Available online at www.sciencedirect.com
ScienceDirect
Journal of Chromatography A, 1158 (2007) 2–14
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Review

Statistical designs and response surface techniques for the optimization of chromatographic systems

Sergio Luis Costa Ferreira [a,*], Roy Edward Bruns [b], Erik Galvão Paranhos da Silva [a], Walter Nei Lopes dos Santos [c], Cristina Maria Quintella [a], Jorge Mauricio David [a], Jailson Bittencourt de Andrade [a], Marcia Cristina Breitkreitz [b], Isabel Cristina Sales Fontes Jardim [b], Benicio Barros Neto [d]

[a] *Universidade Federal da Bahia, Instituto de Química, Campus Universitário de Ondina, Salvador, Bahia 40170-290, Brazil*
[b] *Universidade Estadual de Campinas, Instituto de Química, Campinas, São Paulo 13084-971, Brazil*
[c] *Universidade Estadual de Santa Cruz, Departamento de Ciências Exatas, Ilhéus, Bahia 45650-000, Brazil*
[d] *Universidade Federal de Pernambuco, Departamento de Química fundamental, Recife, Pernanbuco 50740-540, Brazil*

Available online 18 March 2007

Abstract

This paper describes fundamentals and applications of multivariate statistical techniques for the optimization of chromatographic systems. The surface response methodologies: central composite design, Doehlert matrix and Box–Behnken design are discussed and applications of these techniques for optimization of sample preparation steps (extractions) and determination of experimental conditions for chromatographic separations are presented. The use of mixture design for optimization of mobile phases is also related. An optimization example involving a real separation process is exhaustively described. A discussion about model validation is presented. Some applications of other multivariate techniques for optimization of chromatographic methods are also summarized.

Keywords: Review; Multivariate techniques; Doehlert matrix; Box–Benhken design; Central composite design; Chromatographic methods; Optimization

Contents

* Corresponding author. Tel.: +55 71 9117 8046; fax: +55 71 3235 5166.
E-mail address: slcf@ufba.br (S.L.C. Ferreira).

0021-9673/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2007.03.051

1. Introduction

Chromatographic analysis usually involves three steps: sample preparation, compound separation and compound quantification. Of these, the steps of sample preparation and compound separation have been frequently optimized employing multivariate statistical techniques.

The multivariate statistical methods most used in chromatography and indeed in chemistry in general can be conveniently classified according to how one decides which experiments are to be executed. All methods require the user to supply minimum and maximum values for each factor that defines the experimental domain to be investigated during the optimization procedure. The combinations of the different factor levels used to perform the actual experiments are then decided by which multivariate technique is employed.

The most commonly used designs to determine response surfaces are the full and fractional factorial designs and the more complex central composite, Box–Behnken, Doehlert and mixture designs. Although the factorial designs can be used to determine simple response surfaces that are linear in all of the investigated factors, they are normally used to determine which experimental factors are the most important to investigate and which factors do not significantly affect the experimental results. Here their use is discussed as a first stage in a multivariate investigation where a linear response surface is determined. For a two-factor case, the response surface is given by the linear model

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 \quad (1)$$

If the interaction term is negligible the response surface is planar. The more important the interaction term, the greater is the degree of twisting the planar response surface experiences. If the linear model is not sufficient to represent the experimental data adequately, more experiments can be performed in addition to those of factorial design. The central composite design is often formed in this way and its results can be used to determine a quadratic response surface

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 \quad (2)$$

that has curvature and can be used to predict factor levels that produce maximum or minimum response values. The Box–Behnken and Doehlert designs can also be used to determine these kinds of response surfaces and optimize chromatographic factors such as temperature, column characteristics and flow rates. Mixture designs are used to vary proportions of mixture ingredients such as the solvent proportions of a mobile phase. They differ from the other designs that optimize intensive properties like temperature or extensive ones like the total quantity of material used in an experiment.

Rather than executing experiments that have been planned according to a statistical design, optimization can be done by performing experiments that are indicated by a sequential simplex. The sequential simplex can be useful in certain situations, such as instrument optimization when one is only trying to improve a single response and the experiments are very fast. The simplex algorithm even permits automatic optimization that does not necessarily require user intervention. However, most problems in chromatography have multiple responses that need to be simultaneously optimized, like the retention factors of various chromatographic peaks for which a single response function is inadequate. In this case the simplex procedure is not very efficient.

One big advantage of applying the simplex procedure is that the user does not have to understand even basic statistics to do a successful optimization. No decision-making is necessary. After feeding initial factor levels and their proposed changes into the computer, the user simply performs experiments at the factor levels indicated by the simplex algorithm. Three different algorithms can be applied, the basic simplex, the modified simplex and the super-modified simplex. Our aim here is to discuss those methods that are generally more applicable in chromatography. For this reason the interested reader is referred to specialized sequential simplex publications [1–4].

Multivariate optimization of chromatographic systems can be carried out using the following procedure

(i) Choose a statistical design to investigate the experimental region of interest.
(ii) Perform the experiments in random chronological order.
(iii) Perform analysis of variance (ANOVA) on the regression results so that the most appropriate model with no evidence of lack of fit can be used to represent the data. Validation is often not reported in response surface applications even though it is necessary for knowing whether the system is really optimized or not.

Modern commercial statistical computer programs are available to help the research worker carry out each of these steps. A wide variety of designs are presented to the researcher for selection. Options are available for determining the random order for experiment execution. The programs also allow the user to select the models, linear, quadratic and others, he would like to test. After calculating the model coefficients and their standard errors an ANOVA is available to the user to verify the quality of model fit to the data so the researcher can choose the best model to represent the data.

In this section, the experimental designs most frequently used in chemistry for response surface determination are described so the reader can have a basis for choosing designs for his applications. Random execution of experiments is recommended so that an accurate estimation of experimental error is obtained. The regression step does not require user intervention so it is not described here and the reader is referred to basic sources on the subject [5–8] to learn how the computer carries out the calculation. Then the validation of tentative models using ANOVA is detailed since this task requires a decision on the part of the researcher about which models are adequate to represent the data and which models should be rejected because they suffer from significant lack-of-fit to the data.

The principal chemometric tools used for optimization of chromatographic systems are: two-level full factorial, central composite, Box–Behnken, Doehlert and mixture designs [9,10].