ELSEVIER

International Journal of Pharmaceutics 234 (2002) 169–178

international journal of pharmaceutics

www.elsevier.com/locate/ijpharm

Principal component analysis of dissolution data with missing elements

E. Adams [a,*], B. Walczak [b,1], C. Vervaet [c], P.G. Risha [d], D.L. Massart [a]

[a] *Pharmaceutical Institute, Vrije Universiteit Brussel, Laarbeeklaan 103, 1090 Brussel, Belgium*
[b] *Institute of Chemistry, Silesian University, 9 Szkolna street, 40-006 Katowice, Poland*
[c] *Laboratory Pharmaceutical Technology, Ghent University, Harelbekestraat 72, 9000 Gent, Belgium*
[d] *College of Health Sciences, Muhimbili University, Dar es Salaam, Tanzania*

Received 7 September 2001; accepted 20 November 2001

Abstract

The use of principal component analysis (PCA) for incomplete dissolution data sets is examined. The PC space is constructed using a reference set and the test set is projected in that space. Several cases such as a reference set with missing data, an incomplete test set and both sets measured at different time points, are discussed using two examples: one simulation and one obtained from the pharmaceutical practice. From the many possibilities to deal with missing data, the expectation–maximization algorithm in combination with PCA was chosen. The influence on the similarity or f_2 factor is examined too. The sampling with replacement or bootstrap technique, which can be used to obtain confidence limits, can also be used when missing data are present in one of the data sets.

Keywords: Dissolution; Missing data; Expectation–maximization algorithm; Principal component analysis

1. Introduction

The dissolution test for tablets and capsules was introduced in the main pharmacopoeias like the USP and the Ph. Eur. to ensure that the active is released from the galenic form and becomes available for absorption in the gastrointestinal tract. In most cases, the dissolution characteristics of a test batch have to be compared with those of a reference one. A popular, simple way to evaluate the dissolution process is to check whether a minimum percentage of activity is dissolved in a predefined time. However, this single-point approach is only useful for highly soluble and rapidly dissolving drugs because for poorly soluble or slowly dissolving drug products, the dissolution curves of two batches can differ significantly before reaching the same value at the predefined time point. In vivo, this can result in different plasma concentrations. Consequently, it is better to compare the percentages of active dissolved at several time points or even the whole dissolution profiles (FDA Guidance for Industry, 1995, 1997).

* Corresponding author. Tel.: +32-2-477-4723; fax: +32-2-477-4735.
E-mail address: eadams@fabi.vub.ac.be (E. Adams).
[1] On leave.

0378-5173/02/$ - see front matter
PII: S0378-5173(01)00966-8

The FDA recommends the use of the similarity or f_2 factor, which is calculated as:

$$f_2 = 50 \times \log\left\{\left[1 + \frac{1}{n_{t=1}}\sum^{n} w_t(R_t - T_t)^2\right]^{-0.5} \times 100\right\}$$

with R_t and T_t the average percentage dissolved at time t (for $t = 1, 2, \ldots, n$) for the reference and the test batch respectively and w_t the optional weight factor (mostly $w_t = 1$). Two sets are considered equivalent when the f_2 factor is between 50 and 100. A value of 100 is obtained when both batches are identical. The lower limit of 50 was determined empirically by permitting a 10% average difference at any sample time point.

Another interesting tool to deal with multiple time point comparison is principal component analysis (PCA) (Adams et al., 2001). For the data sets studied in Adams et al. (2001), the dissolution profiles with a systematically higher or lower percentage can be recognized along the first principal component (PC1) in the PCA scores plot. Along PC2 the dissolution profiles with a different shape can be seen. A 95% confidence limit was constructed using the bootstrap technique.

When many data have to be measured, the possibility arises that some values are missing. In this case, case deletion and imputation methods are frequently used to obtain a 'complete' data set without missing values. In case deletion, all subjects (here: tablets) with missing values are omitted. It is clear that this approach is inefficient when only a small number of subjects (typically 6 or 12 tablets) are measured since a part of the information is discarded. Imputation methods imply that the missing data are filled in with plausible values. The easiest way is to replace the missing value(s) by the mean for that variable. However, by doing so, the correlation between the data is not respected and the covariance structure can be seriously distorted. Another possibility is the expectation–maximization (EM) algorithm. This can easily be combined with PCA (EM(-PCA)) (Nelson et al., 1996; Grung and Manne, 1998). A method favored by statisticians is the maximum likelihood estimation (MLE) (Duncan, 1986). However, it is complicated and difficult to implement.

In this paper, the performance of the EM(PCA) approach is examined using two examples. The first one is a simulation where some elements of the complete set of dissolution data B used by Adams et al. (2001) are omitted to create an incomplete one. This is done for both the reference and the test set. As a special case of missing data, the fact that the two data sets are measured at different time points is considered. For all three situations, the missing values are replaced with the predicted values obtained by the EM(PCA) algorithm. Next, PCA is performed on the recomposed, originally incomplete data set and the results are compared with the results obtained with the initial, complete data set. The use of the PCA-bootstrap technique in combination with incomplete data sets will also be discussed. Although the EM(PCA) algorithm for missing values is specific to be followed by PCA, the influence on the f_2 factor is examined too.

A second example is obtained from the pharmaceutical practice. In the scope of a comparative study between different generic tablet formulations on the Tanzanian market, dissolution profiles for the same batches were obtained in Tanzania as well as in Belgium. Since both data sets are measured at different time points, the EM(PCA) procedure is applied to be able to compare the dissolution data of both countries.

2. Theory

2.1. Principal component analysis

Principal component analysis (PCA) is a technique that allows to explore multivariate data. Here only the most essential topics will be mentioned. More detailed information can be found in Adams et al. (2001). To perform PCA, singular value decomposition (SVD) on $\mathbf{X}_c$, the column centred matrix of $\mathbf{X}$, is used:

$$\mathbf{X}_c(m \times p) = \mathbf{X} - \bar{\mathbf{X}} = \mathbf{U}(m \times a)\mathbf{\Lambda}(a \times a)\mathbf{V}^{\mathrm{T}}(a \times p)$$
$$= \mathbf{T}(m \times a)\mathbf{V}^{\mathrm{T}}(a \times p)$$

where m is the number of objects (here: the number of tablets measured in the batch), p is the number of original variables (here: the number of

time points) and a is the number of principal components (PCs), with $a = m - 1$ if $m \leq p$ or $a = p$ if $m > p$. **U** is the unweighted (normalised) and **T** the weighted (unnormalised) score matrix. **V** is the loading matrix containing the loadings of the original variables on the different PCs and **Λ** is a diagonal matrix with the singular values λ_j (for $j = 1, 2, \ldots a$) on the main diagonal. Since $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_a$, the first PCs contain the most relevant information, while the remaining PCs contain only noise.

2.2. EM(PCA) algorithm

The expectation–maximization approach is a very general and efficient tool to deal with missing data. The EM algorithm depends on the type of analysis performed afterwards. Since PCA is applied here, the EM(PCA) procedure will be used. This can be summarized as follows:

1. Replace the missing elements in data matrix **X** with their initial estimates, for instance, the mean values for the corresponding variable.
2. Perform PCA of the completed data set.
3. Reconstruct the data matrix **X** with the predefined number of significant principal components. This results in the matrix of the predicted values $\hat{\mathbf{X}}$.
4. Replace the missing elements in the original matrix **X** with their predicted values from $\hat{\mathbf{X}}$ (the observed values of **X** remain unchanged).
5. Repeat steps 2 to 4 till convergence.

The result is the original data matrix **X** in which the missing elements are replaced by the predicted values from the EM(PCA) algorithm. It is important to emphasize that the values for the missing data are optimized in order to be further analyzed by PCA.

The convergence f was calculated as:

$$f = \frac{SS_{\text{miss}}(r) - SS_{\text{miss}}(r-1)}{SS_{\text{miss}}(r-1)}$$

where SS_{miss} (for iteration r) $= \Sigma_{i=1}^{n}(\hat{x}_i)^2$ and c is the estimated value for the missing element x_i.

The EM(PCA) parameters fulfil the least square criterion since the EM(PCA) approach minimizes the sum of the squared residuals: $\min(\Sigma(x_i - \hat{x}_i)^2)$ with $x_i \in \mathbf{X}$ and $\hat{x}_i \in \hat{\mathbf{X}}$

2.3. The bootstrap technique in combination with EM(PCA)–PCA

The use of the sampling with replacement or bootstrap technique to simulate the distribution of the mean in the PC space and the construction of a 95% confidence limit is explained in Adams et al. (2001) for complete data matrices. In the case of missing values in one of the dissolution sets, two possible approaches will be further discussed:

- *EM/BOOT/PCA*: the missing values in matrix $\mathbf{X}_{\text{miss}}$ $(m \times p)$ are first replaced by their values estimated by EM(PCA) to yield a matrix **X** $(m \times p)$. Next, n bootstrap matrices ($\mathbf{X}_1$, $\mathbf{X}_2$,…, $\mathbf{X}_n$) are formed, all with size $(m \times p)$. After computation of the n column mean vectors with size $(1 \times p)$, an $(n \times p)$ matrix is formed followed by PCA.
- *BOOT/EM/PCA*: starting from matrix $\mathbf{X}_{\text{miss}}$ $(m \times p)$ with missing values, first n bootstrap matrices, also with missing values, are formed. Each of the n bootstrap matrices is then completed using EM(PCA). After construction of the $(n \times p)$ matrix of column means, PCA is performed.

2.4. Software

All programs used for the above described methods were written in MATLAB (version 4.0, the MathWorks, Natick, MA).

3. Data

For the first example (data I), the two complete data sets were originally obtained from the literature (Tsong and Hammerstrom, 1994) and correspond also to data B used by Adams et al. (2001). For both the reference and the test batch, twelve units were measured at seven different times (1, 2, 3, 4, 6, 8 and 10 h). The f_2 factor of these complete data sets amounts to 64 so that both batches can be considered as pharmaceutically equivalent.

An incomplete reference data set was created by omitting at random 7 of the 84 values (8%). This resulted in two non-consecutive values missing in

tablet 3, two consecutive in tablet 10 and one in tablets 7, 9 and 11, respectively. An incomplete test set was obtained by omitting at random 8 of the 84 values (10%). For tablets 3 and 10, each time two successive values were missing, for tablet 1 two non-successive and for tablet 2 as well as tablet 4, one value.

To simulate that two sets are measured at different time points, the data at 2, 4 and 8 h for the reference batch and at 1, 3 and 6 h for the test batch were omitted. The values at 10 h were retained for both batches so that they had one measurement time in common.

The data of the second example (data II) were kindly provided by Professor Remon (Ghent University, Ghent, Belgium). Eight different brands of the same drug product were subjected to a dissolution test (according to the USP method, apparatus 2) in both Tanzania and Belgium. For each set, six tablets were measured. The time points used in Tanzania were: 5, 15, 30 and 45 min. In Belgium however, measurements were performed at 5, 10, 15, 20, 25 and 30 min. Remark that none of the 16 data matrices is considered as reference. These data can be analyzed in two ways: (1) the eight brands can be compared with each other; and (2) for each brand the difference between the data obtained in Tanzania and in Belgium can be studied. Since the data of Tanzania and Belgium are measured at different time points, the EM(PCA) algorithm is used.

4. Results and discussion

4.1. Incomplete reference data set (data I)

First, the missing values of the reference data set are replaced by the values predicted by the EM(PCA) algorithm. Since the most relevant information for the evaluation of dissolution curves is found on PC1 and PC2 (Adams et al., 2001), the EM(PCA) values of the missing elements are optimized to fit the PCA model with two PCs. For the initial estimates in EM(PCA), the mean values of the respective variable are used. The value of convergence is set at 10^{-12}. After column centering, the recomposed reference data set is used to construct the PC space, in which the original, complete test set is projected. The PC1/PC2 scores plot is shown in Fig. 1. This figure also contains the scores that were obtained using the initial, complete reference set. As can be seen, similar results are obtained. It is noteworthy that the percentage of explained variance by PC1 and PC2 increases from 90.6% for the initial data set to 91.1% for the one recomposed by EM(PCA). This could be expected because the predicted values for the missing data perfectly fit the model.

The loadings of the original variables (i.e. seven measurement times) on the first two PCs are shown in Fig. 2. In the case that missing values are present in variables with high loadings (like measurement times 1, 2 and 10 h) for objects far away from the rest of the group (like tablets 2 and 9 of the reference batch) it is sometimes possible that the value of convergence is not reached. This is due to the relatively small number of objects in the data set.

The similarity factor (f_2) calculated between the completed reference and original test set is 64, which is the same as obtained with the initial, complete data set.

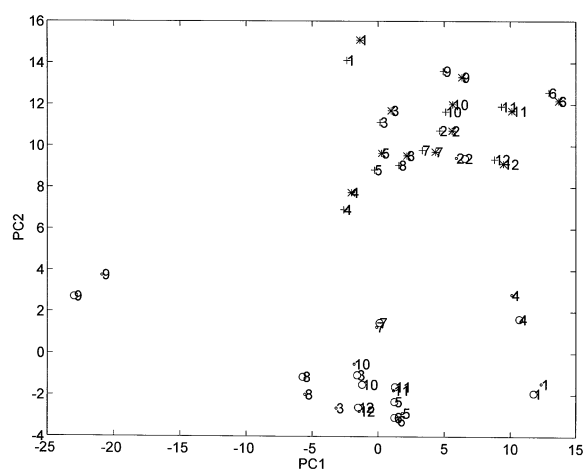

Fig. 1. PCA scores plots of the original dissolution curves and the ones recomposed by EM(PCA) for data I. The (○) represent the original and the (.) the recomposed reference batch. The (∗) and (+) are the projections of the dissolution profiles of the test batch, calculated using the original and the recomposed reference batch, respectively. The figures correspond to the 12 tablets of each batch.

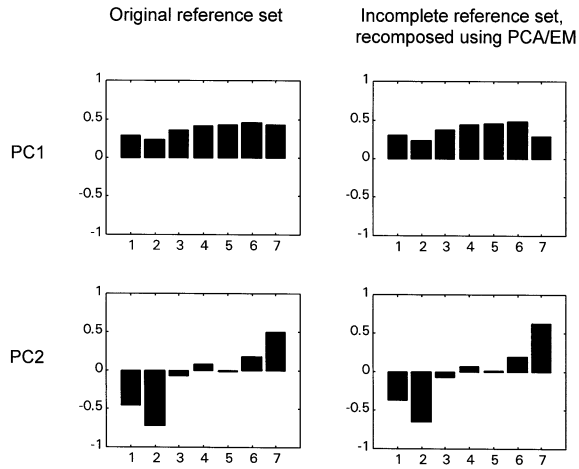

Fig. 2. Loadings of the original variables (i.e. seven measurement times) on PC1 and PC2 for data I.

4.2. Incomplete test data set (data I)

Similar to Section 4.1, the eight missing elements were filled in with the values obtained by the EM(PCA) method. The EM(PCA) values of the incomplete test set were optimized using the PCA model constructed with the (complete) reference set. Afterwards, the initial and the recomposed test set are projected in the PC space built by the reference set. The PC1/PC2 scores plot is shown in Fig. 3. As can be seen, also here comparable results are obtained. When the f_2 factor is computed for the original reference and the recomposed test set, it still is equal to 64, the value which is obtained with the original data.

4.3. Different measurement times

4.3.1. Data I

This special case of missing data was treated in a somewhat different way from the cases described in Sections 4.1 and 4.2. When the reference and the test set are measured at different times, the resulting matrices contain different variables. Since both matrices have to be uniform (contain the same variables), values for the missing variables must be calculated for the reference as well as the test set. However, since some variables are completely missing for the reference matrix, it is not possible to perform properly EM(PCA) and to predict well the scores for the test matrix. For this reason, the matrices of the incomplete reference (12×7) and test set (12×7) are joined together into one matrix (24×7). In this matrix, none of the variables is completely missing. To be able to estimate the correlation structure, it is however necessary that there is still at least one variable in common (here: measurement time 7 (10 h)). As initial estimates in the EM(PCA) procedure, the values linearly interpolated between the observed data are used. The value of convergence is here also set at 10^{-12} and five PCs are used in the optimization step of the EM(PCA) algorithm. To present the results the same way as before, the matrix (24×7) obtained after EM(PCA) is split again into the corresponding matrix for the reference batch (12×7) and the one for the test batch (12×7). Next, PCA is performed after column centering and the recomposed reference set is used to build the PC space wherein the recomposed test set is projected. The PC1/PC2 scores plot is shown in Fig. 4. As can be seen, good results are obtained.

When the similarity factor is calculated for both recomposed batches, 53 is found. Although it still indicates that the batches are pharmaceutically similar, there is an appreciable difference with the

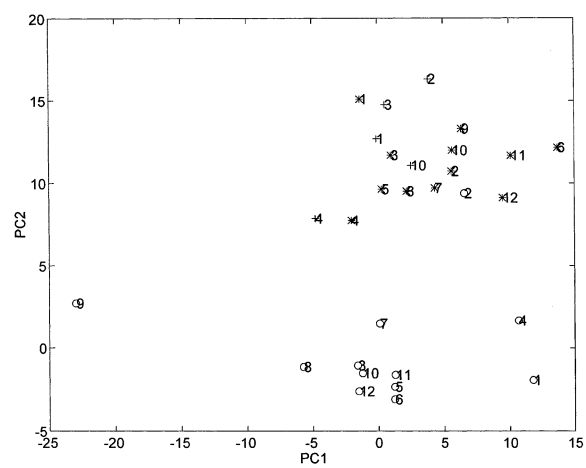

Fig. 3. The PCA scores plots of the dissolution profiles of the reference batch (○), together with the projections of the dissolution profiles of the original test batch (∗) and the test batch recomposed by EM(PCA) (+) for data I.

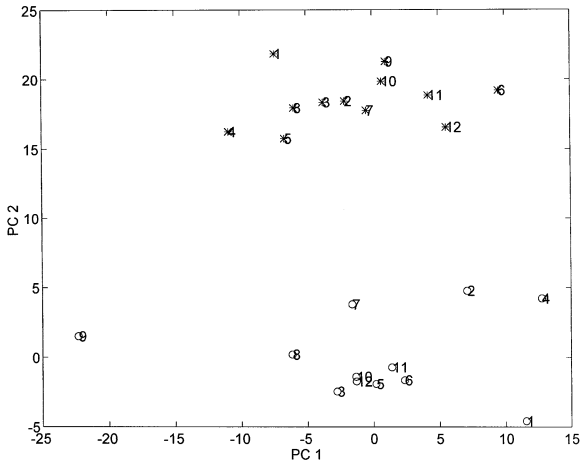

Fig. 4. The PCA scores plots of the dissolution profiles of the reference batch (○), together with the projections of the dissolution profiles of the test batch (∗) for data I. The batches were measured at different time points and were recomposed by EM(PCA).

original value of 64. As already mentioned earlier, the result of the EM(PCA) approach is not meant to be analyzed by the f_2 factor. When only a few values are missing in the data set (as in Sections 4.1 and 4.2) the result is nearly the same, but in the extreme case discussed here, where about half of the data set has to be completed, the difference can become considerable.

4.3.2. Data II

First, the Tanzanian data sets are compared with the corresponding Belgian ones. An analogous procedure as in Section 4.3.1 is used to recompose the data matrices. The only difference is that both sets are used to construct the PC space because no distinction could be made between reference and test set. Although the same batches were measured in Tanzania and in Belgium, some differences between the data of both countries can be observed. A typical example is given in Fig. 5 where the PC1/PC2 scores plot for one of the eight brands is shown. To have an idea about the pharmaceutical relevance of these observed differences, the f_2 factor for each of the eight sets of measurements was calculated. One was below the limit of 50 ($f_2 = 46$), several others were close to it ($f_2 = 53$, 54, 55) and four varied between 65 and 71.

Although a more profound discussion about the differences observed between the Tanzanian and Belgian data is beyond the scope of this paper, some possible causes can be: a difference in storage of the samples (temperature and humidity), small differences in the apparatus used (in spite of the fact that both meet the USP requirements) and a difference in training of the laboratory workers.

A second possibility to analyze data II is the comparison of the eight brands with each other. Because all eight batches from one country are measured at the same time points, there are no missing data so that the EM(PCA) algorithm is not necessary. Since the PCA results can be better visualized in combination with the bootstrap technique, they will be shown in Section 4.4.2.

The pharmaceutical equivalence of the eight brands is evaluated using the f_2 factor. To compare the eight brands two by two, 28 combinations are possible (1–2, 1–3, 1–4, 1–5, 1–6, 1–7, 1–8, 2–3, 2–4,…). Using the original data, in Tanzania only 8 of the 28 f_2 factors are above the limit of 50: brands 1–3 ($f_2 = 73$), 2–5 ($f_2 = 60$), 3–8 ($f_2 = 53$), 4–6 ($f_2 = 68$), 4–7 ($f_2 = 64$), 5–6 ($f_2 = 52$), 5–7 ($f_2 = 61$) and 6–7 ($f_2 = 68$). The lowest similarity factor is even 21 (brands 1–4). The f_2 factors are also computed using the resulting matrices from

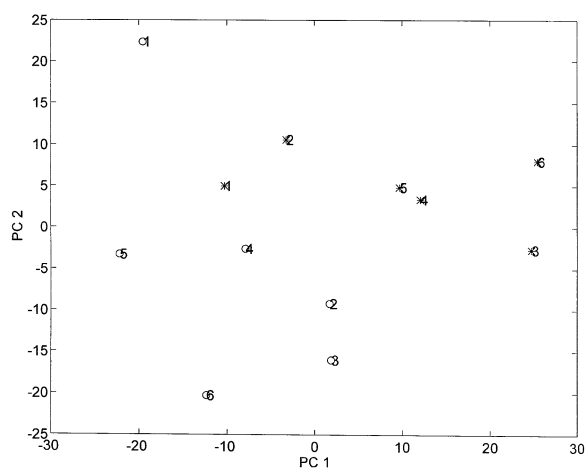

Fig. 5. The PCA scores plots of the dissolution profiles for one of the brands of data II: (○) Belgium and (∗) Tanzania. Since both sets are not measured at the same time points, EM(PCA) is applied.

EM(PCA). After comparison with the f_2 values based on the original data, the maximum difference found with EM(PCA) is 5. In only one case this has an influence on the conclusion, because now 9 of the 28 f_2 values are greater than 50 (the same 8 as above, plus combination 4–5 ($f_2 = 50$)). Using the original Belgian data, 9 of the 28 f_2 factors are above 50: combinations 1–3 ($f_2 = 55$), 1–8 ($f_2 = 58$), 2–7 ($f_2 = 53$), 2–8 ($f_2 = 55$), 3–8 ($f_2 = 50$), 4–6 ($f_2 = 55$), 5–6 ($f_2 = 55$), 5–7 ($f_2 = 70$) and 6–7 ($f_2 = 53$). With the EM(PCA) data, also combination 4–7 has a f_2 factor > 50 ($f_2 = 51$) so that 10 of the 28 brands are considered similar. The maximum difference between the corresponding f_2 factors (without and after EM(-PCA)) is 2.

Beside the considerable differences between the eight brands, the results obtained with the Tanzanian data differ also from those obtained with the Belgian ones. This is not surprising because the comparison of the measurements for the same brands already revealed differences between both countries.

4.4. The bootstrap technique in combination with missing data

4.4.1. Data I

Although PCA is a powerful tool to visualize data, it contains no criteria to decide that two batches are similar or different. A statistically based conclusion can be obtained using the bootstrap technique. When bootstrap samples ($n = 1000$) are generated for both the complete reference and test batch, followed by PCA, a normalized scores plot as shown in Fig. 6 is obtained. The circle represents the 95% confidence limit for the reference batch.

In the case that the reference batch has missing values, the results obtained with BOOT/EM/PCA and EM/BOOT/PCA compared to the original situation, are shown in Fig. 7. For clearness, the picture was simplified compared to Fig. 6 by giving only the contours. Taking into account that bootstrapping does not always give exactly the same result, the differences for the reference batch are negligible. For the test batch, BOOT/EM/PCA gives slightly better results than EM/BOOT/PCA. However, the first approach takes much more time because EM(PCA) is applied to each of the 1000 bootstrapped matrices. When no convergence ($> 10^{-12}$) was obtained within 400 iterations, the result obtained was not taken into account for further calculations. On the other hand, the EM/BOOT/PCA approach has only one EM(PCA) step. This implies that calculations are much faster, but in some cases it is possible

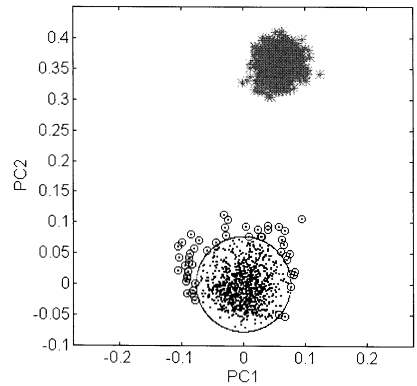

Fig. 6. The PCA normalized scores plot after sampling with replacement ($n = 1000$) for the original reference batch (.) together with the projections of the dissolution curves of the original test set (∗). The circle indicates the 95% confidence limit for the reference set and the 5% omitted objects are indicated by (⊙).

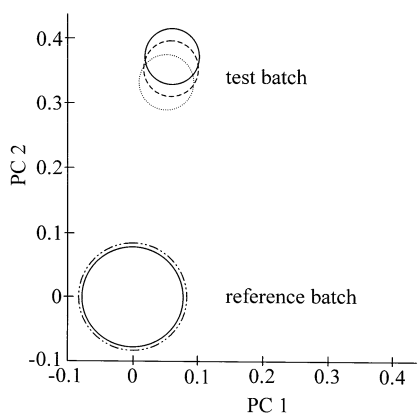

Fig. 7. The PCA normalized scores plots after sampling with replacement ($n = 1000$) for both the complete and incomplete reference batch, together with the projections of the profiles of the test set: (—) represent the results obtained with the complete sets, (···) the ones for the reference set recomposed by EM/BOOT/PCA and (---) by BOOT/EM/PCA.

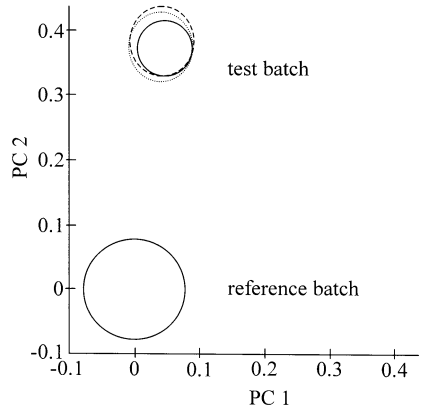

Fig. 8. The PCA normalized scores plots after sampling with replacement ($n = 1000$) for the complete reference batch, together with the projections of the profiles of the test set: (—) represent the results obtained for the complete sets, (···) the ones for the test set recomposed by EM/BOOT/PCA and (---) by BOOT/EM/PCA.

that convergence is slow (> 1000 iterations) or even not reached after 5000 iterations. When more than 1000 iterations are required to obtain convergence in the EM step of EM/BOOT/PCA, BOOT/EM/PCA is preferred since it gives results which are clearly more similar to the original situation with the complete reference batch.

When missing values are present in the test batch, the BOOT/EM/PCA and EM/BOOT/PCA approaches give comparable results as shown in Fig. 8. In both cases, the distribution of the scores was somewhat wider compared to the original, complete test set.

4.4.2. Data II

The BOOT/EM/PCA and EM/BOOT/PCA methods are also applied on data II. The results obtained after BOOT/EM/PCA and EM/BOOT/PCA are similar, but the first method is much slower to perform. When the data of the same batch, measured in Tanzania and in Belgium respectively, are compared, both sets can be clearly distinguished in all of the eight cases. The f_2 factor is less discriminative, since in seven of the eight cases, the respective Tanzanian and Belgian sets are considered as pharmaceutically similar ($f_2 > 50$).

It is also interesting to compare the eight brands with each other using the results measured in Tanzania and in Belgium respectively. As already mentioned before, the dissolution data from one country are measured at the same time points so that the EM(PCA) algorithm is not necessary. However, to check the performance of this algorithm, the data matrices obtained above (after EM/BOOT/PCA for the individual comparison of each batch) were also used here. Fig. 9 shows that the results with the complete sets and those after EM/BOOT/PCA are similar. For the Tanzanian data, brands 1–3, 4–6, 4–7, 5–7 and 6–7 are close to each other. These combinations also have the highest f_2 factors. The other combinations with a f_2 factor above 50 (2–5, 3–8 and 5–6) are somewhat further in the PC-bootstrap plot, but have similar scores on PC1. The Belgian data show similar results for brands 1–8, 2–7, 2–8, 3–8 and 5–7. All have a similarity factor > 50. The other combinations that can be considered pharmaceutically equivalent based on the f_2 factor (1–3, 4–6, 5–6 and 6–7) have also similar scores on PC1 or PC2.

The PC plots also clearly illustrate the differences between the Tanzanian and Belgian data. Beside other brands that overlap, the variability between the dissolution profiles within a batch is greater for the Belgian data. The f_2 factor does not take into account the variability since it is calculated with the mean values.

5. Conclusion

The use of PCA for the evaluation of dissolution profiles in the case of an incomplete reference or test set was examined. The expectation-maximization algorithm in combination with PCA was found to be a good and relatively simple approach to deal with missing data. The EM(PCA) approach is also useful when both dissolution data sets are measured at different time points (except at least one). The bootstrap technique can still be applied in case of data sets with missing elements.

Although it is not the intention to analyze the results from the EM(PCA) algorithm by the simi-

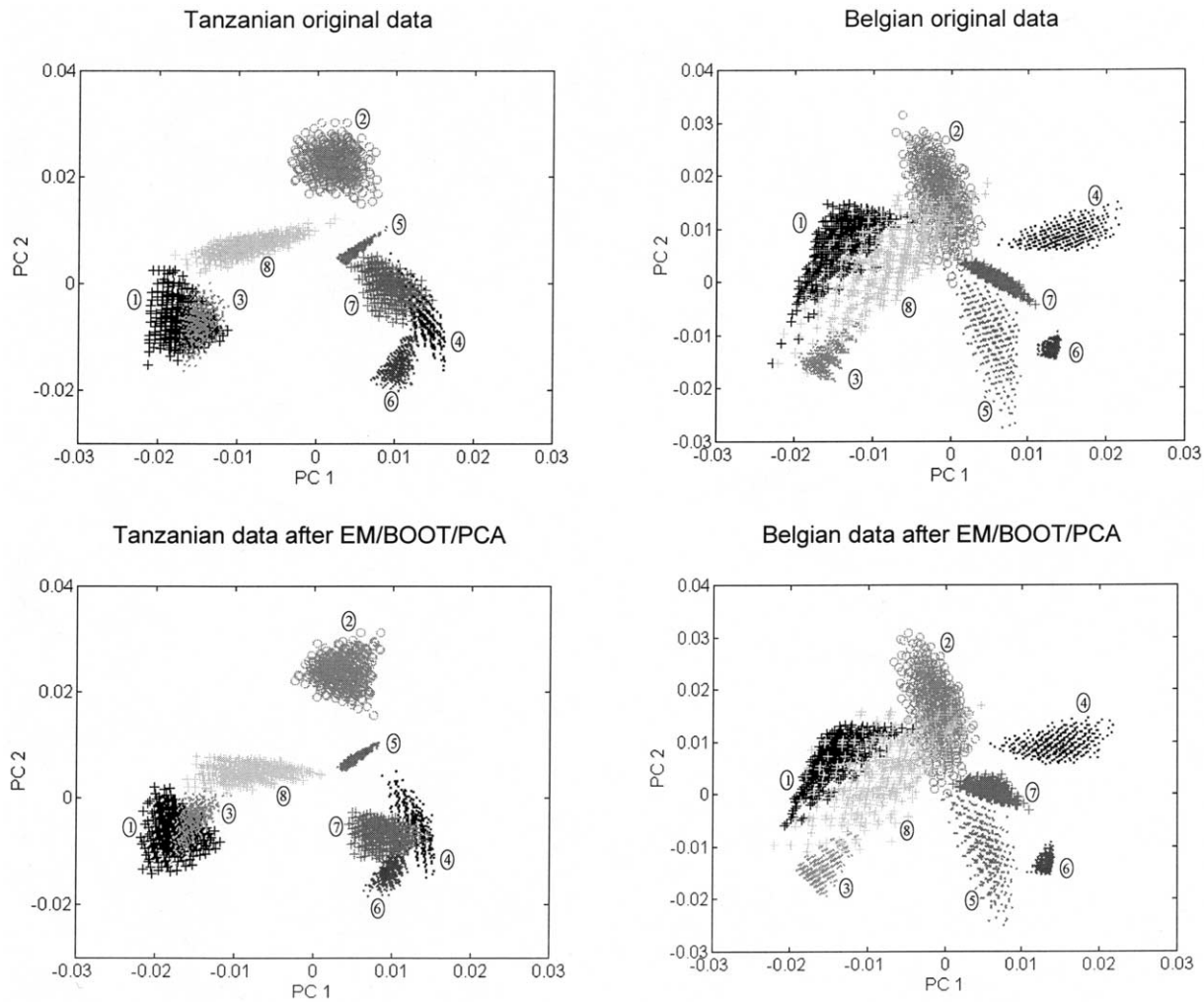

Fig. 9. The PCA normalized scores plot after sampling with replacement ($n = 1000$) for the eight brands of data II.

larity factor, the differences with the original f_2 factor (calculated with the complete data) were, for the cases studied, negligible when only a few values were missing.

Acknowledgements

This research was financed by a post-doctoral fellowship of the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), Brussels.

References

Adams, E., De Maesschalck, R., De Spiegeleer, B., Vander Heyden, Y., Smeyers-Verbeke, J., Massart, D.L., 2001. Evaluation of dissolution profiles using principal component analysis. Int. J. Pharm. 212, 41–53.

Duncan, A.J., 1986. Quality Control and Industrial Statistics. Irwin, Homewood, IL.

FDA Guidance for Industry, November 1995. Immediate Release Solid Oral Dosage Forms — Scale-Up and Postapproval Changes: Chemistry, Manufacturing and Controls; In Vitro Dissolution Testing and In Vivo Bioequivalence Documentation. Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, MD.

FDA Guidance for Industry, August, 1997. Dissolution Testing of Immediate Release Solid Oral Dosage Forms. Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, MD.

Grung, B., Manne, R., 1998. Missing values in principal component analysis. Chemom. Intell. Lab. Syst. 42, 125–139.

Nelson, P.R.C., Taylor, P.A., MacGregor, J.F., 1996. Missing data methods in PCA and PLS: scores calculations with incomplete observations. Chemom. Intell. Lab. Syst. 35, 45–65.

Tsong, Y., Hammerstrom, T., 1994. Statistical issues in drug quality control based on dissolution testing. Proceedings of the Biopharmaceutical Section of the American Statistical Association, pp. 295–300.