



**INTERNATIONAL JOURNAL OF
PHARMACEUTICAL SCIENCES**
[ISSN: 0975-4725; CODEN(USA):IJPS00]
Journal Homepage: <https://www.ijpsjournal.com>



Review Article

A Review On Chemometrics In Pharmaceutical Analysis

Drashti N. Bhalodia*¹, Dharmendrasinh A. Baria²

¹Department of Pharmaceutical Quality Assurance, A.R. College of Pharmacy & G.H. Patel Institute of Pharmacy, Vallabh Vidyanagar

²Department of Pharmaceutical Chemistry, A.R. College of Pharmacy & G.H. Patel Institute of Pharmacy, Vallabh Vidyanagar

ARTICLE INFO

Received: 22 Aug 2024

Accepted: 26 Aug 2024

Published: 01 Sep 2024

Keywords:

Chemometrics, Bilinear models, pharmaceutical science, Analytical data, multivariate analysis, multidisciplinary science.

DOI:

10.5281/zenodo.13625694

ABSTRACT

Chemometrics is the use of mathematical and statistical methods to improve the understanding of chemical information and to correlate quality parameters or physical properties to analytical instrument data. It is a data-driven multidisciplinary science that allows maximum collection and extraction of useful information from the analytical data of numerous application areas such as chemistry, biochemistry, medicine, biology and chemical engineering. This review focuses mainly on numerous chemometric models used such as like parallel factor analysis (PARAFAC), Tucker-3, N-partial least square (NPLS), and bilinear models like principal component regression (PCR) and partial least squares (PLS) and their relevant applications in the field of pharmaceutical sciences. Chemometric approaches can be used to analyze the data obtained from various instruments including near infrared (NIR), attenuated total reflectance Fourier transform infrared (ATR-FTIR), high-performance liquid chromatography (HPLC), and terahertz pulse spectroscopy. The technique has been used in the quality assurance and quality control of pharmaceutical solid dosage forms. Also, this review gives idea about list of software used for multivariate analysis.

INTRODUCTION

Chemometrics is the application of statistical and mathematical methods to analytical data to permit maximum collection and extraction of useful information. The utility of chemometric techniques as tools enabling multidimensional calibration of selected spectroscopic,

electrochemical, and chromatographic methods is demonstrated. Application of this approach mainly for interpretation of UV-Vis and near-IR (NIR) spectra, as well as for data obtained by other instrumental methods, makes identification and quantitative analysis of active substances in

*Corresponding Author: Drashti N. Bhalodia

Address: Department of Pharmaceutical Quality Assurance, A.R. College of Pharmacy & G.H. Patel Institute of Pharmacy, Vallabh Vidyanagar

Email ✉: drashti130701@gmail.com

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.



complex mixtures possible, especially in the analysis of pharmaceutical preparations present in the market. Such analytical work is carried out by the use of advanced chemical instruments and data processing, which has led to a need for advanced methods to design experiments, calibrate instruments, and analyze the resulting data. This review will concentrate on gaining an understanding of how chemometrics can be useful in the modern analytical laboratory. A selection of the most challenging problems faced in pharmaceutical analysis is presented, the potential for chemometrics is considered, and some consequent implications for utilization are discussed. Advances in electronics and computing over the past 30 years have revolutionized the analytical laboratory. Technological developments have allowed instruments to become smaller, faster, and cheaper, while continuing to increase accuracy, precision, and availability. Data analysis methods have also benefitted from advances in technical computing; commercially available mathematical programming packages allow scientists to perform complex calculations with a few simple keystrokes. Furthermore, software sold with many commercial instruments contains automatic data processing algorithms (e.g., Fourier transform analysis, data filtering, and peak recognition). The advances in computing allow researchers to obtain increasing amounts of chemically relevant information from their data; however, this is not always achieved using simple data-processing techniques. Svante Wold first coined the term “kemometri” (“chemometrics” in English) in 1972 by combining the words kemo for chemistry and metri for measure. Presently, the journal *Chemometrics and Intelligent Laboratory Systems* defines chemometrics as “the chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments, and to provide maximum chemical

information by analyzing chemical data”. The field of chemometrics has also benefitted from technological advances in the past 30 years, causing the number of researchers using chemometric methods to grow. Chemometrics has brought us a large number of valuable tools that allow us to resolve and display complex chemical information from often dynamic unstable systems. In general, the main areas of focus for these techniques have been multivariate calibration; pattern recognition, classification, and discriminatory analysis; and reactant and reaction process modeling and monitoring. The purpose of this review is to describe various multivariate calibration chemometrics methods in combination with UV-Vis spectrophotometry, near-IR spectroscopy (NIRS), fluorescence spectroscopy, electro analysis, chromatographic separation, and flow-injection analysis for the analysis of drugs in pharmaceutical preparations. [1] The term “Chemometrics” was coined by Svante Wold in 1972 and is concerned with the application of mathematical and statistical techniques to extract chemical and physical information from complex data with the application of computer science. Various algorithm and analogue ways are available for processing and evaluating the data. They can be promoted to various fields like medicine, pharmacy, food control, environmental monitoring and it is continuing to diverge into new fields such as metabonomic. Chemometrics is applied in areas comprising experimental parameter optimization, data quality improvement, identification and quantification of targeted chemical components, pattern recognition techniques for clustering and classification, multivariate model establishment to correlate chromatographic properties with molecular descriptors and prediction of properties or activities of chemical compounds or technological materials (quantitative structure-activity or structure property relationships) to find out hidden



relationships existing between the available data and the desired information. [2] Chemometrics has been evolving as a sub discipline in chemistry for over 30 years as the need for advanced statistical and mathematical methods has increased with the increasing sophistication of chemical instrumentation and processes. “Chemometrics is a chemical discipline that uses mathematics, statistics, and formal logic (a) to design or select optimal experimental procedures; (b) to provide maximum relevant chemical information by analyzing chemical data; and (c) to obtain knowledge about chemical systems.” In the 1972 review of Statistical and Mathematical Methods in Analytical Chemistry, there are only two areas of active study; “Curve Fitting” and “Statistical Control” where curve fitting is attributed to be the analytical chemists’ area of interest whereas chemical engineers were reported to be primarily concerned with quality control. In 1971, Svante World of the University of Umea in Sweden coined the term “chemometrics” in a grant proposal and shortly after, his collaboration with Bruce Kowalski of the University of Washington, brought the name to the United States. As part of the development of chemometrics as a separate sub discipline, the International Chemometrics Society was formed in 1974. The first paper with chemometrics in the title appeared in 1975. In which Kowalski suggested that chemometrics had developed to the point that it was now a functioning research area in the chemical sciences. He points to the value of pattern recognition and indicates that there were vehicles for publication of research results in journals such as the Journal of Chemical Information and Computer Sciences, but that it would be appropriate for Analytical Chemistry to publish such work. Although there were a limited number of papers published in these journals, there were significant impediments to publishing chemometrics articles as there remained considerable skepticism in the analytical

community as to the need for complex data analysis tools. In the view of many established chemists, the need for complicated data analysis tools was a sign that the proper experiments were not performed, rather than understanding that advanced data analysis was integral to the maximal use of evolving new technologies. The introduction of the section “Computer Techniques and Optimization” in *Analytica Chimica Acta* in 1977 was the first journal publication that was clearly dedicated to this developing area. In 1980, *Analytical Chemistry* changed the name of the review section on Statistical and Mathematical Methods in *Analytical Chemistry* to Chemometrics, and entered the mainstream of the field. Subsequently, in 1982, the separate section in *Analytica Chimica Acta* was terminated because chemometrics had become sufficiently well accepted to eliminate the need for this special attention. Subsequently, two journals dedicated to chemometrics were launched: *Chemometrics and Intelligent Laboratory Systems* and *Journal of Chemometrics*, but now to provide a vehicle for discussion of more of the details of the methods while applications would generally be published in the broader analytical journals. Thus, the critical aspect of chemometric methods is that they had become commonly accepted into the practice of chemistry. There has continued to be the simultaneous development of new analytical instruments that have produced data that demanded new and more effective data analysis methods while the increasing capability of personal computers has permitted more computationally intensive calculations to be performed without the need for access to “super” computers. This combination of developments has opened many new options for data analytical method improvement. Thus, over the intervening years, chemometrics has emerged to have a significant role within analytical chemistry including the incorporation into the operating



systems of a number of commercial analytical instruments. This paper will review the major areas of chemometrics related to analytical chemistry including multivariate calibration, pattern recognition, and mathematical mixture resolution and then highlight some of the new directions that chemometric methods are taking. [3]

Chemometrics [4]

The term chemometrics was first coined in 1971 to describe the growing use of mathematical models, statistical principles, and other logic-based methods in the field of chemistry and, in particular, the field of analytical chemistry. Chemometrics is an interdisciplinary field that involves multivariate statistics, mathematical modeling, computer science, and analytical chemistry. Some major application areas of chemometrics include

- Calibration, validation, and significance testing.
- Optimization of chemical measurements and experimental procedures
- The extraction of the maximum of chemical information from analytical data.

In many respects, the field of chemometrics is the child of statistics, computers, and the “information age.” Rapid technological advances, especially in the area of computerized instruments for analytical chemistry, have enabled and necessitated phenomenal growth in the field of chemometrics over the past 30 years. For most of this period, developments have focused on multivariate methods. Since the world around us is inherently multivariate, it makes sense to treat multiple measurements simultaneously in any data analysis procedure. For example, when we measure the ultraviolet (UV) absorbance of a solution, it is easy to measure its entire spectrum quickly and rapidly with low noise, rather than measuring its absorbance at a single wavelength. By properly considering the distribution of multiple variables simultaneously, we obtain more information than

could be obtained by considering each variable individually. This is one of the so-called multivariate advantages. The additional information comes to us in the form of correlation. When we look at one variable at a time, we neglect correlation between variables, and hence miss part of the picture. A recent paper by Bro described four additional advantages of multivariate methods compared with univariate methods. Noise reduction is possible when multiple redundant variables are analyzed simultaneously by proper multivariate methods. For example, low-noise factors can be obtained when principal component analysis is used to extract a few meaningful factors from UV spectra measured at hundreds of wavelengths. Another important multivariate advantage is that partially selective measurements can be used, and by use of proper multivariate methods, results can be obtained free of the effects of interfering signals. A third advantage is that false samples can be easily discovered, for example in spectroscopic analysis. For any well characterized chemometric method, aliquots of material measured in the future should be properly explained by linear combinations of the training set or calibration spectra. If new, foreign materials are present that give spectroscopic signals slightly different from the expected ingredients, these can be detected in the spectral residuals and the corresponding aliquot flagged as an outlier or “false sample.” The advantages of chemometrics are often the consequence of using multivariate methods. The reader will find these, and other advantages highlighted throughout the book. [4]

What is Chemometrics? [5]

The definition of chemometrics is evident in its name, where chemo– means chemical and metrics means measurement; thus, chemometrics is the study of chemical (and biochemical) measurements and is a branch of analytical chemistry. Examples of chemometric applications include:



- Ensuring that the data we collect is appropriate for our purposes
- Enhancing the quality of an analytical signal by finding ways to minimize the contribution of noise
- Reporting on an experiment in a way that estimates the uncertainty in its results and our confidence in those results
- Building useful models that predict the outcomes of future experiments
- Extracting from chemical data, hidden, but analytically useful information by finding underlying patterns in the data [5]

ORIGIN AND DEVELOPMENT OF CHEMOMETRICS [6,7]

In 1971, a Swedish scientist Svante Wold coined the term “kemometri,” in Swedish and in English it is equivalent to “chemometrics”. 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. [6]

Chemometrics is a branch of science that is used for extraction of the data related to chemical and physical phenomena involved in the manufacturing process by the application of the statistical and mathematical methods. It can be applied in predictive issues solving like predicting the target properties, desired features. Also can be used for the descriptive issue solving like the model composition, identification and understanding. Chemometrics shows its application in the multivariate data collection and analysis. Various algorithms and analogous ways are available for processing and evaluating the data. They can be implemented to various fields, like medicine, pharmacy, food control, and environmental monitoring. Some of the Chemometric models for analysis are mentioned below. [7]

DIFFERENT CHEMOMETRICS MODELS FOR ANALYSIS OF DATA [8]

Chemometrics methods can be categorized in several different ways like clustering, regression and explorative methods. Chemometricians have adopted methods from other research fields such as econometrics and psychometrics where bilinear partial squares and multiway methods, respectively, had been applied and refined. Methods are separated according to how they explore

the data arrays and a distinction can be drawn between bilinear, non-linear and multiway methods as well as between projection, latent variable and factor based methods. However, some methods overlap between the above categorizations.



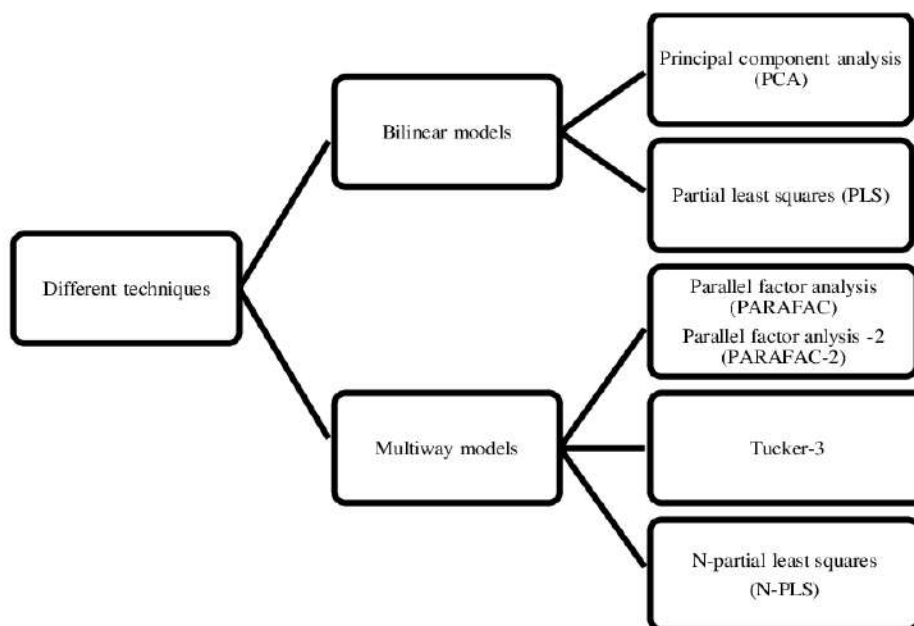


Figure:1 Classification of Chemometrics techniques

Bilinear model:

Bilinearity means the system is linear with relevance to its decomposition, i.e. the system is linear in its estimated parameters. In bilinear models, the knowledge of data is arranged in data matrices so that each horizontal row contains samples and vertical column has variables. Bilinear chemometric techniques include the following:

Principal Component Analysis (PCA)

This is a simple and non-parametric technique used for extracting the relevant information from the data sets on the basis of their affinity and differences. It is widely used in multivariate data analysis. PCA decreases the dimensionality and multivariate data compression in the fields of science. During process monitoring, it can be used to develop a correlation structure between variables and also to examine the changes. Thus it reduces the number of variables in the process. If for a series of sites, or objects, or persons, a variety of variables are measured, then each variable will have a variance, and usually, the variables will be associated with each other that is, there will be statically variance between pairs of variables. Today, PCA is one of the massive utilized

multivariate models because of its wide applicability for multivariate problems. [8-12,28-30]

Special cases of PCA:

The most widely used special cases of PCA are principal component regression (PCR), soft independent modelling of class analogy (SIMCA) and multi-way principal component analysis (MPCA).

Principal component regression (PCR) is a regression analysis technique that is based on principal component analysis (PCA). Typically, it considers regressing the outcome (also known as the response or the dependent variable) on a set of covariates (also known as predictors, or explanatory variables, or independent variables) based on a standard linear regression model, but uses PCA for estimating the unknown regression coefficients in the model. In PCR, instead of regressing the dependent variable on the explanatory variables directly, the principal components of the explanatory variables are used as regressors. One major use of PCR lies in overcoming the multi collinearity problem which arises when two or more of the explanatory variables are close to being collinear. Soft

independent modelling by class analogy (SIMCA) is a statistical method for supervised classification of data. The method requires a training data set consisting of samples (or objects) with a set of attributes and their class membership. The term soft refers to the fact the classifier can identify samples as belonging to multiple classes and not necessarily producing a classification of samples into non-overlapping classes. SIMCA as a method of classification has gained widespread use especially in applied statistical fields such as chemometrics and spectroscopic data analysis.

Multi-way principal components analysis (MPCA) is an efficient tool for reducing higher dimensional data arrays. MPCA allows the detection of spatial and temporal factors of influence and the classification of the parameters can be considered according to these factors. Rene et al., developed a multi-way principal components analysis for a complex data array resulting from physicochemical characterization of natural waters. Molloy et al., applied a multiway principal component analysis for Identification of process improvements in pharmaceutical manufacture. The need for some form of PCA which is weighted according to measurement errors has been recognized for some time. Informally this has been addressed through a variety of scaling procedures, such as range scaling and auto scaling as well as more elaborate schemes. Under the right conditions these can reduce PCA to a maximum likelihood method. However, a number of researchers have developed methods to more rigorously incorporate measurement errors into the modeling process. This is normally done by minimizing the usual weighted residual sum of squares in accordance with some p-dimensional model. Mathematically, if X is an m×n data matrix, this corresponds to minimization of

Equation-1

$$S^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{(x_{ij} - \hat{x}_{ij})^2}{\sigma_{ij}^2}$$

where \hat{x}_{ij} corresponds to the estimated value of the measurement. In the general case where there are no offsets in the model, this is given by

Equation -2

$$\hat{X} = AB$$

where A is m×p and B is p×n. By analogy to PCA, A and B correspond to score and loading matrices, but in equation (2) the individual vectors are not required to be orthogonal. A variety of methods have been devised to obtain A and B through minimization of equation (1) and these differ largely in their representation of the problem, the constraints applied to the solution and their approach to the nonlinear optimization. Gabriel and Zamir⁶ describe a method based on ‘criss-cross regressions’ as a means to obtain lower-rank approximations of the matrix X. Paatero and Tapper have described what they call ‘positive matrix factorization’ (PMF) and have applied this to environmental problems. In addition to satisfying the minimization criterion, PMF also requires A and B to be positive. In this work, MLPCA is presented as an alternative to the above methods. Although MLPCA is similar to these methods in that it seeks to minimize equation (1), it has some important differences. First, it is formulated in terms of singular value decomposition (SVD), which is a very common method for implementing PCA. Second, the standard MLPCA algorithm consists of an alternating least squares procedure which is robust, easy to implement and very efficient compared with conventional gradient search methods. Finally, unlike the methods described in the preceding paragraph, which require measurement errors to be independent, MLPCA allows the inclusion of error covariance in virtually any form. The MLPCA method described here should not be confused with maximum



likelihood common factor analysis (MLCFA) that appears frequently in the literature outside of chemistry. Although the terms are often used interchangeably by chemists, PCA and factor analysis are distinctly different approaches to multivariate analysis. The principles of MLCFA were originally developed by Lawley and Maxwell and later employed in programs such as LISREL. More recently, MLCFA has appeared in the chemical literature with claims that it performs better than PC. However, MLCFA was developed with the intention of finding structural models for random variables. As such, it estimates covariance matrices for random variables and does not generally use information about measurement errors. Another errors-in-variables method that has become popular recently is total least squares (TLS). This method uses SVD for the purpose of developing a regression model and is similar to MLPCA in some ways. However, it is less general in its ability to obtain maximum likelihood estimates of model parameters.

Partial Least Squares (PLS) [8,13-16,31]

Partial least squares (PLS) are a regression method for multivariate data. It is one of the widely implemented methods which describe the relationship between different sets of different observed variables by the means of latent variables. The basic theory of this method is that it modifies relations between sets of the observed variables by a small number of latent variables (not directly noticed or consistent) by assimilate regression, dimension reduction techniques, and modelling tools. The latent variables increase the covariance between the different sets of variables. PLS is identical to canonical correlation analysis (CCA) and can be used as a discrimination tool and dimension reduction method like principal component analysis (PCA). PLS is widely used approach as it can process large chemical data. The determination of flow properties of pharmaceutical powders by near infrared

spectroscopy (NIR) spectroscopy was done using Partial least square technique. Cordeiro et al. conducted multivariate spectroscopic determination of lamivudine- zidovudine associations by partial least square regression (PLS). Regression by means of projections to latent structures (PLS) is today a widely used chemometrics data analytical tool. It applies to any regression problem in industrial research, development, and production (RDP), regardless of whether the data set is short/wide or long/lean, or contains linear or non-linear systematic structure, with or without missing data, and possibly are also ordered in two or more blocks across multiple model layers. PLS exists in many different shapes and implementations. The two-block predictive PLS version is the most often used form in science and technology. These latter are a method for relating two data matrices, X and Y, by a linear multivariate model, but goes beyond traditional regression in that it models also the structure of X and Y. PLS derives its usefulness from its ability to analyze data with many, noisy, collinear, and even incomplete variables in both X and Y. PLS has the desirable property that the precision of the model parameters improves with the increasing number of relevant variables and observations. The regression problem, i.e., how to model one or several dependent variables, responses, Y, by means of a set of predictor variables, X, is one of the most common data-analytical problems in science and technology. Examples in RDP include relating Y = analyte concentration to X = spectral data measured on the chemical samples (Example 1), relating Y = toxicant exposure levels to X = gene expression profiles for rats for the different doses (Example 2), and relating Y = the quality and quantity of manufactured products to X = the conditions of the manufacturing process (Example 3). Traditionally, this modelling of Y by means of X is done using MLR, which works well as long as the X-variables are fairly few and fairly



uncorrelated, i.e., X has full rank. With modern measuring instrumentation, including spectrometers, chromatographs, sensor batteries, and bio-analytical platforms, the X-variables tend to be many and also strongly correlated. We shall therefore not call them "independent", but instead "predictors", or just X-variables, because they usually are correlated, noisy, and also incomplete. In handling numerous and collinear X-variables, and response profiles (Y), PLS allows us to investigate more complex problems than before, and analyze available data in a more realistic way. However, some humility and caution is warranted; we are still far from a good understanding of the complications of chemical, biological, and technological systems. Also, quantitative multivariate analysis is still in its infancy, particularly in applications with many variables and few observations (objects, cases).

Multi-way models [8,17,18,19]

Multi-way models are used when the data is multivariate and linear in more than 2 dimensions. These can be considered to devise a model in n-dimensions so that the system is linear in n-dimensions. A three-linear system is often visualized as a data cube and is called a 3-way data or 3-way array whereas the bilinear system is a rectangular matrix that can be considered as 2-way data. The multi-way modelling originated from psychological data treatment where bilinear data analysing methods were not adequate. Multi-way methods are also applied to process control as well as regression analysis. Some of the advantages of multi-way models are that they have been recognized as useful tools for monitoring batch data since they improve the understanding of the process and summarize its behaviour in a batch wise manner. The methods like multiway principal component analysis (MPCA) and multiway partial least squares (MPLS) improve the process understanding and review its behaviour in a batch-wise manner and are therefore recognized as tools

for monitoring batch data. However, if the initial data contains higher amplitude then it becomes difficult for the models to describe the computed data and therefore multi-way methods that work with three-way or the higher arrangement like parallel factor analysis (PARAFAC) and PARAFAC-2, Tucker-3, and N-partial least squares (N-PLS) are the methods of choice.

Parallel factor analysis (PARAFAC) [8, 20-24]

Parallel factor analysis (PARAFAC) is a disintegration method used for the modelling of three way or higher data and is chiefly intended for data having compatible variable profiles within each batch. The history to interpret PARAFAC is as follows. Cattell (1944) reviewed seven principles for the choice of rotation in component analysis and achieved the principle of "Parallel proportional profiles" as the most basic assumption. This assumption means that the two data matrices with the same variables ought to contain the identical components. By using this assumption as a constraint Harshman (1970) projected a new method to analyse two or more data matrices that contain scores for the same person on the same variables and termed the method as PARAFAC. PARAFAC could be a generalization of the principal component analysis (PCA) projection method for a multi-way array. The PARAFAC model has a second-order advantage, i.e. it can handle interferences in new samples by fitting the new interferent with an extra component.

Parallel factor analysis-2 (PARAFAC-2) [8, 20-24]

PARAFAC-2 is also designed for modelling N-way data but, in contrast to PARAFAC, it handles experiments of different lengths and variable profiles that are deviated or in a different phase. It can handle data variable profiles that are deviated or/are in a different phase. In PARAFAC, trilinearity is a basic situation whereas PARAFAC-2 allows trilinearity. However, it is to



be noted that PARAFAC may be used for fitting nonlinearity to some extent in one mode only in cases where data shifts from linearity are regular. Both the techniques are mainly applied to inspect chemical evidence from experiments that form a 3-way or higher data structure. For example, chromatographic data, fluorescence spectroscopy, temporal varied spectroscopy data with overlapping spectral profiles, and process data.

Tucker-3 [8, 25]

The Tucker-3 method can be used for consolidating and data analysis of the N-way array. The Tucker-3 model consists of storing matrices in n modes, factors that are typically rectangular and a (P, Q, R) dimensional core array G . The Tucker-3 core array differs from the PARAFAC core by having at least one off-diagonal core element as non-zero, whereas the PARAFAC has a so called super-diagonal core array and thus PARAFAC can be expressed as a special case of the Tucker-3 model. This can be used for survey of N-way array data as it consists of N modes of loading matrices. The generalization of the Tucker-3 model, and the fact that it covers the PARAFAC model as a certain case has made it an often used model for decomposition, compression, and interpretation in many applications.

N- Partial least squares (N-PLS) [8, 26]

For handling a multiway data extension of PLS method namely N-PLS was made acquainted that uses dependent and independent variables for finding the latent variables for describing maximal covariance. N-PLS decomposition starts by constructing a distinct PARAFAC-like model for dependent response variables and maximize the covariance between the two matrices. Some researched the convergence of triglyceride in people by utilizing fluorescence spectroscopy within the region 220-900 nm. Nonlinear partial least squares with cubic B-spline-work based nonlinear change was utilized as the chemometric

strategy. Wavelengths within the region of 300-367 nm and 386-392 nm in the first derivative of the original fluorescence spectrum were the enhanced wavelength combination for the prediction model.

Classical Least Squares (CLS) [1,27]

CLS involves the application of multiple linear regression (MLR) to the classical expression of the Beer-Lambert law of spectroscopy:

$$A = KC \text{ (or } dA/d\lambda = KC)$$

where the matrixes A and $dA/d\lambda$ represent the zero-order absorbance and derivative absorbance matrixes, respectively; C is the concentration matrix; and K is the calibration coefficient matrix. The CLS technique can only be applied to systems where every constituent in the sample is known. If there is the possibility of contaminants in the (unknown) samples that were not present in the calibration mixtures, the model will not be able to predict the constituent concentrations accurately.

Inverse Least Squares (ILS) [1,27]

ILS, sometimes known as P-matrix calibration, is called this because it originally involves the application of MLR to the inverse expression of the Beer-Lambert law of spectroscopy:

$$C = PA \text{ (or } C = P \times dA/d\lambda)$$

where the matrixes A and $dA/d\lambda$ represent the zero-order absorbance and derivative absorbance matrixes, respectively; C is the concentration matrix; and P is the calibration coefficient matrix. Since it is not necessary to know the composition of the training mixtures beyond the constituents of interest, the ILS method is better suited to more complex types of analyses not handled by CLS. Disadvantages of the ILS method are that wavelength selection can be difficult and time-consuming, the number of wavelengths used in the model is limited by the number of calibration samples, and a large number of samples are required for accurate calibration.



Principal Component Regression (PCR) and Partial Least Squares Regression (PLS) [1]

Among the different regression methods available for multivariate calibration, the factor analysis-based methods, including PLS and PCR, have received considerable attention in the chemometrics literature. PLS and PCR can be used directly for ill conditioned data by extracting the latent variables (factors). The number of latent variables is lower than the number of objects. These techniques are powerful multivariate statistical tools that have been successfully and widely applied to the quantitative analysis of spectroscopic data because of their ability to overcome problems common to this data such as collinearity, band overlaps and interactions, and the ease of their implementation due to the availability of software. Here, only a brief introduction about PCR and PLS is given as the techniques are routinely used. PCR is a widely used regression model for data having a large degree of covariance in the independent or predictor variables or where ill-conditioned matrixes are present. Instead of regressing the concentrations of a measurement system onto the original measured variables spectrum, PCR implements a principal component analysis (PCA) decomposition of the spectrum X data before regressing the concentrations information onto the principal component scores. Some vectors having small magnitude are omitted to avoid the collinearity problem. PCR solves this by elimination of lower ranked principal components, which in turn reduces noise (error) present within the system. PLS regression is related to both PCR and MLR. PCR aims to find the factors that capture most of the variance within the data before regression onto the concentration variables, whereas MLR seeks a single factor that correlates both the data and their concentrations. PLS attempts to maximize the covariance, thus capturing the variance and correlating the data

together. As PLS searches for the factor space most congruent to both matrixes, its predictions are far superior to PCR. PCR and PLS techniques share many similarities, and the theoretical relationships between them has been covered extensively in the literature. PLS and PCR perform data decomposition into spectral loadings and scores prior to model building with the aid of these new variables. In PCR, the data decomposition is done using only spectral information, while PLS uses spectral and concentration data. Historically, PCR predates PLS. However, since its introduction, PLS appears, by most accounts, to have become the method of choice among chemists. PLS almost always required fewer latent variables than PCR, but this did not appear to influence predictive ability. However, global models, such as PLS, implicitly endeavor to include the variation due to external effects in the model, in much the same way as unknown chemical interferences can be included in an inverse calibration model. Provided the interfering variation is present in the calibration set, an inverse calibration model can, in the ideal case of additivity and linearity, easily correct for the variation due to unknown interferences. It is assumed in global calibration models that the new sources of spectral variation can be modeled by including a limited number of additional PLS factors. Owing to increase in the calibration model's dimensionality, it becomes necessary to measure a large number of samples under changing conditions in order to make a good estimation of the additional parameters. When highly nonlinear effects are present in the spectra, many additional PLS factors are necessary to model the spectral differences, and occasionally, it is not possible to model these spectral differences.

Artificial Neural Networks (ANNs) [1]

The calibration problem is normally thought to be a linear one. However, Long et al. found that the calibration becomes nonlinear when high levels of



noise are present. In this case, an ANN approach can provide better prediction results. ANNs were originally designed to mimic the function of the human brain. They consist of a number of simple processing units (or neurons) linked by weighted modifiable interconnections. ANNs have been developed for quantitative analysis of samples during the last decade. Compared with MLR, ANN is a more flexible modeling methodology, since both linear and nonlinear functions can be used (or combined) in the processing units. This allows more complex relationships between a high dimensional descriptor space and the given retention data, and may lead to better predictive power of the resulting ANN model compared with MLR. However, the major disadvantage of ANN is also directly related to the complex model infrastructure. Compared to MLR, it suffers from the perception of being a “black box” heuristic tool. ANN models generally are more difficult to interpret. Furthermore, to get robust calibrations by ANN, the number of samples must be higher than the number of weights to be estimated, which normally implies the use of a large number of calibration samples. Before building ANNs, methods for reducing the input dimensionality by mathematical preprocessing (fast Fourier transform, PCA, and variance analysis) are often used.

Locally Weighted Regression (LWR) [1]

LWR applies PCR or PLS combined with a weighting scheme such that calibration samples closest to the sample to be predicted are given higher weight. Many variants are possible and the variant that probably should be preferred is one that has PLS for all samples with equal weights as the limiting case. In that case, when the data are linear and not clustered, the local method automatically becomes a global method. LWR was shown to perform well when clustered or nonlinear data were analyzed, but it has not been studied to the same extent as the preceding methods, so fewer diagnostics are available, and it is less well-known in which cases the method should be preferred or avoided. One disadvantage is that it is more time-consuming than other methods, because it requires that several sets of latent variables be determined. There also are indications that it is less robust towards wavelength shifts than other methods.

Radial Basis Functions (RBF) Combined with PLS [1]

RBF-PLS is a local method. It seems to perform particularly well for difficult data structures. Too little is known, however, about its practical use. It is not clear yet in which exact circumstances it works well and what pitfalls may occur.

Table 1: Concepts of the main chemometrics techniques applied in questioned document analysis [31]

Method classification	Multi-variate Analysis	Concept
Unsupervised pattern recognition techniques	PCA	This method projects multivariate data into a smaller space, reducing the spatial dimensionality of the original data set, without affecting the relation between samples. It is a type of controlled loss of information that is compensated by a better understanding of the data inside the data set. This method allows for visualizing and interpreting the differences between the variables and examining the relations that may exist between the samples. The method is also capable to detect samples with an anomalous behavior: the data projection makes the differences evident.
	HCA	This method is originated from numerical taxonomy. It is useful for reducing the dimensionality of large data sets by gathering samples into groups of samples that are most similar to each other. Thus, both internal homogeneity within groups and heterogeneity between groups are maximized. The results are presented in a hierarchical tree denominated dendrogram, and the tree branches show a degree of similarity between the samples. This similarity is calculated over a distance: similar samples from one group have a shorter distance between them when compared to samples from other groups.
Regression methods	PLS	It is a biased method that applies factor analysis, and it is the most popular regression method in chemometrics. The purpose of this technique is to estimate the space of original measures into one of reduced size. The model is built in a single step, in which the information from matrix "X" and values of interest are considered during the data decomposition and compression. Some restriction is imposed on the decomposition of matrix "X" (of samples and variables), directing it to a solution whose target is the property of interest. This is justifiable, if the values of the property of interest are reliable.
	MLR	The general purpose of this method is to quantitate the relationship between several independent or predictor variables and a dependent variable. The MLR model is built with descriptive variables using the least squares methods to minimize the residuals.
Supervised pattern recognition techniques	LDA	It is a linear combination of the sample's set original attributes, characterized by producing the maximum separation between two populations. The main objectives of this method are a) to confirm whether the groups are correctly discriminated, b) to classify unknown observations, and c) to verify which are the most important variables for the discrimination of these groups. It takes a different approach to consider the existence of classes for the data; it involves projecting the data distribution probability on the graphic's axes. Therefore, the method not only maintains but it also highlights a linear separation of the data, if it exists.
	PLS-DA	It is a method that determines which class an unknown sample belongs, based on the information provided to the system. The method applies multivariate regression by partial least squares (PLS), which has been previously discussed. PLS is an inverse calibration method, in which a direct relationship is sought between the instrumental response (matrix X) and the property of interest (matrix Y or vector y). The classification model is built using the same PLS model. However, in PLS-DA the property of interest is a categorical variable that describes the sample's

		class assignment. Generally, a value of 1 is assigned to the class of interest and a value of 0 is assigned to the other class.
	kNN	This deterministic model is defined based on a training set, followed by the classification model construction. When building the model, each sample of the training set is deleted and then classified based on the remaining samples. The distances between the excluded sample and the remaining samples of the training set are calculated in the dimensional space. The excluded sample is then classified according to a majority of “votes” from its closest neighbors, and the sample is attributed to the most voted class. This process is applied to all training samples, with a summary of successful analyzes and errors at the end.
Supervised pattern recognition techniques	SIMCA	This method assumes that the measured values for a group of similar samples will tend towards a uniform and modellable distribution. By increasing the number of samples, the uniform distribution becomes more visible. The probabilistic distribution allows for estimating the degree of certainty in the classification. This model allocates the main component model to be adjusted to each class in the training set, giving rise to a classifier for each one. The number of factors, suitable for modeling each class, can be determined by doing a cross-validation to maximize the predictive capacity of the individual models in relation to the training set. If it is necessary to include a new class, it is possible to build an independent model for each class by adding it to the existing model, without having to repeat the entire modeling process.
Other Methods	MF-ICA	Independent component analysis (ICA) is a computational method for separating a multivariate signal into additive subcomponents. This is done by assuming that the subcomponents are non-Gaussian signals and that they are statistically independent from each other. In MF-ICA we derive an expectation-maximization algorithm, in which the expectation step is performed using different medium field (MF) approaches: variational, linear response, and adaptive TAP (Thouless, Anderson and Palmer). The MF theories produce estimates of later source correlations of increasing quality, needed for the maximization step in the estimate for the multivariate signal and the noise.
	MANOVA	It is a test to perform the relationship analysis between several response variables and a common set of predictors at the same time. It requires continuous response variables and categorical predictors. MANOVA has several important advantages over performing several ANOVAs, one response variable at a time.
	SAM	It is an algorithm that allows for the measurement of spectral similarity between two spectra. This is expressed as a numerical scale (from 0–no similarity to 1–identical spectra). In this approach, selected spectra are treated as vectors in n-dimensional space, in which the number of dimensions is equal to the number of recorded spectral lines. This allows for the calculation of the spectral angle. It is worth mentioning that the SAM method is resistant to illumination variation.
	MCR	The MCR method was first created for process analysis purposes, but nowadays it is also applied to non-evolutionary multicomponent systems. Thus, the MCR method was created and classified as a two-way data analysis method, i.e., a valid method to analyze single data matrices. However, the general applications of MCR relates to the possibility to work with multi-way and multiset data structures, i.e., with several data tables simultaneously

	MCR-ALS	It is an algorithm that solves the MCR basic bilinear model, using a constrained Alternating Least Squares algorithm. The constraints improve the profiles interpretability in pure spectra matrices and the related concentration profiles for each of the compounds in the system.
--	---------	--

Software used for multivariate analysis

Table 2: List of software used for multivariate analysis [31]

Sr. no	Software	Multivariate Analysis
1.	SIMCA 15.0.2	PCA, HCA, OPLS
2.	Analyst® TF 1.6	ULT
3.	Chemostat®	PCA, HCA,
4.	SIMCA 13.0	PLS
5.	SPSS®	PCA, MLR, MANOVA
6.	Unscrambler X 10.4	PCA, LDA, PLSR
7.	Microsoft Excel	PCA
8.	PhotometrixPRO®	PCA, HCA, PLS-DA
9.	MATLAB®	MF-ICA, PLS, PLS-DA, MCR,PP, MCR-ALS
10.	Minitab®	PCA, HCA
11.	SPSS-20®	PCA, HCA, LDA, MLR, PLSR
12.	SPSS-15	PCA,
13.	XLSTAT 2011	PCA, HCA
14.	SPSS-12®	PCA,
15.	SPSS-16®	PCA, DA
16.	N/I	PCA, LS-SVM
17.	Statistica	PCA, HCA, SAM
18.	Unscrambler X 10.5.1	PCA, LDA
19.	Unscrambler X 10.3	PCA, SIMCA, LDA, QDA, DA
20.	JMP Pro 14	PCA, KNN, LDA
21.	Pirouette 3.11	PCA
22.	Unscrambler X 9.8 and 10.1	PCA, SIMCA, PLS-DA, SVM
23.	V-PARVUS 2009 package	PCA
24.	Microsoft Excel and SPSS-20®	PCA, HCA, LDA
25.	Unscrambler X	PCA, MANOVA
26.	Unscrambler v.9.7	PLS

TECHNIQUES USED IN CHEMOMETRICS [7]

Nowadays various spectroscopic techniques like HPLC, NIR, FTIR are being combined with various chemometric models like multivariate analysis methods, PLS, CLS, PCR, and so forth, for the evaluation of different pharmaceutical properties of tablets, powders, granules, and so forth. The most popular technique used is NIR spectroscopy.

APPLICATIONS [8]

Applications of chemometrics in pharmaceutical field:

Chemometrics can be broadly applied to exploratory analysis, regression analysis and classification of data studies. The specific applications of various chemometrics models in pharmaceutical field are as follows.

A Diagnosis and drug synthesis:

- Diagnosis and detection of significant change in patient's condition during medical



treatment and clinical care, prediction of the future medical state of the patient.

- Quality control of laboratory results, standardization and interpretation of laboratory tests in disease monitoring
- Drug synthesis, development and design, tools for drug discovery, SAR, drug mechanism, metabolomics, and proteo-chemometrics.

B. Powder flow properties [8,28]

To conclude the pharmaceutical properties such as mean particle size, angle of repose, tablet porosity, and tablet hardness. NIR spectra of the Antipyrine granules were measured. This was analysed by principal component regression analysis. With the increase in the water amount, the mean particle size of the granules was found to increase from 81 μ m to 650 μ m, and it was possible to make larger spherical granules with small particle size distribution using a high-speed mixer.

C. Water content determination in excipients [8,32]

Water content of hygroscopic pharmaceutical excipients largely affects the manufacturing processes and the performance of the final product. The water content of commonly used tablet disintegrants namely crospovidone, croscarmellose sodium and sodium starch glycolate was studied by simple linear regression.

D. Dissolution studies

Ana Rita explored the application of near-infrared spectroscopy and multivariate data analysis to monitor in-situ and in real-time dissolution tests of an immediate release formulation containing folic acid and four excipients.

E. Tablet-parametric method

Predicted the drug content and hardness of intact tablets of theophylline using artificial neural network and near-infrared spectroscopy. Tanabe et al. employed NIR spectroscopic methods in combination with principal component regression (PCR) analysis for predicting hardness of the

tablet formulations consisting of berberine chloride, lactose, and potato starch. The reflectance NIR spectra of various compressed tablets were used as a calibration set to establish a calibration model to calculate tablet hardness where in the predicted and the actual hardness values exhibited a straight line, an r^2 of 0.925.

F. Formulation development

Formulation and assessment of protein-loaded solid dispersions by nondestructive methods like powder X-ray diffraction (PXRD), near infrared chemical imaging (NIRCI) were performed in combination with principal component analysis and partial least square regression.

G. Salt and polymorph screening

Mixtures of polymorphs of carbamazepine can be detected with PCA score plots and that multivariate regression methods, such as PLS, can be used to estimate and determine the composition of these mixtures.

H. Pharmaceutical analysis

Various chemometric methods in combination with UV-Visible spectrophotometry, NIR spectroscopy, fluorescence spectroscopy, electroanalysis, chromatographic separation, and flow-injection analysis for the analysis of drugs in pharmaceutical preparations have been reported. Chemometrics is of great industrial importance in various chromatographic research areas as a lot of experimental work should be carried out with respect to optimization of different columns, test compounds, mobile phases and their pH, flow rate and peak shape parameters. The chromatographic techniques coupled with chemometric tools provide useful information on separation and elution time. The validation parameters like robustness and ruggedness are also best evaluated with the help of chemometrics. For binary mixture analysis HPLC is combined with different calibrating techniques like PLS, PCR, CLS, and so forth; hence, they are collectively called HPLC-CLS, HPLC-PCR, and HPLC-PLS. Several liquid

chromatographic methods in combination with chemometrics have been applied and reported in various method optimization and validation studies in diversified research areas. Chemometrics has also been widely used in combination with various spectroscopic techniques for the analysis of active molecules in dosage forms, plant materials and biological samples. Chemometrics- assisted simple UV-spectroscopic determination of Carbamazepine in human serum was reported and the results were compared with reference methods.

I. In metabolomics

Further multivariate chemometrics has been used in metabolomics (study of small molecule metabolite profiles) along with GC/MS and NMR techniques to characterize the metabolic profiles of bio fluids, understand the mechanism of pathogenesis and uncover potential biomarkers of disease progression.

J. Chemometric aids in medicine and pharmacy [32,33]

Present automated laboratory instruments in biological/medical research produce a vast volume of measurement data, which are difficult to absorb and interpret. Therefore, a challenging task of powerful mathematical and statistical methods of chemometrics is to reduce them and reveal all useful information. An important role of chemometrics in medicine was known a long time ago, shortly after it was widespread among life sciences. There exist several areas of medicine as well as pharmacy where the aid of chemometrics is essential and indispensable:

- Quality control of laboratory results, laboratory measurement standardization.
- Decision about diagnosis, confirming or excluding a diagnosis.
- Optimal clinical interpretation of laboratory tests in disease monitoring.
- Detection of significant change in patient's condition during medical treatment and

clinical care, prediction of the future medical state of the patient.

- Drug synthesis, development and design, tools for drug discovery, structure-activity relationships, drug mechanism.
- Chemometrics in metabolomics, proteo-chemometrics - a new bioinformatics technology.
- Identification of dopes and toxic substances; expert systems for toxicity classification.
- Prediction of bio-properties of various remedies, herbal medicine, and functional foods.

K. Quality control of laboratory tests, measurement standardization [31]

common way of evaluating the biochemical state of “normality” or “abnormality” is based on the statistically derived reference intervals, which serve as the standards by which the laboratory test results are judged. Careful design of experimental protocol is the key in carrying out any evaluation of clinical diagnostic value. Reference interval development has classically relied on concepts elaborated by the International Federation of Clinical Chemistry Expert Panel on Reference Values during the 1980s. An important part of quality control is statistical comparison of agreement of laboratory tests. Selection of an optimal measurement method is usually based on comparison of a newly developed procedure with the traditional one. In such a case both compared methods represent random variables - as neither of them is error-free. Consequently, the use of ordinary standard least squares regression method is not appropriate since the basic assumption about the error-free independent variable is violated, which may cause critical errors when using standard ways of regression.

L. Medical diagnosis confirmation or prediction, assessment of effectiveness of laboratory tests [31]



Advanced implementation of chemometric and statistical algorithms facilitates clarifying of many important practical applications in medicine. It was found that an appropriate exploitation of multivariate statistics, e.g. the outputs of principal component analysis, techniques of discriminant analysis, logistic regression, together with the standard statistical tools like correlation analysis, analysis of variance (ANOVA) as well as the ROC curves may enable a better medical diagnosis confirmation and/or prediction.

M) Monitoring of the health state of the patients [31]

Further task of chemometrics is its assistance in detecting the changes in patient's condition concerning progress of the disease and reaction to the medical treatment. Another target of chemometrics is determination of patient prognosis – prediction of the future medical state of a patient on the basis of present laboratory test results and clinical care. Several typical examples are shortly discussed in the next paragraphs.

N. Drug design, structure-activity and structure-property relationships [31]

Multivariate chemometric models can be used also for prediction of bio-properties from chemical data obtained either from various instrumental measurements or by calculation of the descriptors derived from the molecular structure. Drug development is a tedious and resource demanding process consisting of several steps. Chemometrics plays important role especially in the starting step – in the search of the compounds with a strong activity against the originators of diseases. In the last decades the QSAR (Quantitative Structure – Activity Relationships) became an important tool in drug discovery and in toxicology. The goal of QSAR is investigation how to predict the biological activity of a set of compounds in demand and elucidation of the specific action of the administered drug; it deals with generalization of relations between the chemical and biological

properties of the examined set of chemically resembling derivatives.

- Recently chemometrics is also used in toxicity assessment and prediction. It plays an important role in expert systems for toxicity classification, in identification of drugs and toxic substances in complex mixtures and in forensic chemistry.
- Further chemometrics can also be applied to the study of bioactive compounds in food. Some available chemometric software's: WINSI, NSAS, Vision, Sesame, SpectroMatrix, Anatec, Delight, GRAMS 386, Unscrambler, MatLab, Pirouette.

CONCLUSION:

- Chemometrics is the data-driven multidisciplinary science using methods frequently employed in core data-analytic disciplines such as multivariate statistics, applied mathematics, and computer science, in order to address problems in chemistry, biochemistry, medical and pharmaceutical fields, biology and chemical engineering.
- Different chemometrics models have been applied for the analysis of data of a particular manufacturing process, quality control test, or an instrumental output data with an aim to achieve maximum accuracy, precision, and robustness in a variety of matrices.
- However, chemometrics has advanced in parallel with advancing analytical instrumentation and computational capability. It has expanded widely from its beginnings in curve fitting and QA into a variety of other areas, including multivariate calibration, pattern recognition, and mixture resolution.
- There is a huge potential for chemometrics in modern analytical and process chemistry as it provides accurate results in a simple and rapid manner.
- One vision of the future is that chemometrics will be at the heart of the automated



laboratory. As instruments acquire data more and more rapidly, there will be the need to optimize the instrumental and interpretative functions.

- Applications, new methodology and software must go hand in hand for quality products in chemometrics.
- Thus, overall discussion provided above chemometrics, its models, application and software used would be of broader interest to the researchers working in the development of analytical methods.

REFERENCES:

1. El-Gindy A, Hadad GM. Chemometrics in pharmaceutical analysis: an introduction, review, and future perspectives. *Journal of AOAC International* 2012;95(3):609-23.
2. Gummadi S, Chandaka PK. Chemometrics approach to drug analysis—An overview. *American Journal of PharmTech. Research* 2019;9:1-3.
3. Hopke PK. The evolution of chemometrics. *Analytica Chimica Acta*. 2003;500(1-2):365-77.
4. Gemperline P: A Text Book of Practical guide to chemometrics. CRC press 2006.
5. Harvey DT: A Text Book of Chemometrics Using R.
6. Kumar N, Bansal A, Sarma GS, Rawal RK. Chemometrics tools used in analytical chemistry: An overview. *Talanta* 2014;123:186-99.
7. Pawar HA, Kamat SR. Chemometrics and its application in pharmaceutical field. *Journal of Physical Chemistry and Biophysics* 2014;4(6):169.
8. Gummadi S, Chandaka PK. Chemometrics approach to drug analysis—An overview. *Am. J. Pharm. Tech. Res.* 2019;9:1-3.
9. Wentzell PD, Andrews DT, Hamilton DC, Faber K, Kowalski BR. Maximum likelihood principal component analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society* 1997;11(4):339-66.
10. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences* 2016 ;374(2065):20150202.
11. Henrion R, Henrion G, Onuoha GC. Multi-way principal components analysis of a complex data array resulting from physicochemical characterization of natural waters. *Chemometrics and intelligent laboratory systems* 1992 ;16(1):87-94.
12. Molloy M, Martin EB. Application of multiway principal component analysis for identification of process improvements in pharmaceutical manufacture. *IFAC Proceedings Volumes* 2013 ;46(32):283-8.
13. Wold S, Eriksson L, Trygg J, Kettaneh N. The PLS method—partial least squares projections to latent structures—and its applications in industrial RDP (research, development, and production). Unea University 2004.
14. Wold S, Ruhe A, Wold H, Dunn, Iii WJ. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing* 1984 ;5(3):735-43.
15. Barker M, Rayens W. Partial least squares for discrimination. *Journal of Chemometrics: A Journal of the Chemometrics Society* 2003;17(3):166-73.
16. Garthwaite PH. An interpretation of partial least squares. *Journal of the American Statistical Association* 1994;89(425):122-7.
17. Bro R. Multiway calibration. multilinear pls. *Journal of chemometrics* 1996; 10(1):47-61.
18. Rinnan Å, Riu J, Bro R. Multi-way prediction in the presence of uncalibrated interferents. *Journal of Chemometrics: A Journal of the Chemometrics Society* 2007; 21(1-2):76-86.



19. Bro R, Kiers HA. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics: A Journal of the Chemometrics Society* 2003; 17(5):274-86.
20. Kiers HA, Ten Berge JM, Bro R. PARAFAC2—Part I. A direct fitting algorithm for the PARAFAC2 model. *Journal of Chemometrics: A Journal of the Chemometrics Society* 1999; 13(3-4):275-94.
21. Bro R, Andersson CA, Kiers HA. PARAFAC2—Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics: A Journal of the Chemometrics Society* 1999 ;13(3-4):295-309.
22. Bro R. PARAFAC. Tutorial and applications. *Chemometrics and intelligent laboratory systems* 1997 ;38(2):149-71.
23. Andersen CM, Bro R. Practical aspects of PARAFAC modeling of fluorescence excitation-emission data. *Journal of Chemometrics: A Journal of the Chemometrics Society* 2003 ;17(4):200-15.
24. Geladi P, Forsström J. Monitoring of a batch organic synthesis by near-infrared spectroscopy: modeling and interpretation of three-way data. *Journal of Chemometrics: A Journal of the Chemometrics Society* 2002 ;16(7):329-38.
25. Smilde AK. Comments on three-way analyses used for batch process data. *Journal of Chemometrics: A Journal of the Chemometrics Society* 2001; 15(1):19-27.
26. Kong X, Zhu W, Zhao Z, Li X, Wang H, Chen R, Chen C, Zhu F, Guo X. Fluorescence spectroscopic determination of triglyceride in human serum with window genetic algorithm partial least squares. *Journal of Chemometrics* 2012; 26(1-2):25-33.
27. Kramer R: A Text Book of Chemometric techniques for quantitative analysis. CRC press; 1998 Jun 10.
28. Gemperline P: A Text Book Practical guide to chemometrics. CRC press; 2006.
29. dos Santos HT, de Oliveira AM, de Melo PG, Freitas W, de Freitas AP. *Chemometrics: theory and application. Multivariate Analysis in Management, Engineering and the Sciences* 2013.
30. VONČINA DB. *Chemometrics in analytical chemistry. Applied Natural Sciences* 2009;211.
31. Gorziza¹ RP, González¹ M, de Carvalho CM, Ortiz RS, Ferrão MF, Limberger RP. Chemometric approaches in questioned documents. *Brazilian Journal of Analytical Chemistry* 2022; 9(34):35-51.
32. Mocák J. *Chemometrics in medicine and pharmacy. Nova Biotechnologica et Chimica.* 2012;11(1):11-25.
33. Singh I, Juneja P, Kaur B, Kumar P. Pharmaceutical applications of chemometric techniques. *International Scholarly Research Notices* 2013;2013(1):795178.

HOW TO CITE: Drashti N. Bhalodia , Dharmendrasinh A. Baria , A Review On Chemometrics In Pharmaceutical Analysis, *Int. J. of Pharm. Sci.*, 2024, Vol 2, Issue 9, 58-78. <https://doi.org/10.5281/zenodo.13625694>



