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.

Contents

* 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

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