A Flow System for Generation of Concentration Perturbation in Two-Dimensional Correlation Near-Infrared Spectroscopy: Application to Variable Selection in Multivariate Calibration

CLAUDETE FERNANDES PEREIRA* and CELIO PASQUINI

Departamento de Química, Universidade Federal da Paraíba, CP 5093, 58051-970, João Pessoa, PB, Brazil (C.F.P.); and Department of Analytical Chemistry, Institute of Chemistry, University of Campinas, POB 6154, 13084-971 Campinas, SP, Brazil (C.P.)

A flow system is proposed to produce a concentration perturbation in liquid samples, aiming at the generation of two-dimensional correlation near-infrared spectra. The system presents advantages in relation to batch systems employed for the same purpose: the experiments are accomplished in a closed system; application of perturbation is rapid and easy; and the experiments can be carried out with micro-scale volumes. The perturbation system has been evaluated in the investigation and selection of relevant variables for multivariate calibration models for the determination of quality parameters of gasoline, including ethanol content, MON (motor octane number), and RON (research octane number). The main advantage of this variable selection approach is the direct association between spectral features and chemical composition, allowing easy interpretation of the regression models.

Index Headings: Flow system perturbation; Two-dimensional correlation spectra; 2D-COS; Near-infrared spectroscopy; Gasoline; Variable selection; Multivariate calibration.

INTRODUCTION

In spectroscopic studies, the establishment of relationships between spectral information and chemical structure or composition of the analyzed samples is sometimes very difficult. This difficulty arrives because, in most of the cases, the spectra are very complex and composed of multiple overlapping features that are difficult to attribute to the observed functionalites. To help solve this problem, many methods have been proposed in the literature, such as Darling–Dennison resonance, deuteration and polarization methods,[1] and derivative pretreatments and partial least squares regression.[2–4]

In recent years, two-dimensional (2D) spectroscopy has gained the attention of the scientific community, aiming at spectral interpretation. This method employs cross-correlation analysis that emphasizes information not easily seen in the original spectrum. Through these 2D spectra it is possible to simplify the analysis and interpretation of complex spectra consisting of overlapping bands and investigate the occurrence of inter- and intra-molecular interactions by correlating absorption bands from different groups or parts of a molecule.[5–7]

The seminal idea of utilizing a second dimension for enhancement of spectral resolution can be traced back to nuclear magnetic resonance (NMR) spectroscopy.[8] Currently, this method can be used for almost every spectroscopy technique. This is possible by using a mathematical formalism proposed in 1993 by Noda[9] that permits two-dimensional correlation spectra to be obtained from any transient or time-resolved spectra having an arbitrary waveform. This method is called generalized 2D correlation spectroscopy (2D-COS).

A 2D experiment can be carried out using a conventional spectrometer and submitting the sample to an external perturbation, as shown in Fig. 1. Thus, an external electrical, thermal, magnetic, chemical, acoustic, or mechanical perturbation is applied to a system and the dynamic fluctuations thus induced are monitored with spectroscopic probes. These fluctuations are manipulated mathematically by using a correlation method and finally a 2D spectrum is obtained.

Among the many types of perturbations utilized, that based on the change of concentration deserves special attention because many processes are influenced by this important chemical variable.[10–20] Usually, this type of perturbation is produced in batch systems for which large amounts of reagent are necessary in a time-consuming procedure to prepare the samples. Flow systems appear as a good alternative to those batch systems. Diewok et al.[21] applied a stopped flow cell to carry out a comparative study between 2D correlation spectroscopy and multivariate curve resolution in analyzing pH-dependent evolving systems monitored by mid-infrared spectroscopy. In 2009, Carmona et al.,[22] described a micro-dialysis flow cell for investigating hydrogen/deuterium (H/D) exchange kinetics in biomolecules using transmission infrared and 2D correlation spectroscopies.

In the present work, a simple flow system is proposed, designed to generate the concentration perturbation necessary to perform two-dimensional correlation near-infrared (NIR) spectroscopy. This system has been evaluated in the production of spectral information (dynamic spectra). The 2D correlation spectra thus obtained were employed to supply information for the selection of relevant variables for multivariate calibration models used for the determination of quality parameters of gasoline, such as ethanol content, motor octane number (MON), and research octane number (RON). The proposed variable selection approach is compared with well-accepted procedures such as Jack-Knife selection, based on the stability of the regression coefficients[3] and the successive projection algorithm (SPA).[23]

EXPERIMENTAL

Reagents and Samples. Analytical grade reagents were used throughout. Typical samples of alcohol-free gasoline produced by petroleum refineries in Brazil were supplied by REPLAN-Petrobras (Campinas, SP). Samples of gasoline with known ethanol content were prepared from three typical gasolines originally free of ethanol and analytical grade absolute alcohol. The sample set was split into calibration and validation sets containing 21 and 7 samples, respectively.

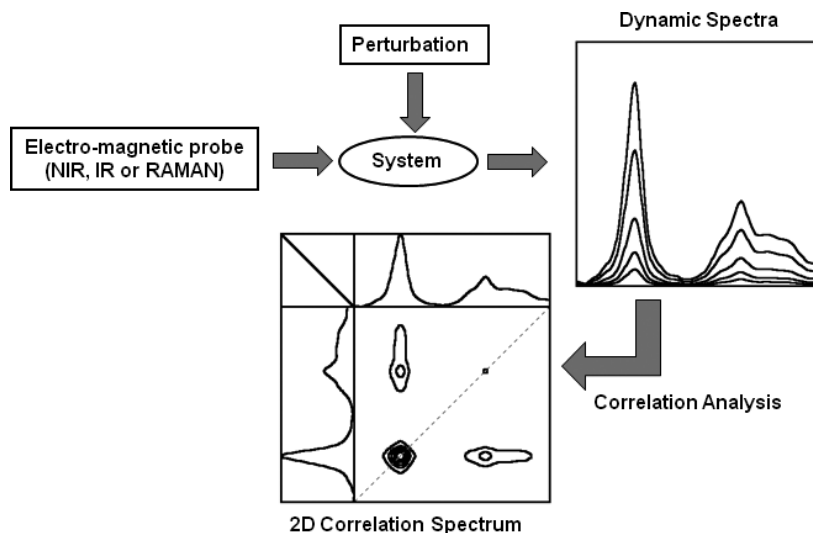* Author to whom correspondence should be sent. E-mail: claudete@quimica.ufpb.br.

FIG. 1.   General scheme of 2D spectroscopy.

Two-hundred and ninety-three (293) synthetic gasoline samples (made by mixing diverse petrochemical residual streams from the petrochemical unit of Petrobrás), were supplied by Braskem S/A (Camaçari, BA), and used for modeling the MON and RON bulk properties. In both cases, the sample set was split into calibration and validation sets containing 196 and 97 samples, respectively.

Flow System for Concentration Perturbation. The proposed flow system for concentration perturbation is shown in Fig. 2. The sample (S) of interest and the species (P) used to produce the concentration perturbation are impelled by a peristaltic pump (C). Two solenoid valves (V1 and V2) control the passage of sample and reagent through the system. For generation of concentration perturbation, the system is initially filled with the sample (V1 is turned on and V2 is turned off). The user requests the start of the perturbation procedure and the computer controlling the system automatically switches V1 off
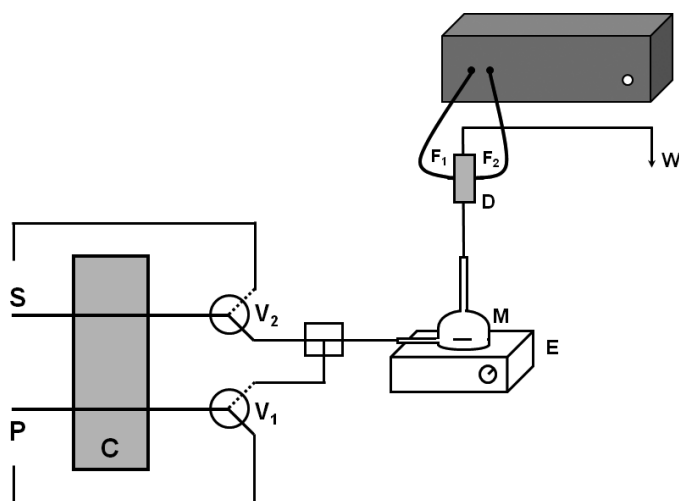

FIG. 2.   Flow system scheme for concentration perturbation. (C) peristaltic pump, (S) sample, (P) perturbing substance, (V$_1$) and (V$_2$) solenoid valves for reagents commutation, (M) stirred mixing chamber, (D) detection cell (optic path approximately 5.0 mm), (E) magnetic stirrer, (F1) and (F2) optical fiber bundles, and (W) waste.

and V2 on and begins spectral data acquisition. A concentration gradient is created inside the magnetically stirred mixing chamber (C) as the sample is continuously replaced by the perturbing species. When the process is finished, the perturbing species occupies the system. The flow rate in this system was fixed at 1.7 mL min^{-1} for each stream and the mixing chamber volume is 1.6 mL.

In the present system, the effect of the concentration perturbation on the sample is monitored by a near-infrared (NIR) probe placed on the detector flow cell (D), which consists of a glass tube with an inner diameter of 5.0 mm and a length of 3.0 cm.

Acquisition of Near-Infrared Spectra. The NIR absorption spectra were obtained using a Brimrose model Luminar 2000 spectrometer operating in the range from 850 to 1800 nm. This instrument is based on an acousto-optic tunable filter (AOTF) monochromator controlled by a computer through the manufacturer's software (Brimrose, Snap 2.03). Each spectrum, containing 475 points equally spaced in wavelength ($\Delta\lambda = 2.0$ nm), is scanned by this instrument in about 80 ms. A pair of low-OH-content optical fiber bundles (200 μm diameter, 1.0 m long), were employed to deliver the monochromatic light selected by the AOTF to the flow cell and to collect the non-absorbed fraction of the radiation that passed through the cell tube (perpendicular to the fluid flow) to return it to the instrument detector. The optical path is roughly determined by the inner diameter of the glass tube (approximately 5.0 mm). Absorbance spectra were obtained by employing the transmittance signal of carbon tetrachloride (CCl_4) as reference. Each spectrum was registered as an average of 50 scans and the spectrometer is programmed to acquire one spectrum every 6 s along the concentration gradient.

The NIR absorption spectra of the samples of synthetic gasoline were obtained by using a batch procedure and by employing a transflectance probe with an effective optical path of 5 mm and with a circular area of 0.5 cm^2 in the range of 850 to 1800 nm.

For measurement of the nine binary mixtures of ethanol and toluene (from 4.0% to 96.0% for each reagent) used for flow system calibration, a transflectance probe with a total optical path of 5 mm and a circular area of 0.5 cm^2 was used.

Calibration of the Flow System. The operation of the flow system generates a nonlinear time-dependent concentration gradient based on the initially pumped sample. The gradient initiates immediately after the valves are switched to computer control.

For simpler calculation of the 2D correlation spectra it is preferable to have evenly spaced sampling. Therefore, the time interval for which a linear gradient of concentration is produced after the valves were switched on was determined.

This operation is called "calibration of the flow system" and is performed by using nine binary solutions containing ethanol and toluene in the range from 4.0% to 96.0% (v/v). These solutions were sequentially pumped through the flow system and, after a steady-state regimen was obtained for each of them, their NIR spectra were acquired. A PLS (partial least squares) regression model was constructed using this spectral data. The pure ethanol and toluene were then pumped in the separate lines (S and P) of the flow system with toluene in the sample line. The system was switched to initiate the concentration gradient and spectra were successively obtained as previously described for six seconds each.

Using the regression model constructed from the set of nine standards, the spectral set registered as a function of time was converted to concentrations of ethanol in toluene. A curve showing the time dependence of the relative concentrations between sample and perturbing stream was obtained. This curve is assumed to be the same for any pair of sample-perturbing stream composition employed in the proposed flow system. This is ensured by the mechanically stirred chamber, which produces a gradient relatively independent of the physico-chemical characteristics of the substances.

A linear range, in which the change in concentration with elapsed time since the valves were switched on is constant, was found between 36 and 72 s (6.7–42.2%, v/v). Therefore, seven sampling spectra, obtained in this time interval, can be employed by the software to generate the 2D spectra.

Data Treatment. All the dynamic spectra obtained using the proposed flow system were preprocessed by offset baseline correction. 2D correlation spectra, from samples submitted to concentration perturbation, were calculated utilizing the software 2D Pocha.[24]

Chemometric treatment of the spectral data was performed using the Unscrambler 9.2 software (CAMO, Norway). This package provides the PLS algorithm used to produce regression models from spectral data.

All regression models were built using first-derivative spectra. These spectra were calculated with a Savitzky–Golay (SG) filter using a second-order polynomial and a five-point window.

The successive projection algorithm (SPA) was employed for variable selection used in the multiple linear regression (MLR) models. The program was written using Matlab 6.1 software.[23]

The statistical significance of differences between the root mean square error of prediction (RMSEP) values was assessed by using *F*-tests at a confidence level of 95%.

RESULTS AND DISCUSSION

Calibration of the Flow System. Figure 3 shows the spectra set obtained for the nine mixtures of ethanol and toluene from 4.0% to 96.0% for each reagent, pumped under steady-state conditions through the flow system. These spectra
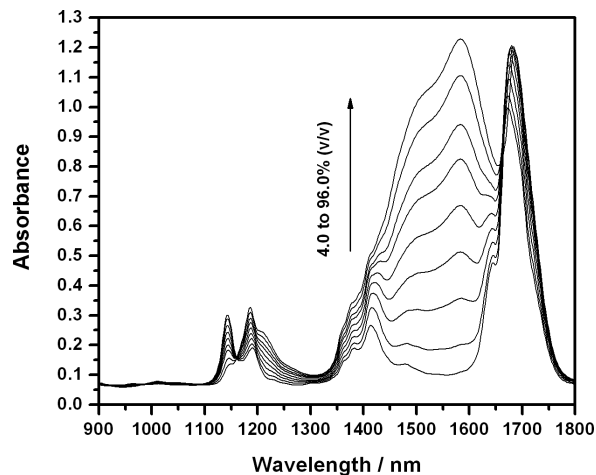
FIG. 3. NIR spectral set obtained for mixtures of ethanol and toluene (4% to 96% v/v).

show spectral variations in the first and second overtones of C–H bonding, which occur between 1600–1900 nm and 1100–1300 nm, respectively. Also, a large variation in the range 1350–1600 nm can be seen and is attributed to the first overtone of combination bands of C–H bonding and to the first overtone of O–H bonding.[25] This set of spectra was utilized to construct a PLS model to estimate the linear range of the concentration gradient obtained with the flow system perturbation. A good correlation coefficient (0.998) and root mean square error of prediction (RMSEP, 1.4 % v/v) were obtained over the full ethanol concentration range (4.0% to 96.0% v/v). These results show the good capability of prediction of the model and it was used to estimate the linear range of the concentration gradient obtained with the flow system. Furthermore, the results for the gradient calibration can be utilized for any pair of sample/perturbing agent as the change of the perturbing species over the sample remains constant and is determined by parameters such as the relative flow ratio and the mixing chamber volume, which are kept constant.

Concentration Perturbation and Two-Dimensional Spectroscopy for Variable Selection Aiming at the Determination of Ethanol in Gasoline. To illustrate the applicability of the flow system for generation of concentration perturbation, it was initially employed to produce 2D correlation spectra for gasoline samples perturbed by ethanol. The aim of this study is the selection of variables to be employed in multivariate regression models for prediction of ethanol content in this type of fuel. The criterion for variable selection is based on the effect of the ethanol on the absorption spectra as determined by synchronous 2D correlation spectra obtained after perturbation of the system (gasoline).

This strategy of using only synchronous 2D correlation spectra was adopted because these spectra show correlations between spectral bands that change in phase with the perturbation applied to the system. On the other hand, the asynchronous 2D correlation spectra give information on bands that change out of phase and whether these changes happen before or after each other. Despite the fact that asynchronous maps can improve the knowledge on a given system, the information can not be useful to variable selection because the multivariate models are constructed (after variable selection) employing spectra of the calibration samples measured in a
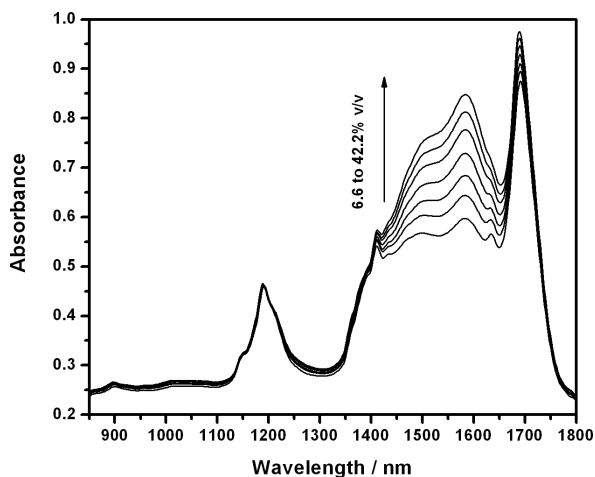
FIG. 4. NIR spectral set obtained for gasoline under ethanol concentration perturbation and used to construct the 2D correlation spectrum.

static way. These measurements will reflect the effect of the analyte as found by using the synchronous 2D correlation spectra.

Gasoline is a complex mixture of many hundreds of different hydrocarbons and their derivatives, most of them found as saturated compounds with four to twelve carbon atoms in the molecule. This composition gives origin to NIR spectra with many bands due to the vibrational modes of the C–H bonds of hydrocarbon molecules. Commercial gasoline in Brazil contains approximately 25% (v/v) anhydrous ethanol. This compound improves the motor octane number (MON) and introduces a partially renewable bio-characteristic to the fuel. The content of ethanol in gasoline is controlled by the Brazilian national fuel authority (Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, ANP).

The spectra set obtained in the linear range of the gradient for ethanol concentration perturbation of a gasoline sample is shown in Fig. 4. An intensification of the band between 1450 and 1650 nm attributed to the first overtone of the O–H bond of ethanol can be seen. On the other hand, the spectral variations influenced by the ethanol concentration in the second overtones of C–H bonding, which occur between 1100 and 1300 nm, are not quite apparent in the original raw data. This fact demonstrates the advantage of using 2D maps for variable selection in multivariate calibration.

Figure 5 depicts the 2D synchronous spectra of ethanol-concentration-dependent spectral variations of gasoline obtained from the data in Fig. 4. The synchronous maps show a number of auto peaks. In the first region (Fig. 5a), there are four auto-peaks that are influenced by the ethanol concentration. These spectral variations are subtle in the original spectra but clear in the 2D correlation spectrum. In the second range (Fig. 5b), a strong auto-peak is observed due to the ethanol O–H bond. The wavelengths associated with these auto-peaks, which are related to ethanol concentration, were used to model the ethanol concentration in gasolines.

The results obtained using the PLS model constructed with the variables selected by the 2D spectrum were compared to PLS models constructed using all spectral variables. Models based on Jack-Knife selection were constructed. Furthermore, an MLR model employing the variables selected by SPA was also constructed. The performances of the resulting models
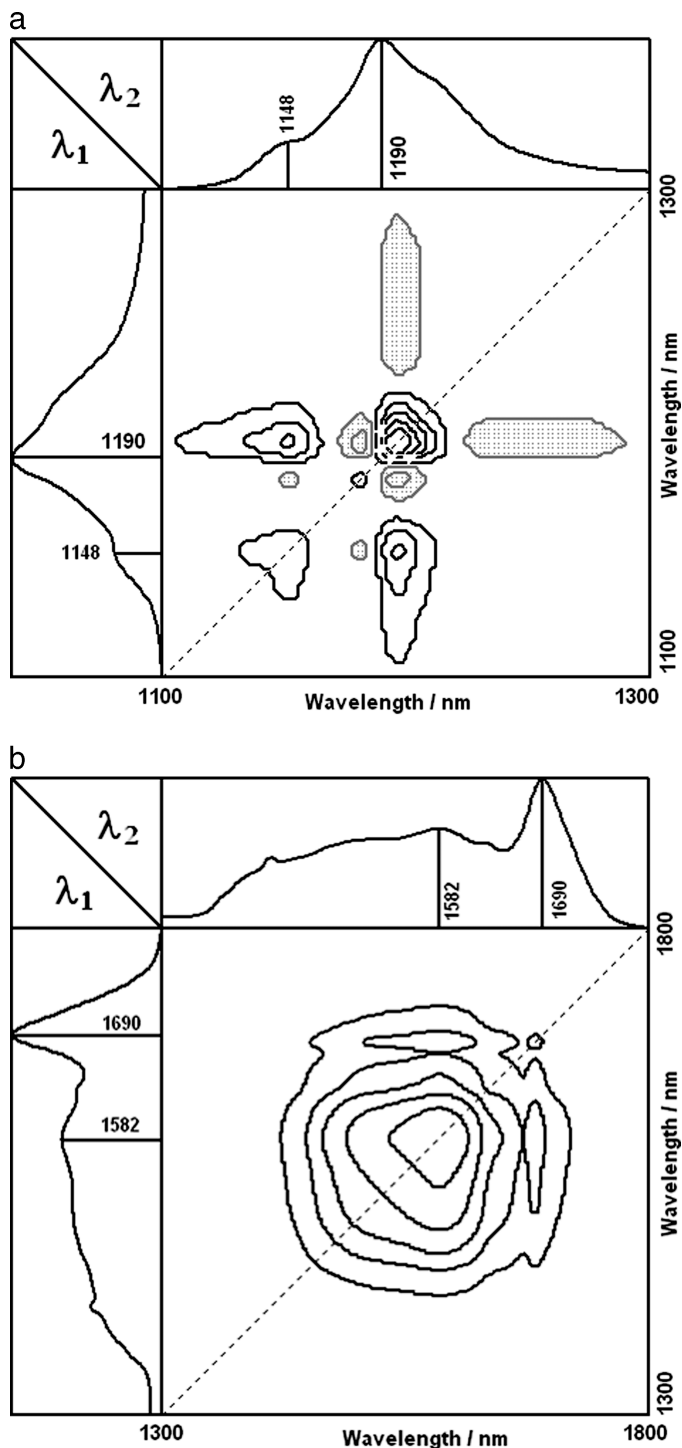

FIG. 5. 2D correlation spectrum (split into two spectral regions) obtained after ethanol concentration perturbation in gasoline. (**a**) Spectral range: 1100–1300 nm, (**b**) spectral range: 1300–1800 nm. Black and gray lines are positive and negative peaks, respectively.

were compared in terms of RMSEP, calculated using external prediction sets that were not included in the modeling stage.

Up to this point it is possible to conclude that the variable selection made with the help of the 2D correlation spectrum obtained after ethanol concentration perturbation can produce a PLS regression model whose performance is statistically undistinguishable from that of the models obtained using other

TABLE I. Overall performance of the multivariate regression models constructed to predict the ethanol content in gasoline.

Model	PLS–2D[a]	PLS–t[b]	PLS–JK[c]	MLR–SPA[d]
RMSEP (% v/v) for uncontaminated gasoline	0.653	0.73	0.73	0.701
RMSEP (% v/v) with 10% contamination by toluene	4.53	5.44	5.44	3.61
Number of spectral variables	131	472	447	4
Number of latent variables	1	1	1	...

[a] Two-dimensional spectroscopy.
[b] All spectral variables.
[c] Jack-Knife selection.
[d] Successive projection algorithm (SPA).

TABLE II. Overall performance of the multivariate regression models constructed to predict the ethanol content in gasoline obtained by the inclusion of contaminated samples in the calibration and prediction sets.

Model	PLS–2D[c]	PLS–t[a]	PLS–JK[c]	MLR–SPA[d]
RMSEP (10% toluene)	1.61	0.77	0.76	2.97
Number of spectral variables	131	476	433	9
Number of latent variables	2	2	2	...

[a] Two-dimensional spectroscopy.
[b] All spectral variables.
[c] Jack-Knife selection.
[d] Successive projection algorithm (SPA).

common variable selection schemes. The advantage of the variable selection method based on 2D spectroscopy is the direct interpretation of the relevance of the variables selected because, in this case, the selection procedure is guided by the influence of the spectral features of the analyte on the final matrix of the sample. The variables selected by other algorithms, based on purely statistical criteria, can also be spectrally interpreted. However, none of them is based directly on the spectral features of the analyte and their effect on the sample matrix as is the 2D approach proposed herein. Therefore, the risk of selecting variables based in some fragile internal correlation is always present when these other methods are employed.

The robustness of the constructed models in relation to severe changes in the gasoline matrix has been evaluated by adding known amounts of toluene, isooctane, and hexane to simulate gasoline adulteration. The objective was to verify which of the proposed calibration schemes (and models produced) are less prone to induce deterioration in the analytical results (ethanol content determinations) when the matrix is changed in relation to the original composition employed in the calibration stage.

Table I shows the RMSEP obtained by the models. The values obtained for samples of non-contaminated gasoline using the different models are very similar and the statistical test can not detect any significant difference at the 95% confidence level.

However, when the toluene content of the gasoline is changed, all models lost their predictive capabilities. This means that the models are influenced more or less at the same level by the matrix effect and that, independent of the variable selection scheme, these models are not robust in relation to matrix composition change.

One way to minimize this type of drawback is to include the variability of the matrix in the data employed for calibration. When this procedure is adopted, the constructed models are more robust. To demonstrate this, PLS and MLR models were constructed employing a calibration set that includes samples with and without toluene in the calibration stage. The validation results of these models, obtained with external sets of samples with and without toluene, are shown in Table II. The models show better prediction ability for the samples whose toluene content was modified. On the other hand, it was verified that the RMSEP are worse than those obtained for models constructed only with gasoline having the original composition. In a general way, PLS–t and PLS–JK are statistically similar and PLS–2D and MLR/SPA models are

different from other models. These models show more susceptibility to the matrix effect (Table II).

Concentration Perturbation and Two-Dimensional Spectroscopy for Variable Selection Aiming at the Determination of Motor Octane Number and Research Octane Number in Synthetic (Manufactured) Gasoline. Many analytical applications of NIR spectroscopy aim to determine properties that are associated not with a single analyte but with the bulk composition of the sample. Examples of such applications are those related to the determination of the motor and research octane number and cetane number of gasoline and diesel fuel, respectively. However, the relative contribution of different species to the property is generally known. This prior information can be employed to select some representative pure substances that can be used in the concentration perturbation scheme in order to search for significant spectral variables to be employed by a multivariate regression model constructed to predict the bulk quality property. For instance, the MON and RON are directly proportional to the hydrocarbon composition of the gasoline. Molecules with linear chains contribute to reduce MON and RON, while branched and aromatic hydrocarbons contribute to increase the values of these properties.

In order to evaluate the proposed system for variable selection, aiming at construction of models for determination of MON and RON in synthetic gasolines, successive perturbations were induced employing pure octane, isooctane, and toluene as model compounds.

The auto-peaks detected in the 2D correlation spectra obtained for the three perturbations allow selection of the spectral regions shown in Table III. Octane concentration perturbation (which reflects a decreasing effect in MON and RON) in gasoline shows an auto-peak of relatively high intensity between 1200 and 1226 nm, which is attributed to methylene groups. For isooctane (known to enhance MON and

TABLE III. Spectral ranges selected by means of the auto-peaks found in two-dimensional correlation spectroscopy for construction of multivariate regression models for prediction of MON and RON in gasoline.

Reagent	Spectral range (nm)
Octane	1188–1200
	1208–1210
	1658–1688
Isooctane	1200–1226
	1658–1696
Toluene	1130–1152
	1172–1234
	1658–1688

TABLE IV. Overall performance of the multivariate regression models for determination of MON in synthetic gasoline.

Model	PLS–2D[a]	PLS–t[b]	PLS–JK[c]	MLR–SPA[d]
RMSEP	0.53	0.55	0.54	0.52
Number of spectral variables	44	476	99	15
Number of latent variables	3	7	4	...

[a] Two-dimensional spectroscopy.
[b] All spectral variables.
[c] Jack-Knife selection.
[d] Successive projection algorithm (SPA).

TABLE V. Overall performance of the multivariate regression models for determination of RON in synthetic gasoline.

Model	PLS–2D[a]	PLS–t[b]	PLS–JN[c]	MLR–SPA[d]
RMSEP	0.67	0.74	0.72	0.69
Number of spectral variables	44	476	97	14
Number of latent variables	3	4	5	...

[a] Two-dimensional spectroscopy.
[b] All spectral variables.
[c] Jack-Knife selection.
[d] Successive projection algorithm (SPA).

RON) concentration perturbation, the 2D spectrum shows an auto-peak of low intensity, between 1186 and 1200 nm, attributed to the methyl group. This compound also produces another auto-peak of low intensity between 1206 and 1212 nm (methylene group). The spectral variations from C–H aromatic bonds were observed in the 2D synchronous spectrum of toluene in gasoline as two auto-peaks with relatively high intensities. The first one, between 1130 and 1152 nm, is attributed to C–H aromatic bonds and the second, between 1172 and 1234 nm, contains absorption regions of methyl and methylene groups.

The spectral region from 1658 to 1688 nm, although it can be attributed to aromatic C–H bonds, was not employed for the construction of multivariate models, since this variation was observed for all three different compounds. Therefore, it was possible to conclude that this spectral region is affected by species that cause increases and reductions of the octane number.

Again, the results obtained by employing 2D spectroscopy were compared with those found by more common variable selection algorithms as described before. Table IV presents the RMSEP results achieved by the four modeling strategies for determination of MON. As can be seen, all the models show similar performances, as proven by the *F*-test at the 95% confidence level.

Some characteristics of these models must be highlighted. The model produced by MLR/SPA could be considered the best for MON prediction because it is the simplest. Furthermore, it uses a smaller number of spectral variables (only 15 variables). However, interpretation of the variables selected is not easy if not supported by data acquired from 2D spectroscopy.

Among the PLS models, that constructed with variables selected with the help of 2D spectroscopy showed an interesting performance. In this case, only 44 spectral variables were used and the number of latent variables (3) was smaller than for the other PLS models. These latent variables, probably, explain the relative spectral contribution of the three model compounds used for concentration perturbation of gasoline.

The MON parameter of gasoline can be determined with high accuracy using any of the four modeling strategies. The RMSEP values obtained are excellent, indicating that NIR can predict motor octane number with an error lower than the reproducibility (2.0) of the standard conventional test (ASTM 2700).[26]

The results for the prediction of the RON of gasoline are shown in Table V. It can be seen that the PLS/2D model yielded the smallest RMSEP for this parameter. However, the *F*-test reveals that there is not a significant difference in comparison to the results obtained for the other strategies of variable selection. The RMSEP values obtained with the four modeling strategies are similar to the reproducibility (0.7) of the standard test (ASTM 2699).[27]

An important characteristic of PLS/2D models produced for MON and RON determination must be emphasized. In both cases, a small number of latent variables was necessary for the PLS models to predict the MON parameter. Therefore, these models, according to the parsimony criterion, must be more robust.

CONCLUSION

A flow system has been proposed and evaluated to automate the production of perturbation by concentration in two-dimensional correlation spectroscopy. The system is robust, operates in a closed environment (avoiding contamination and exposure of the user to hazardous chemicals), employs small volumes of solutions, and allows for rapid access (approximately 1.5 min) to the dynamic spectra necessary to produce the 2D correlation maps. The change of the concentration caused by a given analyte over a sample matrix can be used to extract relevant spectral information directly related to the perturbing species and/or to its interaction with the sample matrix. This last type of information is not available from the simple absorption spectra of a pure substance.

In order to illustrate the usefulness of the proposed system, the concentration perturbation and the resulting 2D spectra were employed to guide the selection of relevant variables to be used in multivariate regression models to predict quality properties of gasoline. In summary, the wavelengths associated with the auto-peaks observed in the 2D maps resulting from concentration perturbations are considered the most relevant source of spectral information for multivariate regression models.

Two applications were evaluated. The first demonstrates the use of the variable selection through 2D spectra for determination of a concentration-dependent parameter (ethanol content in gasoline). The second deals with the determination of bulk quality properties of gasoline (MON and RON).

The results achieved by the model constructed with the variables selected by 2D spectroscopy in predicting the ethanol content in gasoline show a performance indistinguishable from that of other models employing more common algorithms. However, the approach based on 2D spectroscopy provides a direct (physical–chemical based) interpretation of the relevance of the variables employed in the construction of the model.

It was observed that, independent of the variable selection algorithm employed, the regression models based on NIR spectroscopy to predict the ethanol content of gasolines are influenced by alteration of the sample matrix. However, the

inclusion of spectra of samples with similar composition in the regression models minimizes this influence.

The application of the perturbation flow system and 2D correlation spectroscopy to the determination of MON and RON illustrates the potential of using the proposed system and variable selection strategy for the determination of bulk properties of the sample. In this case, the knowledge of the influence of different types of hydrocarbons on the MON and RON helps the selection of the variables employed for multivariate regression models.

The results obtained by the proposed strategy of variable selection for MON and RON determination were compared with other strategies of variable selection and no statistical differences were detected at the 95% confidence level. However, the proposed strategy presents two significant advantages. Interpretation of the selected variables is evident and the regression models could be constructed with a small number of latent variables, adding robustness to the model.

1. J. Workman, Jr. and L. Weyer, *Practical Guide to Interpretative Near-Infrared Spectroscopy* (CRC Press, New York, 2008).
2. P. Williams and K. Norris, *Near-Infrared Technology in the Agricultural and Food Industries* (American Association of Cereal Chemists, St. Paul, MN, 2001), 2nd ed.
3. F. Westad, A. Schmidt, and M. Kermit, J. Near Infrared Spectrosc. **16,** 265 (2008).
4. S. Sasic and Y. Ozaki, Appl. Spectrosc. **54,** 1327 (2000).
5. I. Noda, A. E. Dowrey, C. Marcott, and G. M. Story, Appl. Spectrosc. **54,** 236A (2000).
6. Y. Ozaki, S. Sasic, T. Tanaka, and I. Noda, Bull. Chem. Soc. Jpn. **74,** 1 (2000).
7. C. F. Pereira and C. Pasquini, Quim. Nova **29,** 143 (2006).
8. R. R. Ernst, G. Bodenhausen, and A. Wokaun, *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Clarendon Press, Oxford, 1987).
9. I. Noda, Appl. Spectrosc. **47,** 1329 (1993).
10. W. F. McClure, H. Maeda, J. Dong, Y. Liu, and Y. Ozaki, Appl. Spectrosc. **50,** 467 (1996).
11. N. L. Sefara, N. P. Magtoto, and H. H. Richardson, Appl. Spectrosc. **51,** 536 (1997).
12. Y. Wang, R. Tsenkova, M. Amari, F. Terada, T. Hayashi, A. Abe, and Y. Ozaki, Analusis, **26,** M64 (1998).
13. B. Czarnik-Matusewicz, K. Murayama, R. Tsenkova, and Y. Ozaki, Appl. Spectrosc. **53,** 1582 (1999).
14. Y. Ozaki, K. Murayama, and Y. Wang, Vib. Spectrosc. **20,** 127 (1999).
15. A. Matsushita, Y. Ren, K. Matsukawa, H. Inoue, Y. Minami, I. Noda, and Y. Ozaki, Vib. Spectrosc. **24,** 171 (2000).
16. M. A. Czarnecki, J. Phys. Chem. A **104,** 6356 (2000).
17. B. Czarnik-Matusewicz, K. Murayama, Y. Q. Wu, and Y. Ozaki, J. Phys. Chem. B **104,** 7803 (2000).
18. Y. Wu, K. Murayama, B. Czarnik-Matusewicz, and Y. Ozaki, Appl. Spectrosc. **56,** 1186 (2002).
19. Y. M. Jung, B. Czarnik-Matusewicz, and S. Bin Kim, J. Phys. Chem. B **108,** 13008 (2004).
20. A. Watanabe, S. Morita, and Y. Ozaki, Appl. Spectrosc. **60,** 1054 (2006).
21. J. Diewok, M. J. Ayora-Cañada, and B. Lendl, Anal. Chem. **74,** 4944 (2002).
22. P. Carmona, A. Rodríguez-Casado, and M. Molina, Anal. Bioanal. Chem. **393,** 1289 (2009).
23. M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, and V. Visani, Chemom. Intell. Lab. Syst. **57,** 65 (2001).
24. Y. Ozaki, http://sci-tech.ksc.kwansei.ac.jp/~ozaki/2D-Pocha.htm.
25. J. J. Workman, Jr., Appl. Spectrosc. Rev. **31,** 251 (1996).
26. American Society for Testing and Materials, Annual Book of ASTM Standards, v. 05.01, D 2700 – Standard test method for motor octane number of spark-ignition engine fuel (American Society for Testing and Materials, Philadelphia, 2000).
27. American Society for Testing and Materials, Annual Book of ASTM Standards, v. 05.01, D 2699 – Standard test method for research octane number of spark-ignition engine fuel (American Society for Testing and Materials, Philadelphia, 2000).