Chemometrics and Intelligent Laboratory Systems 102 (2010) 45–52

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Simultaneous optimization of the microextraction of coffee volatiles using response surface methodology and principal component analysis

J.S. Ribeiro [a,b], R.F. Teófilo [a,c], F. Augusto [b], M.M.C. Ferreira [a,*]

[a] Theoretical and Applied Chemometrics Laboratory, Chemistry Institute, University of Campinas, P.O. Box 6154, 13083-970 Campinas, SP, Brazil
[b] Gas Chromatography Laboratory, Chemistry Institute, University of Campinas, P.O. Box 6154, 13083-970 Campinas, SP, Brazil
[c] Chemometrics and Instrumentation Laboratory, Chemistry Department, Federal University of Viçosa, 36570-000 Viçosa, MG, Brazil

ARTICLE INFO

Article history:
Received 3 August 2009
Received in revised form 9 March 2010
Accepted 15 March 2010
Available online 19 March 2010

Keywords:
Solid phase microextraction
Principal component analysis
Multiple response optimization
Coffee volatiles
Experimental design

ABSTRACT

It is well known that no single experimental condition can be found under which the extraction of all the volatile compounds in a gas chromatographic analysis of roasted coffee beans by headspace-solid phase microextraction (HS-SPME) is maximized. This is due to the large number of peaks recorded. In this work, the scores vector of the first principal component obtained from PCA on chromatographic peak areas was used as the response to find the optimal conditions for simultaneous optimization of coffee volatiles extraction via response surface methodology (RSM). This strategy consists in compressing several highly correlated peak areas into a single response variable for a central composite design (CCD). RSM was used to identify an optimal factor combination that reflects a compromise between the partially conflicting behavior of the volatiles groups. This simultaneous optimization approach was compared with the desirability function method. The versatility of the PCA–RSM methodology allows it to be used in other chromatographic applications, resulting in an interpretable procedure to solve new analytical problems.

1. Introduction

The aroma profile is one of the most typical features of food products, in terms of both organoleptic quality and authenticity [1]. Due to the high number of volatile components, the aroma profile represents a "fingerprint" of a product [2]. Aroma compounds are abundantly present in roasted coffee as complex mixtures of volatile components with different functional groups. That is the main reason why roasted coffee volatiles are frequently described in different studies of analytical methods and new extraction materials [3–9].

Solid phase microextraction (SPME) has been shown to be an excellent sampling method, allowing simultaneous extraction and concentration of analytes from sample matrices. This technique makes use of a fused silica optical fiber coated with a thin polymer layer to extract the analytes from a liquid (solution), from the headspace (HS) above a liquid or solid, or from a gaseous phase [10]. The advantages of SPME can be completely and easily exploited in the headspace mode. The enrichment of the analytes is unique in comparison to other HS sample preparation methods [11].

Finding the optimal experimental conditions in SPME is an important task, since the kinetics and thermodynamics of extraction depend on several experimental conditions such as fiber coating, sample concentration, temperature, time and ionic strength, among others [12–14].

From a univariate point of view, certain advances have been made regarding the study of HS-SPME experimental conditions for extraction of volatiles from foods [15–17]. However, with respect to analyses of coffee volatiles, certain experimental variables differ from one paper to another [18–22].

Although several experimental design investigations using HS-SPME can be found in the literature [13,23–27], examples of optimization to improve the extraction of coffee volatiles are scarce and do not employ experimental design [15,28].

Interest in finding the optimal experimental conditions for maximizing more than one peak area is frequent in studies employing chromatographic techniques. However, it is rather difficult to analyze the results obtained from response surface methodologies when several dependent variables (responses) of interest are involved. A large number of the volatiles from roasted coffees appear in larger or smaller amounts in the headspace, depending on the roasting degree and quality of the beans. Unfortunately, the experimental conditions that maximize the extraction of high molar mass volatiles might not be even close to those which optimize the extraction of low molar mass volatiles. In such situations, an extraction optimization for each volatile becomes laborious, complex and not representative for all of them. There are different approaches for multiple response optimization. Frequently, they involve the optimization of one response at a time subject to constraints on the remaining ones, and then finding a set of experimental conditions that in some sense optimizes all responses or, at least, keeps them within desirable ranges [29].

A relatively straightforward approach for optimizing several responses that works well when there are only a few design variables is

* Corresponding author. Tel.: +55 19 3521 3102; fax: +55 19 3521 3023.
E-mail address: marcia@iqm.unicamp.br (M.M.C. Ferreira).

0169-7439/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2010.03.005

to overlay the contour plots for each response [29]. Nevertheless, when there are more than three response variables, overlaying contour plots becomes awkward. Another solution is to formulate and solve a constrained optimization problem which can be accomplished by numerical techniques like nonlinear programming methods [30,31]. Another nonlinear strategy is that based on a neurogenetic approach [13,27].

A useful approach for the simultaneous optimization of several response variables was introduced by Harrington [32] and called the desirability function. This approach was later improved and popularized by Derringer and Suich [33]. In essence, this method transforms a multivariate optimization problem into a univariate one, where all the responses are combined into one measurement, i.e., only one representative response. Among the advantages of using the desirability function, are that responses with different scalings can be compared between themselves, the transformation of different responses into one measurement is rather simple and quick and, lastly, both qualitative and quantitative responses can be taken into account [34].

Other approaches to multiple response analysis found in the literature are dual responses [35], Khuri–Colon distance [36] and those based on the square error loss [37–39], among others.

In this work a quite practical and effective methodology based on PCA was applied as strategy for tackling multiple response optimizations. This procedure was firstly presented by Bratchell [40] and latter demonstrated in five different examples by Carlson et al. [41]. Sandstrom et al. [42] and Ellekjaer et al. [43] have published interesting papers using the same approach. Nowadays, the majority of applications are found in Taguchi design [44–48]. However, the application of this strategy has not yet been used in optimization of experimental conditions for chromatographic analysis, which is explored in this work. To the best of our knowledge, there is no work in literature using RSM and PCA as an optimization methodology for solid phase microextraction operational conditions for coffee volatiles determination.

The aim of this work is to apply a strategy based on central composite design (CCD) and principal component analysis (PCA) to find the operational conditions (extraction temperature, extraction time, and equilibrium time) of the hyphenated method HS-SPME–GC-FID that simultaneously optimizes the amounts of volatile compounds of both low and high molar masses extracted from roasted Arabica coffee. For a comparative analysis, the method of desirability function was also applied.

1.1. Theoretical explanation of multiple response optimization using PCA

Usual chromatographic analyses of natural products, without many clean-up steps, provide chromatograms with several peaks. If some or all of the peak areas are relatively highly correlated, then the original set of correlated responses can be reduced to one or a very few uncorrelated PCA components [44]. The strategy for multiple response analysis employed in this work takes advantage of minimizing the complex analysis of several dependent variables using PCA coupled with experimental design.

The method is based on two principal considerations, as follows.

a) Testing the correlations between relative peak areas, i.e., the raw responses in the experimental design: the Pearson correlation coefficient (r) was used for assessing the degree of linear association between each two variables (peak area vectors) [49]. If peak areas exhibit reasonable mutual correlation, $|r|>0.6$, it can be considered that the corresponding peak responses are related to changes in the system.
b) Applying principal component analysis (PCA) to eliminate redundant information when peak areas are correlated: the most frequent application of this method occurs in situations where the variables are correlated and present redundancies that can be removed together with small variabilities. The aim of PCA is to express the significant information contained in the original variables by a small set of new variables, the principal components [50–53].

Once correlation among the multiple responses is detected, the use of PCA can be recommended and **Y** is replaced by the scores of the first few principal components. In this work, only the first component was used (one column in Fig. 1). Thus, the statistical calculation from experimental design was performed using only the scores of the first principal component as the dependent variable. However, it is very important to verify the variance explained by the first principal component and the correlation between the scores from this component and each original response variable ($\mathbf{y}_i$) used in PCA. If the explained variance and the correlations are satisfactory, the multiple response analysis will be significant and reliable.

1.2. Desirability function

In this methodology [33], the desirable combination of k response variables, each of which depends upon a set of p design variables, is obtained through a desirability function. This function transforms

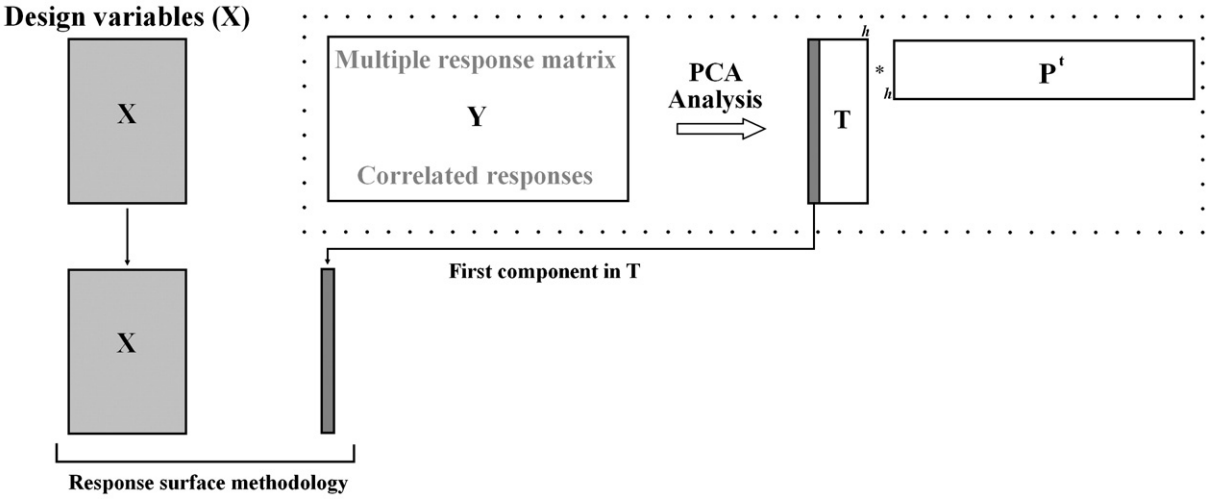

Fig. 1. Scheme showing the calculations used for simultaneous multiple response optimization using principal component analysis and response surface methodology. **T** and **P** are the scores and loadings matrices.

each estimated response variable $\hat{y}_i$, calculated by the fitted response surface associated with the CCD experimental design used in this work, into a desirability value d_i, using the following set of equations:

$$d_i = \begin{cases} 0 & \hat{y}_i \le y_{i\min} \\ \left[\dfrac{\hat{y}_i - y_{i\min}}{y_{i\max} - y_{i\min}}\right] & y_{i\min} < \hat{y}_i < y_{i\max}, \quad \text{for } i = 1, 2, \ldots, k \\ 1 & \hat{y}_i \ge y_{i\max} \end{cases} \quad (1)$$

where the values $y_{i\min}$ and $y_{i\max}$ are the minimum and maximum acceptable value of $\hat{y}_i$, respectively. The values of d_i vary in the interval $0 \le d_i \le 1$, increasing as the desirability of the corresponding response increases.

The individual desirabilities are then combined using the geometric mean (Eq. (2)) to give an overall desirability, D,

$$D = (d_1 \times d_2 \times \cdots \times d_k)^{1/k} \quad (2)$$

which increases as the balance of the properties becomes more favorable. Any existing univariate search technique can be used to optimize D over the independent variable domain (p design variables), resulting in the desirability of the combined response levels.

2. Materials and methods

2.1. Coffee sample

One roasted Arabica coffee sample was used in the chromatographic analyses.

2.2. GC-FID parameters

The analyses were performed on a G-6850 GC-FID system (Agilent, Wilmington, USA) fitted with a HP-5 capillary column (30 m × 0.25 mm × 0.25 μm). Helium (1 mL min^{-1}) was the carrier gas. The oven temperature was programmed as follows: 40 °C → 5 °C/min → 150 °C → 30 °C/min → 260 °C. The injection port was equipped with a 0.75 mm i.d. liner and the injector was maintained at 220 °C in the splitless mode. Under these conditions, no sample carry-over was observed on blank runs conducted between extractions.

2.3. General SPME procedure of sampling and injection

The volatiles extraction was carried out using the HS-SPME technique. In an earlier work [54], different types of SPME fibers were evaluated, taking into consideration polarities, fiber coatings and thickness, with the purpose of identifying which commercial fiber is most suitable for extracting coffee volatiles. Polydimethylsiloxane/divinylbenzene (PDMS/DVB) fibers with 65 μm thickness were chosen. This fiber and the manual holder were purchased from Supelco (Bellefonte, USA). All assays were carried out using 250 mg of ground Arabica roasted coffee and 2 mL of saturated aqueous sodium chloride solution transferred to a septum-sealed glass sample vial (5 mL). The experimental conditions of the assays were those indicated by the experimental design.

2.4. SPME variables

In development and application of the HS-SPME method many aspects (conditions) have to be considered due to various physicochemical properties of the compounds that will be extracted. The salt additives, pH, extraction temperatures, the sample-to-headspace ratio and the time of incubation, for example, are important parameters for achieving the best extraction efficiency [10,14]. However, some of these parameters were not taken into account when designing the present experiments. It is known from the literature that an increase of the ionic strength by adding salt is more effective for the extraction of analytes onto the fiber, because it minimizes the solubility of less polar compounds by forcing them to pass to the vapor phase (salting-out effect) [10,14]. Besides, previous optimization strategies using this variable have shown that super-saturation was the best condition [13,27]. Another variable not considered in this work was the vial size because, usually, the vial is filled to half of its capacity [14,16]. Analyte extraction is improved when the headspace is minimized; however the minimum volume of headspace is limited by the length of the fiber. The pH variable was not included either, because it would be difficult to find an optimum pH value for simultaneous extraction of several volatile compounds with different acid–base properties.

Systematic optimization procedures were carried out by selecting an objective function, which includes the most important factors affecting the microextraction process and investigating the relationship between responses and factors by RSM. Three experimental factors were taken into account in this work: bath temperature (*T*), pre-equilibrium time (*PET*) and extraction time (*Ext*).

2.5. Response surface methodology

Once the instrumental conditions that ensured reasonable responses were established, the optimization procedure was applied in order to find the best experimental conditions for an optimum signal response.

Finding the optimum experimental conditions is more efficient and precise when multivariate statistical techniques are employed since all variables (factors) are simultaneously considered, accompanied with significant experimental savings [55–57]. To perform this task, experimental designs such as response surface methodology are the procedures employed in the majority of optimization studies [23–25,58]. Experimental designs are helpful in determining the effects of individual variables (factors) and interaction among them over the significant responses [25,55].

A central composite design (CCD) with three independent variables was the protocol chosen for carrying out the RSM. The design consisted of a total of 18 experiments: 8 in the factorial points, 6 in the axial points and 4 central points. Other alternatives to the standard CCD could be, for example, the D-optimal, dodecahedron + 1 and dodecahedron + 2 designs performed with 10, 13 and 14 experiments, respectively [59]. The factorial point levels of independent variables investigated were: bath temperature (*T*: 30–50 °C), pre-equilibrium time (*PET*: 5–15 min), extraction time (*Ext*: 10–20 min). These ranges were selected based on prior knowledge about the system under study. All experiments were performed in random order to minimize the effects of uncontrolled

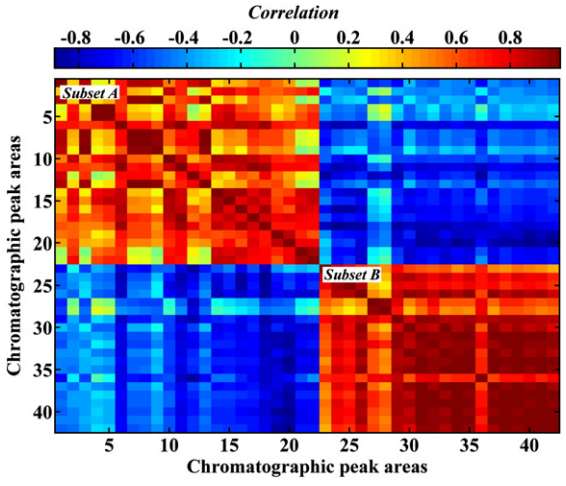

Fig. 2. Correlation map of peak areas.

Table 1
Percent variance described by PCA model in the different subsets.

PC	Subset ***A***		Subset ***B***	
	% Variance captured		% Variance captured	
	Individual	Total	Individual	Total
1	64.51	64.51	81.98	81.98
2	19.67	84.18	8.15	90.13
3	6.82	91.00	6.59	96.72
4	3.80	94.80	1.90	98.62
5	1.63	96.43	0.51	99.12

factors that may introduce a bias into the measurements. For the statistical analysis, the model coefficients were calculated by multiple linear regression and validated by the analysis of variance (ANOVA).

2.6. Software

The data analysis was carried out using Matlab 6.5 (The MathWorks, Co., Natick, MA, USA), Microsoft Excel™ 2003 (The Microsoft, Co, USA) and Statistica 6.0 (The StatSoft, Inc., Tulsa, OK, USA). The algorithms for PCA were made in-house and the experimental design calculations were performed using the spreadsheets presented by Teófilo and Ferreira [55,60]. The desirability calculation was carried out using the software Statistica 6.0.

3. Results and discussions

The initial responses, considered in the statistical treatment and used for building the response surfaces, consisted of the relative peak areas obtained from chromatographic runs as defined by the CCD. A total of 57 peaks covering a wide range of molar masses and distributed in 42 "regions", were selected as initial representative responses. Each "region" was represented by either a single peak or by a group of overlapped peaks. This organization was necessary because the chromatographic separation of adjacent peaks in certain regions was not effective. The area of an individual peak or a group of peaks (region) will be referred as peak area in this work.

The use of these 42 peak areas as responses makes the statistical analysis rather complicated when no treatment with simultaneous responses is used. So, methodologies for multiple responses are necessary in order to make the complex analysis feasible. Thus, the multiple response approach using PCA and desirability function were applied and compared to attain the optimal operational chromatographic conditions.

3.1. RSM–PCA

Since the PCA method groups correlated variables, it is expected that peaks with similar variations, as a function of changes in the experimental conditions of the system, would be correlated. This way, when the correlation matrix of the peak areas was calculated and presented graphically in correlogram format (Fig. 2), a direct correlation among peaks in two quadrants (2nd and 4th as indicated in Fig. 2) could be observed. Correlations in the second quadrant, designated as subset ***A***, account for 24 peaks (22 regions), distributed from the beginning of the chromatographic run up to 8 min. Correlations in the fourth quadrant, designated as subset ***B***, correspond to 33 peaks (20 regions) with retention times between 8 and 19 min.

Correlations between peaks areas from subsets ***A*** and ***B*** are mostly negative (1st and 3rd quadrants in Fig. 2), indicating that the responses of subset ***A*** bring different chemical information from those observed in subset ***B***. Hence, the multiple response analyses using PCA were performed separately for each subset, in order to obtain higher explained variance in the first component for both subsets.

The first PCA components obtained using auto-scaled data from subsets ***A*** and ***B*** explained 64.51 and 81.98% of the data variance, respectively (Table 1). These components showed satisfactory correlation with all peak areas in their respective subsets. The mean correlations

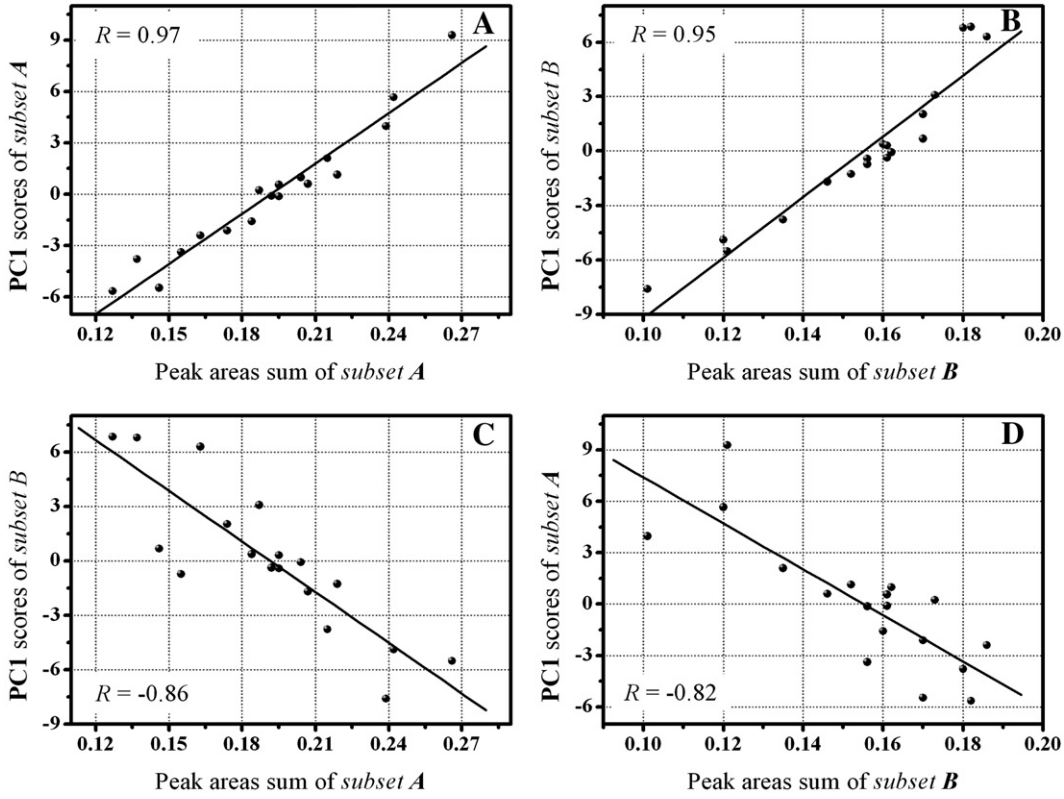

Fig. 3. Correlations between subset peak area sums and the PC1 scores. Correlations between (A) summed peak areas of subset ***A*** and PC1 scores from subset ***A***; (B) summed peak areas of subset ***B*** and PC1 scores from subset ***B***; (C) summed peak areas of subset ***A*** and PC1 scores from subset ***B***; (D) summed and peak areas of subset ***B*** and PC1 scores from subset ***A***.

Table 2
Central composite design for three variables and PC1 scores responses.

	Runs	x_1	x_2	x_3	Resposnses	
					Subset ***A***	Subset ***B***
Factorial points	18	−1	−1	−1	5.66	−4.888
	1	1	−1	−1	−5.46	0.668
	9	−1	1	−1	9.28	−5.525
	13	1	1	−1	−2.12	2.026
	16	−1	−1	1	0.61	−1.693
	4	1	−1	1	−2.40	6.293
	5	−1	1	1	1.14	−1.278
	17	1	1	1	−3.78	6.795
Centre points	3	0	0	0	0.99	−0.076
	7	0	0	0	−3.37	−0.720
	11	0	0	0	−1.57	0.372
	15	0	0	0	0.57	0.305
Axial points	2	−1.682	0	0	3.97	−7.605
$\alpha = 8^{1/4} \approx 1.682$	6	1.682	0	0	−5.64	6.837
	14	0	−1.682	0	−0.13	−0.419
	10	0	1.682	0	−0.09	−0.387
	8	0	0	−1.682	2.10	−3.785
	12	0	0	1.682	0.24	3.082

Experimental domain

Variables	−1.682	−1	0	1	1.682
x_1: Temperature/°C	23	30	40	50	57
x_2: Pre-equilibrium temperature/°C	1.5	5	10	15	18
x_3: Extraction time/min	6.5	10	15	20	23

(±standard deviation) between the peak areas from subset ***A*** and the respective PC1 scores was 0.80 (±0.10) and for subset ***B*** the mean correlation was 0.90 (±0.05).

Another interesting trend noticed is that the summed peak area of both subsets is highly positively correlated to the respective PC1 scores, as can be seen in Fig. 3A and B. On the other hand, a negative correlation is observed between the sum of peak areas from subsets ***A*** and ***B*** and the PC1 scores from subsets ***B*** and ***A***, respectively (Fig. 3C and D). Reasonable explanation for this behavior may lie in the fact that highly correlated peak areas, when compressed by PCA into single latent information, will be exploited similarly in integration where the latent information is basically repeated. In this sense, the sum of peak areas could be also used for multiple response optimizations, but the PCA is preferred since redundant information is removed, the signal/noise ratio is increased, and therefore the quality of the models are improved [41].

Based on the above discussion and the correlations presented in Fig. 3, the use of the first PC scores of each subset as the representative response (Table 2) in the RSM is well supported.

After obtaining the PC scores for subsets ***A*** and ***B***, the central composite design was used to perform the optimization. However, it was verified that quadratic coefficients were not significant and thus, a linear model was fitted using the parsimony principle.

Table 3
Statistical analysis of the model for the subsets ***A*** and ***B***. The coefficients are coded.

	Subset ***A***				Subset ***B***			
	Coefficients	Error	t (3)	p	Coefficients	Error	t (3)	p
Intercept	0	0.48	1×10^{-15}	1	1.1×10^{-15}	0.12	9×10^{-15}	1
T	−3.41*	0.55	6.24	0.008	3.91*	0.14	28.88	9.1×10^{-5}
PET	0.45	0.55	0.83	0.468	0.12	0.14	0.91	0.4285
Ext	−1.09	0.55	2.00	0.139	2.15*	0.14	15.88	0.0005
$T\times$PET	−0.27	0.71	0.38	0.728	0.26	0.18	1.47	0.2381
$T\times$Ext	1.82	0.71	2.55	0.084	0.37	0.18	2.08	0.1287
PET×Ext	−0.98	0.71	1.37	0.265	0.02	0.18	0.14	0.8982

* Significant coefficients using significance level of 0.05 and three degrees of freedom for the t test using the pure error.

Table 4
ANOVA for the two linear models built.

Variation source	SS	*df*	MS	*F*	*p*
*Subset **A***					
Regression	212.94	6	35.49	13.78	0.0002
Residual	28.33	11	2.58		
Lack-of-fit	13.23	8	1.65	0.33	0.9074
Pure error	15.08	3	5.03		
Total SS	241.27	17			
R	0.99				
*Subset **B***					
Regression	274.30	6	45.72	113.48	3.03×10^{-9}
Residual	4.43	11	0.40		
Lack-of-fit	3.84	8	0.48	1.49	0.41
Pure error	0.96	3	0.32		
Total SS	278.73	17			
R	0.99				

SS, sum of squares; *df*, degree of freedom; MS, mean squares.

Table 3 presents the model coefficients for the two linear models built. It can be noticed for subset ***A*** that only the linear effect of temperature (T) was significant and negative, indicating that at lower temperatures a larger amount of volatiles from subset ***A*** was observed. On the other hand, for subset ***B***, the linear effect of the temperature is positive, suggesting that the amount of volatiles for this subset increases with temperature elevation. Another significant effect for subset ***B*** is the linear effect of the extraction time (positive), indicating that the longer the extraction time, the greater is the extracted amount of volatiles from this subset. Consequently, it can be noted

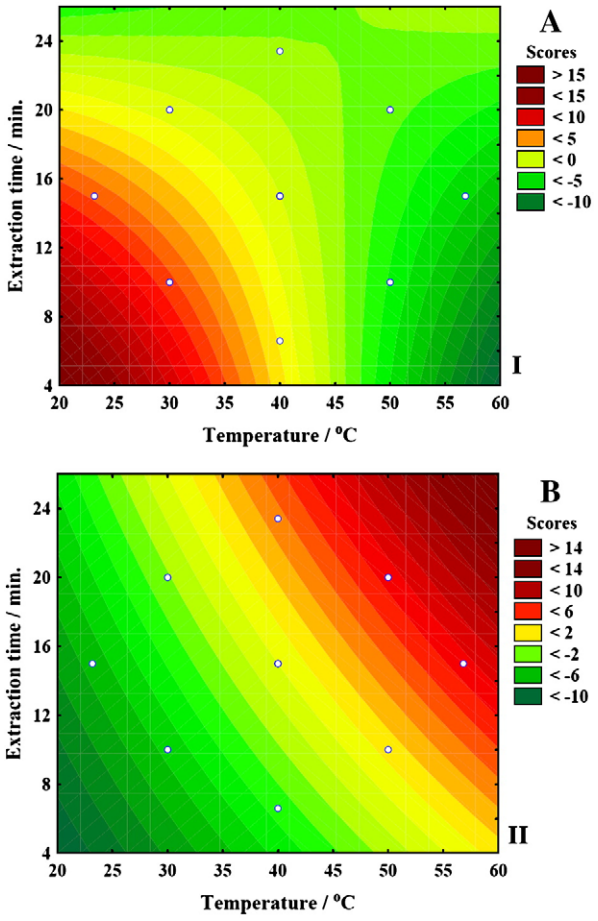

Fig. 4. Response surfaces for subsets ***A*** (**I**) and ***B*** (**II**). Pre-equilibrium time was fixed at 10 min.

from Table 3 that the negative correlation between the two scores is confirmed by the linear effect of temperature in each case.

Analysis of variance (ANOVA) (Table 4) indicates that both regression models are significant ($p<0.05$) and with non significant lack-of-fit ($p>0.05$). These results suggest that the fitted response model can be applied to determine the optimum volatile extraction conditions.

The application of response surface methodology to the PC1 scores of the two subsets ***A*** and ***B*** resulted in two surfaces **I** and **II**, respectively (Fig. 4). They present opposite tendencies: lower bath temperature in **I** and higher bath temperature with longer extraction time in **II**, indicating that it is possible to shift the sorption equilibrium of the system for better extraction of the compounds.

The main physical characteristic of separation in gas chromatography is basically the volatilization of the molecules. So, it is expected that subset ***A*** is composed essentially of compounds with low molar mass, compared with subset ***B*** that is formed by somewhat heavier, less volatile compounds.

At lower temperatures the sorption equilibrium of the system is such that heavier compounds appear in lesser amounts in the headspace compared to the concentration of the lightest compounds and the extraction of the latter becomes more efficient (surface **I** in Fig. 5). On the other hand, the surface response **II** in Fig. 4 indicates that higher temperatures are required to drive the sorption equilibrium in a way to enhance the concentration of heavier compounds in the headspace compared to the lighter compounds. The results might suggest that for temperatures higher than those used in the defined interval of the CCD, the extraction of the heavier compounds could be enhanced. However, higher temperatures are not feasible since the compounds do not stay adsorbed onto the fiber and tend to return to the headspace. Therefore, it is not advisable to perform the extraction at even higher temperatures. According to surface **II**, efficient extraction of the heavier compounds can be achieved without raising the temperature, but by extending the extraction time (also a significant effect). This strategy has been also found in the literature [10,16].

The equilibrium of extracted compounds using a PDMS/DVB fiber can be described by the Langmuir adsorption isotherm. According to this isotherm, the PDMS/DVB fiber possesses a limited number of active sites (pores) at the surface, so that the amount of extracted analytes would be directly proportional to the number of these sites. Therefore, the relationship between the amount of extracted material and its concentration in the sample is fairly linear, except for high concentrations.

Since sorption is a competitive process, molecules with lower affinity for the SPME fiber are substituted by those of higher affinity. At the beginning, lighter molecules (more volatile) quickly adhere to the fiber surface but then, by increasing the extraction time, they are gradually substituted by heavier molecules with better affinity.

In this study, the 42 regions (corresponding to 57 peak areas) were considered in two opposite optimized conditions. But unfortunately, in a real extraction only one condition has to be selected in order to extract efficiently all compounds from subsets ***A*** and ***B*** simultaneously. Investigating surfaces **I** and **II**, and the chromatographic profiles in Fig. 5, it can be seen that, in general, the amount of lighter compounds extracted is higher, compared with that of the heavier compounds that appear at the end of the chromatogram. In this way, temperature values tending to better extraction of the heavier volatile compounds, together with longer extraction times could help in the extraction of the majority of the volatile coffee compounds. The concentration of the lighter compounds would decrease slightly, however, with a significant improvement of the heavier compounds. Based on the above discussion, an intermediate set of conditions was selected, namely temperature of 40 °C and extraction time of 23 min, as being appropriate for extraction of all volatiles.

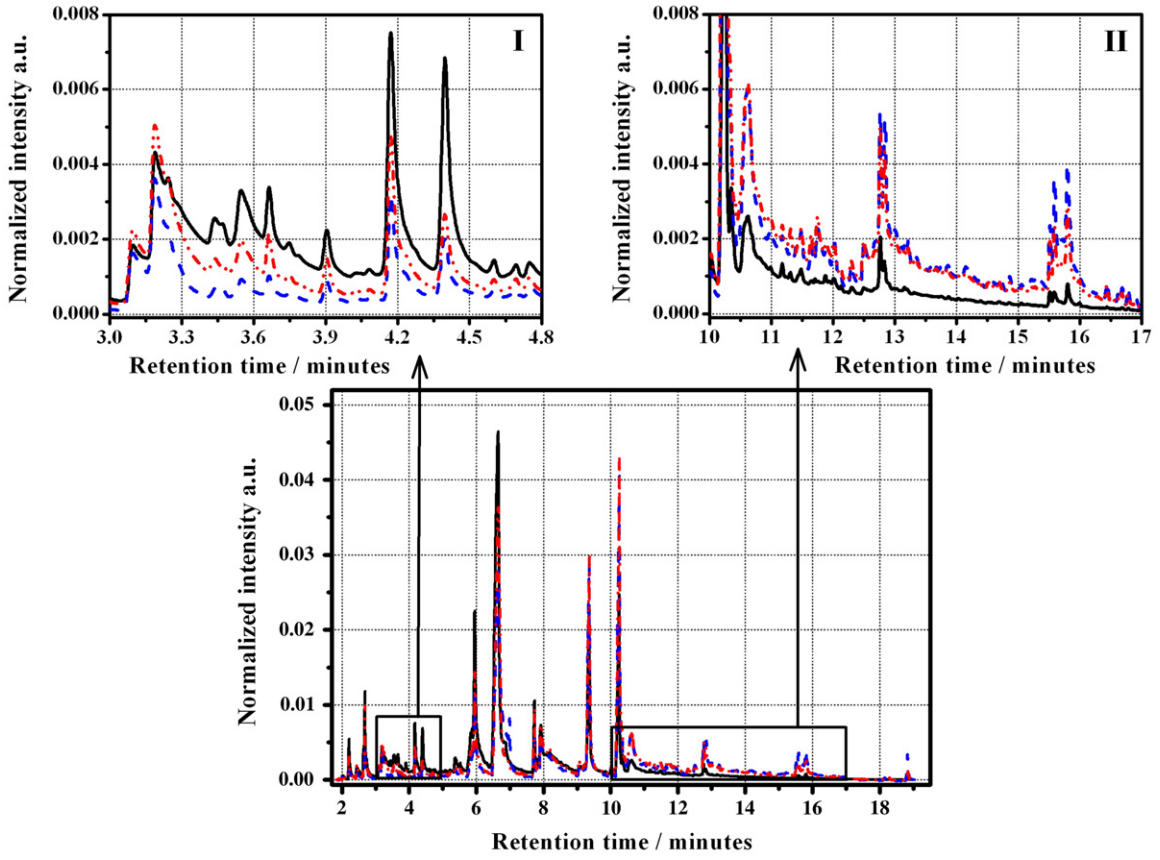

Fig. 5. Typical chromatograms showing enlarged regions corresponding to subset ***A*** (**I**), and subset ***B*** (**II**). Solid line ($T=23$ °C, $Ext=15$ min); dashed line ($T=57$ °C, $Ext=15$ min), dash–dot line ($T=40$ °C, $Ext=23$ min).

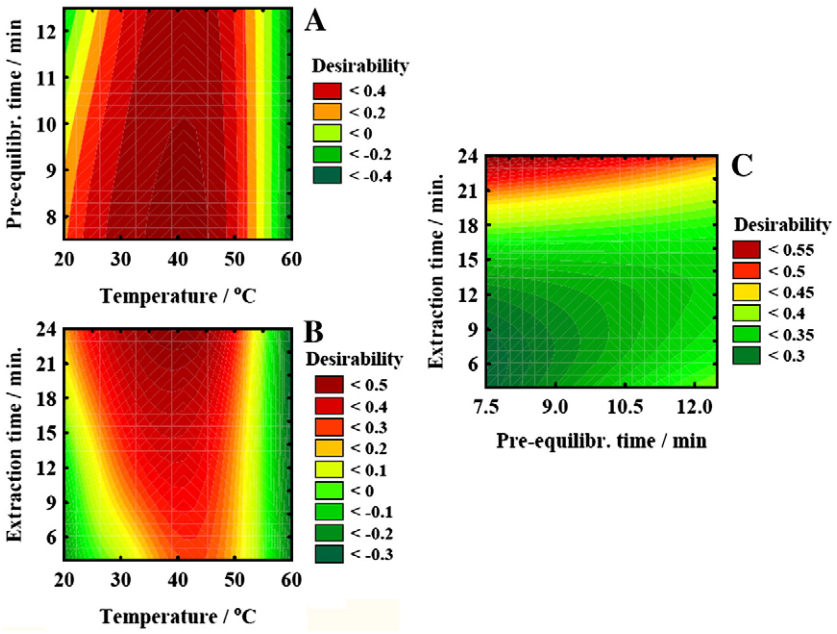

Fig. 6. Response surfaces corresponding to the desirability function when the factors *temperature*, *pre-equilibrium time* and *extraction time* were optimized by analyzing 60 responses simultaneously.

Fig. 5 shows chromatograms obtained using the optimal experimental conditions indicated by the two response surfaces (temperature = 23 °C and extraction time = 15 min for surface **I**; temperature = 57 °C and extraction time = 15 min for surface **II**) and a chromatogram obtained by the suggested optimum experimental condition (T = 40 °C, *Ext* = 23 min). The chromatograms show that low temperatures were more effective in extracting higher amounts of light compounds, as expected. However, the chromatogram obtained from optimum conditions indicated by surface **II**, shows that the extraction of heavy compounds was more intense at higher temperatures. The last chromatogram using the suggested condition shows a balance, leading to satisfactory extraction of lighter compounds and a greater extraction of the heavier compounds.

3.2. Desirability function

Using the desirability function approach, the minimum acceptable peak area was set as $d_i = 0$ (value totally undesirable), the median value was considered as $d_i = 0.5$ and the maximum value as $d_i = 1$ (totally desirable value). These criteria were applied for all 42 peak areas. The individual values of d_i were obtained and then combined into a global function D that was maximized choosing the best conditions of the designed variables.

Fig. 6 shows the contour plots of D for two experimental parameters with the other held at its optimum. According to the response surfaces in Fig. 6 temperature values around 40 °C, equilibrium time ranging from 8 to 10 min and extraction time extending up to 20 min are good experimental conditions for an efficient extraction of all the volatiles regarding the chromatographic peaks selected. These conditions are in good agreement with those suggested by PCA–RSM methodology.

These results indicate that simultaneous optimization of several responses has been efficiently accomplished by both methods. However, the desirability function, although objective and efficient to find the optimal conditions from multiple responses, is not as easily interpretable as the PCA–RMS approach, from the chemical point of view. Besides that, in PCA–RMS one can choose the region of the surface that could be used in the extraction of specific desired compounds.

4. Conclusions

The use of PCA for data compression prior to building the surface responses was of great importance to define the optimum chromatographic conditions for simultaneous extraction of volatile compounds of low and high molar masses from roasted coffee beans. With this strategy, multiple responses could be simultaneously handled without the necessity to use complex methodologies. The high correlation among the chromatographic peak areas, independent of the experimental conditions, makes possible the use of the first component of PCA as the analytical response.

The response surface analyses have indicated the importance of temperature for extraction of different kinds of volatile compounds. Due to the sorption equilibrium of the system, extraction time was another important parameter for extraction of heavier compounds. Similar optimum experimental conditions were obtained by PCA–RSM and the desirability function.

Acknowledgements

This work was supported by grants from CAPES and FAPESP. The authors also acknowledge the Agronomic Institute of Campinas for supplying the Arabica coffee sample and Dr. Carol H. Collins for English revision.

References

[1] M. Careri, A. Mangia, G. Barbieri, L. Bolzoni, R. Virgili, G. Parolari, J. Food Sci. 58 (1993) 968.
[2] L.F. Cuevas-Glory, J.A. Pino, L.S. Santiago, E. Sauri-Duch, Food Chem. 103 (2007) 1032.
[3] M. Ishikawa, O. Ito, S. Ishizaki, Y. Kurobayashi, A. Fujita, Flavour Frag. J. 19 (2004) 183.
[4] C. Bicchi, C. Iori, P. Rubiolo, P. Sandra, J. Agric. Food Chem. 50 (2002) 449.
[5] L. Mondello, A. Casilli, P.Q. Tranchida, P. Dugo, R. Costa, S. Festa, G. Dugo, J. Sep. Sci. 27 (2004) 442.
[6] M. Pardo, G. Sberveglieri, IEEE T. Instrum. Meas. 55 (2002) 1334.
[7] C. Yeretzian, A. Jordan, W. Lindinger, Int. J. Mass Spectrom. 223 (2003) 115.
[8] L.F. Huang, M.J. Wu, K.J. Zhong, X.J. Sun, Y.Z. Liang, Y.H. Dai, K.L. Huang, F.Q. Guo, Anal. Chim. Acta 588 (2007) 216.
[9] M. Poliak, M. Kochman, A. Gordin, A. Amirav, Chromatographia 64 (2006) 487.
[10] C.L. Arthur, J. Pawliszyn, Anal. Chem. 62 (1990) 2145.
[11] D.A. Lambropoulou, I.K. Konstantinou, T.A. Albanis, J. Chromatogr. A 1152 (2007) 70.

[12] H. Lord, J. Pawliszyn, J. Chromatogr. A 885 (2000) 153.
[13] G.A. da Silva, F. Augusto, R.J. Poppi, J. Chromatogr. A 1138 (2007) 251.
[14] J. Pawliszyn, Solid Phase Microextraction: Theory and Practice, Wiley-VCH, New York, 1997.
[15] D.D. Roberts, P. Pollien, C. Milo, J. Agric. Food Chem. 48 (2000) 2430.
[16] H. Kataoka, H.L. Lord, J. Pawliszyn, J. Chromatogr. A 880 (2000) 35.
[17] C.W. Ho, W. Aida, M.Y. Maskat, H. Osman, J. Food Compos. Anal. 19 (2006) 822.
[18] C.G. Zambonin, L. Balest, G.E. De Benedetto, F. Palmisano, Talanta 66 (2005) 261.
[19] M. Akwama, K. Murakami, M. Ikeda, K. Iwatsuki, A. Wada, K. Tokuno, M. Onishi, H. Iwabuchi, J. Food Sci. 72 (2007) C388.
[20] P. Agresti, A.S. Franca, L.S. Oliveira, R. Augusti, Food Chem. 106 (2008) 787.
[21] S. Rocha, L. Maeztu, A. Barros, C. Cid, M.A. Coimbra, J. Sci. Food Agric. 84 (2004) 43.
[22] M. Akiyama, K. Murakami, N. Ohtani, K. Iwatsuki, K. Sotoyama, A. Wada, K. Tokuno, H. Iwabuchi, K. Tanaka, J. Agric. Food Chem. 51 (2003) 1961.
[23] M. Mousavi, E. Noroozian, A. Jalali-Heravi, A. Mollahosseini, Anal. Chim. Acta 581 (2007) 71.
[24] S.Y. Moon, E. Li-Chan, Food Chem. 88 (2004) 141.
[25] M. Polo, C. Garcia-Jares, M. Llompart, R. Cela, Anal. Bioanal. Chem. 388 (2007) 1789.
[26] A. Martinez-Urunuela, J.M. Gonzalez-Saiz, C. Pizarro, J. Chromatogr. A 1056 (2004) 49.
[27] C.H. Kowalski, G.A. da Silva, R.J. Poppi, H.T. Godoy, F. Augusto, Anal. Chim. Acta 585 (2007) 66.
[28] C. Sanz, D. Ansorena, J. Bello, C. Cid, C. J. Agric. Food Chem. 49 (2001) 1364.
[29] R.H. Myers, D.C. Montgomery, Response Surface Methodology, 2 ed, John Wiley & Sons, New York, 2002, p. 798.
[30] E. Del Castillo, D.C. Montgomery, J. Qual. Technol. 28 (1993) 347.
[31] W.M. Carlyle, D.C. Montgomery, G.C. Runger, J. Qual. Technol. 31 (2000) 1.
[32] E.C.J. Harrington, Ind. Qual. Control 21 (1965) 494.
[33] G. Derringer, R. Suich, J. Qual. Technol. 12 (1980) 214.
[34] K. Abu-Izza, L. Garcia-Contreras, D.R. Lu, J. Pharm. Sci. 85 (1996) 572.
[35] R.H. Myers, W.H. Carter Jr., Technometrics 15 (1973) 301.
[36] A.I. Khuri, M. Conlon, Technometrics 23 (1981) 363.
[37] J.J. Pignatiello, IIE Trans. 25 (1993) 5.
[38] A.E. Ames, N. Mattucci, S. MacDonald, G. Szonyi, D.M. Hawkins, J. Qual. Technol. 29 (1997) 339.
[39] G.G. Vining, J. Qual. Technol. 30 (1998) 309.
[40] N. Bratchell, J. Chemom. 3 (1989) 579.
[41] R. Carlson, Å. Nordahl, T. Barth, R. Myklebust, Chemom. Intell. Lab. Syst. 12 (1992) 237.
[42] Sandström, R. Carlson, A.L. Sunesson, J.O. Levin, A.P.F. Turner, Analyst 126 (2001) 2008.
[43] M.R. Ellekjaer, M.A. Ilseng, T. Naes, Food Qual. Prefer. 7 (1996) 29.
[44] C.T. Su, L.I. Tong, Total Qual. Manage. 8 (1997) 409.
[45] J. Antony, Qual. Reliab. Eng. Int. 16 (2000) 3.
[46] C.P. Fung, P.C. Kang, J. Mater. Process. Technol. 170 (2005) 602.
[47] L.I. Tong, C.H. Wang, H.C. Chen, Int. J. Adv. Manuf. Technol. 27 (2005) 407.
[48] H.C. Liao, Int. J. Adv. Manuf. Technol. 27 (2006) 720.
[49] D.C. Montgomery, G.C. Runger, Applied Statistics and Probability for Engineers, 3rd ed, John Wiley & Sons, New York, 2003, p. 171.
[50] H. Martens, T. Naes, T. Multivariate, Calibration, John Wiley & Sons, New York, 1989, p. 419.
[51] K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics: A Practical Guide, John Wiley & Sons, New York, 1998, p. 348.
[52] S. Wold, K. Esbensen, P. Geladi, Chemom. Intell. Lab. Syst. 2 (1987) 37.
[53] R.G. Brereton, Analyst 125 (2000) 2125.
[54] A. Freitas, C. Parreira, L. Vilas-Boas, Chromatographia 54 (2001) 647.
[55] R.F. Teófilo, M.M.C. Ferreira, Quim. Nova 29 (2006) 338.
[56] G.E.P. Box, N.R. Draper, Empirical Model-building and Response Surfaces, John Wiley & Sons, New York, 1987.
[57] R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, John Wiley & Sons, Chichester, 2003, p. 407.
[58] I. Arambarri, M. Lasa, R. Garcia, E. Millan, J. Chromatogr. A 1033 (2004) 193.
[59] A.C. Atkinson, R.D. Tobias, J. Chromatogr. A 1177 (2008) 1.
[60] http://lqta.iqm.unicamp.br.