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].

2. Factorial and central composite designs

Central composite designs (CCD) combine [5–8] two-level full or fractional factorial designs with additional axial or star points and at least one point at the center of the experimental region being investigated. It allows the determination of both linear and quadratic models. The CCD is a better alternative to the full factorial three-level design since it demands a smaller number of experiments while providing comparable results.

In general, a CCD for k factors, coded as $(x_1, \ldots, x_k)$, consists of three parts:

(1) A factorial (or cubic) design, containing a total of n_{fact} points with coordinates $x_i = -1$ or $x_i = +1$, for $i = 1, \ldots, k$;
(2) An axial (or star) part, formed by $n_{\text{ax}} = 2k$ points with all their coordinates null except for one that is set equal to a certain value α (or $-\alpha$);
(3) A total of n_c runs performed at the center point, where, of course, $x_1 = \ldots x_k = 0$.

To build a central composite design, we need to specify each of these three parts. We have to decide how many cubic points to use and where they will be, what will be the value of α, and how many replicate runs should be conducted at the center point. Two designs are presented in Table 1 and Fig. 1. The design on the left accommodates two factors. The first four runs (2^2 factorial design) make up the cubic part, the star design the last four (with $\alpha = \sqrt{2}$) and there are three replicate runs at the center point. The three-factor case on the right has $2^3 = 8$ runs with four center points and six axial points with $\alpha = 1.683$.

Note that the codified values given in Table 1 only specify the relative positions of the experimental design points. In part, this determines the precision of the model coefficients and the values predicted by the model. The researcher defines the actual laboratory values that are linearly related to the codified values.

Table 1
Coded factor levels for central composite designs for two- and three-factor systems

Two-factor		Three-factor		
x_1	x_2	x_1	x_2	x_3
−1	−1	−1	−1	−1
1	−1	1	−1	−1
−1	1	−1	1	−1
1	1	1	1	−1
0	0	−1	−1	1
0	0	1	−1	1
0	0	−1	1	1
−1.414	0	1	1	1
1.414	0	0	0	0
0	−1.414	0	0	0
0	−1.414	0	0	0
		0	0	0
		−1.683	0	0
		1.683	0	0
		0	−1.683	0
		0	1.683	0
		0	0	−1.683
		0	0	1.683

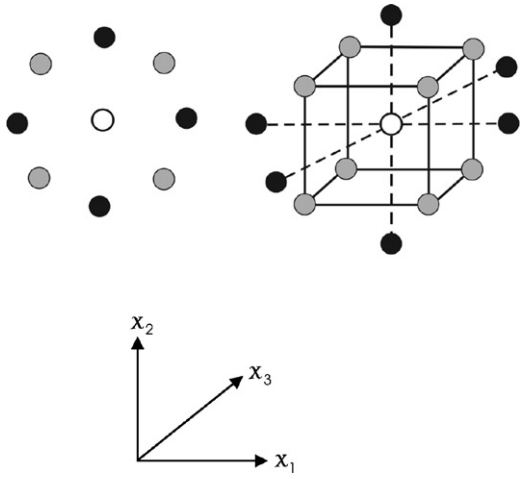

Fig. 1. Central composite designs for two and three factors. The gray dots form the cubic part—the runs of the 2^2 and 2^3 factorial. The black dots represent the star parts.

The cubic points in Table 1 and Fig. 1 are the same as those of full factorial designs although this is not strictly necessary. Depending on the number of factors, it might even be unadvisable, because it would require too many experiments.

The total number of distinct levels in a central composite design is $n_{\text{fact}} + 2k + 1$. Therefore, there are $2^k + 2k + C_0$ total points where C_0 is the number of center points. The complete quadratic model for k factors contains $(k+1)(k+2)/2$ parameters and is given by

$$y = \beta_0 + \sum_i \beta_i x_i + \sum_i \beta_{ii} x_i^2 + \sum_{i<j} \sum_j \beta_{ij} x_i x_j + \varepsilon. \quad (3)$$

With two factors, this model has six parameters. Since the two-factor design in Table 1 has nine different combinations of levels, we could estimate all of the model's parameters using only two cubic points, corresponding to one of the two 2^{2-1} fractions. In such a simple design, the economy is very small and hardly justifies destroying the symmetry of the cubic part, but this procedure – choosing a fractional design instead of a complete factorial to define the cubic points – becomes more attractive as the number of factors increases.

The value of α usually ranges from 1 to $\sqrt{k}$. When $\alpha = \sqrt{k}$, the cubic and axial points are located on the (hyper)surface of a (hyper)sphere, and the design is called spherical. This is the case of the two-factor design in Table 1 where all of the peripheral points are on the circumference of a circle. At the other extreme, when $\alpha = 1$, the axial points are located at the centers of the sides of the square part of the two-factor design and at the center of the faces of the cubic part of the three-factor design. This type of design is advantageous when the experimental space is square or cubical, which occurs in a natural way when the factors are varied independently of one another. It also has the advantage of only requiring three factor levels, which can be a big help if one of the factors is qualitative.

If we choose $\alpha = \sqrt{k}$, the star points will be placed farther from the center point as the number of factors increases. This choice should be made – if it is made at all – with much care, because we run the risk of leaving too much of the intermediate region uninvestigated. With nine factors, for example, α would be equal to 3. The experiments would tell us nothing about the behavior of the response surface in the 1–3 intervals along each axis.

Box and Hunter [11] proposed the rotatability concept as a criterion for choosing the value of α. A design is called rotatable if the variance of its estimates, $V(\hat{y})$, depends solely on the distance from the center point, that is, if the precision of the predicted response is the same for all points situated on a (hyper)sphere around the center of the design. For a design whose cubic portion is a complete factorial or a fractional factorial of resolution V, it can be demonstrated that rotatability is attained when $\alpha = \sqrt[4]{n_{\text{fact}}}$. Even if the resolution is not exactly 5, this expression serves as a guide for choosing an α value, which in any case should also be analyzed for its convenience and feasibility. If, for example, we are interested in investigating the region closer to the faces of the hypercube, then it is better to choose an α value smaller than the rotatable one. On the other hand, it could happen that the chosen α value results in some impracticable runs. In this case, we would have to define new experimental conditions for these runs.

The replicates at the center point have two main objectives: provide a measure of pure error and stabilize the variance of the predicted response. To stabilize the variance, a rule of thumb is to make 3–5 repeated runs if α is close to $\sqrt{k}$, and only one or two if it is close to 1. Between these extremes, 2–4 replicates are indicated. To obtain a more precise estimate of error, the more replicates the better.

Since three distinct parts form central composite designs, we can build them sequentially, according to necessity. If we happen to be in a region of the response surface where curvature is not important, there is no need to fit a quadratic model. The cubic part of the design is sufficient for fitting a linear model, from which we can rapidly move to a more interesting region extending the investigation. Even if we are in doubt about possible curvature, we can test its significance using the runs at the center point. Then, if the curvature is found significant, we can complete the design with the axial points. Actually, we would be performing the runs of the complete design in two blocks—first the cubic and then the axial one.

Suppose that the response values in the axial block contain a systematic error in relation to the response values obtained in the first block. Under certain conditions, this error will not affect the coefficient estimates for the model, that is, the block effect will not confound itself with the effects of the other factors. This will occur if the design blocking is orthogonal, which in turn depends on the α value. Blocking will be orthogonal if

$$\alpha = \sqrt{\frac{n_{\text{fact}}(n_{ax} + n_{\text{c,ax}})}{2(n_{\text{fact}} + n_{\text{c,fact}})}}, \quad (4)$$

where $n_{\text{c,fact}}$ and $n_{\text{c,ax}}$ are the number of runs at the center point in the cubic and axial blocks, respectively [6]. In general, when we make a design with orthogonal blocks we are sacrificing its rotatability, but some designs do exist for which the two conditions are approximately satisfied, and others for which both are exactly satisfied. For example, the two factorial design in Table 1 is both rotatable and orthogonal although the rotatable three-factor design is only approximately orthogonal. More about CCD can be found in several excellent texts on response surface methodology [12–14].

2.1. Application of central composite designs for optimization of chromatographic systems

The central composite design is one of the chemometrics techniques most employed for optimization of chromatographic systems [15,57–90]. It has been used for the determination of the critical conditions of experimental factors during the optimization of extraction steps [15,57–66], derivatization reactions [67–69], separation steps [70–88], quantification processes [89] and also robustness studies [90] involved in chromatographic methods. These methods were performed for determination of organic and inorganic species in samples such as water [15,68,71,72,75,88], drugs [69,73,79,85], biological matrices [59,62,67,74,82,86], wines [70,77], human plasma [63,65,66,78] and others. These applications using several chromatographic techniques for quantification are summarized in Table 2.

Other applications include use of the central composite design for optimizing the preparation of zirconized silica, which was characterized for use as support in HPLC [16].

Another paper proposes the modification of silica with zirconium oxide followed by sorption and thermal immobilization of poly(methyltetradecylsiloxane) (PMTDS). This material was used for preparation of a reversed stationary phase for HPLC. The immobilization step was optimized using a central composite design [17].

An off-line system consisting of ion-pair reversed-phase liquid chromatography was proposed for separating arsenite, arsenate, monomethylarsonate, dimethylarsinate, selenite, selenate, selenocystamine, selenocystine, selenomethionine and selenoethionine. Elution conditions were optimized using the central composite design. The method was applied for the determination of these species in the environment and in mammals, using as analytical technique graphite furnace atomic absorption spectrometry [18]. Another paper by this research group used this system for the separation and determination of monomethylarsonate, dimethylarsinate, selenomethionine and selenite in tap water employing inductively coupled plasma mass spectrometry. The optimization of the separation step was also done using a central composite design [19].

3. Box–Behnken designs

Box–Behnken designs constitute an alternative to central composite designs [20]. They are a class of rotatable or nearly rotatable second-order designs based on three-level incomplete factorial designs. Fig. 2 shows the Box–Behnken design for three factors, corresponding to the coded values listed in Table 3. It is

Table 2
Applications of the central composite design for the optimization of chromatographic methods

Analytes	Samples	Optimization	Chromatographic technique	Ref.
MTBE	Water	Extraction step	GC–FID	[15]
Oleamide and erucamide	Polyethylene films	Extraction step	GC–FID	[57]
Chloroanisoles	Oak barrels	Extraction step	GC–FID	[58]
Cocaine and benzoylecgonine	Coca leaves	Extraction step	GC–FID/CE–UV	[59]
Alkyl- and methoxy-phenolic compounds	Biomass smoke	Extraction step	GC–MS	[60]
Pesticide residues	Soils	Extraction step	GC–MS	[61]
Volatile compounds	Evodia fruits	Extraction step	CG–MS	[62]
1,4-Dihydropyfidines	Human plasma	Extraction step	HPLC	[63]
Antioxidants	Low-density polyethylene	Extraction step	HPLC	[64]
Valsartan and its metabolite	Human plasma.	Extraction step	HPLC	[65]
Prostaglandin E-2	Human plasma	Extraction step	HPLC–MS	[66]
Mercury and methylmercury	Biological matrices	Derivatization step	GC	[67]
Acidic drugs	Sewage water	Derivatization step	GC–MS	[68]
Phosphoric and amino acid group	Pesticides	Derivatization step	GC–MS	[69]
Volatile phenols	Wines	Separation step	GC	[70]
Fuel dialkyl ethers and BTEX	Water	Separation step	GC–FID	[71]
Fluoride, chloride, nitrite, bromide, nitrate and sulphate	Water	Separation step	IC	[72]
2,4,6-Trichloroanisole and guaiacol	Cork stoppers	Separation step	GC–MS	[73]
Organochlorine pesticides	Plant infusions	Separation step	GC–MS	[74]
2,4,6-Trichloroanisole (TCA)	Disinfected water	Separation step	GC–ECD	[75]
Additives	Polyethylene films	Separation step	HPLC	[76]
Biogenic amines	Red wines	Separation step	HPLC	[77]
Citalopram, fluoxetine and paroxetine	Plasma and whole blood	Separation step	HPLC	[78]
Erytromycin	Drugs	Separation step	HPLC	[79]
Oxytetracycline and its impurities	Oytetracycline base	Separation step	HPLC	[80]
Sudan dyes	Colorant Sudan I–IV	Separation step	HPLC	[81]
Organic acid in tobacco	Tobacco	Separation step	AMMS-ICE II	[82]
Chlorophenol isomers	–	Separation step	MEKC	[83]
Nine anthraquinones and bianthrones	Rhubarb	Separation step	MEKC	[84]
Six angiotensin-II-receptor antagonists	Capsules of each compound	Separation step	MEKC	[85]
Triazolopyrimidine sulfoanilide herbicides	Soy milk	Separation step	CE–MS	[86]
Chiral/chlorthalidone	–	Separation step	cLC	[87]
Acrylamide	Drinking water	Separation step	IEC–MS	[88]
Chloropicrin	Soil	Quantification step	GC–MS	[89]
Carboxylic acids	Industrial reaction mixtures	Robustness study	HPLC	[90]

GC–FID: gas chromatography with flame ionisation detection; CE–UV: capillary electrophoresis with UV detection; AMMS-ICE II: anion micromembrane suppressor ion chromatography exclusion; MEKC: micellar electrokinetic capillary chromatography; IPC–ICP-MS: ion-pair chromatography coupled with inductively coupled plasma mass spectrometric detection; IEC–MS: ion-exclusion chromatography–mass spectrometry; MTBE: methyl tert-butyl ether; BTEX: benzene, toluene, ethylbenzene and xylenes; CE–MS: capillary electrophoresis–mass spectrometry; IC: ion chromatography; GC–ECD: gas chromatography with electron-capture detection; cLC: capillary liquid chromatography (cLC).

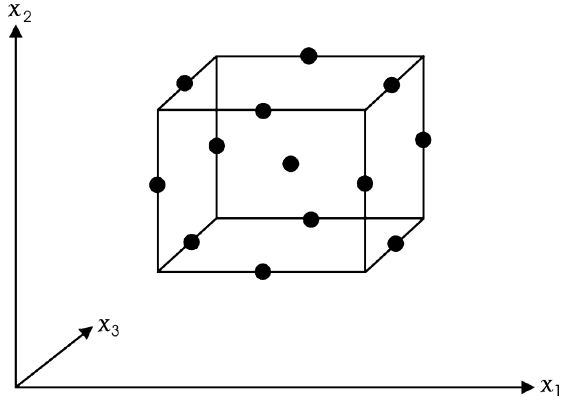

Fig. 2. Points representing the experimental runs of a three-factor Box–Behnken design.

easy to see that this design consists of three parts of four runs. Within each part, two factors are arranged in a full two-level design, while the level of the third factor is set at zero. The points lie on the surface of a sphere centered at the origin of the coordinate system and tangential to the midpoint of each edge of the cube.

Compared to the central composite design, this design has some advantages. The three-factor Box–Behnken design requires only 12 runs plus the replicates at the center point, whereas its central composite counterpart has 14 non-center points. In general the number of experimental points is given by $2k(k-1)+C_0$. Also, each factor is studied at only three levels, an important feature in some experimental situations, as we have already mentioned. On the other hand, using $\alpha = 1$ in a central composite design also results in three levels for each

Table 3
Coded factor levels for a Box–Behnken design for a three-variable system

x_1	x_2	x_3
−1	−1	0
1	−1	0
−1	1	0
1	1	0
−1	0	−1
1	0	−1
−1	0	1
1	0	1
0	−1	−1
0	1	−1
0	−1	1
0	1	1
0	0	0
0	0	0
0	0	0
0	0	0

factor. In most real applications these differences are probably not decisive in determining which design to use, at least for this number of factors. However, since Box–Behnken designs do not contain combinations where all the factors are at their higher or lower levels, they may be useful in avoiding experiments under extreme conditions, for which unsatisfactory results might occur. Conversely, they are not indicated for situations in which we would like to know the responses at the extremes, that is, at the vertices of the cube.

Box–Behnken designs for four and five factors can be arranged in orthogonal blocks, as shown in Table 4. In this table, each (±1, ±1) combination within a row represents a full 2^2 design. Dashed lines separate the different blocks. Because of the block orthogonality, the second-order model can be augmented to include block effects without affecting the parameter estimates, that is, the effects themselves are orthogonal to the block effects. This orthogonal blocking is a desirable property when the experiment has to be arranged in blocks and the block effects are likely to be large.

Table 4
Coded factor levels for Box–Behnken designs for four and five factors

Four-factor				Five-factor				
x_1	x_2	x_3	x_4	x_1	x_2	x_3	x_4	x_5
±1	±1	0	0	±1	±1	0	0	0
0	0	±1	±1	0	0	±1	±1	0
0	0	0	0	0	±1	0	0	±1
				±1	0	±1	0	0
±1	0	0	±1	0	0	0	±1	±1
0	±1	±1	0	0	0	0	0	0
0	0	0	0					
				0	±1	±1	0	0
±1	0	±1	0	±1	0	0	±1	0
0	±1	0	±1	0	0	±1	0	±1
0	0	0	0	±1	0	0	0	±1
				0	±1	0	±1	0
				0	0	0	0	0

3.1. Application of Box–Behnken designs for optimization of chromatographic methods

The Box–Behnken design has not been frequently used for optimization of chromatographic processes. It has been employed for the determination of critical conditions in extraction steps [21,22], derivatization reactions [23,24] and also separation steps [25–27], in chromatographic methods involving PC [25], MECC [26], GC–FID [21], GC–MS [23] and HPLC [22,24,27]. Table 5 summarizes these applications of this design in optimization of chromatographic methods.

4. Doehlert designs

The designs we have discussed so far are all symmetrical—that is, they have the same number of levels for each factor under study. Sometimes, however, it is advantageous to use designs where different factors are studied at different numbers of levels. A simple example is a 2×3 factorial design. Modern statistical design programs provide a wide variety of mixed k-factorial designs of the type $2^p 3^{k-p}$, where $p < k$ and p factors are studied at two levels and the other $k - p$ factors at three levels.

Doehlert designs comprise another class of experimental designs, with which different factors can be studied at different numbers of levels [28–32]. They are attractive for treating problems where specific information about the system indicates that some factors deserve more attention than others, but they also have other desirable characteristics.

All Doehlert designs are generated from a regular simplex, a geometrical figure containing $k + 1$ points, where k is the number of factors. For two factors, the regular simplex is an equilateral triangle. For Doehlert designs of type D-1, which are the most popular, the coordinates of this triangle are those given in the first three lines of Table 6, in coded units. The other runs of the design are obtained by subtracting every run in the triangle from each other, as shown in the table. This corresponds to moving the simplex around the design center.

For any number of factors k, one of the points of the simplex is the origin, and the other k points lie on the surface of a sphere with radius 1.0 centered on the origin, in such a way that the distances between neighboring points are all the same. Each of these points subtracted from the other k points forms k new points, so the design matrix has a total of $k^2 + k + 1$ points. Since the points are uniformly distributed on a spherical shell, Doehlert [29] suggested that these designs be called uniform shell designs. The coordinates of the D-1 designs for three and four factors are presented in Table 7. Note that the design for $k = 3$ is the same as the corresponding Box–Behnken design.

Despite being spherical, Doehlert designs have none of the classical properties of response surface designs. They are neither orthogonal nor rotatable, and the variance of the predicted values is not uniform over the experimental range. However, they have other interesting features that make them advantageous in some scenarios.

Perhaps their most important property is the ability for uniform space-filling, which is unaffected by rotation. This is very

Table 5
Applications of the Box–Behnken design in the optimization of chromatographic methods

Analyte	Sample	Optimization	Chromatographic technique[a]	Ref.
Fatty acid composition	Castor oil	Extraction step	GC–FID	[21]
Organochlorine pesticides	Sediments	Extraction step	HPLC	[22]
Aminoglycoside antibiotics	–	Derivatization reaction	GC–MS	[23]
Aliphatic aldehydes	–	Derivatization reaction	HPLC	[24]
Aminoacids hydroxamates	–	Separation step	PC	[25]
Captopril	Commercial drugs	Separation step	HPLC	[27]
Mixture of sulphonamides, β-lactam antibiotics and dehydrofolate	–	Separation step	MECC	[26]

PC: paper chromatography; MECC: micellar electrokinetic chromatography.
[a] Analytical technique for quantification.

Table 6
Coded factor levels for the two-factor Doehlert D-1 design

Run	x_1	x_2	Subtraction
1[a]	**0.0**	**0.0**	
2[a]	**1.0**	**0.0**	
3[a]	**0.5**	**0.866**	
4[b]	−1.0	0.0	**1–2**
5[b]	−0.5	−0.866	**1–3**
6[b]	−0.5	0.866	**3–2**
7[b]	0.5	−0.866	**2–3**

[a] The runs in bold face are those defining the initial simplex.
[b] The other runs are obtained by subtracting every run from each other.

convenient when one wishes to cover an experimental range, no matter how irregular, with a uniform grid of points. Doehlert designs then can be displaced to more promising regions while preserving some of the runs already carried out. This property is illustrated in Fig. 3. The initial two-factor design is a hexagon, where the letters **BCDEFG** denote its vertices. These points, along with the **A** center point, define five levels for the x_1 variable but only three levels for x_2. If the researchers decide to shift the initial design to higher values of x_1 and lower values of x_2 the **PONLGFM** hexagon can be used, which includes vertices **F** and **G** from the initial configuration. If experimentation is very costly or lengthy, the values already observed for the **F** and **G** vertices can be retained. In case the researchers wish to raise the levels of both x_1 and x_2, displacement to the **BGLKJIH** hexagon is warranted. This time, vertices **B** and **G** belong to the initial hexagon. Finally, if it is decided to shift the levels of only one variable, say x_1, the experimenters might perform new runs only at the **Q**, **R** and **S** vertices to complete the new Doehlert design **AESRQCD**, which has two lower levels along x_1 while keeping the same levels for x_2. Note that the space-filling property of Doehlert designs is clearly illustrated in Fig. 3. No gaps in the experimental region are left as one hexagon substitutes another.

Compared to central composite or Box–Behnken designs, Doehlert designs are more economical, especially as the number of factors increase. The basic hexagon in Fig. 3 has six points

Table 7
Coded factor levels for the three- and four-factor Doehlert D-1 designs

Run	Three-factor			Four factor			
	x_1	x_2	x_3	x_1	x_2	x_3	x_4
1	0	0	0	0	0	0	0
2	1	0	0	1	0	0	0
3	0.5	0.866	0	0.5	0.866	0	0
4	0.5	0.289	0.817	0.5	0.289	0.817	0
5	−1	0	0	0.5	0.289	0.204	0.791
6	−0.5	−0.866	0	−1	0	0	0
7	−0.5	−0.289	−0.817	−0.5	−0.866	0	0
8	0.5	−0.866	0	−0.5	−0.289	−0.817	0
9	0.5	−0.289	−0.817	−0.5	−0.289	−0.204	−0.791
10	−0.5	0.866	0	0.5	−0.866	0	0
11	0	0.577	−0.817	0.5	−0.289	−0.817	0
12	−0.5	0.289	0.817	0.5	−0.289	−0.204	−0.791
13	0	−0.577	0.817	−0.5	0.866	0	0
14				0	0.577	−0.817	0
15				0	0.577	−0.204	−0.791
16				−0.5	0.289	0.817	0
17				0	−0.577	0.817	0
18				0	0	0.613	−0.791
19				−0.5	0.289	0.204	0.791
20				0	−0.577	0.204	0.791
21				0	0	−0.613	0.791

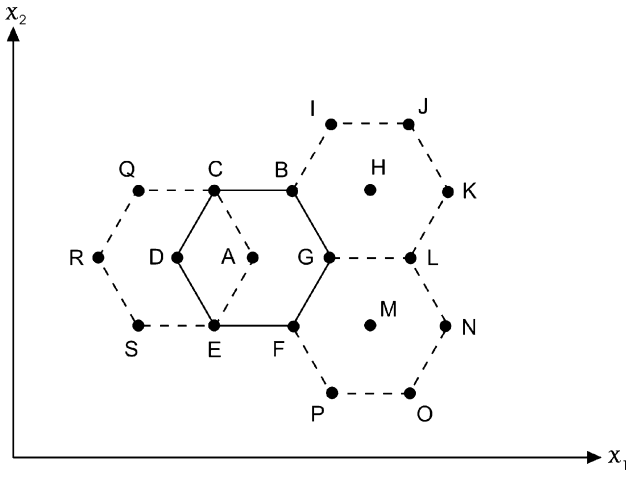

Fig. 3. Hexagonal Doehlert two-factorial design with three possible displacements in the experimental space.

lying on a circumference around the center point, whereas the two-factor central composite design has eight points, also lying on a circumference surrounding its center point. Likewise, the three-factor Doehlert design has 13 points, but the central composite design requires 15. On the other hand, central composite designs are rotatable, a general property that Doehlert designs do not have. Furthermore, since central composite designs consist of factorial and axial blocks, they provide the basis for an efficient sequential strategy. Linear models can be fitted in a first stage, after which the design can be augmented with complementary points, should quadratic models prove necessary. Finally, using full designs, Doehlert or otherwise, to fit second-order models is hardly practicable for more than four factors, since a five factor quadratic model has twenty coefficients to be determined and it is unlikely that all factors will be relevant. Fractional factorial screening designs to discriminate between inert and relevant factors should always be applied before higher order designs when many factors are being investigated.

Another very interesting feature of Doehlert designs is the possibility of introducing variations in new factors during the course of an experimental study, without losing the runs already performed. Sometimes we might wish to study first – say – the two factors that seem more promising, analyze the results, and only then introduce variation in a third factor, then in a fourth, and so on. With Doehlert D-1 designs this is possible, provided that all potential factors of interest are introduced in the experiments right from start, set at their average levels (that is, zero in coded units).

For example, let us say that there are four factors of potential interest. We can begin with the two-factor design defined in Table 7, taking care to keep the levels of factors 3 and 4 fixed at zero in all runs. Then, when we wish to study the influence of the two factors that have been kept fixed, we only have to add to the initial design the runs corresponding to the rest of the rows in the four-factor design in the Table 7.

As can be seen in Table 7, a Doehlert design of type D-1 with three or more factors always has one factor at five levels (the first one), one factor at three levels (the last), and the others all at seven levels. Two other Doehlert design types, D-2 and D-3, can be generated by different simplexes and result in different level distributions.

4.1. Application of Doehlert design for optimization of chromatographic systems

Hu and Massart first proposed the use of Doehlert designs for optimization of methods by HPLC. Several examples of applications were given [33]. Table 8 summarizes applications of Doehlert matrix for the optimization of chromatographic methods.

5. Mixture designs

Mixture designs [34] differ from those discussed up to this point since the properties of mixtures depend on ingredient proportions, x_i, and not on their absolute values. As such these proportions are not independent variables since

$$\sum_{i=1}^{q} x_i = 1 \quad \text{for} \quad i = 1, 2, \ldots q. \tag{5}$$

As a consequence, mixture models have mathematical expressions that are different from those involving independent variables,

$$\hat{y} = \sum_{i}^{q} b_i x_i + \sum_{i \neq j}^{q} \sum_{j}^{q} b_{ij} x_i x_j + \sum_{i \neq j}^{q} \sum_{j \neq k}^{q} \sum_{k}^{q} b_{ijk} x_i x_j x_k + \ldots, \tag{6}$$

most noticeably the absence of the constant b_0 term. Experimental designs can be made for any number of components but investigation of three-component systems is the most common.

A simplex centroid design with axial points, presented in the concentration triangle shown in Fig. 4, is especially useful for ternary studies. The component proportions of this design are given in the middle columns of Table 9. Each point at a vertex of the triangle represents a pure component or a mixture

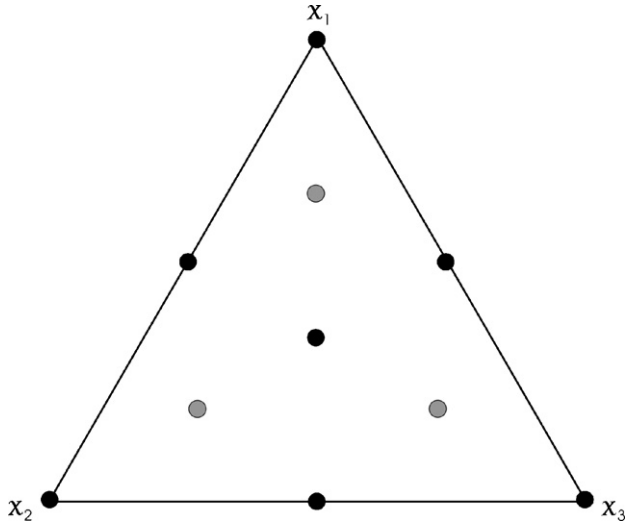

Fig. 4. A simplex centroid design with axial points for a three component mixture system.

Table 8
Application of Doehlert design for the optimization of chromatographic methods

Analytes	Samples	Optimization	Chromatographic technique	Ref.
Oxadiazon	Ground water, agricultural soil, must, wine and human urine samples	Extraction step	GC–MS	[91]
Chloroanisoles	Cork taint of red wine.	Extraction step	GC–ECD	[92]
Pesticides	*Passiflora alata* infuses	Extraction step	GC–ECD	[93]
Chloroanisoles cork	Cork	Extraction step	GC–TOF-MS	[94]
Heterocyclic Aromatic Amines	Meat	Extraction step	HPLC	[95]
Biogenic amines	–	Derivatization reaction	HPLC	[96]
Hydrocarbons	Gasoline	Separation step	CG	[97]
Aromatic compounds	Petroleum cuts	Separation step	HPLC	[98]
Cephalothin and its related substances	–	Separation step	HPLC	[99]
Methylxanthines	–	Separation step	HPLC	[100]
Mono-, di-, and polyaromatic compounds	Middle distillate	Separation step	HPLC	[101]
Sodium cefazolin and related substances	–	Separation step	HPLC	[102]
Cholesterol	Reference materials	Quantification[a]	CG	[103]

Chloroanisoles: 2,4-dichloroanisole, 2,6-dichloroanisole, 2,4,6-trichloroanisole and pentachloroanisole; Pesticides: chlorotalonil, methyl parathion, malathion, α-endosulfan and β-endosulfan; GC–TOF-MS: gas chromatography with time-of-flight mass spectrometry.
[a] Estimation of the optimal amount of internal standard.

of components. The points centered on each leg of the triangle represent 1:1 binary mixtures of the components or mixtures of their neighboring vertex points. The point in the center of the triangle represents a 1:1:1 ternary mixture of the three pure components or mixtures represented at the vertices. The axial points contain a 2/3 portion of one of the ingredients and 1/6 portions of the other two.

This simplex centroid with axial point design is important since it permits the evaluation and validation of linear, quadratic and special cubic models. The special cubic model for a ternary system

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \quad (7)$$

has seven terms. The first three represent the linear model, which is only valid in the absence of interaction effects between components, i.e. ideal solutions in physical chemistry. The next three terms represent synergic or antagonistic binary interaction effects for all possible pairs of components and, along with the linear terms, forms the quadratic model. The last term represents a ternary interaction effect and is usually important for systems having maximum or minimum values in the interior of the concentration triangle.

Table 9
Coded pseudocomponent proportions for mobile phases of a simplex centroid design with axial points and their resolution factor response values

Mixture	x_{ACN}	x_{MeOH}	x_{THF}	Resolution
1	1	0	0	0.99, 1.07
2	0	1	0	5.31, 5.64
3	0	0	1	4.12, 4.34
4	0.5	0.5	0	3.79, 3.98
5	0.5	0	0.5	3.88, 4.07
6	0	0.5	0.5	5.85, 6.16
7	0.333	0.333	0.333	5.22, 5.21
8	0.667	0.167	0.167	3.42, 3.50
9	0.167	0.667	0.167	5.80, 5.81
10	0.167	0.167	0.667	4.84, 4.83

As stated earlier, each vertex can represent a chemical mixture. In this case, the optimization involves investigating mixtures of mixtures. In liquid chromatography, these chemical mixtures and/or pure components can be chosen to have similar chromatographic strengths or other properties that might aid optimization. The example discussed in a later section shows how this can be done.

Often it is not of interest or even possible to investigate the entire range of proportion values, 0–100%, of the mixture components. Many mixture optimization problems require the presence of all ingredients to form a satisfactory product. In these cases it is convenient to define pseudocomponents. Consider a mixture for which the proportions of each vertex component have to obey non-zero lower limits, which we shall generically call a_i. Obviously the sum of all these limits must be less than one, otherwise the mixture would be impossible to prepare. Considering the general case of a mixture containing q components,

$$0 \le a_i \le 1 \quad \text{and} \quad \sum_{i=1}^{q} a_i < 1. \quad (8)$$

The levels of the mixture components in terms of pseudocomponents, denoted by x_i, are given by the expression

$$x_i = \frac{c_i - a_i}{1 - \sum_{i=1}^{q} a_i} \quad \text{for} \quad i = 1, 2 \ldots q \quad (9)$$

which is a kind of variable coding like we have used for the designs for independent variables.

5.1. Model validation

The ANOVA table for model validation is the same for all the above experimental designs and is presented in Table 10.

Table 10
Analysis of variance table for the least squares fit of a model that is linear in its parameters[a]

Source of variation	Sum of squares	Degr. freedom	Mean square
Regression	$SS_R = \sum_i^m \sum_j^{n_i} (\hat{y}_i - \bar{y})^2$	$p-1$	$MS_R = [SS_R/(p-1)]$
Residual	$SS_r = \sum_i^m \sum_j^{n_i} (y_{ij} - \hat{y}_i)^2$	$n-p$	$MS_r = [SS_r/(n-p)]$
Lack of fit	$SS_{lof} = \sum_i^m \sum_j^{n_i} (\hat{y}_i - \bar{y}_i)^2$	$m-p$	$MS_{lof} = [SS_{lof}/(m-p)]$
Pure error	$SS_{pe} = \sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y}_i)^2$	$n-m$	$MS_{pe} = [SS_{pe}/(n-m)]$
Total	$SS_T = \sum_i^m \sum_j^{n_i} (y_{ij} - \bar{y})^2$	$n-1$	

% explained variation: SS_R/SS_T
Maximum % explainable variation: $[(SS_T - SS_{pe})/SS_T]$

[a] n_i: number of replicates at the i^{th} level; m: number of distinct levels of the independent variables; $n = \sum n_i$ = total number of observations; p: number of parameters in the model.

The total data variance is divided into two main contributions, the sum of squares explained by the regression, SS_R, and the residual sum of squares, SS_r. Both summations are taken over all the experimental design levels, $i = 1,2 \ldots m$ and all the replicates performed at each level, $j = 1,2 \ldots n_i$. SS_R is a sum of squares of differences between values predicted by the regression and the grand average of all the response values and has $p-1$ degrees of freedom where p is the number of coefficients in the model. SS_r is a sum of squares of differences or residuals between all the experimental values and the predicted values from the model. It has $n-p$ degrees of freedom where n is the total number of experimental data used to determine the model. Large SS_R and small SS_r values tend to occur for models that accurately describe the experimental data. Their sum is equal to SS_T, the total sum of squares of differences between the experimental values and the grand average of the data set. This sum, of course, has $n-1$ degrees of freedom since it represents the total variance in the data. The SS_R/SS_T ratio represents the fraction of explained variation and is commonly represented as R^2, the coefficient of determination that varies between 0 and 1. If pure error exists it is impossible for R^2 to actually attain 1. Although this coefficient is a measure of how close the model fits the data, it cannot be used to judge model lack of fit because it does not take into account the numbers of degree of freedom for model determination. A related statistic

$$R_a^2 = [1 - (1 - R^2)\{(n-1)/(n-p)\}] \quad (10)$$

makes an adjustment for the varying numbers of degrees of freedom in the models being compared. Draper and Smith [35], however, caution against its use in comparing models obtain from different data sets. Model quality can only be rigorously judged if the SS_r is decomposed into two contributions, the lack-of-fit and the pure error sums of squares, SS_{lof} and SS_{pe}. The latter is a sum of squares of differences between all the individual experimental values and the average of the experimental values at the same level. It has $n-m$ degrees of freedom where m is the number of distinct levels in the experimental design. The SS_{lof} is a sum of squares of differences between the values predicted at each level and the average experimental value at that level and has $m-p$ degrees of freedom. Regression lack of fit is determined performing an F-test by comparing the SS_{lof}/SS_{pe} ratio with the tabled F value for $m-p$ and $n-m$ degrees of freedom at the desired confidence level, usually 95%. If the calculated quotient is greater than the tabled value there is evidence of model lack of fit and the model must be discarded. If not, the model can be accepted at this confidence level as providing an adequate representation of the data. Regression significance can be tested by comparing the calculated SS_R/SS_r value with the tabled F-distribution value for $p-1$ and $n-p$ degrees of freedom. The regression is significant if the calculated value is greater than the tabled one. Note that this last F-test is only valid for models for which there is no evidence of lack of fit. Finally, since the regression model does not explain experimental error the maximum percentage of explainable variation is given by $[(SS_T - SS_{pe})/SS_T] \times 100\%$. Besides ANOVA model validation should also include an analysis of coefficient/standard error ratios and residual plots. An example of model validation of a chromatographic system is given in the next section. Although experimental design applications are not uncommon in the chemical literature, not all of them include information on model validation.

5.2. Optimization example

In this section, the statistical analysis for model determination and validation is presented. The strategy is the same for analyzing the results of all the above experimental designs although the form of the model depends on whether the variables are independent, as for the central composite, Box–Behnken and Doehlert designs or dependent on one another as for the mixture designs.

Chromatographic peak separation is very sensitive to mobile phase composition. Here peak separation results of a simplex centroid design with axial points are analyzed with emphasis on choosing and validating the correct model. Table 9 contains the acetonitrile, x_{ACN}; methanol, x_{MeOH}; and tetrahydrofuran, x_{THF}, proportions of the mobile phase mixtures of the design along with their corresponding resolution values. First, a linear model was adjusted to the experimental data of only the simplex centroid points. The axial points were temporarily removed from the data set for model determination so they could be used for model validation. The linear model is given by the equation

$$\hat{y} = \underset{(\pm 0.02)}{1.667} x_{ACN} + \underset{(\pm 0.02)}{6.035} x_{MeOH} + \underset{(\pm 0.02)}{5.075} x_{THF}. \quad (11)$$

The standard errors were calculated from a pooled estimate of the experimental error obtained from the seven duplicate experiments of the simplex centroid design. The model was tested for lack-of-fit using the ANOVA results in the upper portion of Table 11. The MS_{lof}/MS_{pe} is 61.46 which is larger than the tabled value of $F_{4,7} = 4.12$ at the 95% confidence level. Commercial programs often present this result expressed as a p-value. In this case $p = 0.000016$, which is the area to the right of the $F_{cal} = 61.46$ value under the $\nu_1 = 4$, $\nu_2 = 7$ F distribution. Since $p < 0.05$ the linear model presents evidence of lack of fit at this level and must be rejected.

The poor agreement between the experimental values and model predictions may be due to the existence of interaction effects between mixture components. So a quadratic model was adjusted to the data

$$\hat{y} = \underset{(\pm 0.12)}{1.013} x_{ACN} + \underset{(\pm 0.12)}{5.458} x_{MeOH} + \underset{(\pm 0.12)}{4.213} x_{THF} + \underset{(\pm 0.50)}{2.865} x_{ACN}x_{MeOH} + \underset{(\pm 0.50)}{5.715} x_{ACN}x_{THF} + \underset{(\pm 0.50)}{4.945} x_{MeOH}x_{THF}. \quad (12)$$

The ANOVA for this model is included in Table 11. The MS_{lof}/MS_{pe} ratio is only 3.12 which is smaller than the 95% confidence value of $F_{1,7} = 5.59$ ($p = 0.12$). Since this model does not present evidence of lack of fit, its regression significance can be tested using the MS_R/MS_r ratio. This value of 213.98 is much larger than the tabled $F_{5,8}$ value of 3.69 indicating a highly significant regression ($p < 10^{-6}$). Furthermore, since the calculated value is more than 10 times the tabled value, we can expect the regression model to be useful for quantitative predictions [5].

Table 11
ANOVA tables for linear and quadratic models

Variation	SS	df	MS	F-test quotient
Linear model				
Regression	26.3462	2	13.1731	24.16
Residual	5.9977	11	0.5452	–
Lack of fit	5.8316	4	1.4579	61.46
Pure error	0.1661	7	0.0237	–
Total	32.3439	13	2.4880	
Quadratic model				
Regression	32.1038	5	6.4208	213.98
Residual	0.2401	8	0.0300	–
Lack of fit	0.0740	1	0.0740	3.12
Pure error	0.1661	7	0.0237	–
Total	32.3439	13		

The results for the axial point experiments in Table 9 were used to further validate the model. Predicted results from Eq. (12) for the three axial points are 3.37, 5.54 and 5.15 compared with their corresponding experimental average values of 3.46, 5.81 and 4.84, respectively. The average error between the predicted and experimental averages of 0.22 is close to the pooled experimental error of the simplex centroid design, 0.15.

An even more accurate model is expected using both the simplex centroid design and axial point results to determine a new model. This results in the equation

$$\hat{y} = \underset{(\pm 0.13)}{1.031} x_{ACN} + \underset{(\pm 0.13)}{5.527} x_{MeOH} + \underset{(\pm 0.13)}{4.130} x_{THF} + \underset{(\pm 0.60)}{3.061} x_{ACN}x_{MeOH} + \underset{(\pm 0.60)}{5.605} x_{ACN}x_{THF} + \underset{(\pm 0.60)}{4.937} x_{MeOH}x_{THF} \quad (13)$$

that has coefficient values very similar to those of the model in Eq. (12). Furthermore, the special cubic model was adjusted to all the data in Table 9 in an unsuccessful attempt to obtain an even better fit. Although it does not suffer from lack of fit, as can be expected, the coefficient of the special cubic term was not significant at the 95% confidence level.

It remains to interpret the quadratic model in terms of the effects the different mobile phase components have on resolution. Of the linear effects, methanol is more efficient at separating these two peaks followed closely by THF while acetonitrile is not particularly effective. Furthermore, there are three synergic binary interaction effects that are significant at the 95% confidence level since the standard error of ±0.60 is at least five times smaller than the binary interaction coefficients. Of these synergic interactions, the ones between acetonitrile and THF and the other between methanol and THF are the most important.

The contour plot for the response surface is shown in Fig. 5. It is clearly seen that the optimum separation factor occurs close to the 1:1 binary mixture of methanol and THF, i.e. the experimental point that has the largest average value in Table 9. In fact, the optimum mixture is a little richer in methanol, about 60% methanol and 40% THF. Certainly, the large synergic interaction between methanol and THF as well as the large linear effects of methanol and THF relative to the small one for acetonitrile helps understand why this solution gives the best separation for this pair of peaks. The above example involves the modeling of only one elementary response function. A real application would involve the simultaneous optimization of a number of elementary response functions. Evidence so far seems to indicate that it is more efficient to model each elementary response separately and then apply a multi-criteria decision-making procedure such as the Derringer and Suich desirability function rather than combine several elementary responses into a more complex objective function that could be used to determine a single response surface [36,37].

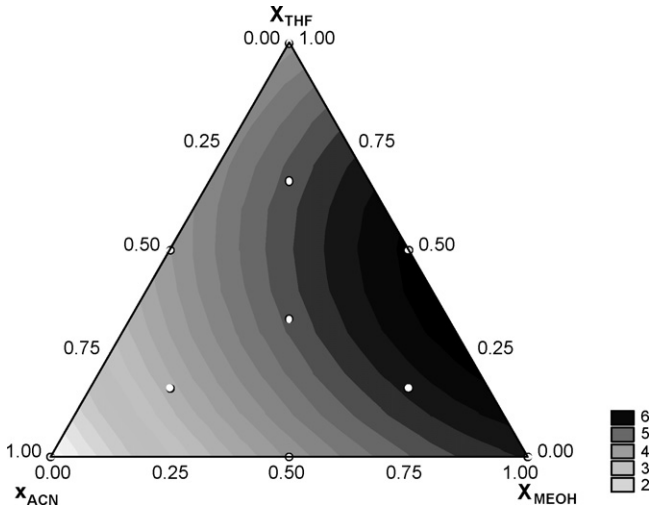

Fig. 5. Mixture response contour map of the separation factor of two component peaks for different acetonitrile, methanol and tetrahydrofuran concentrations.

5.3. Application of mixture design for optimization of chromatographic methods

The mixture design was employed for optimization of the mobile phase of an HPLC method proposed for the separation and determination of benzodiazepines: triazolam, de-moxepam, oxazepam, flunitrazepam, nitrazepam, clon-azepam, diazepam, ethyl loflazepate and nordazepam [38].

Vuorela and co-workers developed a computer program for the mobile phase optimization of a HPLC method. A desirability function technique combined with the mixture design was employed to enhance the quality of HPLC separations. The use of statistical models to predict the behaviour of retention times (t(R)) and band broadening at the different eluent compositions obtained by prisma method was examined for dansyl amides and coumarins [39].

6. Other experimental designs used for the optimization of chromatographic methods

In chromatography, the two-level full factorial has been often used for the optimization of the extraction steps and also for the determination of the experimental conditions for separation processes [40–50].

The full three-level factorial design (3^k) can be used also to obtain quadratic models and determine critical conditions. This design has been little used because it requires quite a number of experiments if the factor number is higher than 2 [51]. Some authors have employed this design for optimization of chromatographic methods [52–56].

7. Conclusion

Multivariate statistical techniques applied to chromatographic methods have been basically employed for optimization of the steps of sample preparation and compound separation. Of the response surface techniques, the Box–Behnken design, the Doehlert matrix and the central composite design, the latter has been the one more frequently used for the optimization of chromatographic methods.

Statistical mixture designs are recommended for the optimization of mobile phase, where the proportions of the components determine chromatographic peak separation and not their total amounts. Model validation using F-distribution tests, model coefficient standard errors, residual plots and the adjusted R_a^2 statistic is recommended if one must guarantee that the actual optimum experimental conditions have been found.

The two-level full factorial design has often been used for the preliminary evaluation of those experimental factors that are important for the optimization of chromatographic systems.

Acknowledgements

The authors are grateful for Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal do Ensino Superior (CAPES), and Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for financial support and fellowship. S.L.C.F., J.B.A., B.B.N, R.E.B., C.M.Q. and J.M.D. acknowledge senior research scholarships from CNPq. M.C.B. thanks Fapesp for a master's fellowship.

References

[1] S.N. Deming, in: B.R. Kowalski (Ed.), Chemometrics, Mathematics and Statistics in Chemistry, Reidel, Dordrecht, 1981.
[2] J.A. Nelder, R. Mead, Comput. J. 7 (1965) 308.
[3] W. Spendley, G.R. Hext, F.R. Himsworth, Technometrics 4 (1962) 441.
[4] M.W. Routh, P.A. Swartz, M.B. Denton, Anal. Chem. 49 (1977) 1422.
[5] G.E.P. Box, J.S. Hunter, W.G. Hunter, Statistics for Experimenters, second ed., Wiley-Interscience, New York, 2005.
[6] D.C. Montgomery, Design and Analysis of Experiments, fourth ed., Wiley, New York, 1997.
[7] R.E. Bruns, I.S. Scarminio, B.B. Neto, Statistical Design—Chemometrics, Elsevier, Amsterdam, 2006.
[8] G.E.P. Box, K.B. Wilson, J. R. Statist. Soc. B13 (1951) 1.
[9] Y. Vander Heyden, C. Perrin, D.L. Massart, in: K. Valkó (Ed.), Handbook of Analytical Separations, 1, Elsevier, Amsterdam, 2000, Ch. 6.
[10] A.M. Siouffi, R. Phan-Tan-Luu, J. Chromatogr. A 892 (2000) 75.
[11] G.E.P. Box, J.S. Hunter, Ann. Math. Stat. 28 (1957) 195.
[12] G.E.P. Box, N.R. Draper, Empirical Model-Building and Response Surfaces, Wiley, New York, 1987.
[13] R.H. Myers, D.C. Montgomery, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, second ed., Wiley, New York, 2002.
[14] J. Goupy, Plans d'Expériences pour Surfaces de Résponse, Dunod, Paris, 1999.
[15] J. Dron, R. Garcia, E. Millan, J. Chromatogr. A 963 (2002) 259.
[16] A.M. Faria, D.R. Magalhães, K.E. Collins, C.H. Collins, Anal. Chim. Acta 550 (2005) 137.
[17] A.M. Faria, K.E. Collins, C.H. Collins, J. Chromatogr. A 1122 (2006) 114.
[18] B. Do, S. Robinet, D. Pradeau, F. Guyon, J. Chromatogr. A 918 (2001) 87.
[19] B. Do, S. Robinet, D. Pradeau, F. Guyon, Analyst 126 (2001) 594.
[20] G.E.P. Box, D.W. Behnken, Technometrics 2 (1960) 195.
[21] C. Turner, L.C. Whitehand, T. Nguyen, T. McKenon, J. Agric. Food Chem. 52 (2004) 26.
[22] M. Gferer, E. Lankmayr, Anal. Chim. Acta 533 (2005) 203.
[23] M. Preu, D. Guyot, M. Petz, J. Chromatogr. A 818 (1998) 95.

[24] A. Stafiej, K. Pyrzynska, A. Ranz, E. Lankmayr, J. Biochem. Biophys. Methods 69 (2006) 15.
[25] F.H. Walters, H.C. Qiu, Anal. Lett. 25 (1992) 1131.
[26] M.E.P. Hows, D. Perrett, J. Kay, J. Chromatogr. A 768 (1997) 97.
[27] T. Mirza, H.S.I. Tan, J. Pharm. Biomed. Anal. 25 (2001) 39.
[28] D.H. Doehlert, Appl. Statist. 19 (1970) 231.
[29] D.H. Doehlert, V.L. Klee, Discrete Math. 2 (1972) 309.
[30] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics. Part A, Elsevier, Amsterdam, 1977.
[31] S.L.C. Ferreira, W.N.L. dos Santos, C.M. Quintella, B.B. Neto, J.M. Bosque-Sendra, Talanta 63 (2004) 1061.
[32] B.B. Neto, I.S. Scarminio, R.E. Bruns, Quim. Nova 29 (2006) 1401.
[33] Y. Hu, D.L. Massart, J. Chromatogr. 485 (1989) 311.
[34] J.A. Cornell, Experiments with Mixtures: Designs, Models and the Analysis of Mixture Data, Wiley, New York, 1990.
[35] N.R. Draper, H. Smith, Applied Regression Analysis, Wiley, New York, 1981.
[36] G. Derringer, R. Suich, J. Qual. Technol. 12 (1980) 214.
[37] M.C. Breitkreitz, I.C.S.F. Jardim, R.E. Bruns, Presented at the 10th International Conference on Chemometrics in Analytical Chemistry, Aguas de Lindoia, 10–15 September 2006, abstract P083.
[38] P.F. Vanbel, B.L. Tilquin, P. Schoenmakers, Chemom. Intell. Lab. Syst. 35 (1996) 67.
[39] K. Outinen, H. Haario, P. Vuorela, M. Nyman, E. Ukkonen, H. Vuorela, Eur. J. Pharm. Sci. 6 (1998) 197.
[40] A. Padilla, M. Palma, C.G. Barroso, J. Chromatogr. A 1091 (2005) 83.
[41] K. Novotna, J. Havlis, J. Havel, J. Chromatogr. A 1096 (2005) 50.
[42] N. Jakobsson, M. Degerman, B. Nilsson, J. Chromatogr. A 1099 (2005) 157.
[43] M. Polo, M. Llompart, C. Garcia-Jares, G. Gomez-Noya, M.H. Bollain, R. Cela, J. Chromatogr. A 1124 (2006) 11.
[44] M.D. Cantu, D.R. Toso, C.A. Lacerda, F.M. Lancas, E. Carrilho, M.E. Queiroz, Anal. Bioanal. Chem. 386 (2006) 256.
[45] V. Yusa, A. Pastor, M. Guardia de la, Anal. Chim. Acta 565 (2006) 103.
[46] A.K. Dutta, B.A. Avery, C.M. Wyandt, J. Chromatogr. A 1110 (2006) 35.
[47] V. Casas, M. Llompart, C. Garcia-Jares, R. Cela, T. Dagnac, J. Chromatogr. A 1124 (2006) 148.
[48] L. Hlabangana, S. Hernandez-Cassou, J. Saurina, J. Chromatogr. A 1130 (2006) 130.
[49] O. Anurukvorakun, W. Suntornsuk, L. Suntornsuk, R. Yekkala, S. Vandenwayenberg, J. Hoogmartens, E. Adams, J. Chromatogr. A 1134 (2006) 326.
[50] R. Yekkala, S. Vandenwayenberg, J. Hoogmartens, E. Adams, J. Chromatogr. A 1134 (2006) 56.
[51] W.N.L. dos Santos, F. de S. Dias, M.S. Fernandes, M.V. Rebouças, M.G.R. Vale, B. Welz, S.L.C. Ferreira, J. Anal. At. Spectrom. 20 (2005) 127.
[52] S. Hillaert, Y. Vander Heyden, W. Van den Bossche, J. Chromatogr. A 978 (2002) 231.
[53] S. Hillaert, W. Van den Bossche, J. Chromatogr. A 979 (2002) 323.
[54] V. Harang, S.P. Jacobsson, D. Westerlund, Electrophoresis 25 (2004) 1792.
[55] N.M. de Aragao, M.C.C. Veloso, M.S. Bispo, S.L.C. Ferreira, J.B. de Andrade, Talanta 67 (2005) 1007.
[56] J.B. de Andrade, E.T. Souza, F.M. Rodrigues, C.C. Martins, F. Oliveira Santos, P.A.P. Pereira, Microchem. J. 82 (2006) 142.
[57] A. Garrido-Lopez, V. Esquiu, T.M. Tena, J. Chromatogr. A 1124 (2006) 51.
[58] C. Pizarro, J.M. Gonzalez-Saiz, N. Perez-del-Notario, J. Chromatogr. A 1132 (2006) 8.
[59] A. Brachet, S. Rudaz, L. Mateua, P. Christen, J.L. Veuthey, J. Sep. Sci. 24 (2001) 865.
[60] F.J. Conde, A.M. Afonso, V. Gonzalez, J.H. Ayala, Anal. Bioanal. Chem. 385 (2006) 1162.
[61] C. Goncalves, J.J. Carvalho, M.A. Azenha, M.F. Alpendurada, J. Chromatogr. A 1110 (2006) 6.
[62] F. Pellati, S. Benvenuti, F. Yoshizaki, D. Bertelli, M.C. Rossi, J. Chromatogr. A 1087 (2005) 265.
[63] A.B. Baranda, N. Etexbarria, R.M. Jimenez, R.M. Alonso, Talanta 67 (2005) 933.
[64] M.S.D. Garcia, J.M. Lopez, R. Bouza, M.J. Abad, E.G. Soto, M.V.G. Rodriguez, Anal. Chim. Acta 521 (2004) 179.
[65] G. Iriarte, N. Ferreiros, I. Ibarrondo, R.M. Alonso, M.I. Maguregi, L. Gonzalez, R.M. Jimenez, J. Sep. Sci. 29 (2006) 2265.
[66] P. Araujo, L. Froyland, J. Choromatogr. B 830 (2006) 212.
[67] R. Rodil, A.M. Carro, R.A. Lorenzo, M. Abuin, R. Cela, J. Chromatogr. A 963 (2002) 313.
[68] I. Rodriguez, J.B. Quintana, J. Carpinteiro, A.M. Carro, R.A. Lorenzo, R. Cela, J. Chromatogr. A 985 (2003) 265.
[69] C.D. Stalikas, G.A. Pilidis, J. Chromatogr. A 872 (2000) 215.
[70] R. Castro Mejias, R. Natera Marin, M.D. Garcia Moreno, C. Garcia Barroso, J. Chromatogr. A 995 (2003) 11.
[71] I. Arambiri, M. Lasa, R. Garcia, E. Millan, J. Chromatogr. A 1033 (2004) 193.
[72] M. Nowak, A. Seubert, J. Chromatogr. A 855 (1999) 91.
[73] O. Ezquerro, A. Garrido-Lopez, M.T. Tena, J. Chromatogr. A 1102 (2006) 18.
[74] M.V.N. Rodrigues, F.G.R. Reyes, V.L.G. Rehder, S. Rath, Chromatographia 61 (2005) 291.
[75] P.B.M. Pinheiro, J.C.G. Esteves da Silva, Anal. Bioanal. Chem. 382 (2005) 341.
[76] A. Garrido-Lopez, M.T. Tena, J. Chromatogr. A 1099 (2005) 75.
[77] N. Garcia-Villar, J. Saurina, S. Hernandez-Cassou, Anal. Chim. Acta 575 (2006) 97.
[78] L. Kristoffersen, A. Bugge, E. Lundanes, L. Slordal, J. Chromatogr. B 734 (1999) 229.
[79] V. Harang, D. Westerlund, Chromatographia 50 (1999) 525.
[80] J. Diana, G. Ping, E. Roets, J. Hoogmartens, Chromatographia 56 (2002) 313.
[81] Y.P. Zhang, Y.J. Zhang, W.J. Gong, A.I. Gopalan, K.P. Lee, J. Chromatogr. A 1098 (2005) 183.
[82] J.S. Qiu, X.H. Jin, J. Chromatogr. A 950 (2002) 81.
[83] M.R. Hadjmohammadi, F. Safa, J. Sep. Sci. 27 (2004) 997.
[84] C.H. Kuo, S.W. Sun, Anal. Chim. Acta 482 (2003) 47.
[85] S. Hillaert, T.R.M. De Beer, J.O. De Beer, W. Van den Bossche, J. Chromatogr. A 984 (2003) 135.
[86] J. Hernandez-Borges, M.A. Rodriguez-Delgado, F.J. Garcia-Montelongo, A. Cifuentes, J. Sep. Sci. 28 (2005) 948.
[87] D. Mangelinas, M. Maftouh, D.L. Massart, Y.V. Heyden, J. Pharm. Biomed. Anal. 35 (2004) 789.
[88] S. Cavalli, S. Polosello, G. Sacan, J. Chromatogr. A 1039 (2004) 155.
[89] B. Muir, W.A. Carrick, D.B. Cooper, Analyst 127 (2002) 1198.
[90] E. Destandau, J. Vial, A. Jardy, M.C. Hennion, D. Bonnet, P. Lancelin, Anal. Chim. Acta 572 (2006) 102.
[91] A. Navalon, A. Prieto, L. Araujo, J.L. Vílchez, J. Chromatogr. A 946 (2002) 239.
[92] A. Martínez-Uruñuela, J.M. González-Sáiz, C. Pizarro, J. Chromatogr. A 1056 (2004) 49.
[93] G.A. da Silva, F. Augusto, R.J. Poppi, J. Chromatogr. A 1138 (2007) 251.
[94] E. Carasek, E. Cudjoe, J. Pawliszyn, J. Chromatogr. A 1138 (2007) 10.
[95] L. Cardenes, A. Martin-Calero, J. Ayala, V. Gonzalez, A. Afonso, Anal. Lett. 39 (2006) 405.
[96] R. Romero, M.G. Bagur, M. Sanchez-Vinas, D. Gazquez, Chromatographia 51 (2000) 404.
[97] J.P. Durand, S. Gautier, E. Robert, M.C. Guilhem, R. Phan-Tan-Luu, J. High Resolut. Chromatogr. 20 (1997) 289.
[98] R. Fraile, V. Sanchez, J. High Resolut. Chromatogr. 16 (1993) 169.
[99] L. Ding, K. Ni, Y. Chen, Y. Hu, R. Yu, SEPU 11 (1993) 293.
[100] P. Araujo, Trends Anal. Chem. 19 (2000) 524.
[101] R. M'Hamdi, D. Thiébaut, M. Caude, E. Robert, C. Grand, J. High Resolut. Chromatogr. 21 (1997) 94.
[102] L. Ding, K. Ni, Y. Chen, Y. Hu, S. Qiu, R. Yu, Yaowu Fenxi Zazhi 15 (1995) 3.
[103] P. Araujo, F. Couillard, E. Leirnes, K. Ask, A. Bøkevoll, L. Frøyland, J. Chromatogr. A A1121 (2006) 99.