analysis system", *SPIE Instrumentation in Astronomy III*, **172**, 1979, 422-428.

(See also other references of Tyson/Jarvis and Jarvis/Tyson.)

66. J.F. Jarvis and J.A. Tyson, "Performance verification of an automated image cataloging system", *SPIE Vol. 264 Applications of Digital Image Processing to Astronomy*, 222-229, 1980.

67. J.F. Jarvis and J.A. Tyson, "FOCAS - Faint object classification and analysis system", *The Astronomical Journal*, **86**, 1981, 476-495.

(A hyperplane separation surface is determined in a space defined by 6 parameters used to characterise the objects. This is a 2-stage procedure where the first stage is that of training, and the second stage uses a partitioning clustering method.)

68. H.T. MacGillivray, R. Martin, N.M. Pratt, V.C. Reddish, H. Seddon, L.W.G. Alexander, G.S. Walker, P.R. Williams, "A method for the automatic separation of the images of galaxies and stars from measurements made with the COSMOS machine", *Monthly Notices of the Royal Astronomical Society*, **176**, 265-274, 1976.

(Different parameters are appraised for star/galaxy separation. Kurz — see reference above under Cluster Analysis — lists other parameters which have been used for the same objective.)

69. M.L. Malagnini, "A classification algorithm for star-galaxy counts", in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 1983, pp. 69-72.

    (A linear classifier is used and is further employed in the following reference.)

70. M.L. Malagnini, F. Pasian, M. Pucillo and P. Santin, "FODS: a system for faint object detection and classification in astronomy", *Astronomy and Astrophysics*, **144**, 1985, 49-56.

71. "Recommendations for Guide Star Selection System", private notes, GSSS Group, Space Telescope Science Institute, Baltimore, 1984.

    (A Bayesian approach, using the IMSL subroutine library — see below — is employed in the GSSS system. Documentation will follow on this, in the future.)

72. W.J. Sebok, "Optimal classification of images into stars or galaxies — a Bayesian approach", *The Astronomical Journal*, **84**, 1979, 1526-1536.

    (The design of a classifier, using galaxy models, is studied in depth and validated on Schmidt plate data.)

73. J.A. Tyson and J.F. Jarvis, "Evolution of galaxies: automated faint object counts to 24th magnitude", *The Astrophyiscal Journal*, **230**, 1979, L153-L156.

    (A continuation of the work of Jarvis and Tyson, 1979, above.)

74. F. Valdes, "Resolution classifier", *SPIE Instrumentation in Astronomy IV*, **331**, 1982, 465-471.

    (A Bayesian classifier is used, which differs from that used by Sebok, referenced above. The choice is thoroughly justified. A comparison is also made with the hyperplane fitting method used in the FOCAS system – see the references of Jarvis and Tyson. It is concluded that the results obtained within the model chosen are better than a hyperplane

based approach in parameter space; but that the latter is computationally more efficient.)

5 Discriminant Analysis: General

75. S.-T. Bow, *Pattern Recognition*, Marcel Dekker, New York, 1984.

(A textbook detailling a range of Discriminant Analysis methods, together with clustering and other topics.)

76. C. Chatfield and A.J. Collins, *Introduction to Multivariate Analysis*, Chapman and Hall, London, 1980.

(An excellent introductory textbook.)

77. E. Diday, J. Lemaire, J. Pouget and F. Testu, *Eléments d'Analyse de Données*, Dunod, Paris, 1982.

(Describes a large range of methods.)

78. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

(Excellent treatment of many image processing problems.)

79. R.A. Fisher, "The use of multiple measurements in taxonomic problems", *The Annals of Eugenics*, **7**, 179-188, 1936.

(Still an often referenced paper; contains the famous Iris data.)

80. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1972.

81. D.J. Hand, *Discrimination and Classification*, Wiley, New York, 1981.

(A comprehensive description of a wide range of methods; very recommendable.)

82. International Mathematical and Statistical Library (IMSL), Manual sections on ODFISH, ODNORM.

(A useful range of algorithms is available in this widely used subroutine library.)

83. M. James, *Classification Algorithms*, Collins, London, 1985.

(A very readable introduction.)

84. M.G. Kendall, *Multivariate Analysis*, Griffin, London, 1980 (2nd ed.).

(Dated in relation to computing techniques, but exceptionally clear and concise in its treatment of many practical problems.)

85. P.A. Lachenbruch, *Discriminant Analysis*, Hafner Press, New York, 1975.

86. J.L. Melsa and D.L. Cohn, *Decision and Estimation Theory*, McGraw–Hill, New York, 1978.

(A readable decision theoretic perspective.)

87. J.M. Romeder, *Méthodes et Programmes d'Analyse Discriminante*, Dunod, Paris, 1973.

(A survey of commonly–used techniques.)

88. Statistical Analysis System (SAS), SAS Institute Inc., Box 8000, Cary, NC 27511-8000, USA; Manual chapters on STEPDISC, NEIGHBOUR, etc.

(A range of relevant algorithms is available in this, — one of the premier statistical packages.)

6 Principal Components Analysis: Astronomy

PCA has been a fairly widely used technique in astronomy. The following list does not aim to be comprehensive, but indicates instead the types of problems to which PCA can be applied. It is also hoped that it may provide a convenient entry–point to literature on a topic of interest. References below are concerned with stellar parallaxes; a large number are concerned with the study of galaxies; and a large number relate also to spectral reduction.

89. A. Bijaoui, "Application astronomique de la compression de l'information", *Astronomy and Astrophysics*, **30**, 199-202, 1974.

90. A. Bijaoui, SAI Library, Algroithms for Image Processing, Nice Observatory, Nice, 1985.

    (A large range of subroutines for image processing, including the Karhunen–Loève expansion.)

91. P. Brosche, "The manifold of galaxies: Galaxies with known dynamical properties", *Astronomy and Astrophysics*, **23**, 259-268, 1973.

92. P. Brosche and F.T. Lentes, "The manifold of globular clusters", *Astronomy and Astrophysics*, **139**, 474-476, 1984.

93. V. Bujarrabal, J. Guibert and C. Balkowski, "Multidimensional statistical analysis of normal galaxies",*Astronomy and Astrophysics*, **104**, 1-9, 1981.

94. R. Buser, "A systematic investigation of multicolor photometric systems. I. The UBV, RGU and *uvby* systems.", *Astronomy and Astrophysics*, **62**, 411-424, 1978.

95. C.A. Christian and K.A. Janes, "Multivariate analysis of spectrophotometry". *Publications of the Astronomical Society of the Pacific*, **89**, 415-423, 1977.

96. C.A. Christian, "Identification of field stars contaminating the colour–magnitude diagram of the open cluster Be 21", *The Astrophysical Journal Supplement Series*, **49**, 555-592, 1982.

97. T.J. Deeming, "Stellar spectral classification. I. Application of component analysis", *Monthly Notices of the Royal Astronomical Society*, **127**, 493-516, 1964.

    (An often referenced work.)

98. T.J. Deeming, "The analysis of linear correlation in astronomy", *Vistas in Astronomy*, **10**, 125-, 1968.

    (For regression also.)

99. G. Efstathiou and S.M. Fall, "Multivariate analysis of elliptical galaxies", *Monthly Notices of the Royal Astronomical Society*, **206**, 453-464, 1984.

100. S.M. Faber, "Variations in spectral–energy distributions and absorption–line strengths among elliptical galaxies", *The Astrophysical Journal*, **179**, 731-754, 1973.

101. M. Fofi, C. Maceroni, M. Maravalle and P. Paolicchi, "Statistics of binary stars. I. Multivariate analysis of spectroscopic binaries", *Astronomy and Astrophysics*, **124**, 313-321, 1983.

     (PCA is used, together with a non-hierarchical clustering technique.)

102. M. Fracassini, L.E. Pasinetti, E. Antonello and G. Raffaelli, "Multivariate analysis of some ultrashort period Cepheids (USPC)", *Astronomy and Astrophysics*, **99**, 397-399, 1981.

103. M. Fracassini, G. Manzotti, L.E. Pasinetti, G. Raffaelli, E. Antonello and L. Pastori, "Application of multivariate analysis to the parameters of astrophysical objects", in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 21-25, 1983.

104. P. Galeotti, "A statistical analysis of metallicity in spiral galaxies", *Astrophysics and Space Science*, **75**, 511-519, 1981.

105. A. Heck, "An application of multivariate statistical analysis to a photometric catalogue", *Astronomy and Astrophysics*, **47**, 129-135, 1976.

     (PCA is used, along with regression and discriminant analysis.)

106. A. Heck, D. Egret, Ph. Nobelis and J.C. Turlot, "Statistical confirmation of the UV spectral classification system based on IUE low–dispersion spectra", *Astrophysics and Space Science*, **120**, 223-237, 1986.

     (Many other articles by these authors, which also make use of PCA, are referenced in the above.)

107. S.J. Kerridge and A.R. Upgren, "The application of multivariate analysis to parallax solutions. II. Magnitudes and colours of comparison stars", *The Astronomical Journal*, **78**, 632-638, 1973.

     (See also Upgren and Kerridge, 1971, referenced below.)

108. J. Koorneef, "On the anomaly of the far UV extinction in the 30 Doradus region", *Astronomy and Astrophysics*, **64**, 179-193, 1978.

(PCA is used for deriving a photometric index from 5-channel photometric data.)

109. M.J. Kurtz, "Automatic spectral classification", PhD Thesis, Dartmouth College, New Hampshire, 1982.

110. F.T. Lentes, "The manifold of spheroidal galaxies", *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 73-76, 1983.

111. D. Massa and C.F. Lillie, "Vector space methods of photometric analysis: applications to O stars and interstellar reddening", *The Astrophysical Journal*, **221**, 833-850, 1978.

112. D. Massa, "Vector space methods of photometric analysis. III. The two components of ultraviolet reddening", *The Astronomical Journal*, **85**, 1651-1662, 1980.

113. B. Nicolet, "Geneva photometric boxes. I. A topological approach of photometry and tests.", *Astronomy and Astrophysics*, **97**, 85-93, 1981.

(PCA is used on colour indices.)

114. S. Okamura, K. Kodaira and M. Watanabe, "Digital surface photometry of galaxies toward a quantitative classification. III. A mean concentration index as a parameter representing the luminosity distribution", *The Astrophysical Journal*, **280**, 7-14, 1984.

115. S. Okamura, "Global structure of Virgo cluster galaxies", in O.-G. Richter and B. Binggeli (eds.), Proceedings of ESO Workshop on The Virgo Cluster of Galaxies, ESO Conference and Workshop Proceedings No. 20, 201-215, 1985.

116. D. Pelat, "A study of H I absorption using Karhunen–Loève series", *Astronomy and Astrophysics*, **40**, 285-290, 1975.

117. A. W. Strong, "Data analysis in gamma-ray astronomy: multivariate likelihood method for correlation studies", *Astronomy and Astrophysics*, **150**, 273-275, 1985.

(The method presented is not linked to PCA, but in dealing with the eigenreduction of a correlation matrix it is clearly very closely related.)

118. B. Takase, K. Kodaira and S. Okamura, *An Atlas of Selected Galaxies*, University of Tokyo Press, VNU Science Press, 1984.

119. D.J. Tholen, "Asteroid taxonomy from cluster analysis of photometry", PhD Thesis, University of Arizona, 1984.

120. A.R. Upgren and S.J. Kerridge, "The application of multivariate analysis to parallax solutions. I. Choice of reference frames", *The Astronomical Journal*, **76**, 655-664, 1971.

(See also Kerridge and Upgren, 1973, referenced above.)

121. J.P. Vader, "Multivariate analysis of elliptical galaxies in different environments", *The Astrophysical Journal*, **306**, 390-400, 1986.

(The Virgo and Coma clusters are studied.)

122. C.A. Whitney, "Principal components analysis of spectral data. I. Methodology for spectral classification", *Astronomy and Astrophysics Supplement Series*, **51**, 443-461, 1983.

123. B.C. Whitmore, "An objective classification system for spiral galaxies. I. The two dominant dimensions", *The Astrophysical Journal*, **278**, 61-80, 1984.

7 Principal Components Analysis: General

124. T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, New York, 1984 (2nd ed.).

(For inferential aspects relating to PCA.)

125. C. Chatfield and A.J. Collins, *Introduction to Multivariate Analysis*, Chapman and Hall, London, 1980.

(An excellent introductory textbook.)

126. R. Gnanadesikan, *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York, 1977.

(For details of PCA, clustering and discrimination.)

127. M. Kendall, *Multivariate Analysis*, Griffin, London, 1980 (2nd ed.).

(Dated in relation to computing techniques, but exceptionally clear and concise in its treatment of many practical problems.)

128. L. Lebart, A. Morineau and K.M. Warwick, *Multivariate Descriptive Statistical Analysis*, Wiley, New York, 1984.

(An excellent geometric treatment of PCA.)

129. F.H.C. Marriott, *The Interpretation of Multiple Observations*, Academic Press, New York, 1974.

(A short, readable textbook.)

8 Regression: Astronomy

Regression analysis, and fitting problems, have always been central in the physical sciences. The following selection of references in this area will therefore simply indicate the range of possible applications, and in some cases will additionally illustrate where regression and fitting might profitably complement other multivariate statistical techniques.

130. R.L. Branham Jr., "Alternatives to least-squares", *The Astronomical Journal*, **87**, 928-937, 1982.

131. R. Buser, "A systematic investigation of multicolor photometric systems. II. The transformations between the UBV and RGU systems.", *Astronomy and Astrophysics*, **62**, 425-430, 1978.

132. C.R. Cowley and G.C.L. Aikman, "Stellar abundances from line statistics", *The Astrophysical Journal*, **242**, 684-698, 1980.

133. M. Crézé, "Influence of the accuracy of stellar distances on the estimations of kinematical parameters from radial velocities", *Astronomy and Astrophysics*, **9**, 405-409, 1970.

134. M. Crézé, "Estimation of the parameters of galactic rotation and solar motion with respect to Population I Cepheids", *Astronomy and Astrophysics*, **9**, 410-419, 1970.

135. T.J. Deeming, "The analysis of linear correlation in astronomy", *Vistas in Astronomy*, **10**, 125, 1968.

136. H. Eichhorn, "Least-squares adjustment with probabilistic constraints", *Monthly Notices of the Royal Astronomical Society*, **182**, 355-360, 1978.

137. H. Eichhorn and M. Standish, Jr., "Remarks on nonstandard least-squares problems", *The Astronomical Journal*, **86**, 156-159, 1981.

138. J.R. Kuhn, "Recovering spectral information from unevenly sampled data: two machine-efficient solutions", *The Astronomical Journal*, **87**, 196-202, 1982.

139. J.R. Gott III and E.L. Turner, "An extension of the galaxy covariance function to small scales", *The Astrophysical Journal*, **232**, L79-L81, 1979.

140. A. Heck, "Predictions: also an astronomical tool", in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 1983, pp. 135-143.

    (A survey article, with many references. Other articles in this conference proceedings also use regression and fitting techniques.)

141. A. Heck and G. Mersch, "Prediction of spectral classification from photometric observations — application to the $uvby\beta$ photometry and the MK spectral classification. I. Prediction assuming a luminosity class", *Astronomy and Astrophysics*, **83**, 287-296, 1980.

    (Stepwise multiple regression and isotonic regression are used.)

142. W.H. Jefferys, "On the method of least squares", *The Astronomical Journal*, **85**, 177-181, 1980.

143. W.H. Jefferys, "On the method of least squares. II.", *The Astronomical Journal*, **86**, 149-155, 1981.

144. M.O. Mennessier, "Corrections de précession, apex et rotation galactique estimés à partir de mouvements propres fondamentaux par une méthode de maximum vraisemblance", *Astronomy and Astrophysics*, **17**, 220-225, 1972.

145. M.O. Mennessier, "On statistical estimates from proper motions. III.", *Astronomy and Astrophysics*, **11**, 111-122, 1972.

146. G. Mersch and A. Heck, "Prediction of spectral classification from photometric observations — application to the $uvby\beta$ photometry and the MK spectral classification. II. General case", *Astronomy and Astrophysics*, **85**, 93-100, 1980.

147. J.F. Nicoll and I.E. Segal, "Correction of a criticism of the phenimenological quadratic redshift-distance law", *The Astrophysical Journal*, **258**, 457-466, 1982.

148. J.F. Nicoll and I.E. Segal, "Null influence of possible local extragalactic perturbations on tests of redshift-distance laws", *Astronomy and Astrophysics*, **115**, 398-403, 1982.

149. D.M. Peterson, "Methods in data reduction. I. Another look at least squares", *Publications of the Astronomical Society of the Pacific*, **91**, 546-552, 1979.

150. I.E. Segal, "Distance and model dependence of observational galaxy cluster concepts", *Astronomy and Astrophysics*, **123**, 151-158, 1983.

151. I.E. Segal and J.F. Nicoll, "Uniformity of quasars in the chronometric cosmology", *Astronomy and Astrophysics*, **144**, L23-L26, 1985.

9 Regression: General

152. P.R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill, New York, 1969.

(A very recommendable text for regression and fitting, with many examples.)

153. N.R. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, 1981 (2nd ed.).

154. B.S. Everitt and G. Dunn, *Advanced Methods of Data Exploration and Modelling*, Heinemann Educational Books, London, 1983.

(A discursive overview of topics such as linear models and analysis of variance; PCA and clustering are also covered.)

155. D.C. Montgomery and E.A. Peek, *Introduction to Linear Regression Analysis*, Wiley, New York, 1982.

156. G.A.F. Seber, *Linear Regression Analysis*, Wiley, New York, 1977.

157. G.B. Wetherill, *Elementary Statistical Methods*, Chapman and Hall, London, 1967.

(An elementary introduction, with many examples.)

10 Other Statistical Methods: Astronomy

We have not sought to focus on the application of statistics, *tout court*, in astronomy in this bibliography. However some of the varied studies, listed below, constitute valuable background or survey material.

158. D. Clarke and B.G. Steward, "Statistical methods of stellar photometry", *Vistas in Astronomy*, **29**, 27-51, 1986.

159. H. Eelsalu, *Theoretical Foundations of Stellar Statistics*, Academy of Sciences of the Estonian S.S.R., 1982.

(A monograph on a general theory of stellar statistical data.)

160. E.D. Feigelson and P.I. Nelson, “Statistical methods for astronomical data with upper limits. I. Univariate distributions”, *The Astrophysical Journal*, **293**, 192-206, 1985.

(Survival analysis is used for left-censored data. See also Isobe et al. below.)

161. A. Heck, J. Manfroid and G. Mersch, “On period determination methods”, *Astronomy and Astrophysics Supplement Series*, **59**, 63-72, 1985.

162. Isobe, T., E.D. Feigelson and P.I. Nelson, “Statistical methods for astronomical data with upper limits. II. Correlation and regression”, *The Astrophysical Journal*, 1986 (in press).

(Survival analysis is used on data with upper limits.)

163. D.G. Kendall, “Mathematical statistics in the humanities, and some related problems in astronomy”, in A.C. Atkinson and S.E. Fienberg (eds.), *A Celebration of Statistics*, Springer-Verlag, New York, 1985, pp. 393-408.

(Problems relating to testing for one-dimensionality and for alignments — of importance in quasar astronomy — are overviewed, and some other relevant references are to be found in this paper.)

164. J.V. Narlikar, “Statistical techniques in astronomy”, *Sankha: The Indian Journal of Statistics, Series B, Part 2*, **44**, 125-134, 1982.

(A range of astronomical problems with statistical solutions are presented.)

165. M.E. Özel and H. Mayer-Haßelwander, “Application of bootstrap sampling in γ-ray astronomy: time variability in pulsed emmission from Crab pulsar”, in V. Di Gesù, L. Scarsi, P. Crane, J.H. Friedman and S. Levialdi (eds.), *Data Analysis in Astronomy*, Plenum Press, New York, 1985, pp. 81-86.

166. J. Pelt, “Phase dispersion minimization methods for estimation of periods from unequally spaced sequences of data” in *Statistical Methods in Astronomy*, European Space Agency Special Publication SP-201, 37-42, 1983.

167. J. Pfleiderer and P. Krommidas, "Statistics under incomplete knowledge of data", *Monthly Notices of the Royal Astronomical Society*, **198**, 281-288, 1982.

168. J.D. Scargle, "Studies in astronomical time series analysis. I. Modelling random processes in the time domain", *The Astrophysical Journal Supplement Series*, **45**, 1-71, 1981.

169. J.V. Wall, "Practical statistics for astronomers. I. Definitions, the normal distribution, detection of signal", *Quarterly Journal of the Royal Astronomical Society*, **20**, 130-152, 1972.