Talanta 123 (2014) 186–199

Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

Review

Chemometrics tools used in analytical chemistry: An overview

Naveen Kumar, Ankit Bansal, G.S. Sarma, Ravindra K. Rawal *

Department of Pharmaceutical Analysis, Indo-Soviet College of Pharmacy (ISFCP), Moga 142001, India

ARTICLE INFO

Article history:
Received 7 November 2013
Received in revised form
2 February 2014
Accepted 3 February 2014
Available online 14 February 2014

Keywords:
Singular value decomposition
Evolving factor analysis
Chemometrics
Rank annihilation
Multivariate calibration
Principal component

ABSTRACT

This article presents various important tools of chemometrics utilized as data evaluation tools generated by various hyphenated analytical techniques including their application since its advent to today. The work has been divided into various sections, which include various multivariate regression methods and multivariate resolution methods. Finally the last section deals with the applicability of chemometric tools in analytical chemistry. The main objective of this article is to review the chemometric methods used in analytical chemistry (qualitative/quantitative), to determine the elution sequence, classify various data sets, assess peak purity and estimate the number of chemical components. These reviewed methods further can be used for treating *n*-way data obtained by hyphenation of LC with multi-channel detectors. We prefer to provide a detailed view of various important methods developed with their algorithm in favor of employing and understanding them by researchers not very familiar with chemometrics.

© 2014 Elsevier B.V. All rights reserved.

Contents

1. Introduction . . . 187
2. Origin and development of chemometrics . . . 187
3. Chemometrics (multivariate data analysis) tools . . . 187
   3.1. Multivariate regression method . . . 187
      3.1.1. Multiple linear regressions . . . 188
      3.1.2. Multivariate calibration methods . . . 188
   3.2. Multivariate decomposition methods . . . 189
      3.2.1. Principal component analysis (PCA) . . . 190
      3.2.2. Parallel factor analysis (PARAFAC) . . . 191
   3.3. Hierarchical cluster analysis (HCA) . . . 191
   3.4. Pattern recognition methods . . . 191
      3.4.1. Soft independent modeling of class analogy (SIMCA) . . . 191
      3.4.2. Linear discriminate analysis (LDA) . . . 192
   3.5. Chemometric resolution methods (CRMs) . . . 192
      3.5.1. Rank annihilation factor analysis (RAFA) . . . 192
      3.5.2. Generalized rank annihilation method (GRAM) . . . 193
      3.5.3. Evolving factor analysis (EFA) . . . 193
      3.5.4. Heuristic evolving latent projections (HELP) . . . 193
      3.5.5. Subwindow factor analysis (SFA) . . . 194
      3.5.6. Window factor analysis (WFA) . . . 194
      3.5.7. Alternating least square (ALS) . . . 195
      3.5.8. Orthogonal projection analysis (OPA) . . . 195
4. Application of chemometric tools in analytical chemistry . . . 196

* Corresponding author. Tel.: +91 9878 882520; fax: +91 1636 239515.
E-mail address: rawal.ravindra@gmail.com (R.K. Rawal).

http://dx.doi.org/10.1016/j.talanta.2014.02.003
0039-9140 © 2014 Elsevier B.V. All rights reserved.

5. Conclusion . . . 198
Acknowledgments . . . 198
References . . . 198

1. Introduction

Analytical procedure is a powerful tool that has the potential to increase the efficiency of audits since it is a relatively low-cost procedure that seems to have considerable value in identifying errors or irregularities and in guiding audits. Analytical result is one of the keys to guarantying the quality of products. These enable us to verify the stability and purity of pharmaceutical and food products. Hence, we say that analytical procedures must comply with the audits, if the consistency of their results is assured [1].

In the present era analysts have increasing interest in the chromatographic and spectroscopic analyses of complex mixtures such as drugs, herbal medicines, food and blood plasma samples. To achieve the best separation quality in support of ensuing qualitative or/and quantitative analysis, chromatographic and spectroscopic conditions should be optimized according to the particular analytical objectives. So the important and significant factors like solvent, mobile phase, pH and column temperature affect the outcome of analysis and should be defined. In addition an experimental procedure also approaches the chemical reality of a sample. The next step involves determining the resolution or separation quality measured in terms of response or optimization function, which is followed by the development of a mathematical or statistical model that describes the relationship between analytical (spectroscopic/chromatographic) parameters and the responses of the designed experiments and at last predict optimal separation conditions. Few times it might be repeated and the predicted optimal conditions will be modified and validated to achieve desirable analytical results [2].

Analytical results involve resolution, asymmetry ratio, peak purity, precision, accuracy, robustness and so on. For evaluating these parameters from complex mixtures, a practical response is most likely to be multi-criterion, i.e., one weighting gives several different criteria. So, some response functions like derringer's response function are used traditionally [3,4].

Obviously, high-quality analytical results can rarely be obtained by optimizing an inappropriate response function not well related to the actual separation quality and special analytical objectives. Many traditional response functions involve single-response detectors. Traditional response function may fail to characterize some crucial aspects of separation quality and encounter some difficulties in practical use. However, now recent approaches involve different hyphenations, e.g. HPLC-DAD, GC–MS and LC–NMR that can basically increase the available information, and are also useful for qualitative and qualitative analyses. With the aid of the spectral information in hyphenated instruments, greatly enhanced performance is seen in terms of the elimination of instrumental interferences, retention time shift correction, selectivity, chromatographic separation abilities and measurement precision; however, data coming from these instruments is very complex and difficult to resolve or interpret [5].

Therefore to achieve these results from our instrument we require an appropriate mathematical and statistical approach using a suitable response function. The proper use of such a model with information obtained from hyphenations will cast new light on the evaluation of analytical data. Therefore, the hyphenated technique is further combined with chemometric approaches to develop a clear picture of herbal fingerprint.

Thus, our main aim is to discuss the various chemometrics tools with respect to regular analysis to improve the quality of analytical determinations of complex samples by fulfilling the performance criteria. As we know there is a wide gap between analysts and chemometricians. Thus, many times, they are not able to use analytical instruments. In the present review we have discussed analytical process hyphenated with multivariate analysis and the application of various tools of multivariate analysis in analytical chemistry.

2. Origin and development of chemometrics

In 1971, a Swedish scientist Svante Wold coined the term “kemometri,” in Swedish and in English it is equivalent to “chemometrics” [6]. The science of chemometrics can briefly be described as the interaction of certain mathematical and statistical methods in chemical measurement processes. It has been developed as a consequence of the change in the data obtained with the emergence of new analytical techniques as well as microprocessors. During 1986–1987 two journals – named “Chemometrics and Intelligent Laboratory Systems” and “Journal of Chemometrics” – were published.

The breakthrough in chemometrics came in the 21st century by various software development companies, which promoted equipment intellectualization and offered new methods for the construction of new and high-dimensional hyphenated equipment. This hyphenated equipment has opened many new options for data analytical method improvement. Now, chemometrics plays a major role in analytical chemistry [7].

3. Chemometrics (multivariate data analysis) tools

Multivariate data analysis involves the analysis of data consisting of numerous variables measured from a number of samples. The aim of multivariate data analysis is to determine all the variations in the data matrix study. Thus, chemometric tools try to find the relationships between the samples and variables in a given data set and convert into new latent variables. Multivariate data analysis is mainly classified into multivariate regression and multivariate calibration methods based on complexity of the data estimated.

Multiple linear regression is widely applied for solving various types of problems in one or few component analyses; however, in many cases the involvement of multiple variables' interaction of analytes with each other, especially in herbal medicines (HMs), leads to quantify error. Therefore, in those cases biased regression methods can provide better results [8]. These methods are commonly known as multivariate calibration methods. Therefore, in this article we discuss both multiple linear regression and multivariate calibration methods with their corresponding examples.

3.1. Multivariate regression method

In study of most of the chemicals, the concentration of one or more analytes which has to be determined is based on measured properties of the system [9]. For example, linear regression equation between two variables, concentration and absorbance, for the spectrophotometric determination of X analyte at λ

wavelength can be defined by the following equation:

$$A_{Xi} = b_{Xi}C_X + a_{Xi} \quad (1)$$

where C_X and A_{Xi} are the concentration and absorbance of the analyte X at λ wavelength, respectively, b_{Xi} is the slope of the linear regression equation and a_{Xi} is the intercept of the regression model [10]. Absorption of electromagnetic radiation at λ can be related to concentration through the Beer–Lambert law.

$$\log Io/I = \alpha cl \quad (2)$$

where I is the intensity of light at λ passing through a sample of length l, Io is the light intensity incident on the sample, c is he concentration and α is the absorptivity of the sample at the specific wavelength. Mixture containing multiple analytes becomes more complicated and to resolve these components we need a regression model. Although there are different types of models, generally linear, interaction and quadratic models are most often used [11]. A linear model is

$$Y = \varepsilon 1c1l + \varepsilon 2c2l + \cdots + \varepsilon pcpl + E \quad (3)$$

where Y is the response of a particular sample at λ, ε_p is the molar extinction coefficient for the pth sample, c_p is the corresponding concentration and E is the spectral error. If ε is measured at several wavelengths, then the equation obtained is

$$Y_j(\lambda i) = \sum_{k=1}^{p} \varepsilon_k c_{kj} l + E_y \quad (4)$$

where i is the index for wavelengths, j is the index for samples, p is the number of components to be determined and E_y is spectral errors. In matrix notation, Eq. (4) can be written as

$$Y = C*K + E_Y \quad (5)$$

where Y is the $j \times i$ matrix of calibration spectra, C is the $j \times l$ matrix of component concentration, K is the $l \times i$ matrix of absorptivity-path length products, and E_Y is the $j \times i$ matrix of spectral errors.

3.1.1. Multiple linear regressions

3.1.1.1. Tri-linear regression-calibration (TLRC). It is a calibration model used to determine the mixture having three analytes with the condition that they are not interfering with each other. For example, we have three analytes (X, Y and Z) that are measured at three wavelength sets ($\lambda_i = 1$, 2 and 3). The following equations can be written for a three-component analysis:

$$A_{mix1} = b_{X1}C_X + b_{Y1}C_Y + b_{Z1}C_Z + a_{XYZ1} \quad (6a)$$

$$A_{mix2} = b_{X2}C_X + b_{Y2}C_Y + b_{Z2}C_Z + a_{XYZ2} \quad (6b)$$

$$A_{mix3} = b_{X3}C_X + b_{Y3}C_Y + b_{Z3}C_Z + a_{XYZ3} \quad (6c)$$

where A_{mix1}, A_{mix2} and A_{mix3} represent the absorbance of the mixture of X, Y and Z analytes at three-wavelength sets, $b_{X1,\ 2\ \text{and}\ 3}$, $b_{Y1,\ 2\ \text{and}\ 3}$ and $b_{Z1,\ 2\ \text{and}\ 3}$ are the slopes of linear regression equations of X, Y and Z, respectively, and a_{XYZ1}, a_{XYZ2} and a_{XYZ3} are the sums of intercepts of linear regression equations at the three wavelengths.

Eq. (6) in matrix notation is represented as

$$\begin{bmatrix} A_{mix1} \\ A_{mix2} \\ A_{mix3} \end{bmatrix} = \begin{bmatrix} b_{X1} & b_{Y1} & b_{Z1} \\ b_{X2} & b_{Y2} & b_{Z2} \\ b_{X3} & b_{Y3} & b_{Z3} \end{bmatrix} * \begin{bmatrix} C_X \\ C_Y \\ C_Z \end{bmatrix} + \begin{bmatrix} a_{XYZ1} \\ a_{XYZ3} \\ a_{XYZ3} \end{bmatrix} \quad (7)$$

This equation can also be written as

$$(A_{mix} - a_{XYZ})_{3*1} = K_{3*3} \cdot C_{3*1} \quad (8)$$

The matrix, b, corresponding to the slope values of linear regression equations is called the matrix, K, which is expressed as

$$K = \begin{bmatrix} b_{X1} & b_{Y1} & b_{Z1} \\ b_{X2} & b_{Y2} & b_{Z2} \\ b_{X3} & b_{Y3} & b_{Z3} \end{bmatrix} \quad (9)$$

In case, for the calculation of the concentration of the analytes, X, Y and Z in ternary mixture, the matrix $(A_{mix} - a_{XYZ})_{3*1}$ is multiplied by the inverse $(K^{-1})_{3*3}$ of the matrix K_{3*3} and it can be written as

$$C_{3*1} = (k^{-1})_{3*3}(A_{mix} - a_{XYZ})_{3*1} \quad (10)$$

This procedure is the mathematical basis of the TLRC method for multi-component analysis. The developed calibration model can be applied easily to the resolution of the three-component mixtures. However, this method is not appropriate to resolve mixtures with more than three components in a pharmaceutical mixture [12,13].

3.1.1.2. Multi-linear regression-calibration (MLRC). In case the absorbance values of a mixture of three or more analytes are measured at n wavelengths ($\lambda_i = 1, 2, \ldots, n$), the following set of equations can be written for a multi-component analysis:

$$\begin{aligned} A_{mix1} &= b_{X1}C_X + b_{Y1}C_Y + \cdots + b_{M1}C_M + a_{XY\ldots m1} \\ A_{mix2} &= b_{X2}C_X + b_{Y2}C_Y + \cdots + b_{M2}C_M + a_{XY\ldots m2} \\ &\cdots \\ &\cdots \\ A_{mixn} &= b_{Xn}C_X + b_{Yn}C_Y + \cdots + b_{Mn}C_M + a_{XY\ldots mn} \end{aligned} \quad (11)$$

where A_{mix1}, A_{mix2}, … and A_{mixn} represent the absorbances of the mixture of X, Y, … and m analytes at the selected wavelength set, (λ_i–λ_n); $b_{X1,\ 2,\ \ldots,\ n}$, $bY_{1,\ 2,\ \ldots,\ n,\ \ldots}$ and $b_{M1,\ 2,\ \ldots,\ n}$ are the slopes of n linear regression equations of m components and $a_{X\ldots m1}$, $a_{XY\ldots m2,\ldots}$ and $a_{XY\ldots mn}$ are the sums of intercepts of linear regression equations at n wavelengths.

In the matrix terms, the multi-equation system in Eq. (11) can be formulated as follows:

$$\begin{bmatrix} A_{mix1} \\ A_{mix2} \\ \vdots \\ \vdots \\ A_{mix3} \end{bmatrix} = \begin{bmatrix} b_{X1} & b_{Y1} & \ldots & \ldots & b_{m1} \\ b_{X2} & b_{Y2} & \ldots & \ldots & b_{m2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{Xn} & b_{Yn} & \ldots & \ldots & b_{mn} \end{bmatrix} * \begin{bmatrix} C_X \\ C_Y \\ \vdots \\ \vdots \\ C_{mM} \end{bmatrix} + \begin{bmatrix} a_{XY\ldots m1} \\ a_{XY\ldots m2} \\ \vdots \\ \vdots \\ a_{XY\ldots mn} \end{bmatrix} \quad (12)$$

Thus Eq. (12) can also be written as

$$(A_{mix} - a_{XY\ldots.m})_{n*1} = K_{n*m} \cdot C_{m*1} \quad (13)$$

The matrices C_{m*1} are calculated from Eq. (13) as follows:

$$C_{m*1} = [(K^T)_{m*n}(K)_{n*m}]^{-1} * [K^{-1}]_{m*n} * (A_{mix} - a_{XY\ldots m})_{n*1} \quad (14)$$

In this case, the MLRC model contains the use of linear algebra, also known as matrix mathematics. This calibration model can be applied for the multi-resolution of a multi-component mixture system containing m compounds. But it has limited applicability on biological and herbal mixtures [12–14].

3.1.2. Multivariate calibration methods

Multivariate calibration methods have been widely used because it analyses those analytes that interact with each other. In these methods, first we require the preparation of the training set from which a series of properties has been measured and then

the prediction set is made in which the training set is used to determine the concentration of the components of unknown mixtures from their spectral data [15].

3.1.2.1. Classical least squares (CLS). This method assumes Beer's law model with the absorbance at each frequency being proportional to the component concentrations. In matrix notation, Beer's law model for m calibration standards containing l chemical components with the spectra of n digitized absorbances is given by

$$A = C{*}K + E_A \tag{15}$$

where A is the $m \times n$ matrix of calibration spectra, C is the $m \times l$ matrix of component concentration, K is the $l \times n$ matrix of absorptivity-path length products, and E_A is the $m \times n$ matrix of spectral errors.

The classical least squares solution during calibration is given as follows:

$$K = (C^T C)^{-1} C^T {*} A \tag{16}$$

where K represents the matrix of pure component spectra at unit concentration and unit path length.

Analysis based on the spectrum of unknown components concentration (samples) is given as follows:

$$c_0 = (KK^T)^{-1} K{*}A \tag{17}$$

where c_0 is the vector of predicted concentrations and K^T is the transpose of the matrix K [16,17]. CLS is a linear least square method and its main disadvantages are limitations in the matrix shapes that linear models can assume over long ranges, possibly poor extrapolation properties, and its sensitivity to outliers.

3.1.2.2. Inverse least squares (ILS). This method treats these concentrations as a function of absorbance. The inverse of Beer's law model for m calibration standards with spectra of n digitized absorbance is given by

$$C = A{*}P + E_c \tag{18}$$

where C and A are as before, P is the $n \times l$ matrix of unknown calibration coefficient relating the l component concentrations of the spectral intensities, and E_c is the $m \times l$ vector of errors. The inverse least square solution during calibration for P is

$$P = (A^T A)^{-1} A^T {*} C \tag{19}$$

In this method the concentration of the analyte in the unknown sample is given as

$$c_0 = a^T {*} P \tag{20}$$

where c_0 and a represent the concentration and spectrum of the unknown analyte, respectively. Since in ILS the number of frequencies cannot exceed so that the total number of calibration mixtures is used, and stepwise multiple linear regressions have been used for the selection of frequencies [17–19].

3.1.2.3. Partial least square regression (PLSR). PLSR is used to analyze strongly collinear and noisy data with numerous X variables (independent variables) and also simultaneously model the several response variables, i.e. Y (dependent variables). MLR in which modeling of Y by means of X is done as long as when data is few and fairly uncorrelated. However, in modern instrumentation only X variables are in larger numbers and also strongly correlated so that they are usually noisy and incomplete [20,21]. PLSR allows one to inspect more complex problems by handling numerous and collinear X variables and response variables Y and analyze data in a more rational way.

Mean centering or scaling of both X and Y data matrix is performed in PLSR so that it is fitted in such a way that it describes the variance of X and Y. PLSR is a maximum covariance method, because the main aim of PLS is to predict the y-variables from the x-variables. PLSR finds out the new variables for both X and Y matrices, i.e. X-scores (T) and Y-scores (U), respectively [22].

X-scores estimate the linear combination of variable x_k with coefficient of weight (W^*)

$$T = X \cdot W^* \tag{21}$$

However, the weight W can be transformed to W^* which is directly related to X [23].

From Eq. (21), W^* can be written as

$$W^* = W(P^T W)^{-1} \tag{22}$$

The PLSR model can be supposed to consist of an outer relation and an inner relation where the outer relation describes the X and Y matrices individually while the inner relation links the two matrixes together. The outer relations are given by the following equations:

$$X = T \cdot P^T + E \tag{23}$$

$$Y = U \cdot C^T + F \tag{24}$$

where P^T is the loading matrix of the X space, C^T is the loading matrix of the Y space. E and F are the residual matrices of the X and Y spaces, respectively.

X scores (T) are also good predictors for Y variables, i.e. correlated according to the following equation:

$$Y = T \cdot C^T + G \tag{25}$$

By combining Eqs. (21) and (25) we can write

$$Y = XW^* C^T + G = XB + G \tag{26}$$

where

$$B = W^* C^T \tag{26a}$$

where B represents the PLSR coefficient and G is the residual matrix. The prediction of y-variables of new samples is determined by Eq. (26).

By putting the value of W^* from Eq. (22) in Eq. (26a)

$$B = W(P^T W)^{-1} C^T \tag{27}$$

The part which is not explained by the model is called residuals. It is useful in determining model applicability which is indicated by residual value. Large residual value indicates that the model is poor. When the first PLS component has been calculated, then further one can be calculated based on the residual matrices. This process continues until we achieve approximately 99 of the explained variance. The number of significant PLS components in a calibration model can be decided by means of cross-validation [24–27]. The main limitation in this method is the preparation of calibration as well as the prediction of the set and the employing of human decision for selecting the number of factors. In spite of the wide applicability of this method, it is not applicable to complex herbal mixtures.

3.2. Multivariate decomposition methods

These methods are useful to resolve multivariate data by reducing their dimensionality into a number of correlated variables, called the latent variables (principal components).

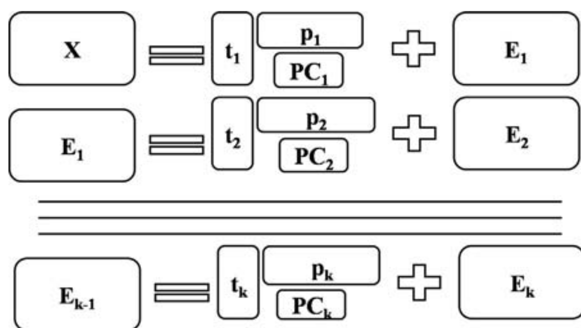

Fig. 1. Representing schematic description of data matrix X into various PCs.

These methods are applied by decomposing the data into lesser dimensions, i.e. unfolding of three-way data is likely to be slicing up the three-dimensional data cube into two-dimensional tables. And then further placing these tables side by side to each other creates a large two-dimensional data matrix. These methods are further explained in the following sections given below.

3.2.1. *Principal component analysis (PCA)*

The central idea of PCA is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, which are uncorrelated and ordered so that the first few components retain most of the variation in all the original variables [28–30].

PCA is based on eigenvector decomposition of the covariance matrix of the process variables. Here, we will use that rows of a data matrix X correspond to samples while columns correspond to variables [31,32]. For a data matrix X with m rows and n columns, the covariance matrix of X is defined as

$$COV(X) = (X^T X)/(m-1) \tag{28}$$

If the columns of data matrix X are made 'mean centered' and 'autoscaled' then Eq. (28) gives the correlation matrix of X. PCA decomposes the data matrix X as the sum of the outer product of vectors score t (i.e. contain information regarding the interaction of the samples to each other) and loading p (i.e. eigenvectors of the covariance matrix) plus a residual matrix E

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_k p_k^T + E \tag{29}$$

Here k must be less than or equal to the smaller dimension of X. Fig. 1 represents the decomposition of data matrix by using PCA, where X is decomposed into score and loadings up to the desired response.

$$COV(X) p_i = \lambda_i p_i \tag{30}$$

where λ_i is the eigenvalue associated with the eigenvector p.

Score vector t_i (orthogonal set) is the linear combination of the original X data which is defined by p_i (orthonormal set), calculated by the following equation:

$$X p_i = t_i \tag{31}$$

The λ_i in Eq. (30) describes the amount of variance present in the t_i p_i pair. The t_i p_i pairs are in descending order of λ_i and the first pair captures the largest amount of information of any pair in the process. It is also proved that the t_i, p_i pair captures the greatest amount of variation in the data that is possible to capture with a linear factor and then each next pair captures the greatest possible amount of variance remaining at this step. Further, PCA describes data by using much fewer variables compared to the original data [33].

It is also possible to calculate a lack of model fit statistic Q, for the ith sample in data matrix X, i.e. x_i:

$$Q = x_i (I - P_k P_k^T) x_i^T \tag{32}$$

where P_k is the matrix of the first k loadings vectors retained in the PCA model and I is the identity matrix of appropriate size (n by n). The Q statistic measures the amount of variation in each sample not captured by the k principal components retained in the model and also tells how well each sample conforms to the PCA model.

Hotelling's T^2 statistics measures the variation present within the PCA model. T^2 is the sum of normalized squared scores defined as

$$T^2 = x_i P \lambda^{-1} P^T x_i^T \tag{33}$$

The matrix λ^{-1} is a diagonal matrix containing the inverse eigenvalues associated with the k eigenvectors (principal components) retained in the model.

PCA based on singular value decomposition (SVD) [30,34] is explained using the matrix X and it can be written as

$$X = UDV^T + E \tag{34}$$

where U contains the same column vectors as does t (score), V^T is identical to P^T (loading) but normalized to length one and D is a diagonal matrix. These diagonal elements of D are the square roots of the eigenvalues of $X^T X$.

Once the PC model has been developed for a "training matrix" then it is fitted to the model, giving scores for the new objects or loadings for the new variables. The formula for a new object x is as follows:

$$t = xP \tag{35}$$

PCA based on non-linear iterative partial least squares (NIPALS) [34–36] algorithm is explained using matrix X which is scaled for each dimension.

Score vector t is selected from the column of matrix X with the largest variance and then the loading vector p^T is calculated using the following equation:

$$p^T = t^T X / t^T t \tag{36}$$

Now p is normalized to unit length by multiplying a constant factor c.

$$c = 1/\sqrt{p^T p} \tag{37}$$

Then the new score vector is calculated for the ith element using the following equation:

$$t = Xp/p^T p \tag{38}$$

Check the overlapping among different data using the sum of squared differences between all elements in two consecutive score vectors. If the data meets then continue with Eq. (38), otherwise find other score values with the second largest variance. If overlap has not been reached in the maximum number of iterations, i.e. 20, then there is no strongly preferred direction of maximum variance.

For calculating residual we use the following equation:

$$E = X - tp^T \tag{39}$$

PCA is widely applicable in herbal and biological mixtures and easily resolved data using latent variables, but it also employs the selection of a number of components which is human decision-based and thus leads to chances of error. Dimension reduction can only be achieved if the original variables were correlated. If the

original variables were uncorrelated, PCA does nothing, except for ordering them according to their variance.

3.2.2. Parallel factor analysis (PARAFAC)

Parallel factor analysis (PARAFAC) is a decomposition method for three-way arrays that can be seen as a simplification of bilinear PCA to higher order arrays. This multi-way method originates from psychometrics [37,38]. In this, data decomposition is made into tri-linear components. Each component consists of one score vector and two loading vectors instead of one score and one loading as in PCA.

The PARAFAC model for three-way array is given by three loading matrices, A, B and C, with elements a_{if}, b_{jf}, and c_{kf}. The PARAFAC model of a three-way array minimizes the sum of the squares of the residuals e_{ijk} for F components [39,40].

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{40}$$

For the fth column of loading matrix equation can be written as

$$X = \sum_{f=1}^{F} a_f b_f c_f \tag{41}$$

where a_f, b_f and c_f are the fth columns of the loading matrices A, B and C, respectively.

Alternating least squares (ALS) finds out the solution of the PARAFAC model sequentially assuming two modes of loading which are known and then estimates the parameters for the unknown set [41].

The additional advantage of PARAFAC over PCA is to eliminate the rational problem. By using the right number of components in PARAFAC where tri-linear data is needed, we can find the true spectra with appropriate signal-to-noise ratio. Hence, loadings obtained with PARAFAC can be interpreted directly [42]. Furthermore, in the PARAFAC algorithm all the components are calculated simultaneously which is not possible in PCA because it requires a number of steps. It is the most advanced method used for the resolution of three-dimensional data obtained from different hyphenated techniques and the method has wide applicability in herbal mixtures, impurity profiling, etc.

3.3. Hierarchical cluster analysis (HCA)

Cluster analysis is a multivariate technique that arranges components on the basis of their characteristics. It classifies components on the basis of their similarity in space. As a result, cluster exhibits high homogeneity in the intergroup and high heterogeneity among different groups [43–45].

Hierarchical agglomerative clustering provides good similarity correlation of data of one sample to the entire data of other samples. It is further expressed graphically as a picture of different groups and their proximity by high reduction in dimensionality of the original data. The HCA dendrogram is based on the assumption that the sample having similar values numerically is close to each other in space so that the sample with high similarity has high proximity. In cluster analysis similarity is measured on the basis of distance between different data points; so Euclidean distance (ED) and Mahalanobis distance (MD) are used to measure the distance between clusters.

For two clusters r and s having a number of components n_r and n_s, respectively, ED and MD are calculated as

$$ED_{rs} = \sqrt{(x_r - x_s)(x_r - x_s)^T} \tag{42}$$

$$MD_{rs} = \sqrt{(x_r - x_s)V^{-1}(x_r - x_s)^T} \tag{43}$$

where V is the covariance matrix.

The basic process of HCA is explained using N data sets to be clustered. Assign each data set as a cluster so there are N clusters and the distance between different clusters represents similarity between data sets. Then find the closest pairs of cluster and merge them into a single cluster and then finally compute the distance between new clusters that has to be formed by merging. Finally, clustering tree of size N is obtained. The analysis is not stable when cases are dropped; this occurs because the selection of a case (or merger of clusters) depends on the similarity of one case to the cluster. Dropping one case can drastically affect the course in which the analysis progresses.

3.4. Pattern recognition methods

3.4.1. Soft independent modeling of class analogy (SIMCA)

Soft independent modeling of class analogy (SIMCA) [46–48] is a supervised classification technique that uses PCA or PLS for classification by creating confidence region around each class using residuals of the samples in the calibration set. SIMCA is a pattern recognition technique in which new objects are projected as a member of a particular class based on their Euclidian distance from its particular principle component space. Euclidian distance does not exceed a confidence limit from a particular principle component. SIMCA can also be treated as a multivariate outlier test because it checks outliers in the space of the selected PCs. On the basis of residual variance of the objects in the training set, residual standard deviation (s_0) is calculated as

$$s_0 = \sqrt{\frac{\sum_{k=1}^{m}\sum_{i=1}^{n} e_{ik}^2}{(m-1-a)(m-a-1)}} \tag{44}$$

where e_{ik}^2 is the residual of object, k is the calibration set for variable i, m is the number of observations in the calibration set, n is the number of variables and a is the number of principal components. The number of degrees of freedom given in the equation is used in case when the number of observations is less than the number of variables, i.e. $m \le n$.

The residuals of the training set follow normal distribution and the F-test may be used to describe the critical value of Euclidean distances of the objects towards the model.

$$s_{crit} = \sqrt{F_{crit} s_0^2} \tag{45}$$

where F_{crit} is the tabulated value for the specific degree of freedom at a significant level. For a new object x_j^{new} that belongs to a certain class having residual vector e_{ji}^{new}, the residual standard deviation (s_j) is calculated as

$$s_j = \sqrt{\frac{\sum_{i=1}^{n} (e_{ji}^{new})^2}{(m-1-a)}} \tag{46}$$

By investigating whether the residual variance s_j is significantly different from the pooled residual variance s_0 of the model or not. This is done by calculating the F_j^{new} and comparing it with the tabulated critical value F_{crit}.

$$F_j^{new} = s_j^2 / s_0^2 \tag{47}$$

Mahalanobis distance (MD) proposed by Hawkins is used instead of Euclidean distance for multivariate outlier tests. Further descriptions of SIMCA can be found elsewhere. An attractive quality of SIMCA is that a principal component mapping of the data has occurred. Hence, samples that may be described by spectra are mapped onto a much lower dimensional subspace for arrangement. If a sample is comparable to the other samples in

the class, it will lie closer to them in the principal component map defined by the samples indicating that class [49,50].

3.4.2. Linear discriminate analysis (LDA)

LDA is a supervised pattern recognition method, which is used in cases where class variance is asymmetrical [51,52]. This method maximizes the ratio between both variances compared to the within-group variance. It searches for a linear function of the variable in multivariate space. When the number of variables is larger than the number of observations in multi-dimensional data, then we cannot use LDA directly. In that case first PCA is employed for data compression to transform the original data set comprising of large number of inter-correlated variables into a reduced new set of variables. LDA makes a discriminate function for each group according to the following equation:

$$F(G_i) = k_i + \sum_{j=1}^{n} w_{ij} p_{ij} \quad (48)$$

where i represents the number of groups (G), k_i is the constant of each group, n is the number of parameters used to classify a data set into a given group and w_j is the weight coefficient assigned by LDA to a selected parameter (p_i).

3.5. Chemometric resolution methods (CRMs)

The main goal in the analysis of any multi-component system is getting useful information from the raw experimental data and knowledge of the number of chemical components in the analyzed sample or certain important retention regions. It is an important analytical objective especially in herbal drugs. The common purpose of all CRMs is providing the linear model of individual component contributions using solely the raw experimental measurements [53–55]. CRMs decompose mathematically a mixed response from instrumental data into the pure contributions due to each of the components in the system. Further, this response is organized in the matrix X containing raw information about all the components present in the data set. Resolution methods allow the decomposition of data matrix X into the dot product of two data matrices C and S^T associated with the row and the column direction of the data matrix X, respectively, and each of them includes the pure response profiles of the n components. In matrix notation, the expression valid for all resolution methods is

$$X = C*S^T + E \quad (49)$$

where X $(r \times c)$ is the original data matrix, C $(r \times n)$ and S^T $(n \times c)$ are the matrices containing pure response profiles related to the data variation in the row and in the column direction, respectively, and E $(r \times c)$ is the residual variation of the data set that is not related to any chemical contribution.

Parameters r and c represent the number of rows and the number of columns of the original data matrix, respectively, and n is the number of chemical components. We will give a schematic representation in Fig. 2 on how the various MCR methods work. In resolution methods, there is no need of any previous knowledge of chemical or mathematical expression to analyze the data set. It makes these methods useful in routine analysis. However, initial information that can be obtained from instrumental analysis influences positively the resolution of the system. Thus, this information can be used to build good initial estimates of concentration profiles and responses.

In many CRMs, preliminary analysis derived from principal component analysis (PCA) is one of the most basic and widely used chemometric tools discussed above to find the number and direction of the relevant sources of variation in a bilinear data set. The information provided by PCA is essential in the resolution process and in many resolution methods to determine the total number of chemical components, which basically affect the ambiguity of the final solutions. Therefore, our main aim is to discuss the various important resolution methods which minimize the source of uncertainty in the analytical results and to make these methods familiar with regular analysis.

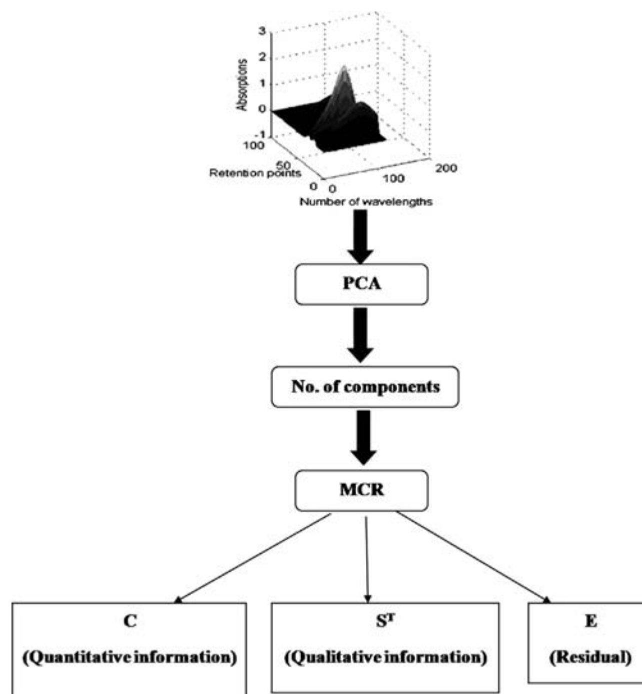

Fig. 2. Flow chart of the application of multivariate curve resolution methods. Examples of 3D chromatogram obtained from LC-DAD and by using PCA and various MCR tools resolute the pure spectral and chromatographic profile.

3.5.1. Rank annihilation factor analysis (RAFA)

Rank annihilation factor analysis (RAFA) [56,57] is the first method used for the decomposition of three-way data. RAFA is based on rank analysis for two-way bilinear matrix and based on the principle that the rank of pure analyte is one. It is used to analyze the given component quantitatively in the presence of other or possibly unknown two-way bilinear matrix. RAFA involves two bilinear data sets, first is the calibration set M_{std} and the second is the sample set M_u. By measuring a mixture that contains a known amount of analytes, a calibration set is prepared and then the analyte has to be quantified in the sample set which contains measurements of the sample of interest. After applying principal component analysis (PCA) on data, rank R_u and R_s for data matrix M_u and M_s, respectively, are obtained. Data matrix M_s has 1 component with rank R_s=1 and data matrix R_u is theoretically equal to $1+n_i$ (the number of interfering compounds). Then the rank of E residual is calculated as

$$E = M_u - \beta M_s \quad (50)$$

where β is equal to the ratio of the concentration of the kth component in the standard solution to its value in the mixture. Therefore, the rank of residual matrix E is $R_u - 1$, i.e. it only shows interfering compounds. Matrices M_u and M_s are written as

$$M_s = x_k y_k^T \quad (51)$$

$$M_u = \sum_{k=1}^{N} x_k y_k^T \quad (52)$$

where x_k is the column vector containing x_i and y_k is a column vector containing y_j for component k. The data matrix of samples is

composed of N components. The concentration of the analyte in the standard solution is c_s and in the sample is βc_s. To find out β an iterative procedure is used which plots the eigenvalues. Ideally eigenvalues show the number of components N but in real situations the exact number of components is not known for sure. It becomes minimal when β exactly compensates for the signal of the analyte in the sample.

The concentration c_k of the kth constituent is now computed as

$$c_k = c_k^0/\beta \tag{53}$$

3.5.2. Generalized rank annihilation method (GRAM)

GRAM [58,59] is a non-iterative method from which simultaneous quantification of all the analytes present in the sample can be carried out using one bilinear matrix from a mixture of standards (calibration set), i.e., one standard for each analyte. It can determine the relative concentration of each analyte in a sample (unknown mixture) from which one can easily determine the actual concentration. GRAM involves two bilinear data sets: first is a calibration set M_{std} containing spectra at each retention time of analytes in the standard mixture and the second is a sample set M_u containing spectra at each retention time of analytes in the unknown sample. Both matrices M_{std} and M_u are written as

$$M_{std} = XC_oY^T \tag{54}$$

$$M_u = XC_uY^T \tag{55}$$

where X (J_1*K) and Y (J_2*K)matrices in which J_1 relate to the normalized chromatographic profile, J_2 contains spectral data and K represents the total number of components. C_o and C_u are diagonal matrices representing the concentration of analytes. The next step is to perform singular value decomposition (SVD) to calculate the number of significant factors, i.e. the total number of analytes

$$M = UDV^T \tag{56}$$

where U contains the same column vectors as does J_1, V^T is identical to J_2 but normalized to length 1 and D is a diagonal matrix. Using eigenvalue-eigenvector equation for obtaining the eigenvalue matrix

$$(S^{-1}VM_uU^T)Z = Z\beta \tag{57}$$

Here Z is the eigenvector matrix and β is the matrix of eigenvalues. After this, chromatographic profile and spectral profile are calculated.

$$X = U\beta Z \tag{58}$$

$$Y = V(Z^{-1})^T \tag{59}$$

The concentration c_k of the kth constituent is now computed as

$$c_k = c_k^0/\beta_k \tag{60}$$

Here β_k is the eigenvalue of the kth analyte, i.e., analyte of interest. The assignment of eigenvalue correspondence to analyte of interest is done by the hit and trail method. This can be done by calculating the correlation coefficient between the spectrum of analyte in standard and sample. Finally, the eigenvalue is calculated that is associated with the spectrum of the highest correlation.

3.5.3. Evolving factor analysis (EFA)

EFA [60,61] is a non-iterative method based on time-dependent rank analysis i.e., each row is associated with an increase in rank by one. However, this method is based on the first in first out assumption, i.e. the compound that started to elute first will disappear first. EFA includes evolving PCA in two directions along the retention point R_t, namely forward and backward. Eigenvalue from forward PCA shows the retention points where chemical components begin to appear while backward PCA indicates the retention points where chemical components begin to disappear; by combining information from these data we find out the elution sequence of each of the components by this method.

The collected information is in the form of two different regions, i.e. concentration region and zero concentration region. Concentration window, i.e. the region where each compound exists, is defined by performing process in both forward and backward directions, i.e. starting from the first spectrum to the last one and the last one to the first, respectively. Zero concentration windows are the regions where nothing elutes out from a particular component.

This method can be understood by considering data matrix X having n number of chemical components containing true concentration profile C which is a linear combination of abstract chromatogram U and spectral profile S which is a linear combination of abstract spectrum V. Both U and V can be determined by performing singular value decomposition (SVD).

$$X = UDV^T + E \tag{61}$$

where U contains the same column vectors as does t (score), V^T is identical to P^T (loading) but normalized to length one and D is a diagonal matrix. These diagonal elements of D are the square roots of the eigenvalues of X^TX.

After this, we have to find out the transformation matrix T using Eq. (62); it is used to determine the individual concentration profile

$$C = UT \tag{62}$$

Then the specific component concentration profile c_i is determined by the following equation:

$$c_i = U{*}t_i (i = 1, 2, \ldots, n) \tag{63}$$

Further, this procedure is repeated to determine n components, and at the end, the concentration profile is used to resolve the spectral profile.

$$S = (C^TC)^{-1}C^TX \tag{64}$$

By taking advantage of the evolutionary feature of chromatographic separation, this method may become the first method used in a sophisticated manner to obtain the elution sequence and it is also used as a main step for many self-modeling curve resolution (SMCR) methods till now. As in EFA one can analyze the progressively increasing sub-matrix; Keller and Massart [62] suggested a method in which a window of fixed size is used and moves along the retention points to perform PCA, called the fixed-size moving window evolving factor analysis (FSMWEFA).

Eigen structure tracking analysis (ETA), which was proposed by Toft and Kvalheim [63,64], is very similar to FSMWEFA. In this, it starts with a window of size 2 and the size is progressively increased by 1 up to the size when the window exceeds the number of components in the examined window by one. Local noise can be determined from the last evolving eigenvalue. ETA can reveal the noise pattern as well as the elution sequence.

3.5.4. Heuristic evolving latent projections (HELP)

HELP [65,66] is a method which is used to resolve two-way bilinear multi-component data into spectra and chromatograms of pure constituents. This method has a feature visual interface from

latent variable projection graph and also provides information on the rank of the data matrix. It employs the formation of eigenvalue plot which is plotted between logarithms of eigenvalues against retention time. In this method one can systematically calculate the eigenvalue for m sub-matrices containing spectra collected at retention points 1, 1 to 2, 1 to 3,…,1 to $m-1$, 1 to m, where m is the total number of retention points. Then all the eigenvalues obtained are joined by a line. In the next step, zero concentration regions and selective concentration region of the compound are determined from the eigenvalue plot. Selective concentration region is defined as the region in which only a single component elutes out, i.e. having rank equal to one.

This method can be understood by considering data matrix X having n number of chemical components containing true concentration profile C which is a linear combination of abstract chromatogram U and spectral profile S which is a linear combination of abstract spectrum V. Both U and V can be determined by performing singular value decomposition (SVD).

$$X = UDV^T + E \tag{65}$$

where U contains the same column vectors as does t (score), V^T is identical to P^T (loading) but normalized to length one and D is a diagonal matrix. These diagonal elements of D are the square roots of the eigenvalues of X^TX.

Then we have to find out the transformation matrix T which is used to determine the individual concentration profile

$$C = UT \tag{66}$$

In order to determine t_i for individual component i, both zero concentration region and selective concentration region are incorporated in Eq. (55):

$$c_{sel+zero,i} = U_{sel+zero,i} * t_i (i = 1, 2, \dots, n) \tag{67}$$

where subscripts $sel+zero$, i imply the use of the selective region plus the zero concentration region for component i. Then t_i can be written as

$$t_i = (U^T{}_{sel+zero,i} * U_{sel+zero,i})^{-1} U^T{}_{sel+zero,i} * c_{sel+zero,i} \tag{68}$$

Evolving latent projection (ELP) graph gives us information about the selective region of the individual component in wavelength as well as retention time space. Such a region can be identified by passing straight lines in ELP graphs. After determining t_i, the next step is to determine the elution pattern of a particular component by Eq. (66). This procedure is repeated to determine n components and at the end the concentration profile is used to resolve the spectral profile using Eq. (64).

3.5.5. Subwindow factor analysis (SFA)

Subwindow factor analysis (SFA) [67] is a method which is introduced directly as a solution to extract component spectra from overlapping structures obtained from a hyphenated instrument without resolving the concentration profiles. It is based on the theorem [68] "If the concentration window of the analyte for every interferents has a subwindow where the interferent is absent. Now it is possible to calculate the spectrum of the analyte". Further, one has to select subwindows ≥ 2 where only spectral component corresponding to particular analyte exists and it is followed by identification of that component.

SFA is employed to determine a selective region, i.e. subwindow in which analyte appears without interferents. It involves determining three more subwindows, namely left subwindow, right subwindow and middle subwindow. Left subwindow is the one in which only left interferents are present, i.e. interfering compound which begins to elute before the selective analyte and appears in a chromatogram to the left of the selective analyte. In the same way the right subwindow is one in which only the right interferents are present, i.e. interfering compound which continues to subsequently elute with selective analyte and appears in a chromatogram to the right of the selective analyte. A subwindow in which both interferents are present is called a middle subwindow. The first step in the SFA is the recognition of subwindows which is similar to other window-based methods and is done by any one of the methods discussed in this article (EFA, HELP, WFA, OPA). The difference of SFA from other methods is the combination of elution limits in windows or subwindows.

This method is discussed by considering data matrix X in which rank analysis is performed that gives the number of chemical components of the left and right subwindows, say m and n, respectively. Then the total number of chemical components in both subwindows is $m+n-1$ because the analyte is common in both. Now the procedure is used to determine the vectors corresponding to the largest singular values which represent the chemical information related to singular-value decomposition. Further, common vector v corresponding to analyte in both subspaces can be determined using the following equation:

$$v = Ma = Nb \tag{69}$$

where M and N are matrices with m and n columns, respectively, but in practice Ma is not equal to Nb and one has to find out a and b which minimize the squared norm E

$$E = \|Ma - Nb\|^2 = a^T M^T Ma + b^T N^T Nb - 2a^T M^T Nb \tag{70}$$

where $a^Ta = b^Tb = 1$ and M^TM and N^TN are identity matrices. Then we can write

$$E = 2(1 - a^T M^T Nb) \tag{71}$$

If a and b are left and right singular values associated with the largest singular value d_i of data matrix M^TN then we can write

$$E = 2(1 - d_1) \tag{72}$$

The singular value of d lies in the range 0–1. By determining vector v one can ensure that only this vector is common in both regions or if there is no common vector present then d_1 will be significantly less than 1. Similarly if there is more than 1 common vector then d_2 will be close to 1.

Sometimes it is also possible that the rank is less than the number of components present, i.e. subwindow limit is very narrow. In such type of cases we have to select the outer limits of both windows wider than that suggested by rank analysis and change the inner limits till good fit is obtained.

3.5.6. Window factor analysis (WFA)

WFA [69,70] is a self-modeling technique developed for the resolution of multivariate data matrix obtained from evolutionary process. In this method we find that chemical components have single unique maxima in the evolutionary process by specifying the window, i.e., the region which is indigenous to a chemical component along the evolutionary axis. After this, the concentration profile of the component is directly obtained without recourse to any information concerning the other components.

Let X represent a data matrix where each column is a spectrum recorded at different retention points during an evolutionary process. Window factor analysis (WFA) assumes that the total spectra S consist of a linear sum of abstract spectrum s. If there are n components, then

$$X = \sum_{i=1}^{n} X_i = \sum_{i=1}^{n} s_i c_i^T \tag{73}$$

where s_i and c_i are vectors representing the spectra and concentration profile of component i, respectively, and X_i is a matrix representing the contribution of i to the data matrix.

Using SVD, data matrix X can be decomposed into a product of spectral profile S which is a linear combination of abstract spectrum V and concentration profile C which is a linear combination of abstract chromatogram U which are multiplied together to reproduce the data within experimental uncertainty:

$$X = UDV^T \tag{74}$$

where D is a diagonal matrix. These diagonal elements of D are the square roots of the eigenvalues of X^TX.

Suppose X_0 is a sub-matrix of X and is obtained by removing all data columns containing signals of component n. If all other components except for component n exist outside the window then SVD of matrix X_0 will yield a matrix S_o containing $n-1$ orthonormal spectral vectors V_j^0 and a matrix C_0 containing $n-1$ concentration profile vectors U_j^o that are mutually orthogonal, so that

$$X_O^J = U_j^o D V_j^{0T} \tag{75}$$

A matrix X_j with columns proportional to the real concentration of the jth compound can be calculated as follows [68–70]:

$$X_j = (I - V^{0T}V^0)X = b{*}c_j{*}s_j^{0T} \tag{76}$$

where I is the identity matrix of dimension n, b is constant, and c_j and s_j are concentration and spectral profile of jth component, respectively. X_j represents an average concentration profile of jth component and once the concentration profile for the component is determined then the true spectral profile with least square is determined.

3.5.7. Alternating least square (ALS)

Multivariate curve resolution (MCR) supported by the alternating least squares (ALS) [24,71–73] optimization algorithm has been revealed as a powerful tool for resolving two- and three-way data matrices. This method is easily adapted for tri-linear data sets of different complexities and provides the least square solutions. This method is used to improve the initial estimation of either spectral or concentration profile. This method is explained by using data matrix X which are obtained by analyzing a mixture of different concentrations C and X_i for the ith component:

$$X_i = C_i S^T + E_i \quad i = 1, 2, 3, \ldots, i \tag{77}$$

If the total number of rows and columns of spectral profile are the same for all the samples, then all the samples are estimated simultaneously.

$$X = \begin{matrix} X_1 \\ X_2 \\ - \\ - \\ X_i \end{matrix} = \begin{matrix} C_1 \\ C_2 \\ - \\ - \\ C_i \end{matrix} S^T + E \tag{78}$$

At starting point, first we have to determine the number for co-eluted components in a particular peak cluster by principal component analysis (PCA). Afterwards one has to determine the initial estimation of spectral or concentration profile of analytes by one of the best suited methods discussed above in this article (EFA, WFA, SFA, and OPA) before ALS constrained optimization starts. In ALS, 4 constraints – named as non-negativity [54,74], unimodality [75,76], selectivity [65,77] and normalization – have been applied. These constraints are applied to avoid the presence of rotational and intensity uncertainties.

Non-negativity constraint is applied to both concentration and spectral profiles because chemical concentrations and spectra are only defined to be positive or zero. Unimodality is a constraint which is frequently applied to force chromatographic elution profiles of matrix X into a single peak shape. This constraint is applied in the case of co-eluted components that have similar spectra. Then it avoids the appearance of elution profiles with double peaks. Selectivity constraint is applied either to concentration or spectral profiles. Finally, a normalization constraint is applied to fix scale indeterminacy during ALS resolution.

In this way, all the components in a particular chromatographic run are assumed to have the same relative signal contribution in the spectral domain. All the differences in concentration of the components in the mixtures will be expressed in the area of elution profiles. Outcomes from ALS optimization are the estimations of C, S^T and E matrices. This gives us MCR-ALS resolved elution profiles, pure species spectra and residuals, respectively, fitted by the constrained ALS optimization procedure.

3.5.8. Orthogonal projection analysis (OPA)

OPA [78,79] is a stepwise process and selects one key variable in each step. This method calculates dissimilarity based on the mathematical concept of orthogonalization [80,81]. The method is based on the fact that pure spectra are extreme spectra and will cover the mixture of spectra. It compares each spectrum with one or more than one reference spectra and searches for least correlation. The first dissimilar plot represents a comparison of each spectrum with the average spectrum. The principle of OPA is as follows: the instrument produces a data matrix, X (mn), where the m rows are spectra measured at regular time intervals and the n columns are chromatograms measured at different wavenumbers. The data matrix X is bilinear, i.e., it can be decomposed into the product of the individual concentration matrix (C) and the pure compound spectra matrix (S).

The dissimilarity of the ith spectrum, d_i, is defined as the determinant of the dispersion matrix of Y_i. In general, matrices Y_i consist of one or more reference spectra and the spectrum measured at the ith analysis time.

$$d_i = \det(Y_i^T {*} Y_i) \quad i = 1, 2, \ldots, m \tag{79}$$

A dissimilarity plot is then obtained by plotting the dissimilarity values against the analysis time. The spectrum having the highest dissimilarity value is the least correlated with the mean spectrum, and is then selected x_{s1}. Then the dissimilarity of each individual spectrum of X with respect to x_{s1} is calculated. As before, Y_i consists of one reference spectrum which is the mean spectrum of matrix X and each x_i. Then, the mean spectrum is substituted by x_{s1}, which is the most dissimilar spectrum in the dispersion matrix Y_i of dimensions $(m(n{*}2))$ and the spectrum most dissimilar with x_{s1} is selected as x_{s2} and a second dissimilarity plot is obtained by applying Eq. (79). The spectrum most dissimilar with x_{s1} is selected as (x_{s2}) and added to matrix Yi (m ($n{*}3$)). Then this process continues until each spectrum contained in X is compared with the spectra already selected and includes the most dissimilar one into matrix Y_i. Then each step is plotted and visualized and the random profile indicates the number of spectra equal to the number of components.

In short, we can say that the procedure consists of three steps: (1) comparison of each spectrum of X with all spectra already selected by applying Eq. (79), (2) plotting of the dissimilarity plot and (3) selection of the spectrum with the highest dissimilarity value by including it as reference in matrix Y_i.

Then, find out concentration by least-squares procedure using the following equation:

$$C = X{*}S{*}(S^T S)^{-1} \quad (80)$$

where S represents the spectral matrix. Concentration profile is obtained from least-squares by applying constrain of non-negativity and unimodality. A new set of spectra (matrix S) is obtained by the following equation:

$$S = X^T{*}C{*}(C^T C)^{-1} \quad (81)$$

Then, the sum of square of residual SSR is calculated:

$$R = X - C{*}S^T \quad (82)$$

$$SSR = \sum_{i=1}^{m} \sum_{j=1}^{n} r_{ij}^2 \quad (83)$$

where r_{ij} is the difference between the measured and the predicted absorbance values at time i and wavelength j. This process continues until the relative difference in the SSR between two successive iterations is lower than a pre-defined threshold.

4. Application of chemometric tools in analytical chemistry

Analytical application of chemometric technique is given in Table 1, which shows use of data from multivariate techniques generated with the aid of analytical techniques. This data is a compilation of the literature available in Scopus [82] and various research works were carried out using chemometric tools. This comprehensive review focuses on the application of various chemometric techniques from its advent to the present era. In this article we have tried to incorporate maximum research work dealing with multivariate data analysis in the field of analytical chemistry (i.e. plant and herbal medicines and mixtures/miscellaneous compounds) and classified

Table 1
Chemometric methods used in analytical techniques classified on the basis of herbal medicines and mixtures of miscellaneous compounds.

Analytical technique	Chemometric method	Plant and herbal medicine	Mixtures and miscellaneous compounds
UV-spectrometry	TLRC		[83]
	MLRC		[83]
	CLS		[83,84,86–88,92,96,99,100,102–104]
	ILS		[19,84–88,99,102–104]
	PCR		[19,84–86,88,89,92–96,99–104,106]
	PLSR		[84,88–106,145]
HPLC-DAD	PCA	[107–110,112,113,115,146–151]	[120]
	CLS, PLS, PCR		[154]
	PLS	[107,112]	
	PARAFAC	[136,152]	[119,135,155–157]
	HCA	[108,110,125]	[147]
	SIMCA	[112,129]	
	RAFA		[123,124,126,127,158]
	GRAM		[121,122,135,159–161]
	EFA		[138,151]
	FSMW-EFA		[130]
	WFA		[138]
	OPA		[130,138]
	MCR-ALS	[136]	[120,128,130,131,135,155,162]
	PCA-LLS	[153]	
LC–MS	PARAFAC		[155]
	MCR-ALS		[132,133]
LC–NMR	EFA, SFA, ITTFA, MCR-ALS, OPA, AUTOWFA		[139]
	PCA	[163]	
UPLC-DAD	PCA, HCA	[111,134]	
UPLC–MS	PCA		[164]
GC–MS	PCA	[117]	[165]
	GRAM		[141]
	PARAFAC		[140,165]
GC–GC–MS	PARAFAC		[137]
	PLS-DA		[166]
HS-SPME–GC–MS	PCA	[114]	
NIR	PLS, PCA		[118]
	SIMCA		[50]
FT-IR	PLS-DA		[167,168]
	PCA		[116,169]
Raman spectroscopy	PCA, SIMCA		[170]
TLC-FD	MCR-ALS		[171]
Spectrofluorimetry	PLS		[172]
	PARAFAC		[173]

CLS: classical least square; EFA: evolving factor analysis; FSMW-EFA: fixed size moving window evolving factor analysis; FT-IR: Fourier transform infrared; GC–MS: gas chromatography–mass spectrometry detection; GC × GC–MS: two-dimensional gas chromatography–mass spectrometry detection; GRAM: generalized rank annihilation method; HCA: hierarchical cluster analysis; HPLC-DAD: high performance liquid chromatography, with diode array detection; HS-SPME–GC–MS: head space solid-phase microextraction coupled with gas chromatography mass spectrometry; ILS: inverse least square; ITTFA: iterative target transformation factor analysis; LC–MS: liquid chromatography mass spectrometry; LC–NMR: liquid chromatography nuclear magnetic resonance; LLS: local least square; MCR-ALS: multivariate curve resolution by alternating the least squares approach; MLRC: multi-linear regression-calibration; NIR: near infrared; PCA: principle component analysis; PCR: principle component regression; PLSR: partial least square regression; PARAFAC: parallel factor analysis; PCA: principal components analysis; PLS: partial least square; PLS-DA: partial least square discriminate analysis; OPA: orthogonal projection analysis; RAFA: rank annihilation factor analysis; SFA: sub-window factor analysis; SIMCA: soft independent modeling of class analogy; TLC-FD: thin layer chromatography with fluorescence detection; TLRC: tri-linear regression-calibration; UPLC-DAD: ultra-performance liquid chromatography with diode array detection; UPLC–MS; ultra-performance liquid chromatography with mass spectrometry; WFA: window factor analysis.

them according to the analytical and chemometric techniques used. This section describes them on the basis of obtained analytical results.

TLRC and MLRC methods are used for the multi-resolution of a ternary mixture [83]. These methods have been proved as useful with very simple mathematical contents for multi-resolution of the three-component mixture systems and show accuracy in the range 99–101% for a mixture of highly overlapping spectra. Various approaches like CLS, ILS, PCR and PLSR have been used for the resolution of multi-component mixture using a UV spectrophotometer and their results are comparable with the HPLC methods and it was found that the accuracy is in the range of 98–103% [19,84–106]. A six-component mixture is resolved by PLS and PCR and the obtained results show good agreement with the HPLC method [95].

Two-wavelength HPLC fingerprinting was applied to the quality assessment for 46 Cassia seed samples [107]. Based on this method, the roasted COS, raw COS, roasted CTS, and raw CTS samples were discriminated by PCA calculation. Moreover, the PLS prediction models produced satisfactory results on the test set.10 triterpenoid acids were simultaneously determined by the HPLC-DAD method in fruits of *Ziziphus jujuba* [108] and the method was successfully applied to 42 samples. HPLC fingerprint analysis with chemometric methods was used for the purpose of species differentiation, quality evaluation and consistency check of *Radix Paeoniae* collected from different sources [109]. The chemometric methods including HCA and PCA proved satisfactory for matching and discrimination of *Artemisia selengensis* [110] and *Rhizoma Coptidis* samples [111]. HCA, PCA, PLS-DA and SIMCA were able to classify samples of *Ganoderma lucidum* successfully in accordance with the province of origin [112]. HPLC-PCA assay is used to differentiate e*notoginseng* root extract from the extract of other plant parts of *notoginseng* and also from the extract of Asian or American ginseng plant parts [113].

HS-SPME/GC–MS analysis of the volatile profile of Balsamic Vinegar of Modena coupled with statistical data analysis by means of PCA have shown this method to be useful to discriminate Balsamic Vinegar of Modena samples with different maturation and aging characteristics [114]. *Cnidiummonnieri* fruits obtained from different regions of China have been clustered reasonably into different groups based on the coumarin content by principal component analysis (PCA) and cluster analysis [115]. Adulteration of cod liver oil with selected vegetable oils (CaO, CO, SO, and WO) can be monitored with FTIR spectroscopy and by using PLS, normal spectra can successfully be used to detect the level of oil adulterants [116]. *Scutellariabarbata* from different origins was studied by GC–MS and the results were evaluated by principal component analysis and it was found that this method was useful for the discrimination of its adulterants [117]. NIR spectroscopy combined with multivariate analysis was used for the simultaneous quantification of α-phosphatidylcoline and cholesterol in liposome [118]. PARAFAC was used as a technique for resolving partly separated peaks of lidocaine and prilocaine into their pure chromatographic, spectral and concentration profiles [119]. HPLC-DAD combined with MCR-ALS is used for impurity profiling in order to obtain a good estimate of the content or relative response factors of small chromatographic impurity peaks without knowledge of their molar absorption coefficients and without any pre-calibration [120]. GRAM was proved to be very useful for quantifying contaminants in complex marine samples [121]. GRAM was used to quantify aromatic sulfonates in water with HPLC when interfering components co-eluted with the analytes of interest. Concentrations of analyte of interest were determined more quickly because a complete resolution is not required [122]. The RAFA procedure was used to estimate the model parameters in a complex gray chemical system when there was not enough information about the whole process [123]. Mean centering of ratio spectra using the spectrum of absorbing reagent as divisor by combining with RAFA is used to detect the contribution of one chemical component in an unknown sample [124]. The HPLC-DAD method was developed to evaluate the quality of *Receptaculum Nelumbinis* (dried receptacle of Nelumbonucifera) through establishing chromatographic fingerprint and simultaneous determination of five flavonol glycosides, including hyperoside, isoquercitrin, quercetin-3-O-β-D-glucuronide, isorhamnetin-3-O-β-D-galactoside and syringetin-3-O-β-D-glucoside [125]. RAFA was used for quantitative analyses of multi-component fluorescence data as acquired in the form of an excitation-emission matrix (EEM) by the video fluorimeteris demonstrated using 10 different samples of a six-component poly-nuclear aromatic hydrocarbon solution whose constituents have a wide range of fluorescence quantum efficiencies and spectral overlaps [126]. Simultaneous multi-component rank annihilation method was applied to a set of six-component poly-nuclear aromatic hydrocarbon [127]. Liquid – solid extraction coupled to LC-DAD was applied for quantization of co-eluted organophosphorus pesticides: fenitrothion, azinphos-ethyl, diazinon, fenthion and parathion-ethyl and data evaluation by Multivariate self-modeling curve resolution and this shows improved resolution of the co-eluted organophosphorus insecticides and their quantization at trace level [128]. PCA, SIMCA and HCA were applied on HPLC fingerprint of *Epimediumwushanenseas* in order to identify and distinguish their secondary metabolites [129].

HPLC-DAD was used for the analysis of a tetracycline hydrochloride sample and the data obtained is resolved by OPA and FSMW-EFA into pure spectra and individual concentration profiles. This method was successfully applied for the detection of impurities in the sample [130]. Use of MCR-ALS is evaluated in the analysis of complex biocide environmental sample mixtures by LC-DAD and Multivariate Curve Resolution has been shown to be a powerful Chemometrics tool to solve the lack of chromatographic resolution and strong co-elution problems encountered in the analysis of complex biocide mixtures in environmental samples [131]. Resolution and quantitative determination of a mixture of co-eluted pesticides (carbofuran, propoxur and pirimicarb) in LC–MS by MCR are shown. Carbofuran and pirimicarb both exhibit similar mass spectrum and co-eluting is distinguished by the proposed MCR method [132]. Mixtures of multiple biocide compounds were simultaneously analyzed in standard mixtures and in environmental samples with little sample pretreatment using LC-DAD; all biocide compounds were properly resolved by MCR-ALS and quantitatively analyzed with the estimated errors always below 20% [133]. Ultra-performance liquid chromatography–quadrupole time-of-flight mass spectrometry (UPLC-QTOF/MS) and multivariate statistical analysis were used to investigate the processing technology of Loquat (*Eriobotrya japonica*) leaf (pipaye, PPY). The differences in samples processed under different methods were revealed by unsupervised principal component analysis (PCA) [134].

GRAM, PARAFAC and MCR-ALS were able to quantify overlapped and highly drifted chromatographic profiles. Such profiles can be found in the determination of compounds at very low concentrations in natural samples [135]. MCR-ALS is evaluated in the analysis of nine phenolic acids, both in standards mixture samples and in strawberry juice samples, by LC-DAD [136].

In [137], 100% of 12 selected trimethylsilyl derivatized organic acid metabolites in human infant urine were located with the Dot Map algorithm. Vanillic acid (TMS) was located by Dot Map, but also exhibited overlap with other organic acids. The presence of vanillic acid (TMS) was confirmed by PARAFAC to yield pure component information suitable for subsequent quantification. In [138], OPA is used for the determination of the number of compounds present in a multi-component system. The performance of the OPA algorithm is compared with that of two window-based self-modeling curve

resolution approaches: EFA and WFA. The results obtained with the EFA method are slightly better than those with OPA.

Multivariate curve resolution methods are classified and the need to check the applicability of various curve resolution methods to data obtained with different types of instruments is highlighted [139]. It is demonstrated that LC–NMR data can be resolved if NMR peak cluster information is utilized. PARAFAC is proposed in [140], for the alignment of LC–MS data in the chromatographic direction, with moderate shifts present in the data. In [141] a semi-automated approach is proposed for the resolution and quantification of unresolved target-analyte with GC–MS. In this paper two methods were utilized to correct retention-time shifts after the GRAM method was applied. When an environmental sample of 1,4-dithiacyclohexane was analyzed, the GRAM calculated concentration was evaluated with an error of about 10%.

From Table 1 we concluded that applicability of chemometrical approaches is increased rapidly. There are thousands of research papers across these fields using chemometric-like data processing methods during the period 2008–2013. There has been an increased use of multivariate methods across broad ranges of scientific disciplines, and these tools have become rather standard training for most graduate-level scientific disciplines.

In the past, one had to code these algorithms in order to use them; however, this has now become basic because of more powerful computer software packages that are available today which provide the option of a standard form of the algorithm. Multivariate analysis methods are now standard for basic and applied research.

The application of hyphenated techniques with the aid of chemometrics for fingerprint analysis can be very useful and the data collected are analyzed in the form of two-dimensional matrices. The results obtained from matrices allow a better discrimination of the samples depending on the two variables measured. Chemometrics also allows the utilization of more than one analytical method in order to find the correlation between different variables. Chemometrics is an efficient and powerful tool for the quality control of different herbs and plants [142,143]. In [144] Lavine describes the importance of chemometrics in regular analysis and also cited various articles published during 2008–2009.

5. Conclusion

To evaluate the fingerprint of complex herbal products, hyphenation with chromatography offers a powerful tool for separating the individual components. After the analysis is accomplished then for extracting useful information that resides within the generated data, various data-handling methods are used. Therefore, we stressed upon various multivariate methods that are used to extract the information in the data. First various multivariate regression methods are discussed that are used to explain the variation in data set by latent variables. These methods are very useful and even necessary in all branches of chemistry and the applicability of chemometrics to very complicated problems makes the life of a chemometrician very challenging and they seem to work at the border of impossible. Then various multivariate resolution methods are explained which provide a linear model for individual component contributions using only the raw experimental measurements.

The advancement of multivariate methods is continuous, fast and efficient. With the improvements in exploratory tools, adaptability for analysis of complex data, quality control of herbal drugs and assessment of the results, it can be envisioned to be an important increase in the range of application of these methods and a more generalized and standardized use of multivariate methods by the analytical community.

Acknowledgments

We are grateful to Shri Parveen Garg (Chairman), I.S.F. College of Pharmacy, Ferozepur Road, Moga 142001, India, for the necessary support and motivation. The authors gratefully acknowledge PSCST of Punjab (project PSO/POS/80/9334) for the financial support and ISFCP, Moga for supporting this work.

References

[1] M. Ortiz, L. Sarabia, J. Chromatogr. A 1158 (2007) 94–110.
[2] L. Xu, L.J. Tang, C.B. Cai, H.L. Wu, G.L. Shen, R.Q. Yu, J.H. Jiang, Anal. Chim. Acta 613 (2008) 121–134.
[3] A. Siouffi, R. Phan-Tan-Luu, J. Chromatogr. A 892 (2000) 75–106.
[4] P. Vanbel, J. Pharm. Biomed. Anal. 21 (1999) 603–610.
[5] Y.Z. Liang, P. Xie, K.C. Chan, J. Chromatogr. B 812 (2004) 53–70.
[6] R. Kiralj, M. Ferreira, J. Chemometr. 20 (2006) 247–272.
[7] D. Jing, H. Linfang, J. Med. Plants Res. 5 (2006) 4001–4008.
[8] L.S. Ramos, K.R. Beebe, W.P. Carey, E. Sanchez, M. Erickson, B.E. Wilson, L.E. Wangen, B.R. Kowalski, Anal. Chem. 58 (1986) 294–315.
[9] P.K. Hopke, Anal. Chim. Acta 500 (2003) 365–377.
[10] E. Dinç, J. Pharm. Biomed. Anal. 33 (2003) 605–615.
[11] K. Wiberg, Stockholm, Multivariate Spectroscopic Methods for the Analysis of Solutions, Kent Wiberg, Sweden, 2004.
[12] E. Dinç, A. Özdemir, Chem. Pharm. Bull. 53 (2005) 899–906.
[13] E. Dinc, A. Ozdemi, Die Pharmazie-Int. J. Pharm. Sci. 59 (2004) 700–705.
[14] F.G.K. Baucke, R. Naumann, C. Alexander-Weber, Anal. Chem. 65 (1993) 3244–3251.
[15] P. Geladi, Spectrochim. Acta B 58 (2003) 767–782.
[16] S. Frans, J. Harris, Anal. Chem. 57 (1985) 2680–2684.
[17] E.V. Thomas, Anal. Chem. 66 (1994) 795–804.
[18] P. Geladi, B.R. Kowalski, Anal. Chim. Acta 185 (1986) 1–17.
[19] E. Dinc, E. Baydan, M. Kanbur, F. Onur, Talanta 58 (2002) 579–594.
[20] G. Musumarra, M. Fichera, Chemometr. Intell. Lab. 44 (1998) 363–372.
[21] S. Wold, M. Sjöström, L. Eriksson, Chemometr. Intell. Lab. 58 (2001) 109–130.
[22] D.M. Haaland, E.V. Thomas, Anal. Chem. 60 (1988) 1193–1202.
[23] H. Swierenga, P. De Groot, A. De Weijer, M. Derksen, L. Buydens, Chemometr. Intell. Lab. 41 (1998) 237–248.
[24] R. Manne, Chemometr. Intell. Lab. 2 (1987) 187–197.
[25] U. Indahl, J Chemometr. 19 (2005) 32–44.
[26] D.S. Cao, Q.S. Xu, Y.Z. Liang, X. Chen, H.D. Li, J Chemometr. 24 (2010) 584–595.
[27] S. García-Muñoz, M. Polizzi, Chemometr. Intell. Lab. 114 (2012) 116–121.
[28] S. Wold, K. Esbensen, P. Geladi, Chemometr. Intell. Lab. 2 (1987) 37–52.
[29] R. Tauler, E. Casassas, J Chemometr. 3 (1989) 151–161.
[30] J. Camacho, J. Picó, A. Ferrer, Chemometr. Intell. Lab. 100 (2010) 48–56.
[31] B.M. Wise, N.B. Gallagher, J. Proc. Control 6 (1996) 329–348.
[32] A. Duarte, S. Capelo, J. Liq. Chromatogr. Relat. Technol. 29 (2006) 1143–1176.
[33] U. Depczynski, V. FrostK. Molt, Anal. Chim. Acta 420 (2000) 217–227.
[34] J. Mandel, Am. Stat. 36 (1982) 15–24.
[35] S. Wold, M. Sjöström, Chemometr. Intell. Lab. 44 (1998) 3–14.
[36] I. Jolliffe, Principal Component Analysis, Wiley Online Library, U.K., 2005.
[37] R. Bro, Chemometr. Intell. Lab. 38 (1997) 149–171.
[38] B.D. Marx, P.H.C. Eilers, B. Li, Chemometr. Intell. Lab. 109 (2011) 120–130.
[39] J.D. Carroll, J.J. Chang, Psychometrika 35 (1970) 283–319.
[40] J.B. Kruskal, Linear Algebra Appl. 18 (1977) 95–138.
[41] J.M.F. ten Berge, Least Squares Optimization in Multivariate Analysis, DSWO Press, Leiden University Leiden, The Netherlands, 1993.
[42] O. Abollino, M. Malandrino, A. Giacomino, E. Mentasti, Anal. Chim. Acta 688 (2011) 104–121.
[43] D.L. Massart, B.G.M. Vandeginste, L. Buydens, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier Science, Amsterdam, The Netherland, 1997.
[44] B.G.M. Vandeginste, D.L. Massart, Handbook of Chemometrics and Qualimetrics: Part B, Elsevier Science, Amsterdam, The Netherlands (1998) 1–713.
[45] K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics: A Practical Guide.
[46] R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Wiley, West Sussex, England, 2003.
[47] S. Wold, Pattern Recognition 8 (1976) 127–139.
[48] W.J. Dunn , S. Wold, Chemometric Methods in Molecular Design, VCH Publisher, New York (2008) 179–193.
[49] B. Mertens, M. Thompson, T. Fearn, Analyst 119 (1994) 2777–2784.
[50] R. De Maesschalck, A. Candolfi, D. Massart, S. Heuerding, Chemometr. Intell. Lab. 47 (1999) 65–77.
[51] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, San Diego, CA, USA, 1990.
[52] S. Axler, Linear Algebra Done Right, Springer, San Francisco, CA, USA, 1997.
[53] A. De Juan, R. Tauler, Anal. Chim. Acta 500 (2003) 195–210.
[54] W.H. Lawton, E.A. Sylvestre, Technometrics 13 (1971) 617–633.
[55] E. Sylvestre, W. Lawton, M. Maggio, Technometrics 16 (1974) 353–368.
[56] C. Ho, G. Christian, E. Davidson, Anal. Chem. 50 (1978) 1108–1113.
[57] A. Lorber, Anal. Chem. 57 (1985) 2395–2397.

[58] N. Faber, L. Buydens, G. Kateman, J. Chemometr. 8 (2005) 147–154.
[59] E. Sanchez, B.R. Kowalski, Anal. Chem. 58 (1986) 496–499.
[60] M. Maeder, Anal. Chem. 59 (1987) 527–530.
[61] C.J. Xu, J.H. Jiang, Y.Z. Liang, Analyst 124 (1999) 1471–1476.
[62] A.C. Whitson, M. Maeder, J. Chemometr. 15 (2001) 475–484.
[63] H. Keller, D. Massart, Anal. Chim. Acta 246 (1991) 379–390.
[64] J. Toft, O.M. Kvalheim, Chemometr. Intell. Lab. 19 (1993) 65–73.
[65] O.M. Kvalheim, Y.Z. Liang, Anal. Chem. 64 (1992) 936–946.
[66] Y.Z. Liang, O.M. Kvalheim, H.R. Keller, D.L. Massart, P. Kiechle, F. Erni, Anal. Chem. 64 (1992) 946–953.
[67] R. Manne, H. Shen, Y. Liang, Chemometr. Intell. Lab. 45 (1999) 171–176.
[68] R. Manne, Chemometr. Intell. Lab. 27 (1995) 89–94.
[69] E.R. Malinowski, J. Chemometr. 6 (2005) 29–40.
[70] W. Den, E.R. Malinowski, J. Chemometr. 7 (2005) 89–98.
[71] R. Tauler, I. Marqués, E. Casassas, J. Chemometr. 12 (1998) 55–75.
[72] R. Tauler, A. Smilde, B. Kowalski, J. Chemometr. 9 (2005) 31–58.
[73] R. Tauler, Chemometr. Intell. Lab. 30 (1995) 133–146.
[74] O.S. Borgen, B.R. Kowalski, Anal. Chim. Acta 174 (1985) 1–26.
[75] B.G.M. Vandeginste, W. Derks, G. Kateman, Anal. Chim. Acta 173 (1985) 253–264.
[76] S. Wold, Technometrics 20 (1978) 397–405.
[77] P. Geladi, S. Wold, Chemometr. Intell. Lab. 2 (1987) 273–281.
[78] F.C. Sanchez, J. Toft, B. Van den Bogaert, D. Massart, Anal. Chem. 68 (1996) 79–85.
[79] F.C. Sanchez, B. Vandeginste, T. Hancewicz, D. Massart, Anal. Chem. 69 (1997) 1477–1484.
[80] F.C. Sanchez, M. Khots, D. Massart, Anal. Chim. Acta 290 (1994) 249–258.
[81] F.C. Sanchez, M. Khots, D. Massart, J. De Beer, Anal. Chim. Acta285 (1994) 181–192.
[82] w.s. ⟨http:/www.scopus.com/⟩, Scopus, An abstract and citation database of research literature and quality, 2013.
[83] E. Dinç, J. Pharm. Biomed. Anal. 33 (2003) 605–615.
[84] E. Dinç, Ö. Üstündağ, Il Farmaco 58 (2003) 1151–1161.
[85] E. Dinç, D. Baleanu, F. Onur, J. Pharm. Biomed. Anal. 26 (2001) 949–957.
[86] E. Dinç, C. Yücesoy, F. Onur, J. Pharm. Biomed. Anal. 28 (2002) 1091–1100.
[87] E. Dinç, İ.M. Palabıyık, Ö. Üstündağ, F. Yurtsever, F. Onur, J. Pharm. Biomed. Anal. 28 (2002) 591–600.
[88] E. Dinç, Ö. Üstündağ, J. Pharm. Biomed. Anal. 29 (2002) 371–379.
[89] A. El-Gindy, S. Emara, A. Mostafa, IL Farmaco 60 (2005) 269–278.
[90] M.S. Cámara, C. Mastandrea, H.C. Goicoechea, J. Biochem. Biophys. Methods 64 (2005) 153–166.
[91] P.C. Damiani, A.C. Moschetti, A.J. Rovetto, F. Benavente, A.C. Olivieri, Anal. Chim. Acta 543 (2005) 192–198.
[92] G.M. Hadad, A. El-Gindy, W.M.M. Mahmoud, Spectrochim. Acta A 70 (2008) 655–663.
[93] Y. Ni, Y. Wang, S. Kokot, Food Chem. 109 (2008) 431–438.
[94] A. El-Gindy, F. El-Yazby, A. Mostafa, M.M. Maher, J. Pharm. Biomed. Anal. 35 (2004) 703–713.
[95] A. El-Gindy, S. Emara, A. Mostafa, J. Pharm. Biomed. Anal. 41 (2006) 421–430.
[96] Y. Ni, Y. Wang, S. Kokot, Spectrochim. Acta A 70 (2008) 1049–1059.
[97] J. Ghasemi, A. Niazi, S. Ghobadi, J. Chin. Chem. Soc. Taipei 52 (2005) 1049–1054.
[98] N. Ramos Martos, A. Molina Díaz, A. Navalón, I. De Orbe Payá, L. Capitán Vallvey, J. Pharm. Biomed. Anal. 23 (2000) 837–844.
[99] D. Nagavalli, V. Vaidhyalingam, A. Santha, A.S.K. Sankar, O. Divya, Acta Pharm. 60 (2010) 141–152.
[100] M.C.F. Ferraro, P.M. Castellano, T.S. Kaufman, J. Pharm. Biomed. Anal. 34 (2004) 305–314.
[101] A. Ruiz MedinaM. Fernandez de CordovaA. Molina-Diaz, J. Pharm. Biomed. Anal. 21 (1999) 983–992.
[102] A.S.K. Sankar, T. Vetrichelvan, D. Venkappaya, Acta Pharm. 61 (2011) 283–296.
[103] K. VIPUL, M. Rajshree, Anal. Sci. 23 (2007) 445–451.
[104] K. Kaur, A.K. Malik, B. Singh, M. Godarzi, Thai J. Pharm. Sci. 33 (2009) 123–136.
[105] M. Luis, J. Fraga, F. Jiménez, A. Jiménez, J. Arias, Talanta 53 (2001) 761–770.
[106] R. Bhaskar, M.K. Sagar, V. Saini, K. Bhat, J. Anal. Sci. Methods Instrum. 2 (2012) 161–166.
[107] Y.H. Lai, Y.N. Ni, S. Kokot, Chin. Chem. Lett. 21 (2010) 213–216.
[108] S. Guo, J. Duan, Y. Tang, S. Su, E. Shang, S. Ni, D. Qian, J. Pharm. Biomed. Anal. 49 (2009) 1296–1302.
[109] S. Xu, L. Yang, R. Tian, Z. Wang, Z. Liu, P. Xie, Q. Feng, J. Chromatogr. A 1216 (2009) 2163–2168.
[110] L. Peng, Y. Wang, H. Zhu, Q. Chen, Food Chem. 125 (2011) 1064–1071.
[111] W.J. Kong, Y.L. Zhao, X.H. Xiao, C. Jin, Z.L. Li, Phytomedicine 16 (2009) 950–959.
[112] Y. Chen, S.B. Zhu, M.Y. Xie, S.P. Nie, W. Liu, C. Li, X.F. Gong, Y.X. Wang, Anal. Chim. Acta 623 (2008) 146–156.
[113] H.L. Ma, M.J. Qin, L.W. Qi, G. Wu, P. Shu, Biomed. Chromatogr. 21 (2007) 931–939.
[114] M. Cirlini, A. Caligiani, L. Palla, G. Palla, Food Chem. 124 (2011) 1678–1683.
[115] F. Yan, Z. Liang, C. Jianna, W. Zhengtao, X. Losahan, Z. Zhengxing, Talanta 53 (2001) 1155–1162.
[116] A. Rohman, Y.B. Che Man, Vib. Spectrosc. 55 (2011) 141–145.
[117] R. Pan, F. Guo, H. Lu, W. Feng, J. Pharm. Biomed. Anal. 55 (2011) 391–396.
[118] A. Porfire, I. Tomuta, L. Tefas, S.E. Leucuta, M. Achim, J. Pharm. Biomed. Anal. 63 (2012) 87–94.
[119] K. Wiberg, S.P. Jacobsson, Anal. Chim. Acta 514 (2004) 203–209.
[120] K. Wiberg, J. Chromatogr. A 1108 (2006) 50–67.
[121] R. Gimeno, E. Comas, R. Marcé, J. Ferré, F. Rius, F. Borrull, Anal. Chim. Acta 498 (2003) 47–53.
[122] E. Comas, R.A. Gimeno, J. Ferré, R.M. Marcé, F. Borrull, F.X. Rius, J. Chromatogr. A 988 (2003) 277–284.
[123] H. Abdollahi, A. Golshan, Anal. Chim. Acta 693 (2011) 26–34.
[124] M. Bahram, M. Mabhooti, Anal. Chim. Acta 639 (2009) 19–28.
[125] Y.-B. Wu, L.-J. Zheng, J. Yi, J.-G. Wu, T.-Q. Chen, J.-Z. Wu, Int. J. Mol. Sci. 14 (2013) 1999–2010.
[126] C. Ho, G. Christian, E. Davidson, Anal. Chem. 52 (1980) 1071–1079.
[127] C.N. Ho, G.D. Christian, E. Davidson, Anal. Chem. 53 (1981) 92–98.
[128] R. Tauler, S. Lacorte, D. Barceló, J. Chromatogr. A 730 (1996) 177–183.
[129] L. Wang, X. Wang, L. Kong, Biochem. Syst. Ecol. 40 (2012) 138–145.
[130] K. De Braekeleer, A. De Juan, D. Massart, J. Chromatogr. A 832 (1999) 67–86.
[131] E. Peré-Trepat, A. Hildebrandt, D. Barceló, S. Lacorte, R. Tauler, Chemometr. Intell. Lab. 74 (2004) 293–303.
[132] J. SalauM. HoningR. TaulerD. Barcelo, J. Chromatogr. A 795 (1998) 3–12.
[133] E. Peré-Trepat, S. Lacorte, R. Tauler, J. Chromatogr. A 1096 (2005) 111–122.
[134] L. Wu, X. Jiang, L. Huang, S. Chen, PLoS ONE 8 (2013) e64178.
[135] E. Comas, R.A. Gimeno, J. Ferré, R.M. Marcé, F. Borrull, F.X. Rius, J. Chromatogr. A 1035 (2004) 195–202.
[136] S. Mas, G. Fonrodona, R. Tauler, J. Barbosa, Talanta 71 (2007) 1455–1463.
[137] A.E. Sinha, J.L. Hope, B.J. Prazen, E.J. Nilsson, R.M. Jack, R.E. Synovec, J. Chromatogr. A 1058 (2004) 209–215.
[138] F. Cuesta Sanchez, S. Rutan, M. Gil Garcia, D. Massart, Chemometr. Intell. Lab. 36 (1997) 153–164.
[139] M. Wasim, R.G. Brereton, J. Chromatogr. A 1096 (2005) 2–15.
[140] D. Bylund, R. Danielsson, G. Malmquist, K.E. Markides, J. Chromatogr. A 961 (2002) 237–244.
[141] C.G. Fraga, J. Chromatogr. A 1019 (2003) 31–42.
[142] H.A. Gad, S.H. El-Ahmady, M.I. Abou-Shoer, M.M. Al-Azizi, Phytochem. Anal. 24 (2012) 1–24.
[143] C. Tistaert, B. Dejaegher, Y.V. Heyden, Anal. Chim. Acta 690 (2011) 148–161.
[144] B. Lavine, J. Workman, Anal. Chem. 82 (2010) 4699–4711.
[145] A. Niazi, A. Azizi, Turk. J. Chem. 32 (2008) 217–228.
[146] C.J. Xu, Y.Z. Liang, F.T. Chau, Y.V. Heyden, J. Chromatogr. A 1134 (2006) 253–259.
[147] A. Detroyer, V. Schoonjans, F. Questier, Y. Vander Heyden, A. Borosy, Q. Guo, D. Massart, J. Chromatogr. A 897 (2000) 23–36.
[148] Y. Ni, L. Zhang, J. Churchill, S. Kokot, Talanta 72 (2007) 1533–1539.
[149] C.Z. Wang, M. Ni, S. Sun, X.L. Li, H. He, S.R. Mehendale, C.S. Yuan, J. Agric. Food Chem. 57 (2009) 2363–2367.
[150] B. Schmidt, J.W. Jaroszewski, R. Bro, M. Witt, D. Stærk, Anal. Chem. 80 (2008) 1978–1987.
[151] K. Wiberg, M. Andersson, A. Hagman, S.P. Jacobsson, J. Chromatogr. A 1029 (2004) 13–20.
[152] F. Gong, B.T. Wang, Y.Z. Liang, F.T. Chau, Y.S. Fung, Anal. Chim. Acta 572 (2006) 265–271.
[153] B.Y. Li, Y. Hu, Y.Z. Liang, P.S. Xie, Y.P. Du, Anal. Chim. Acta 514 (2004) 69–77.
[154] E. Dinç, A. Özdemir, H. Aksoy, Ö. Üstündağ, D. Baleanu, Chem. Pharm. Bull. 54 (2006) 415–421.
[155] R. Gargallo, R. Tauler, F. Cuesta-Sanchez, D. Massart, Trends Anal. Chem. 15 (1996) 279–286.
[156] K.J. Johnson, S.L. Rose-Pehrsson, R.E. Morris, Energy Fuels 18 (2004) 844–850.
[157] M. Zougagh, A. Ríos, M. Valcárcel, Anal. Chim. Acta 539 (2005) 117–124.
[158] H. Abdollahi, F. Nazari, Anal. Chim. Acta 486 (2003) 109–123.
[159] E. Comas, J. Ferré, F. Xavier Rius, Anal. Chim. Acta 515 (2004) 23–30.
[160] L.S. Ramos, E. Sanchez, B.R. Kowalski, J. Chromatogr. A 385 (1987) 165–180.
[161] E. Sanchez, L.S. Ramos, B. Kowaiski, J. Chromatogr. A 385 (1987) 151–164.
[162] S. Lacorte, R. Tauler, J. Chromatogr. A 697 (1995) 345–355.
[163] C. Daolio, F.L. Beltrame, A.G. Ferreira, Q.B. Cass, D.A.G. Cortez, M. Ferreira, Phytochem. Anal 19 (2008) 218–228.
[164] X.L. Cheng, F. Wei, X.Y. Xiao, Y.Y. Zhao, Y. Shi, W. Liu, P. Zhang, S.C. Ma, S.S. Tian, R.C. Lin, J. Pharm. Biomed. Anal. 62 (2011) 191–195.
[165] G.M. Escandar, P.C. Damiani, H.C. Goicoechea, A.C. Olivieri, Microchem. J. 82 (2006) 29–42.
[166] T. Gröger, R. Zimmermann, Talanta 83 (2011) 1289–1294.
[167] A. Rohman, Y. Che Man, Middle-East J. Sci. Res. 7 (2011) 726–732.
[168] N. Mobaraki, B. Hemmateenejad, Chemometr. Intell. Lab 109 (2011) 171–177.
[169] H.L. Huang, P. Antonelli, J. Appl. Meteorol. 40 (2001) 365–388.
[170] M.L. O'Connell, A.G. Ryder, M.N. Leger, T. Howley, Appl. Spectrosc. 64 (2010) 1109–1121.
[171] A. de Juan, S.C. Rutan, R. Tauler, D.L. Massart, Chemometr. Intell. Lab. 40 (1998) 19–32.
[172] I. Duran Martin-Meras, M.I. Rodriguez-Caceres, M.C. Hurtado-Sanchez, Anal. Sci. 27 (2011) 745–749.
[173] P. Zhao, W. Yutian, S. Xiaoqing, Am. J. Eng. Technol. Res. 11 (2011) 1717–1719.