Multiple imputation and maximum likelihood principal component analysis of incomplete multivariate data from a study of the ageing of port

Keywords: Missing values; Principal components analysis; Multiple imputation; Maximum likelihood principal components analysis; Port; Ageing; Colour; Phenolic composition

P. Ho *, M.C.M. Silva, T.A. Hogg

Universidade Católica Portuguesa, Escola Superior de Biotecnologia, Rua Dr. António Bernardino de Almeida, Porto 4200-072, Portugal

Abstract

A multivariate data matrix containing a number of missing values was obtained from a study on the changes in colour and phenolic composition during the ageing of port. Two approaches were taken in the analysis of the data. The first involved the use of multiple imputation (MI) followed by principal components analysis (PCA). The second examined the use of maximum likelihood principal component analysis (MLPCA). The use of multiple imputation allows for missing value uncertainty to be incorporated into the analysis of the data. Initial estimates of missing values were firstly calculated using the Expectation Maximization algorithm (EM), followed by Data Augmentation (DA) in order to generate five imputed data matrices. Each complete data matrix was subsequently analysed by PCA, then averaging their principal component (PC) scores and loadings to give an estimation of errors. The first three PCs accounted for 93.3% of the explained variance. Changes to colour and monomeric anthocyanin composition were explained on PC1 (79.63% explained variance), phenolic composition and hue mainly on PC2 (8.61% explained variance) and phenolic composition and the formation of polymeric pigment on PC3 (5.04% explained variance). In MLPCA estimates of measurement uncertainty is incorporated in the decomposition step, with missing values being assigned large measurement uncertainties. PC scores on the first two PCs after multiple imputation and PCA (MI + PCA) were comparable to maximum likelihood scores on the first two PCs extracted by MLPCA.

1. Introduction

Many methods are now available for the analysis of multivariate data in food research applied to areas of exploratory data analysis, classification, calibration and prediction [1–5]. However, incomplete data matrices may sometimes arise from experiments, as the result of insufficient sampling, errors in measurements or during data acquisition. Standard statistical packages have been designed to analyse only complete data and missing data procedures available are crude and unable to handle data sets with a high per-

* Corresponding author. Tel.: +351-22-5580043; fax: +351-22-5090351.

E-mail address: peter@esb.ucp.pt (P. Ho).

centage of incomplete cases [6]. Two main procedures for dealing with missing values, deletion and mean substitution, are commonly found in these programs. With listwise or casewise deletion, variables or cases are removed. However, this method of handling missing data is far from ideal as important information for subsequent analysis of the data might be lost. Pairwise deletion is used when calculating correlation or covariance matrices. In the case of correlations, the correlations between each pair of variables are calculated from all cases having complete data for those two variables. However, a systematic bias may result from a "hidden" systematic distribution of missing data, causing different correlation coefficients in the same correlation matrix [7]. In mean substitution, all missing values of a variable are replaced with the mean for that variable. However, this method is clearly not appropriate for data that is time dependent. These and other disadvantages of using deletion or mean substitution have been discussed recently [8,9]. Modern methods for handling missing data are based on maximum likelihood estimation and Monte Carlo Markov Chain methods [6,9–14]. One such method, known as multiple imputation, consists of generating $m > 1$ plausible missing values, thus producing m apparently complete data matrices which can then be analysed by the complete data methods. The results are then combined by simple rules [15] to produce overall estimates and standard errors that reflect missing data uncertainty [6].

In a study of the changes to colour and phenolics during the initial stages of ageing port [16], a multivariate data matrix containing a number of variables with missing values was obtained as certain measurements were not taken or not calculated due to measurement errors. The objective of that study was to examine the relationships among variables that are commonly used to examine colour and phenolics in wine [17–21]. Therefore, before analysing the data matrix by principal components analysis (PCA), a method had to be used to calculate the missing values. Multiple imputation was chosen as one of the methods for handling these missing values. We then compare this approach with the use of maximum likelihood principal component analysis (MLPCA) [22]. MLPCA is a decomposition method similar to PCA, and its ability in analysing incomplete data matrices has already been demonstrated [23].

2. Experimental

2.1. Data matrix

The incomplete data matrix was obtained from a study comparing changes to colour and phenolic composition during the ageing of port [16]. Ports were aged in wood, stainless steel and glass for a period of 11 months in a controlled temperature environment at 18°C. Spectrophotometric measurements (replicates, $n = 4$) were made on wine; samples taken at intervals of 0, 13, 34, 77, 110, 198, 255 and 311 days. The incomplete data matrix was composed of 26 colour and phenolic variables as follows: (1) 14 original colour and phenolic variables [17,18]: wine colour, WC; wine colour density, CD1; wine hue, HUE; polymeric pigment colour, PPC; wine colour in acid, WCA; anthocyanin colour, AC; non-coloured anthocyanin, NA; anthocyanin colour in acid, ACA; total monomeric anthocyanins, TAC; total phenolics, TP; chemical age at wine pH, CAW; chemical age in acid, CAA; degree of ionisation, α; natural degree of ionisation, α'. (2) Six modified colour variables [19]: wine colour density or Glories' Index, CD2; chemical age index (I), CA(I); chemical age index (II), CA(II); chemical age index (III), CA(III); degree of coloration, $A_{3.7}$; colour synergism factor, $S_{3.7}$. (3) Other variables [20,21,37]: total phenolics as gallic acid equivalents, TPGAE.; formation of brown pigments as the absorbance at 420 nm, BI, monomeric anthocyanins, TMA; polymeric anthocyanins, TPA; total anthocyanins, TAC2; total anthocyanins, TAC3.

2.2. Multiple imputation of missing values

Multiple imputation (MI) is based on three main assumptions: a probability model on complete data (observed and missing values), a prior distribution reflecting the uncertainty of the parameters for the imputation model and that the data is said to be *missing at random* (MAR) [6,9]. The program NORM was used for this study, which performs multiple imputation under a multivariate normal model [24]. NORM, together with other three other probability models for multiple imputation, are freely available as a set of S-PLUS libraries at the authors website (http://www.stat.psu.edu/~jls/misoftwa.html). Basically,

NORM works in two steps by using: (1) the EM algorithm for efficient estimation of parameters (mean, variances, covariances or correlations); (2) data augmentation for generating multiple imputations of missing values. As neither the EM algorithm nor the data augmentation have been described in chemometrics literature so far, a brief description will be presented here. Computational routines used in NORM have been described elsewhere [25].

The Expectation–Maximization (EM) algorithm is a general method for obtaining maximum likelihood estimates of parameters in problems with incomplete data. The EM algorithm was first described by Dempster et al. [10] in the late 1970s. A number of extension and variations of the EM algorithm have since been developed, improving the convergence of these EM-type algorithms [9,26–29]. In an incomplete data matrix, we will have both the observed data, Y_{obs}, missing data, Y_{mis}, and a vector of parameters, $\boldsymbol{\theta}$. Complete data, Y_{com}, can therefore be defined as $Y_{com} = (Y_{obs}, Y_{mis})$. With the complete data log-likelihood function, $L(\boldsymbol{\theta}) = f(Y_{com} \mid \boldsymbol{\theta})$ and the observed data log-likelihood function, $L(\boldsymbol{\theta}) = f(Y_{obs} \mid \boldsymbol{\theta})$, the expected complete data log-likelihood function can be defined as $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}') = E\{\ln[\, f(Y_{com} \mid \boldsymbol{\theta})] Y_{obs}, \boldsymbol{\theta}'\}$ [30]. The EM algorithm starts at some value of $\boldsymbol{\theta}$ and alternates between two steps [5,30]:

1. Expectation step (E-step): Computing $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$ as a function of $\boldsymbol{\theta}$;
2. Maximization step (M-step): Find $\boldsymbol{\theta}^{(t+1)}$ that maximizes $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)})$

The log-likelihood function $L(\boldsymbol{\theta})$ increases with each iteration of the EM algorithm until converging to a local or global maximum [10]. The rate of convergence is directly related to the amount of unobserved or missing information in a data matrix, i.e. slower convergence with greater amount of missing data [31]. In NORM, a ridge prior can be selected, which stabilize the estimation of parameters, to solve the problem of slow convergence of the EM algorithm. Selection of a ridge prior tends to shrink estimated correlations toward zero [25]. The degree of shrinkage is specified by the hyperparameter, a positive real number which corresponds roughly to the number of prior observations being introduced, i.e. the higher the value for the hyperparameter, the greater the shrinkage [24].

Data augmentation (DA) is an iterative process that alternately fills in the missing data and makes inferences about the unknown parameters, but unlike the EM algorithm, this is done in a stochastic or random fashion [6]. DA first performs a random imputation of missing data under assumed values of the parameters, and then draws new parameters from a Bayesian posterior distribution based on the observed and imputed data [6]. Starting at some value of $\boldsymbol{\theta}$, each iteration of the DA algorithm of Tanner and Wong [32] alternates between two steps [33]:

1. Imputation step (I-step): Draws $Y_{mis}^{(t+1)} \sim P(Y_{mis} \mid Y_{obs}, \boldsymbol{\theta}^{(t)})$
2. Posterior step (P-step): Draws $\boldsymbol{\theta}^{(t+1)} \sim P(\boldsymbol{\theta} \mid Y_{obs}, \boldsymbol{\theta}^{(t+1)})$

The procedure of alternately simulating missing data and parameters creates a Markov chain that eventually stabilizes or converges in distribution [33]. A more complete explanation of the principal behind multiple imputation, the EM and DA algorithms have been reviewed elsewhere [25].

Missing values were calculated by multiple imputation (MI) using the program NORM [24] according to Fig. 1. The original 26 variable data matrix (with replicates) was first divided into three data matrices according to the three ageing methods (wood, stainless steel, glass) (step I), in order to preserve any relationships or associations between variables. Data for all variables were then transformed to approximately normal before imputation using a logit transformation function and then transformed back to their original scale after imputation. The logit or logistic transformation is defined as:

$$\log it(x) = \log\left(\frac{x}{(1-x)}\right)$$

First approximations of missing values by multiple imputation (MI) were generated for every data matrix using the EM algorithm (step II). The convergence criterion, which was the maximum relative parameter change in the value of any parameter from one cycle to the next, was set at 0.0001. A hyperpa-

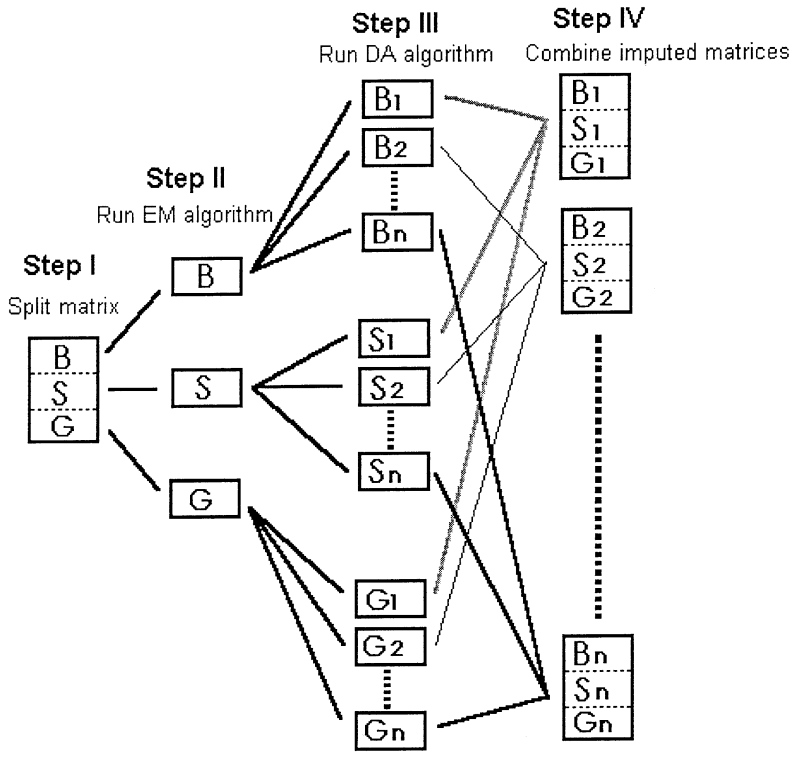

Fig. 1. Procedure for multiple imputation of missing values. Data were from wines aged in barrels (B), stainless steel (S), glass (G). *n* is the total number of data matrices.

rameter of 1 was also set for the ridge prior to improve the convergence of the EM algorithm. After obtaining these initial estimates, the DA algorithm was then used to generate five imputed data matrices for every data matrix, using a ridge prior with a hyperparameter of 1 (step III). Series plots for single parameters (means, variances and covariances) and for the worst linear function of the parameters were used to access the convergence of the DA algorithm. The 15 data matrices were then reorganised so that each of the final five imputed data matrices, containing a single replicate imputation of every missing value, were composed of samples from all three ageing methods (step IV).

2.3. Principal components analysis

Principal components analysis (PCA) by single value decomposition (SVD) was conducted using the statistical program R [34] on the five imputed data matrices (replicates were averaged before analysis) after standardisation. PC scores and loadings obtained from each imputed data matrix were averaged to give an estimation of the standard error of the mean (SEM), in order to account for missing value uncertainty in the principal components analysis.

2.4. Maximum likelihood principal component analysis

Maximum likelihood principal component analysis (MLPCA) was conducted using the MLPCA routine of Andrews and Wentzell [23], which was performed using Matlab (The Mathworks, Natick, MA). MLPCA is a more direct approach than using multiple imputation followed by PCA. MLPCA is similar to PCA as it performs a PCA-like decomposition of data. However, unlike PCA which assumes equal error variances for all measurements, the MLPCA algorithm incorporates variance information in the decomposition step [23]. This is done by using a matrix of standard deviations instead of variances, with

missing values being assigned a value of zero. A subroutine in the MLPCA algorithm then calculates the variances, where very large variances are assigned for missing values compared to observed data (in this study a value of 10^{10} was used). One of the major differences between PCA and MLPCA is that solutions for MLPCA are not nested, and the dimensionality of the subspace (the rank estimate, p) must be specified before running the algorithm. A number of different values of p principal components were examined to access the convergence of the MLPCA algorithm. The following parameters need to be set when running MLPCA: (1) a matrix of observations where each case were measurement means values, (2) a matrix of standard deviations associated with the observations, with a value of zero set for missing measurements; (3) the model dimensionality, ranging from $p = 1$ to 5, were used in this study. Maximum likelihood scores were obtained from maximum likelihood projections, which weights the direction of the projection in proportion to the magnitude of the measurement error variances [23]. The theoretical basis behind MLPCA has already been described [22] and has already been applied to the problem of incomplete data matrices and multivariate calibration [23,35].

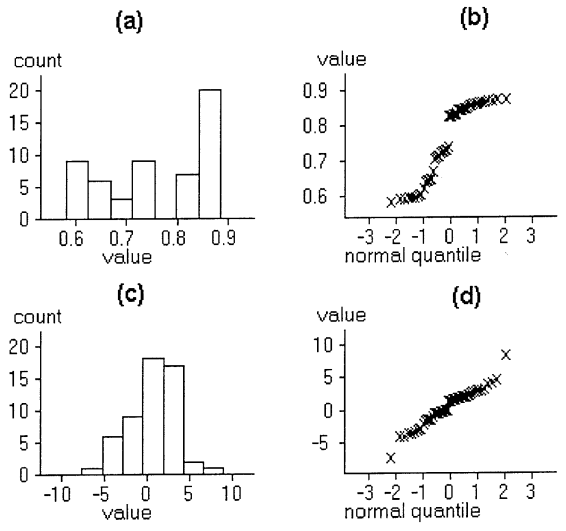

Fig. 2. Transformation of data using the logit transformation function for variable CA(I) (21.74% of missing data). Density histogram (a) and normal probability plot (b) before transformation and density histogram (c) and normal probability plot (d) after transformation.

3. Results and discussion

3.1. Multiple imputation

Before applying multiple imputation (MI) to calculate the missing values, density histograms and normal probability plots were drawn to examine if variables were normally distributed. A logit transformation function was then used to transform the data, as it was found to give the best approximation of variables to normality (Fig. 2). Setting the convergence criteria to 0.0001 and using a ridge prior of 1 resulted in the EM algorithm converging after 140, 79 and 147 iterations for barrel, stainless steel and glass data matrices, respectively. The EM estimates were then used as starting values for data augmentation (DA). The number of iterations needed for the convergence of the DA algorithm were calculated from the convergence behaviour of the EM algorithm for each data matrix, i.e. enough cycles between each imputed data matrix were used to ensure statistical independence. Rapid convergence of the DA algorithm, determined by examining series plots for individual variables and the worst linear function of the parameters (Fig. 3), was achieved for all data matrices. Missing values were then imputed for each data matrix and five complete data matrices were obtained.

3.2. Principal components analysis

Each complete data matrix after multiple imputation was analysed individually by PCA, producing individual matrices of principal component (PC) loadings and PC scores. A single matrix of mean estimates for loadings and one for scores were subsequently calculated, together with values of the standard errors of the mean (SEM). Typical SEM values for elements in both the mean PC loadings matrix and the mean PC scores matrix were less than 0.02. The lowest SEM values, many as low as 0.001, came from loadings and scores on the first principal component (PC1). The highest SEM values for loadings and scores were found with higher principal components (PCs) and were no higher than 0.1. As these higher

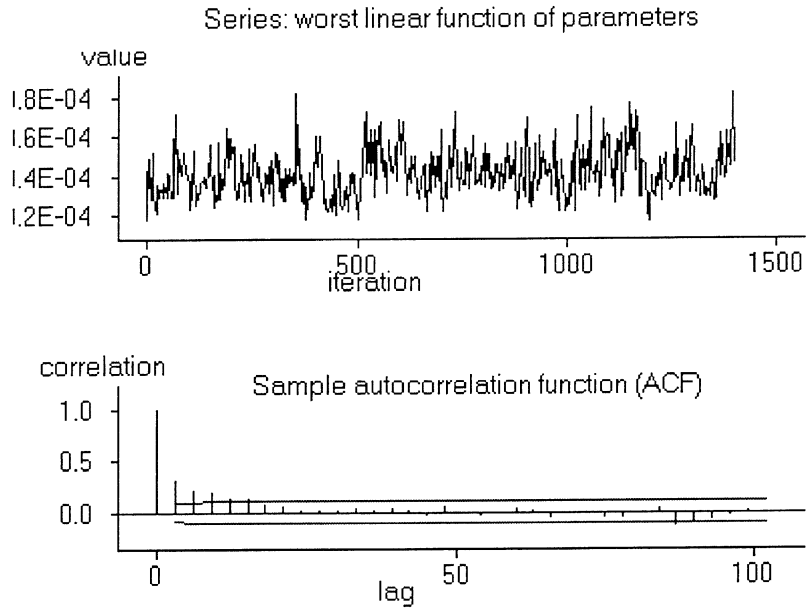

Fig. 3. Series plots for accessing convergence behaviour of the data augmentation algorithm.

PCs are not included in a PC-model, and SEM values for the first few PCs were very low, it was reasonable to conclude that imputed missing values did not seem to bias the principal components analysis.

The selection of the appropriate number of principal components for the 26 variable PC-model was based on the rule of eigenvalues greater than 1.0, the scree test and the proportion of variance explained [36]. The first three principal components selected accounted for 93% of the explained variance in the PC-model (Table 1). Correlations between a number of variables were very high, which could be seen by the loadings plot for the first two principal components (PC1 vs. PC2) and the first and third principal components (PC1 vs. PC3) (Figs. 4 and 5). For example, high correlations were found for colour density measurements (CD vs. CD2), total monomeric anthocyanins (TAC vs. TMA) and chemical age indices (CAW vs. CA(I)). These high correlations found among many variables suggest that a reduction in the number of variables before PCA of the data could be done. Selection of variables and a more detailed study of differences in the changes of colour and phenolics during the ageing of port under the conditions mentioned are the subject of another paper [16] and hence, only a brief discussion of the results will follow.

Table 1
Estimated average eigenvalues and the proportion of variance from imputed data matrices

PC	Eigenvalues	SEM[a]	% Variance	% Cumulative
1	20.5312	0.0535	79.63	79.63
2	2.2206	0.0118	8.61	88.24
3	1.3007	0.0245	5.04	93.29
4	0.9724	0.0129	3.77	97.06
5	0.4416	0.0130	1.71	98.77
6	0.1747	0.0031	0.68	99.45
7	0.1415	0.0068	0.55	100

[a] Standard error of the mean of eigenvalues, $n = 5$.

Two main groups of variables were found to have high loadings on the first principal component (PC1). The first group which described various colour mechanisms occurring in port, such as browning, anthocyanin equilibrium, co-pigmentation and the increasing importance of oligomeric and polymeric pigments with ageing, was loaded positively on PC1. Variables describing colour in terms of mainly monomeric anthocyanin concentration were all loaded negatively on PC1. These two groups of colour variables had almost no contribution on the second principal component (PC2), with only total phenolics and hue giving negative loadings of any

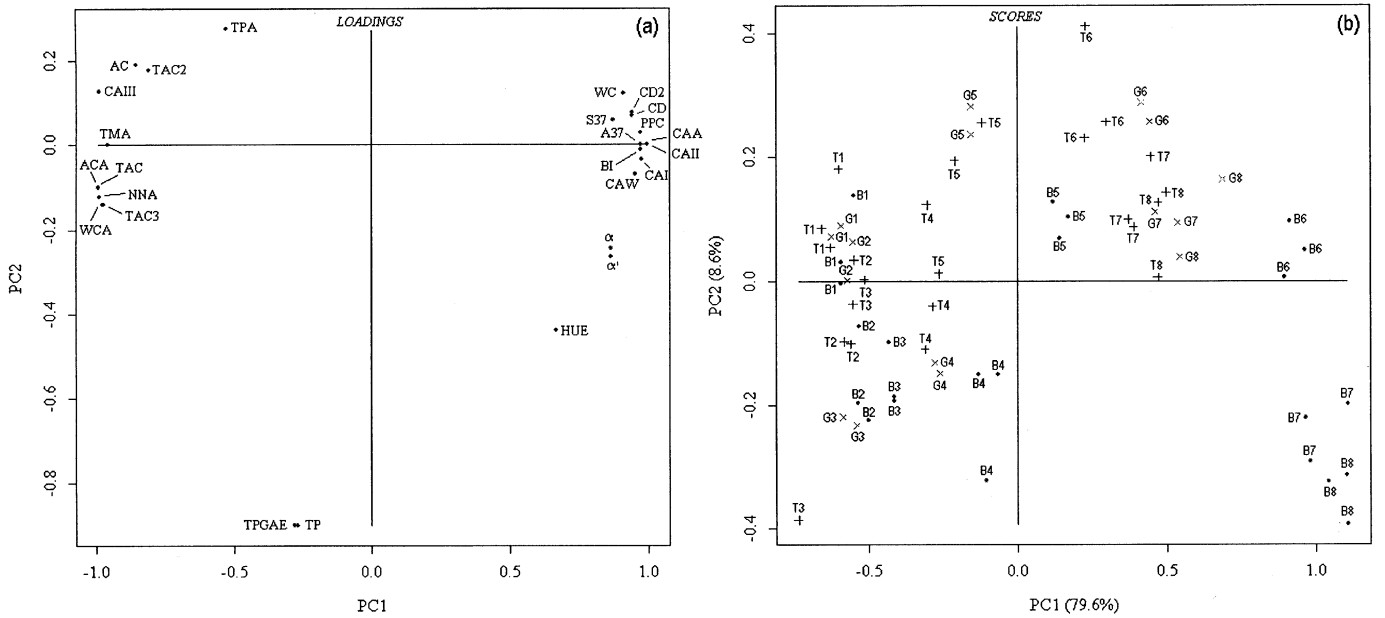

Fig. 4. Principal components analysis biplots (PC1 vs. PC2). (a) Loadings plot, (b) scores plot. Legends—B: aged in barrels; T: aged in stainless steel; G: aged in glass. Number indicate time of ageing (days), 1: initial time; 2: 13; 3: 34; 4: 77; 5: 110; 6: 198; 7: 255; 8: 311.

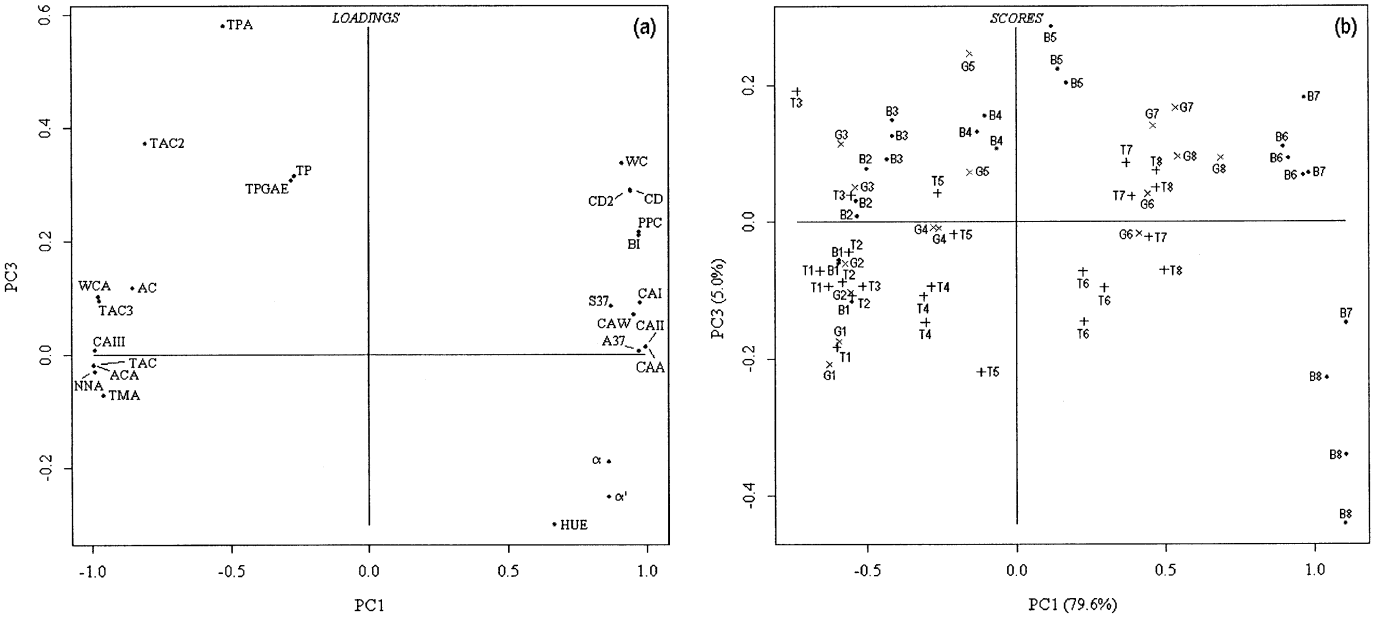

Fig. 5. Principal components analysis biplots (PC1 vs. PC3). (a) Loadings plot, (b) scores plot. Abbreviations as in Fig. 4.

significance. On the other hand, variables describing polymeric pigments (PPC and TPA) had positive loadings on the third principal component (PC3). These patterns suggest that three main principal components factors describe colour and phenolic compositional changes in ports ageing: chemical mechanisms and monomeric anthocyanins composition on PC1, changes mainly to phenolic composition and hue on PC2, and polymeric pigment formation on PC3.

3.3. Maximum likelihood principal component analysis

Recently, Andrews and Wentzell [23], used a new method known as maximum likelihood hood principal component analysis (MLPCA) in handling incomplete data matrices. They demonstrated that principal components extracted by MLPCA retains much of the original information with 10% of missing data and that MLPCA projections were comparable to PCA for uncensored data. The incomplete data matrix consists of 64 cases, after taking the means of replicate measurements, by 26 variables. The percentages of missing data in the data matrix ranged from 1% up to 14% for some of the variables. A number of different values of the rank estimate, p, were examined. The MLPCA algorithm converged in all cases from $p = 1$ to 5, but not for higher values. The reasons for the algorithm not converging is not known, but as only three principal components (PC) were sufficient to explain most of the variance in the principal components analysis (PCA) conducted previously, it was decided that results for $p = 5$ would only be examined. Therefore, five maximum likelihood principal components (MLPCs) were extracted from the data matrix, and from this maximum likelihood, scores were calculated. Fig. 6 shows a plot of maximum likelihood scores on the first two MLPCs from the MLPCA. The results are almost similar to those obtained using multiple imputation followed by PCA (MI + PCA). The MLPCA scores plot (Fig. 6) was in fact a mirror image of the PCA scores plot (Fig. 4b). The projection of data on the second MLPC

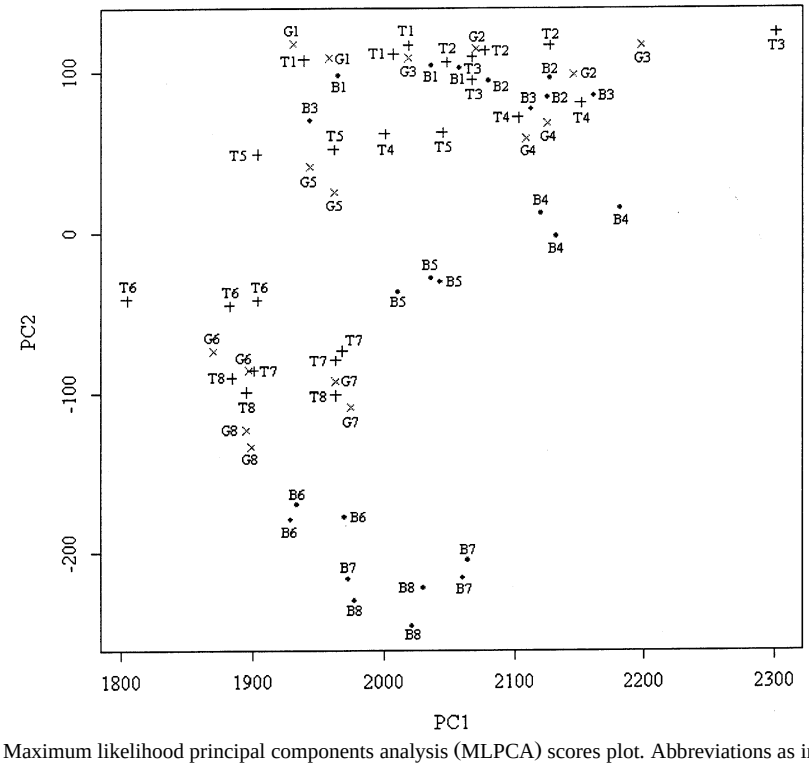

Fig. 6. Maximum likelihood principal components analysis (MLPCA) scores plot. Abbreviations as in Fig. 4.

for MPLCA, with increasing negative values, was similar to that of the first principal component (PC1) for PCA, with increasing positive values, which showed increasing storage time of ports. Therefore, results from both methods were comparable to one another. Ports aged in wood were well separated from ports aged in stainless steel and glass after 77 days of storage, forming identical groupings of samples of the same age.

4. Conclusions

Multiple imputation is an effective method in handling missing values from an incomplete data matrix. The generation of *m* number of complete data matrices can be analysed by any number of statistical methods available and standard errors can be combined, giving an assessment of missing value uncertainty. Multiple imputation of the incomplete data matrix from the study of ageing port, gave very low values of the standard errors of the mean (SEM) for scores and loadings on the first three principal components after a PCA, suggesting that missing values were well estimated. Maximum likelihood scores from the first two principal components from MLPCA were similar to those from PCA, giving comparable interpretation of the results. Multiple imputation should be the preferred method for handling missing values when the data matrix is to be analysed by more than one statistical method. However, if the objective is to conduct a PCA-like analysis of the data, MLPCA provides a more convenient and simpler approach.

Acknowledgements

The authors gratefully acknowledge Ramos-Pinto, Lda for providing the port and the wooden barrels. The author P. Ho would also like to thank J.L. Schafer and G. King for their valuable comments on how to handle missing data and P.D Wentzell on the use of the MLPCA algorithm. The author P. Ho was financed from a PRAXIS XXI grant (BD/13825/97) from the Fundação para a Ciência e a Technologia (FCT).

References

[1] J.R. Piggot (Ed.), Statistical Procedures in Food Research, Elsevier, Barking, 1986.
[2] A.V.A. Resurreccion, Food Technol. 42 (1988) 128–134.
[3] T. Aishima, T.S. Nakai, Food Rev. Int. 7 (1991) 33–101.
[4] E.V. Thomas, Anal. Chem. 66 (1994) 795A–804A.
[5] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge Univ. Press, Cambridge, 1996.
[6] J.L. Schafer, M.K. Olsen, Multivariate Behav. Res. 33 (1998) 545–571.
[7] Statsoft, Electronic Statistic Textbook, Statsoft, Tulsa, OK, 1997. http://www.statsoft.com/textbook/stathome.html.
[8] A.C. Acock, Fam. Sci. Rev. 10 (1997) 76–93.
[9] G. King, J. Honaker, A. Joseph, K. Scheve, Analyzing incomplete political science data: an alternative algorithm for multiple imputation, Proceedings of the 1998 Annual Meeting of the American Political Science Association, Boston, USA, 1999. http://gking.harvard.edu/files/evil.pdf, [August 19].
[10] A.P. Dempster, N.M. Laird, D.B. Rubin, J. R. Stat. Soc., Ser. B 39 (1977) 1–38.
[11] J.L. Arbuckle, Amos for Windows: Analysis of moment structures (Version 3.5), SmallWaters, Chicago, IL, 1995. http://www.smallwaters.com/amos/.
[12] J.L. Schafer, Some improved procedures for linear mixed models,http://www.stat.psu.edu/ ~ jls/improve.pdf, [June 28].
[13] C. Liu, J. Mult. Anal. 69 (1999) 206–217.
[14] M.C. Neale, Mx: Statistical Modeling, 4th edn., Department of Psychiatry, Box 710 MCV, Richmond, VA 23298, 1999. http://views.vcu.edu/mx/.
[15] D.B. Rubin, Multiple Imputation for Nonresponse in Surveys, Wiley, New York, 1987.
[16] P. Ho, M.C.M. Silva, T.A. Hogg, J. Sci. Food Agric., submitted for publication.
[17] T.C. Somers, M.E. Evans, J. Sci. Food Agric. 28 (1977) 279–287.
[18] M.G. Jackson, C.F. Timberlake, P. Bridle, L. Vallis, J. Sci. Food Agric. 29 (1978) 715–727.
[19] T.C. Somers, E. Vérette, Phenolic composition of natural wine types, in: H.F. Linskins, J.F. Jackson (Eds.), Modern Methods of Plant Analysis, Wine Analysis vol. 6, Springer-Verlag, New York, 1988, pp. 219–257, New Series.
[20] J. Bakker, P. Bridle, C.F. Timberlake, Vitis 25 (1986) 40–52.
[21] G.K. Niketić-Aleksić, G. Hrazdina, Lebensm.-Wiss. Technol. 5 (1972) 163–165.
[22] P.D. Wentzell, D.T. Andrews, D.C. Hamilton, K. Faber, B.R. Kowalski, J. Chemom. 11 (1997) 339–366.
[23] D.T. Andrews, P.D. Wentzell, Anal. Chim. Acta 350 (1997) 341–352.
[24] J.L. Schafer, NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2. Software for Windows 95/98/NT,http://www.stat.psu.edu/ ~ jls/misoftwa.html.
[25] J.L. Schafer, Analysis of Incomplete Multivariate Data, Chapman & Hall, London, 1997.

[26] X.L. Meng, D.B. Rubin, Biometrika 80 (1993) 267–278.
[27] C. Liu, D.B. Rubin, Biometrika 81 (1994) 633–648.
[28] X.L. Meng, D.A. van Dyk, J. R. Stat. Soc., Ser. B 59 (1997) 511–567.
[29] C. Liu, D.B. Rubin, Y. Wu, Biometrika 85 (1998) 755–770.
[30] J.C. Pinheiro, C. Liu, Y. Wu, J. Comput. Graphical Stat. (2000) submitted for publication.
[31] C. Fraley, Comput. Stat. Data Anal. 31 (1999) 13–26.
[32] M.A. Tanner, W.H. Wong, J. Am. Stat. Assoc. 82 (1987) 528–550.
[33] J.L. Schafer, D.B. Rubin, Multiple imputation for missing-data problems, Short course presented at Joint Statistical Meetings, Aug. 12, Dallas, TX, Cosponsored by the Survey Research Methods Section and the Biometrics Section, American Statistical Association, 1998. http://www.stat.psu.edu/ ~ jls/aug98.pdf.
[34] R. Ihaka, R. Gentleman, J. Comput. Graphical Stat. 5 (1996) 299–314.
[35] P.D. Wentzell, D.T. Andrews, Anal. Chem. 69 (1997) 2299–2311.
[36] J.R. Piggot, K. Sharman, Methods to aid interpretation of multidimensional data, in: J.R. Piggot (Ed.), Statistical Procedures in Food Research, Elsevier, Barking, 1986, pp. 181–232.
[37] J. Oszmianski, T. Ramos, M. Bourzeix, Am. J. Enol. Vitic. 39 (1988) 259–262.