Development of Calibration Models for Estimation of Monomer Concentration by Raman Spectroscopy During Emulsion Polymerization: Facing the Medium Heterogeneity

Marlon M. Reis, Pedro H. H. Araújo, Claudia Sayer, Reinaldo Giudici

Universidade de São Paulo, Escola Politécnica, Department of Chemical Engineering, Caixa Postal 61548, CEP 05424–970, São Paulo, SP, Brazil

Received 21 May 2003; accepted 1 February 2004
DOI 10.1002/app.20474
Published online in Wiley InterScience (www.interscience.wiley.com).

ABSTRACT: This work compares different calibration models for the estimation of monomer concentrations by Raman spectroscopy during semicontinuous emulsion copolymerization reactions. The limitations of these models are discussed in terms of a complex reaction, namely the copolymerization of vinyl acetate and butyl acrylate, whose monomers present overlapping Raman spectra, especially the C=C stretching band. Additionally, the copolymerization was monitored in a spectroscopic setup arranged for fast spectral acquisition, which resulted in a low signal-to-noise ratio. These realistic conditions for in-line monitoring of emulsion copolymerization, i.e., considerable noise level in the spectra and medium heterogeneity, are discussed in the context of different approaches for adjusting the calibration model and the ensuing model limitations. It was verified that combining data obtained during reactions with synthetic samples is interesting from the statistical point of view, since in this way it is possible to produce data sets with a wide range of variation, allowing the accurate estimation of statistical parameters. These parameters are of major importance for process variables and product property estimations, especially if they are to be used for process control and decision making purposes. © 2004 Wiley Periodicals, Inc. J Appl Polym Sci 93: 1136–1150, 2004

Key words: Raman spectroscopy; process monitoring; emulsion polymerization; partial least squares; copolymerization; sensors

INTRODUCTION

Emulsion polymerization consists of the conversion, by free-radical polymerization, of monomer(s) in an aqueous dispersion into a stable dispersion of polymer particles. At the beginning of the reaction, a very small amount of the monomer can be found in solution; some is dissolved in the micelles, but most is present in the form of monomer droplets. As the reaction proceeds, polymer particles are nucleated and the reaction occurs mainly in the polymer phase. At the end of the reaction, the residual monomer is found in the polymer particles.[1] Emulsion polymerization is widely used in industry to produce latices for a variety of applications (such as latex paints, adhesives, coatings, binders in paper and textile products, and synthetic rubber). As has been pointed out in the literature, the control of emulsion polymerization reactors is a particularly difficult task, especially given the lack of robust on-line and in-line measurements for the important product properties and key process variables.[2] Among these properties, the on-line quantification of the residual monomer concentration during the polymerization process is still a challenge. In the last decade, Raman spectroscopy has emerged as an important tool for the estimation of monomer concentration during polymerization. In applications to emulsion polymerization, Raman spectroscopy has the additional advantage of the weak Raman scattering of the water and the much stronger bands due to the vinyl groups present in most of the monomers used in emulsion polymerization.[3,4] In emulsion polymerization, the variability of the process due to temperature changes, the heterogeneous medium, different particle sizes, etc., are additional important problems that must be overcome. In particular, in emulsion polymerization, medium heterogeneity (i.e., monomer droplets, micelles and polymer particles) changes significantly during the process.

Raman spectroscopy has been widely used on a laboratory scale for polymer analysis. Recently, the advent of fiber optics technology has enhanced the applicability of Raman spectroscopy to process monitoring, making it possible to take a monitoring beam, i.e., the laser, to the process and bring the system

Correspondence to: R. Giudici (rgiudici@usp.br).

response, i.e., the scattered light, back to the instrument.[5–12] This combination of technologies, Raman spectroscopy plus optical fibers, permits *in-situ* and in-line measurements and thus process monitoring in real time. In this case, several problems associated with sampling in an industrial reactor can be avoided. On the other hand, however, the "sample" cannot be "prepared" or conditioned for the analysis (e.g., filtered, cooled). In addition to the variability normally found in industrial processes, changes in medium heterogeneity occur during industrial emulsion polymerization processes. As all of these factors may affect the Raman measurements, process monitoring using this spectroscopic technique presents a series of challenges. Thus, in addition to hardware developments, process monitoring by spectroscopic techniques requires the elaboration of reliable calibration models for correlation of the sensor measurements, i.e., the *spectra*, with the desired properties, even in the presence of process variability.

The calibration model used to estimate monomer concentrations by Raman spectroscopy during emulsion polymerization must be based on spectra that represent the actual behavior of the system during the polymerization process. These spectra can be obtained in basically two ways: by collecting the spectra of samples taken during the reaction process and subsequently quantifying the monomer concentration by a reference method (e.g., gas chromatography), or, alternatively, by mimicking the behavior of the process with appropriate synthetic samples. The first approach, which uses samples from actual reactions, usually produces good results but may not be the best choice for incorporating process variability, since each new reaction may be slightly different and the model would not be able to predict these changes. The second approach is thus interesting since it is possible to produce synthetic samples that mimic different reaction behaviors and generate more robust calibration models without performing reactions. Although the second way may be more interesting, it may be difficult, sometimes impossible, to mimic all the stages of the reaction.

The development of an adequate calibration model is crucial for Raman reaction monitoring. Thus, the present work focuses on a description of a calibration model for emulsion polymerization monitoring based on Raman spectra collected from synthetic samples. The limitations of this model are discussed in terms of a complex reaction in which the two comonomers exhibit overlapping Raman spectra and spectral acquisition is limited to a small number of scans (with a concomitantly lower signal-to-noise ratio).

Emulsion polymerization

In an emulsion polymerization, an aqueous dispersion of monomer(s) is converted by free-radical polymerization into a stable dispersion of polymer particles. A typical emulsion recipe consists of the dispersing medium (e.g., water), monomer(s), a water-soluble initiator, and an emulsifier. Emulsion polymerization is a typical multiphase system in which four phases may be present: monomer droplets, micelles, polymer particles, and the aqueous phase.

The emulsion polymerization mechanism can be divided into three stages. In the first stage, most of the monomer is dispersed in monomer droplets or micelles and there are few polymer particles. In the second stage, micelles have been depleted and part of monomer is found in the swollen polymer particles and part is found in the monomer droplets. Finally, in the last stage, no more monomer droplets are present and the residual monomer swells the polymer particles.[1]

FT-Raman spectroscopy

Raman spectroscopy is a light scattering based technique. The light (radiation) scattering can be elastic, without changes in the energy of the scattered light, or inelastic with loss or gain of energy. The elastic scattering is known as Rayleigh scattering and the inelastic is known as Raman scattering. The classical theory of light scattering from molecules describes the molecule–radiation interaction by means of the oscillating dipole moment induced in the molecule by the presence of an incident radiation field. The Raman measurements discussed here are obtained by means of Fourier-transform Raman spectroscopy (FT-Raman).[3,4]

For isotropic samples, such as a liquid, the Raman intensity can be written in terms of the mean polarizability and the anisotropy. In general terms, the intensities of Raman bands can be expressed by an equation analogous to the Beer–Lambert law, as given by eq. (1).[3]

$$I_\nu = cI_0VK_\nu \tag{1}$$

where I_ν is the Raman intensity of band ν, I_0 is the intensity of the exciting radiation, V is the volume of sample illuminated by the source and viewed by the spectrometer, c is the sample concentration, and K_ν is a constant characteristic for each band.

Linear regression models

Raman scattering intensity versus monomer concentration

Equation (1) shows that the intensity of Raman scattering at frequency ν_1 is linear with the concentration of the active compound; for example, for sample i of compound A at concentration $c_{A,i}$, as shown by eq. (2):

$$I_{\nu_1,\mathrm{A},i} = c_{\mathrm{A},i}\phi_{\nu_1,\mathrm{A}} \quad (2)$$

where $\phi_{\nu_1,A} = I_0 V K_{\nu_1,A}$, corresponding to eq. (1).

The simplest way to perform quantification with Raman spectroscopy is to build a linear model (or a calibration curve) from a known data set. A linear model corresponding to eq. (2) is given by eq. (3).

$$I_{\nu_1,\mathrm{A},i}(\phi_{\nu_1,\mathrm{A}})^{-1} = c_{\mathrm{A},i} \quad (3)$$

In this way, the calibration data set (the data set with known concentrations) is used to calculate $(\phi_{\nu_1,A})^{-1}$ and, for unknown samples, the Raman intensity at the frequency ν_1 is measured and used to quantify the concentration of the compound A in the unknown sample according to eq. (3).

The Raman scattering intensity for a sample with more than one active compound contributing to the signal at frequency ν_1, for example, compounds A, B, and C, is given by eq. (4) or, in matrix form, by eq. (5).

$$I_{\nu_1,i} = c_{\mathrm{A},i}\phi_{\nu_1,\mathrm{A}} + c_{\mathrm{B},i}\phi_{\nu_1,\mathrm{B}} + c_{\mathrm{C},i}\phi_{\nu_1,\mathrm{C}} \quad (4)$$

$$I_{\nu_1,i} = \begin{pmatrix} c_{\mathrm{A},i} & c_{\mathrm{B},i} & c_{\mathrm{C},i} \end{pmatrix} \begin{pmatrix} \phi_{\nu_1,\mathrm{A}} \\ \phi_{\nu_1,\mathrm{B}} \\ \phi_{\nu_1,\mathrm{C}} \end{pmatrix} \quad (5)$$

While for the single compound case, the linear model is obtained by taking the inverse of $(\phi_{\nu_1,A})$, in the three-compound case, it is necessary to use intensities at more than one frequency, as shown in eq. (6).

$$\begin{pmatrix} I_{\nu_1,i} & I_{\nu_2,i} & I_{\nu_3,i} \end{pmatrix} = \begin{pmatrix} c_{\mathrm{A},i} & c_{\mathrm{B},i} & c_{\mathrm{C},i} \end{pmatrix} \times \begin{pmatrix} \phi_{\nu_1,\mathrm{A}} & \phi_{\nu_2,\mathrm{A}} & \phi_{\nu_3,\mathrm{A}} \\ \phi_{\nu_1,\mathrm{B}} & \phi_{\nu_2,\mathrm{B}} & \phi_{\nu_3,\mathrm{B}} \\ \phi_{\nu_1,\mathrm{C}} & \phi_{\nu_2,\mathrm{C}} & \phi_{\nu_3,\mathrm{C}} \end{pmatrix} \quad (6)$$

The matrix on the right hand side of eq. (6) can be inverted, which results in:

$$\begin{pmatrix} I_{\nu_1,i} & I_{\nu_2,i} & I_{\nu_3,i} \end{pmatrix} \begin{pmatrix} \phi_{\nu_1,\mathrm{A}} & \phi_{\nu_2,\mathrm{A}} & \phi_{\nu_3,\mathrm{A}} \\ \phi_{\nu_1,\mathrm{B}} & \phi_{\nu_2,\mathrm{B}} & \phi_{\nu_3,\mathrm{B}} \\ \phi_{\nu_1,\mathrm{C}} & \phi_{\nu_2,\mathrm{C}} & \phi_{\nu_3,\mathrm{C}} \end{pmatrix}^{-1} = \begin{pmatrix} c_{\mathrm{A},i} & c_{\mathrm{B},i} & c_{\mathrm{C},i} \end{pmatrix} \quad (7)$$

Linear models

In general, eq. (7) can be written in terms of a linear model, as given in eq. (8).

$$\mathbf{Y} = \mathbf{XB} \quad (8)$$

where:

$$\mathbf{Y} = \begin{pmatrix} c_{A,1} & c_{B,1} & c_{C,1} \\ c_{A,2} & c_{B,2} & c_{C,2} \\ \vdots & \vdots & \vdots \\ c_{A,n} & c_{B,n} & c_{C,n} \end{pmatrix} \quad (9)$$

$$\mathbf{X} = \begin{pmatrix} I_{\nu_1,1} & I_{\nu_2,1} & I_{\nu_3,1} \\ I_{\nu_1,2} & I_{\nu_2,2} & I_{\nu_3,2} \\ \vdots & \vdots & \vdots \\ I_{\nu_1,n} & I_{\nu_3,n} & I_{\nu_3,n} \end{pmatrix} \quad (10)$$

$$\mathbf{B} = \begin{pmatrix} \phi_{\nu_1,\mathrm{A}} & \phi_{\nu_2,\mathrm{A}} & \phi_{\nu_3,\mathrm{A}} \\ \phi_{\nu_1,\mathrm{B}} & \phi_{\nu_2,\mathrm{B}} & \phi_{\nu_3,\mathrm{B}} \\ \phi_{\nu_1,\mathrm{C}} & \phi_{\nu_2,\mathrm{C}} & \phi_{\nu_3,\mathrm{C}} \end{pmatrix}^{-1} \quad (11)$$

For the calibration data set (a data set with known concentrations) the model described by eq. (8) has, as the unknown, the matrix **B**, whose estimation is discussed in the following.

The simplest way of estimating **B** is via eq. (12):

$$\mathbf{B} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \quad (12)$$

Equation (12) is the least-squares solution of eq. (8), i.e., the solution for a minimization problem where the function $f(\mathbf{B}) = \| \mathbf{Y} - \mathbf{XB} \|^2$ must be minimized ($\| \mathbf{x} \|^2$ the is sum of the squares of the **x** elements).

Equation (12) is valid if **X** is a full rank matrix,[13] otherwise the inverse $(\mathbf{X}^T\mathbf{X})^{-1}$ does not exist. This condition is not always attained and an alternative way of calculating **B** must be used.

Projection based linear models

When the inverse $(\mathbf{X}^T\mathbf{X})^{-1}$ does not exist, i.e., when the matrix **X** is not full rank, an alternative for estimating **B** is the projection of **X** in an orthogonal basis set, for example, as given in eq. (13).

$$\mathbf{T} = \mathbf{XP} \quad (13)$$

where

$$\mathbf{P}^T\mathbf{P} = \mathbf{I}, \quad (14)$$

and the number of columns of **T** and **P** is set equal to the rank of matrix **X, T** is called the matrix of *scores* and **P** are the *loadings*,[13] and the decomposition $\mathbf{X} = \mathbf{TP}^T$ corresponds to the so-called principal component analysis (PCA).

Once **X** has been projected, **B** is estimated as shown by eq. (15).

$$\mathbf{B} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y} \quad (15)$$

The estimate of the concentration of a new sample (new spectrum) is obtained as given in eq. (16).

$$\hat{\mathbf{Y}}_{new} = \mathbf{X}_{new}\mathbf{PB} \tag{16}$$

where $\mathbf{X}_{new}$ and $\hat{\mathbf{Y}}_{new}$ are the new spectra and the corresponding estimated concentration matrix.

This method is called principal component regression (PCR).[13,14] Besides PCR being an elegant algebraic approach, in some cases the estimation of **T** by using information from **Y** produces a better approximation. One method widely used to estimate **T** employing information from **Y** is partial least squares (PLS),[14] described here in terms one of its variations, the SIMPLS, developed by de Jong.[15] The SIMPLS estimation of **B** is given by eq. (17).

$$\mathbf{B}_{SIMPLS} = \mathbf{RQ}^T \tag{17}$$

where

$$\mathbf{Q} = (\mathbf{Y} - 1_n\bar{\mathbf{y}}^T)^T(\mathbf{X} - 1_n\bar{\mathbf{x}}^T)\mathbf{R} \tag{18}$$

$\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ are column vectors with average values of the columns **X** and **Y**, respectively, and $\mathbf{1}_n$ denotes a column vector ($n \times 1$).

The estimation of **R** is not discussed here, but it is important to emphasize that it depends on **X** and **Y**. For a new sample (new spectrum) the estimation of the concentration is given by eq. (19).

$$\mathbf{y}_{new}^T = \bar{\mathbf{y}}^T + (\mathbf{x}_{new}^T - \bar{\mathbf{x}}^T)\mathbf{B}_{SIMPLS} \tag{19}$$

In more general terms, eq. (6) can be written as in eq. (20), which represents a linear model for spectra collected at n wavelengths.

$$(I_{\nu_1,i} \quad I_{\nu_2,i} \quad I_{\nu_3,i} \quad \cdots \quad I_{\nu_n,i}) = (c_{A,i} \quad c_{B,i} \quad c_{C,i}) \times \begin{pmatrix} \phi_{\nu_1,A} & \phi_{\nu_2,A} & \phi_{\nu_3,A} & \cdots & \phi_{\nu_n,A} \\ \phi_{\nu_1,B} & \phi_{\nu_2,B} & \phi_{\nu_3,B} & \cdots & \phi_{\nu_n,B} \\ \phi_{\nu_1,C} & \phi_{\nu_2,C} & \phi_{\nu_3,C} & \cdots & \phi_{\nu_n,C} \end{pmatrix} \tag{20}$$

If the spectra at unit concentration are known for all compounds, e.g., for A, B, and C, the concentration of the compounds in an unknown sample can be obtained directly as shown in eqs. (21–25).

$$\mathbf{X}_{Gen} = \mathbf{YS} \tag{21}$$

$$\mathrm{X}_{Gen,i} = (I_{\nu_1,i} \quad I_{\nu_2,i} \quad I_{\nu_3,i} \quad \cdots \quad I_{\nu_n,i}), \tag{22}$$

$$\mathbf{S} = \begin{pmatrix} \phi_{\nu_1,A} & \phi_{\nu_2,A} & \phi_{\nu_3,A} & \cdots & \phi_{\nu_n,A} \\ \phi_{\nu_1,B} & \phi_{\nu_2,B} & \phi_{\nu_3,B} & \cdots & \phi_{\nu_n,B} \\ \phi_{\nu_1,C} & \phi_{\nu_2,C} & \phi_{\nu_3,C} & \cdots & \phi_{\nu_n,C} \end{pmatrix}$$

$$\mathbf{X}_{Gen}\mathbf{S}^T = \mathbf{YSS}^T \tag{23}$$

$$\mathbf{X}_{Gen}\mathbf{S}^T(\mathbf{SS}^T)^{-1} = \mathbf{Y}(\mathbf{SS}^T)(\mathbf{SS}^T)^{-1} \tag{24}$$

$$\mathbf{X}_{Gen}\mathbf{S}^T(\mathbf{SS}^T)^{-1} = \mathbf{Y} \tag{25}$$

where $\boldsymbol{X}_{Gen,i}$ corresponds to row i of matrix $\mathbf{X}_{\mathrm{Gen}}$.

Equation (25) represents an alternative way of performing quantification, which is referred to by some authors as classical least squares.[6]

Raman spectra of the monomers

The monomers used in the copolymerization monitored by the models discussed in this work are vinyl acetate and butyl acrylate. Figure 1 shows the Raman spectra of the pure monomers.

FT-Raman is an attractive method for monitoring the polymerization of these monomers because the bands due to C=C stretching, which disappear during the polymerization reaction, are a strong scattering group.[16,17] Nevertheless, since the spectral bands for vinyl acetate and butyl acrylate almost overlap, it is necessary to use multivariate models for monomer quantification.

EXPERIMENTAL

The spectra used to develop a model for the polymerization process monitoring must mimic the three stages or intervals described above. The acquisition time must be small to make possible realistic in-line monitoring and to minimize the variation in Raman intensity as the compounds' concentrations change during the reaction.

Spectra

Raman spectra were collected with a FRA 106/S FT-Raman accessory attached to an IFS 28/N spectrometer from Bruker equipped with a quartz beamsplitter; the laser wavelength was 1,064 nm; the acquisition mode was set to double sided (forward-backward) and the correlation mode was set to *Full I Gram Length*, where all points of the new interferogram are compared. In case of deviations, the area that diverges is replaced by the corresponding part of the average spectrum. Scans are discarded only if they contain more than 10 defective areas or if the number of defective points exceeds one-eighth of the total number of interferogram points. A phase resolution equal to 32 was used, a phase correction mode equal to

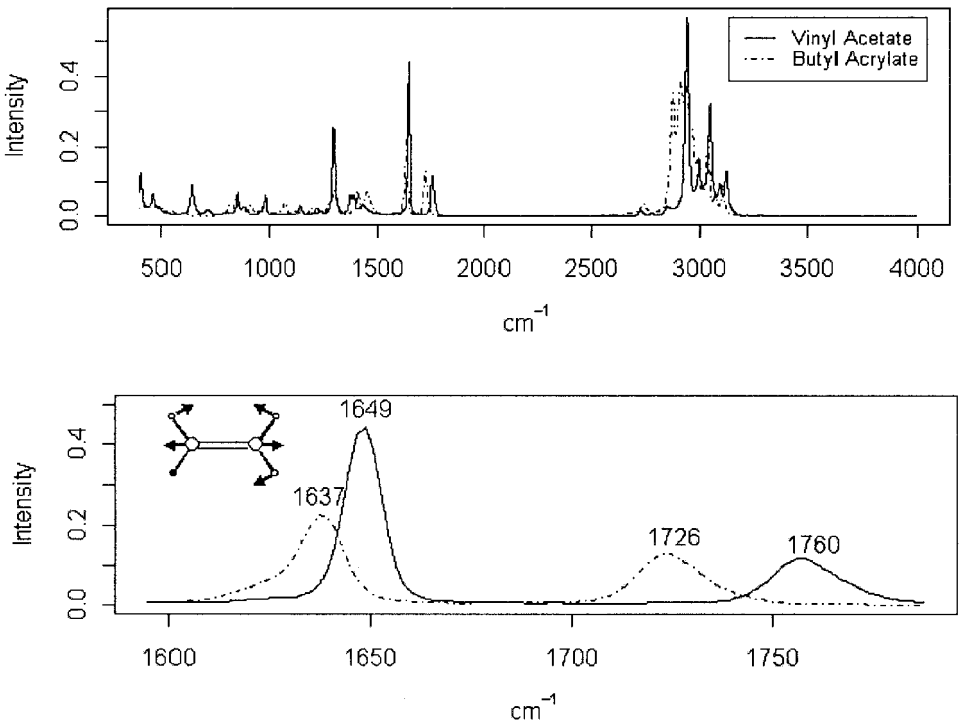

Figure 1 Raman spectra of vinyl acetate and butyl acrylate. The bands due to C=C stretching (1,637 and 1,649 cm^{-1}) and C=O stretching (1,726 and 1,760 cm^{-1}) are shown at the bottom.

power spectrum, the apodization function was the Norton-Beer, and the error filling factor was set equal to 4. Each spectrum is an average of 8 scans with a resolution of 8 cm^{-1} and laser power equal to 450 mW. Samples were analyzed in 8-mm-diameter glass tubes.

All calculations were performed in the R-language (http://www.r-project.org) with our own SIMPLS routine.

Synthetic samples

In this work, several synthetic samples were obtained by dispersing a known weight of monomer into a monomer-free polymer emulsion, previously polymerized from the same monomer. These samples were then used to build the calibration model.

Development of the calibration model

The calibration models were built from spectra collected from the synthetic samples and the PLS was applied as the calibration model. Normalized spectra were evaluated in building the models. The normalization treatment corresponds to dividing each element of the spectrum by the element corresponding to wavenumber 427 cm^{-1}.

The set of synthetic samples was prepared by adding monomers and water to a polymer emulsion (latex). Three kinds of latices were used and designed as: L1(a), L2(d), and L3(b), with solids contents (vinyl acetate/butyl acrylate copolymer) of 46.78, 11.29, and 20.66%, respectively. The bottles containing the synthetic samples were closed tightly to avoid any monomer losses and agitated periodically during 2 hours before spectra acquisition to allow the monomer to swell polymer particles. The samples are described in Table I.

Reaction samples were analyzed as soon as collected from the reactor by Raman spectroscopy and by the reference method (gas chromatography).

RESULTS AND DISCUSSION

The first step of monitoring a polymerization reaction is the development of a calibration model. As mentioned before, PLS models were built using spectra collected from the samples obtained by dispersing a known concentration of monomers, i.e., vinyl acetate and butyl acrylate, in polymer emulsions. In this case, three different latices, with different polymer particle sizes and solids contents, were used for the sample preparation. Two types of PLS model were developed, one to estimate the total monomer content and the other specific for each of the two monomers. The results for the model building with normalized spectra collected from the samples described in Table I are shown in Figure 2, for six-latent-variable PLS models.

TABLE I
Synthetic Samples

	VA(g)	BA(g)	Latex(g)	Water(g)	
1#	0.0108	0.1002	8.0116	2.0018	L1 (a) Dp: 148.7 nm VA/BA: 80/20
2#	0.0626	1.1012	8.0180	2.0010	
3#	0.3009	0.0395	8.0104	2.0033	
4	1.2001	0.0612	8.0008	2.0154	
5#	0.0106	0.0912	8.0186	2.0055	
6	0.0204	0.2023	8.0033	2.0309	L2(d) Dp: 144.4 nm VA/BA: 85/15
7#[a]	0.0747	1.7106	8.0040	2.0113	
8#	0.4020	0.0825	8.0098	2.0491	
9[a]	1.6012	0.0511	8.0115	2.0022	
10#	0.0925	0.0138	8.0383	2.0373	
11	0.0142	0.3049	8.0218	2.0290	L3(b) Dp: 217.1 nm VA/BA: 80/20
12	0.698	1.5007	8.0342	2.0096	
13#	0.1027	0.0730	8.0392	2.0981	
14	2.0064	0.0535	7.9996	2.0561	
15	0.0527	0.0532	8.0172	2.0008	
16#	0.1015	0.3061	8.0008	2.0000	L1(a)
17#	0.8022	0.5019	8.0068	2.0119	
18#	1.0025	1.0045	8.0140	2.0175	
19	2.2086	2.8042	8.0002	2.0059	
20#	0.4007	0.3059	8.0006	2.0080	L2(d)
21#	0.51	0.5306	8.0371	2.0163	
22[a]	1.8018	1.3422	8.0217	2.0036	
23[a]	2.671	2.2778	8.0072	2.0849	
24#	0.2046	0.2026	8.0031	2.0318	L3(b)
25	0.1052	0.7046	8.0103	2.0794	
26[a]	1.3125	1.3307	8.0019	2.0049	
27[a]	2.7005	2.1033	8.0082	2.0030	
28	0	0	8.0036	2.0082	L1 (a)
29	0	0	8.0753	2.0490	L2(d)
30	0	0	8.0356	2.0249	L3(b)
31#	0.0396	0.5036	4.0049	6.0156	L1(a)
32#[a]	0.0818	2.4008	4.0152	6.0190	
33#	0.8018	0.0707	4.0070	6.0024	
34	2.2025	0.0905	4.0007	6.0186	
35[a]	0.0498	0.8082	4.0207	6.0398	L2(d)
36#[a]	0.09	2.2187	4.0624	6.0412	
37[a]	1.0204	0.0427	4.0843	6.0771	
38#[a]	2.6017	0.0114	4.3547	6.0407	
39	0.04	0.6060	4.0071	6.0263	L3(b)
40#[a]	0.0946	2.8152	4.0094	6.0658	
41	0.416	0.0231	4.0646	6.0972	
42#[a]	2.4062	0.0954	4.0687	6.0344	

[a] Samples that might have monomers droplets; Dp, Average polymer particle diameter; VA/BA, polymer composition, VA vinyl acetate and BA butyl acrylate;# sample used for the hybrid model (see Results).

The results shown in Figure 2 do not exhibit a good fit, one possible explanation being the medium heterogeneity, as illustrated in Figure 3.

Figure 3 suggests that the amount of monomer irradiated by the laser depends on the number and size of the polymer particles. Thus, the spectra is not directly proportional to the amount of monomer added to the polymer emulsion, but rather depends on the dispersion of monomers in the polymer particles and the number of polymer particles, as well as the presence of monomer droplets inside the volume irradiated by the laser beam. The data described in Table I are thus difficult to calibrate since three different latices were used.

To quantify the amount of monomer in the emulsion correctly, the spectra of pure monomers, latices, and water were collected. Equation (25) was applied using the spectra of monomers, water, and latices, i.e., these spectra were used as rows of the matrix **S** in eq. (25). The concentration obtained with this equation, i.e., **Y**, was used in the PLS model fitting. The results for the six-latent-variable PLS model is shown in Figure 4.

The results in Figure 4 indicate a good fit, suggesting that this model can be used for the quantification of the monomers throughout a polymerization reaction. The copolymerization is described in the Appendix and named NR1.

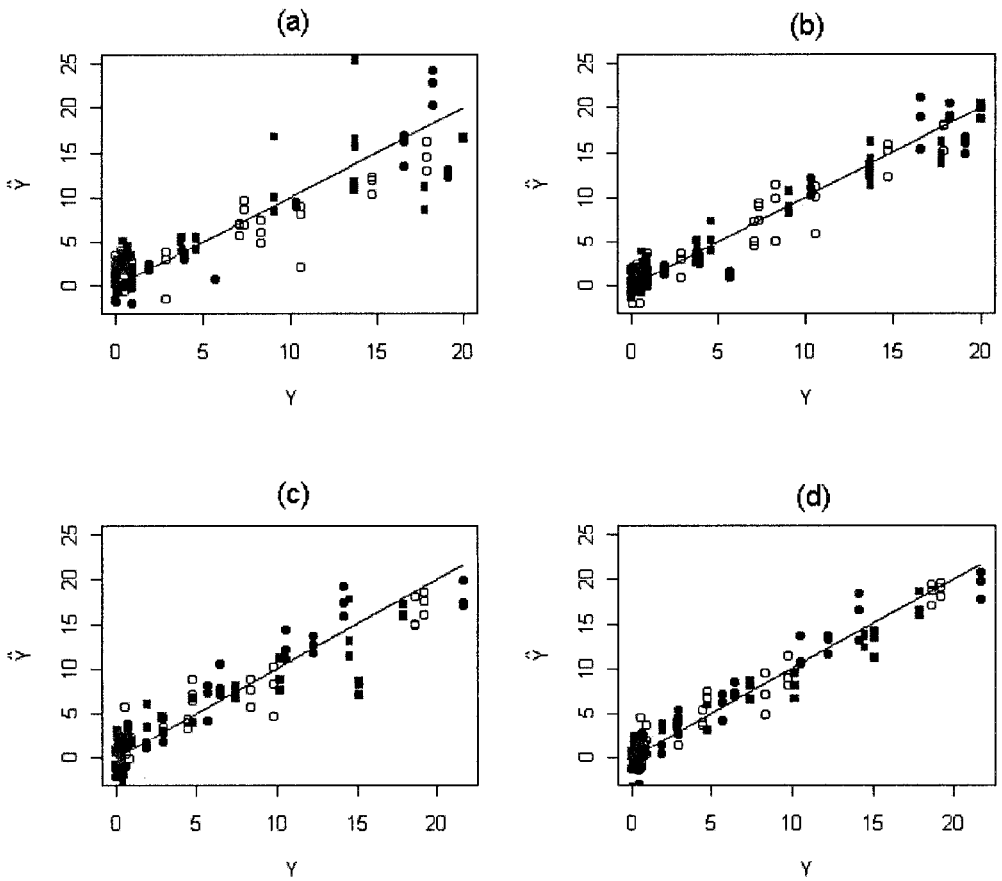

Figure 2 Predicted ($\hat{\mathbf{Y}}$) versus actual ($\mathbf{Y}$) concentration of monomers, in (a) and (b) for vinyl acetate using the model fit for both monomers and the model fit only for prediction of vinyl acetate, respectively; (c) and (d) are equivalent to (a) and (b), but for predictions of butyl acrylate. Squares, open circles, and filled circles correspond to the type of the latex used in the sample preparation, i.e., L1(a), L2(d), and L3(b), respectively.

The data obtained during the reaction were used to test four models, two models built using the original concentration (i.e., the known amount of monomers added to the polymer emulsions), the model built using the concentration found by eq. (25), and, finally, a model based on spectra obtained during the reaction for samples with monomer concentrations quantified by gas chromatography. The results for the four models for the quantification of vinyl acetate and butyl acrylate during the reaction are illustrated in Figure 5.

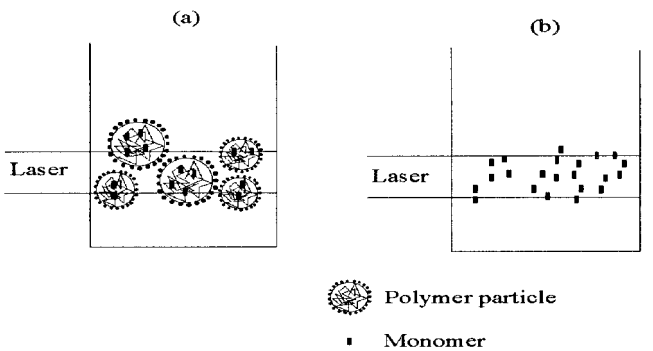

Figure 3 Illustration of the volume illuminated by the laser in (a) a heterogeneous medium and (b) a homogeneous medium.

Figure 5 shows that the model developed with the concentrations found by eq. (25) does perform better than the models built with the monomer concentrations added to the emulsion. The results shown in Figure 5 suggest that the models developed using the original monomer concentrations overestimate the predictions. This overestimation might be expected, since the model is built to quantify a monomer concentration that was not really measured. This overestimation can also be noted in Figure 6, where the concentration found by eq. (25) is plotted versus the original concentration added to the emulsion, which is in general larger than the concentration found by eq. (25). As shown in Figure 6, there are some samples that are quite different from the others. These samples might have monomer droplets in the emulsion, as indicated in Table I.

Figure 7 illustrates the overestimation by two models, i.e., univariate linear models, fitted for the same set of independent variables, x, but for two different

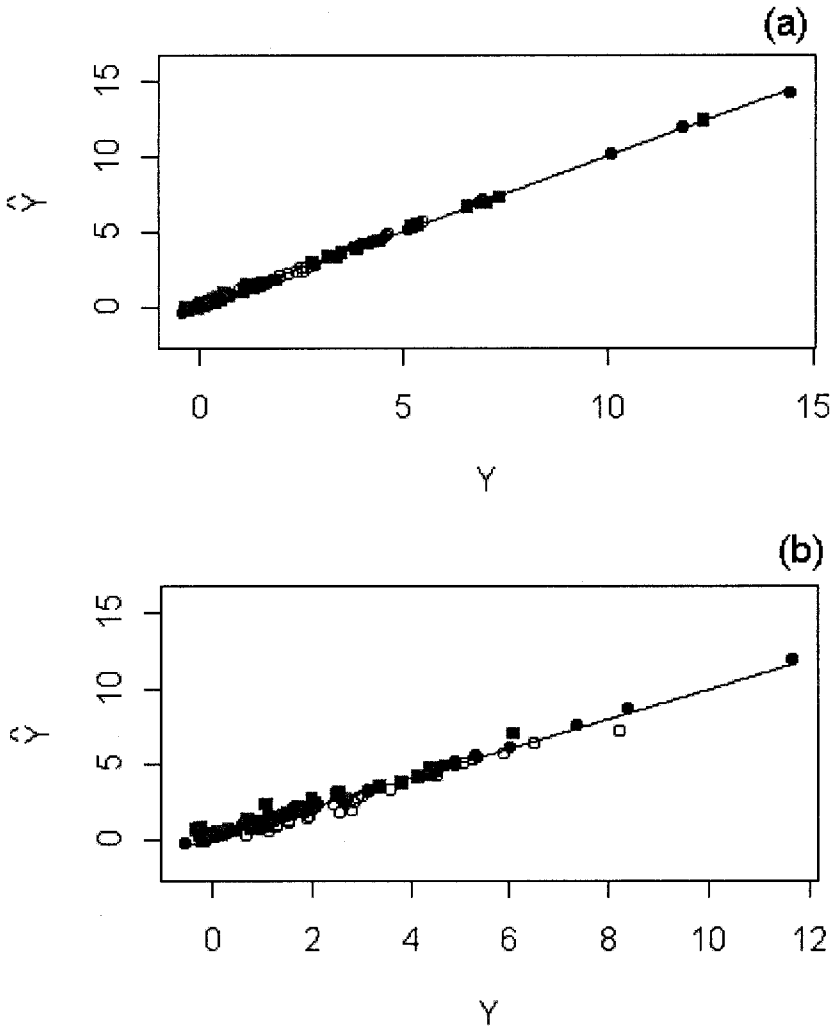

Figure 4 Predicted ($\hat{\mathbf{Y}}$) versus actual ($\mathbf{Y}$) concentration of monomers for (a) vinyl acetate using the model fit with monomer concentrations given by eq. (25); (b) is equivalent to (a), but for predictions of butyl acrylate. Squares, open circles, and filled circles correspond to the type of latex used in the sample preparation, i.e., L1(a), L2(d), and L3(b), respectively.

sets of dependent variables, y, for Range 1 and Range 2. In this case, if a model is fitted with Range 1 for the dependent variable and then used for predictions of samples that lie at Line 2, it results in overestimated predictions. In other words, if the monomer concentration inside the volume irradiated by the laser is not the same as that calculated for the amount of monomer added to the polymer emulsion, there are two concentration ranges: one that was measured by the laser and other calculated from the amount of monomer added to the polymer emulsion. If the last concentration range is used in the model fitting, then overestimated predictions result.

The model based on monomer concentrations found by eq. (25) produced good predictions for vinyl acetate, as shown in Figure 5. On the other hand, the predictions for butyl acrylate are not good. To understand why the same model is able to predict the data for vinyl acetate, but not for the butyl acrylate concentration, a hybrid model was built. For this model, spectra collected from some of the samples described in Table I (denoted with #) and from samples of the reaction were employed. The results for this hybrid model (a 10-latent-variable PLS model) are very good for the vinyl acetate quantification and a better approximation for butyl acrylate than the other models, as shown in Figure 8.

The hybrid model predictions for the butyl acrylate concentration are closer to the reference method, but still exhibit large variation in the predictions. An explanation for this may be the low concentration of butyl acrylate compared to the level of noise present in the spectra. Although the noise level is an important factor in model efficiency, the results for the model fit with only the reaction data, which have the same noise level as the other data, are quite good. Thus, PCA[13] was applied to two matrices: one for all the spectra of the samples described in Table I plus the spectra from

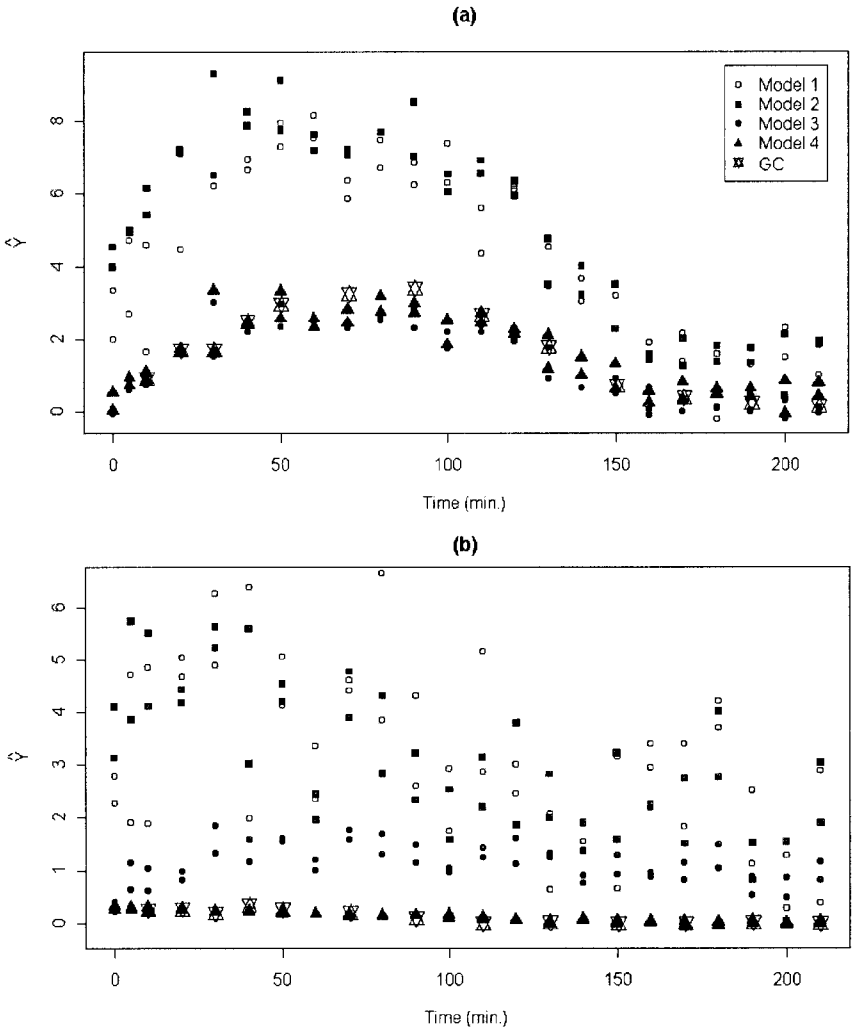

Figure 5 (a) Predicted ($\hat{\mathbf{Y}}$) vinyl acetate concentration during polymerization NR1, duplicate analysis of each sample. Model 1 corresponds to the model fit for both monomers with the original monomer concentration; Model 2 is similar to Model 1 but fit only for prediction of the corresponding monomer; Model 3 corresponds to the model for both monomers fit with the concentration given by eq. (25); Model 4 is the model fit only with data from the reaction. (b) Predicted ($\hat{\mathbf{Y}}$) butyl acrylate during polymerization NR1, duplicate analysis of each sample. GC corresponds to the monomer concentration found by gas chromatography.

the reaction and another with the spectra used in the hybrid model (all the spectra from the reaction were used). It was verified that components 5 and 6 (with minor importance in the description of the data) are those responsible for the difference between the spectra of samples from the reaction and the synthetic samples.

The PCA of the data used for the hybrid model shows that the principal components 2 and 4 were among the most important for the separation of the samples from the beginning of the reaction, also described in Figure 9.

The loadings for the principal components 2 and 4, Figure 10, identify the more important spectral regions for the separation shown in Figure 9. These spectral regions are indicated in the polymer spectra shown in Figure 11.

Figure 10 suggests that the positive values of the second principal component for the samples corresponding to 0–30 min of reaction described in Figure 9 are due to the increase of intensity in the spectral region denoted with squares, filled circles, and stars and a decrease in intensity in the region denoted with open circles. The spectral region denoted with squares presents strong scattering from vinyl acetate and the polymer. The spectral region for the full circles has strong Raman scattering from vinyl acetate and a medium Raman scattering signal from butyl acrylate. Finally, the spectral region indicated by stars corresponds to C=C stretching. The increase in the scattering intensity in these three regions and the decrease in the intensity in the region related to the polymer (open circles) suggest that the second principal component is sensitive to the ratio of monomers/polymer, in other

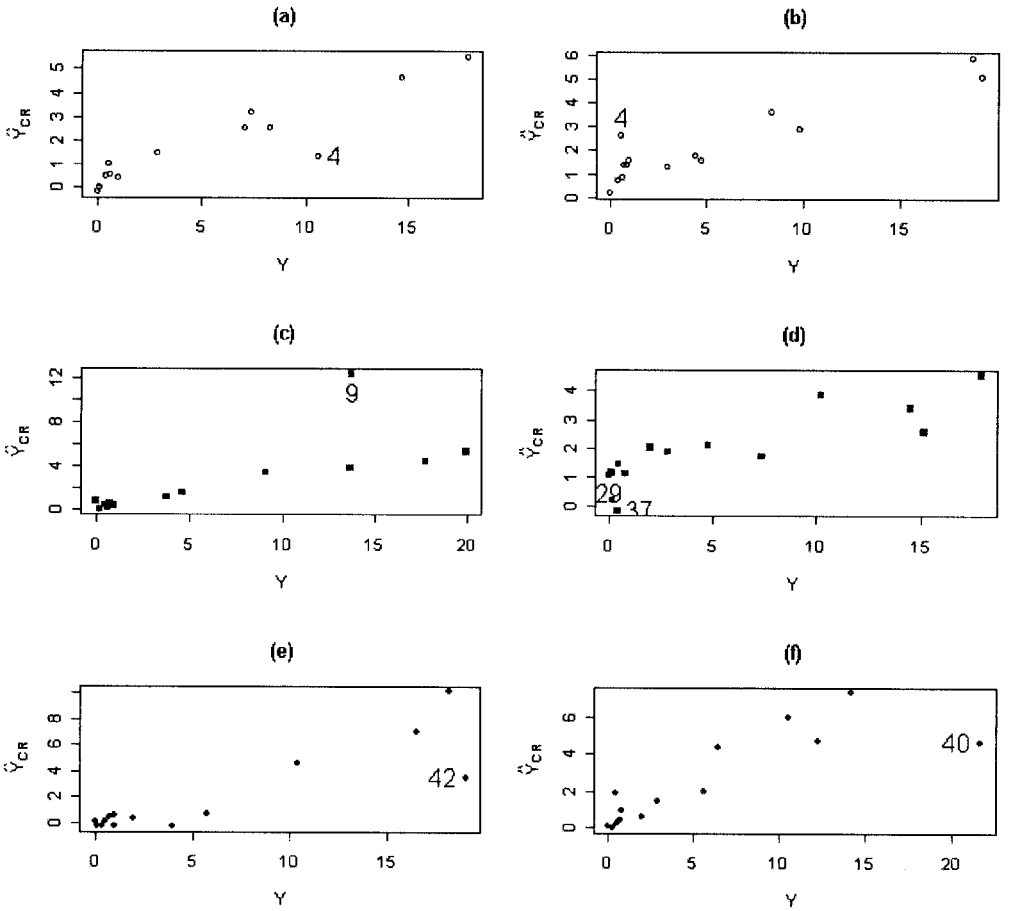

Figure 6 Predicted ($\hat{\mathbf{Y}}_{CR}$) monomer concentrations by eq. (25) versus original monomer concentration; in (a), (c), and (e) for vinyl acetate in synthetic samples prepared with the latices L1(a), L2(d), and L3(b), respectively; in (b), (d), and (f) for butyl acrylate for samples prepared with the latices L1(a), L2(d), and L3(b), respectively. The numbers of some of the samples correspond to those in Table I.

words, it is favored by a high concentration of monomers and a low concentration of polymer. The fourth-principal component is dominated by the spectral regions denoted with filled circles and squares. Since the spectral region corresponding to the filled circles presents Raman scattering for both monomers and that corresponding to squares only for vinyl acetate, the samples of the reaction from 0 to 30 min would appear to have more vinyl acetate than butyl acrylate. In short, this PCA analysis indicates that the samples from 0 to 30 min of reaction represent a situation in which there is a higher concentration of monomers, especially vinyl acetate, than polymer. This situation is more difficult to be simulated by synthetic samples since, in general, there is a phase separation between monomers and the polymer particles at high monomer concentrations. This conclusion is also confirmed by the GC data, but PCA is important because it shows that there are no synthetic samples that mimic this reaction stage. PCA provides information about the spectral space and aids in verifying whether the differences between spectra collected from synthetic samples and during the reactions are due to changes in the spectra of the compounds or due to variations in the concentration of these compounds.

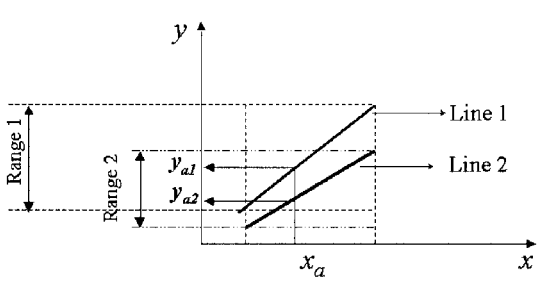

Figure 7 Illustration of the overestimation due to a model fit with a range of concentration wider than the real one.

One question that still remains to be answered is: Why are the predictions in polymerization monitoring with the model based on synthetic samples worse for butyl acrylate? As discussed, there is a stage during the reaction that is not possible to mimic by synthetic samples, which means that the predictions by the

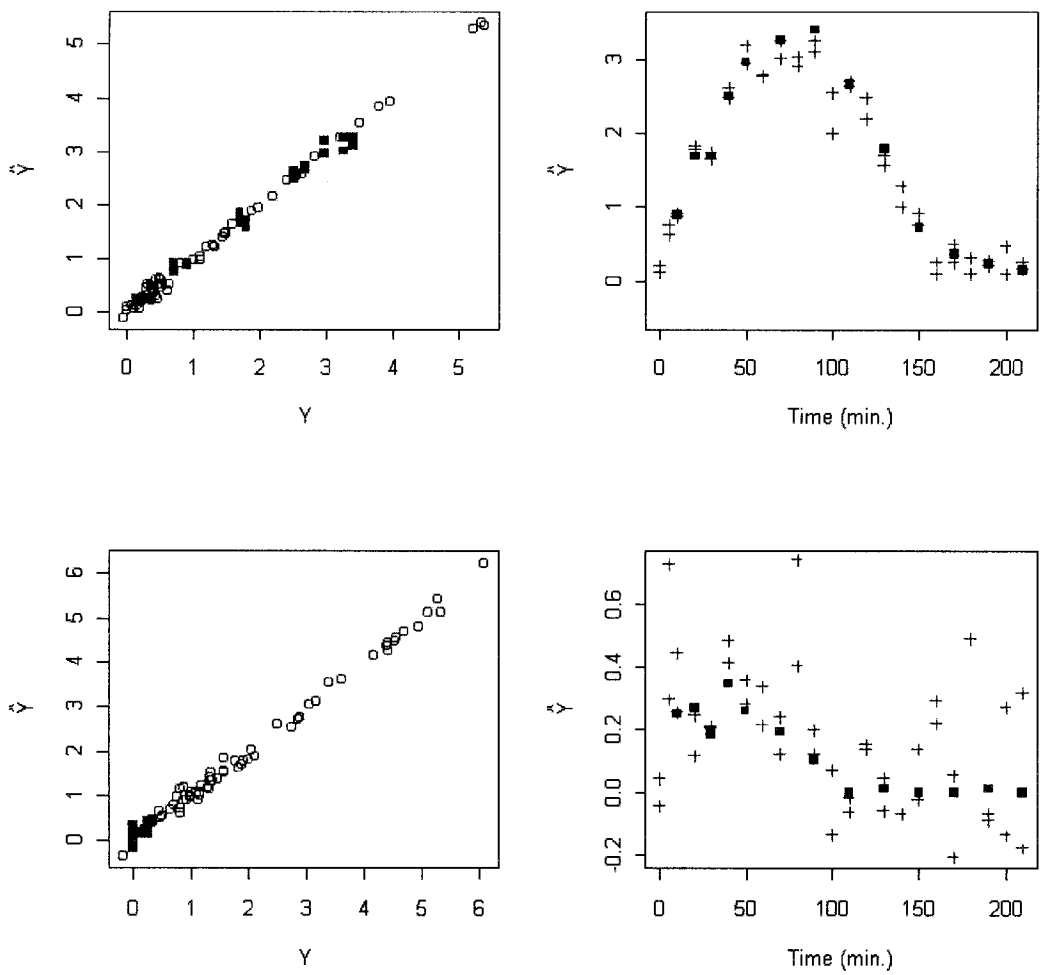

Figure 8 Predicted ($\hat{\mathbf{Y}}$) versus actual monomer concentration in (a) and (c) for the hybrid model fit. In (b) and (d), predicted ($\hat{\mathbf{Y}}$) monomer concentration during the polymerization NR1 for vinyl acetate and butyl acrylate, respectively, with duplicate analysis of each sample. Squares correspond to the GC quantification and crosses to the PLS model predictions.

model based on these samples are extrapolations. Figure 12 illustrates two extrapolation cases, where the open circles correspond to predictions by the model fit with filled circles. The dashed line represents what would be the real model ("ideal model"), while the solid line represents the model based on the data measured in a limited range (calibration region) used for the predictions outside the calibration range. Case (a) is related to the predictions for butyl acrylate and case (b) is related to those for vinyl acetate. Because butyl acrylate is more reactive than vinyl acetate, its concentration is smaller during the reaction and the situation illustrated in Figure 12 could prevail, consequently resulting in poorer predictions for butyl acrylate.

The mathematical treatment of the problem illustrated in Figure 12 is not easily done within the framework of a PLS model, both because it is a multivariate model and because the information of "Y" is present in the model fitting. On the other hand, this does give some insight into the origin of bad predictions by the model and can be used as a guide for the development of new models.

Based on the results discussed up to now, a new polymerization reaction, named NR2, was performed, the description of the which is given in the Appendix. The latex resulting from this new reaction, with 25% solids content, was used to create new synthetic samples, which are described in Table II. A new model was then fit using the data from the reaction, with monomer concentrations quantified by GC and corresponding spectra collected from synthetic samples. Each spectrum is an average of 32 scans with resolution of 8 cm^{-1} and laser power equal to 450 mW, the number of scans being increased to increase the signal-to-noise ratio; 8-mm-diameter glass tubes were used for the synthetic samples. Two new models were developed: one specific for vinyl acetate and the other for butyl acrylate. The results for these two models are presented in Figure 13, where it can be seen that the models present good predictions. The model for vinyl

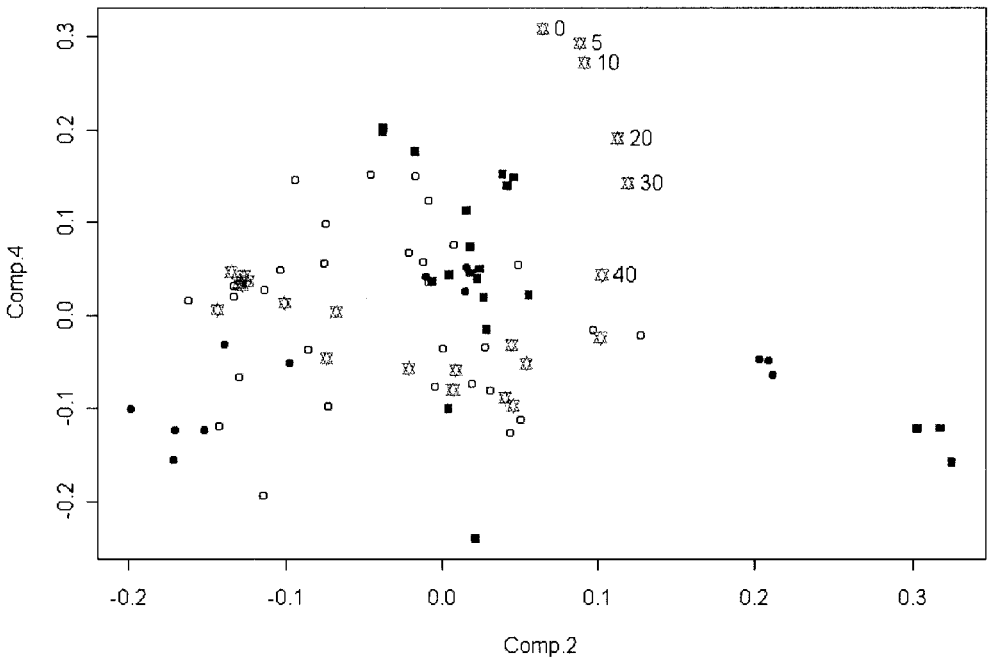

Figure 9 Normalized scores for principal component analysis applied to the spectra of some synthetic samples (denoted by # in Table I) and samples collected during the polymerization NR1. The numbers denote the time in minutes for the reaction. Stars correspond to samples from the reaction and the others to synthetic samples.

acetate used six latent variables and that for butyl acrylate used eight latent variables.

The two new models were tested against the data from polymerization NR1 (the average spectrum for each duplicate was used) and the results are shown in Figure 14. These predictions are quite similar to those obtained by the hybrid model, which used spectra from synthetic samples and from the reaction NR1, showing that mixing data collected from the reaction and from synthetic samples in the fit of

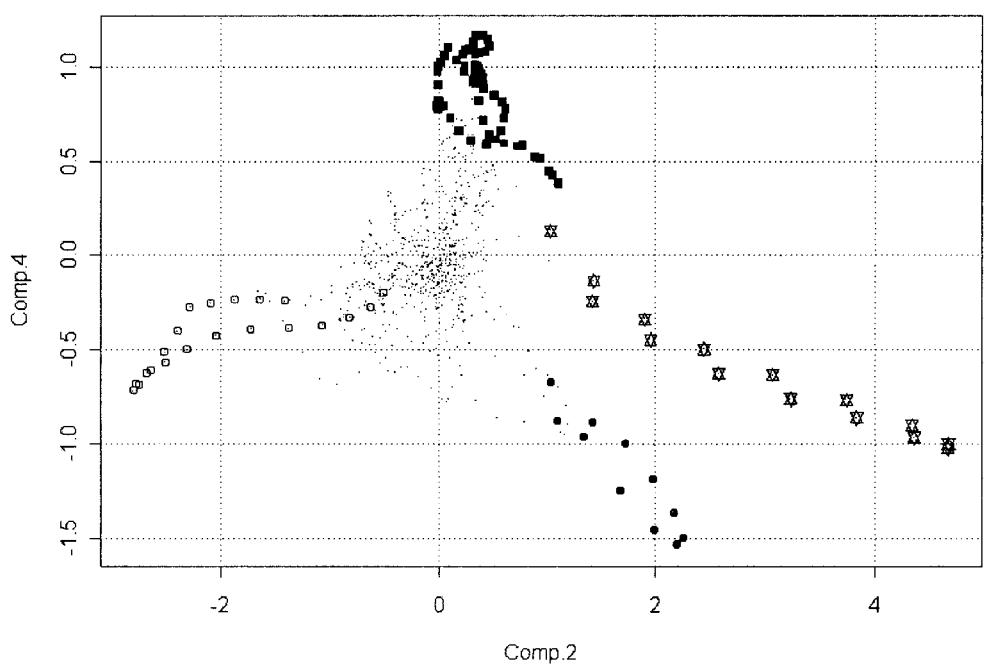

Figure 10 Normalized loadings for principal component analysis applied to the spectra of some synthetic samples (denoted by # in Table I) and samples collected during the polymerization NR1.

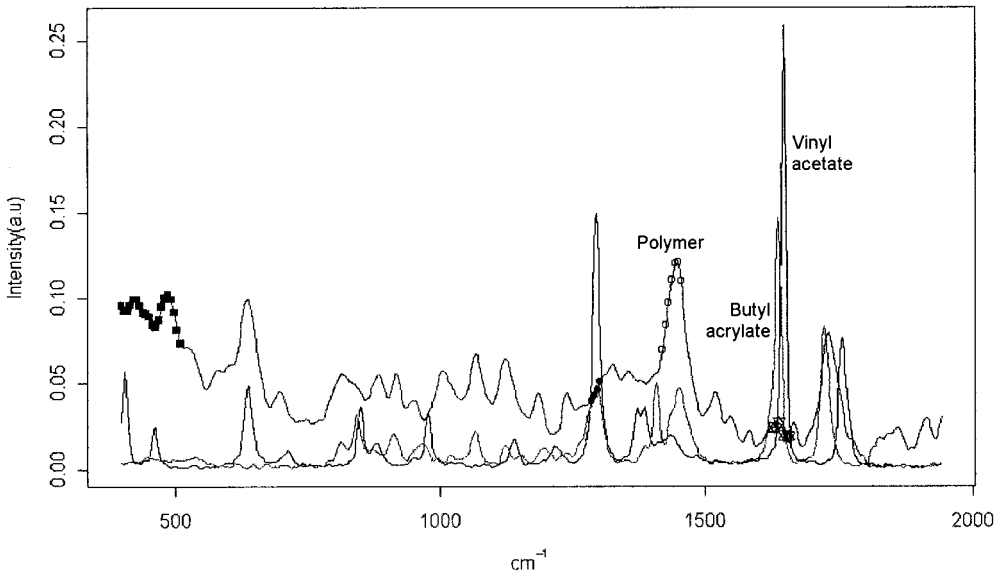

Figure 11 Monomer and polymer spectra; the stars, squares, open circles, and filled circles correspond to the loadings described in the legend to Figure 10.

the calibration model is a good alternative for overcoming the extrapolation problems.

CONCLUSION

As discussed in the Introduction, process variability affects the calibration model predictions and is a problem to be overcome in emulsion polymerization monitoring. Thus, the calibration model used to quantify the monomer concentration by Raman spectroscopy during emulsion copolymerization must be built with spectra that represent the actual behavior occuring during the polymerization process. These spectra can be obtained in two ways: (1) by collecting spectra from samples taken during the reaction process and subsequently quantifying the monomer concentration by a reference method such as GC or (2) by mimicking the behavior of the process with synthetic samples. This work tested these two possibilities for model fitting by using spectra from synthetic samples and those collected during a reaction. The comonomers studied, vinyl acetate and butyl acrylate, have overlapping spectra, especially in the region due to C=C stretching. Additionally, the copolymerization was monitored in a spectroscopic setup arranged for fast spectral acquisition, which resulted in a decrease of the signal-to-noise ratio. Under these realistic conditions, different approaches to the calibration model fitting as well as limitations of the models were examined. The results shown here can be useful for planning reactions, as well as synthetic samples for fitting

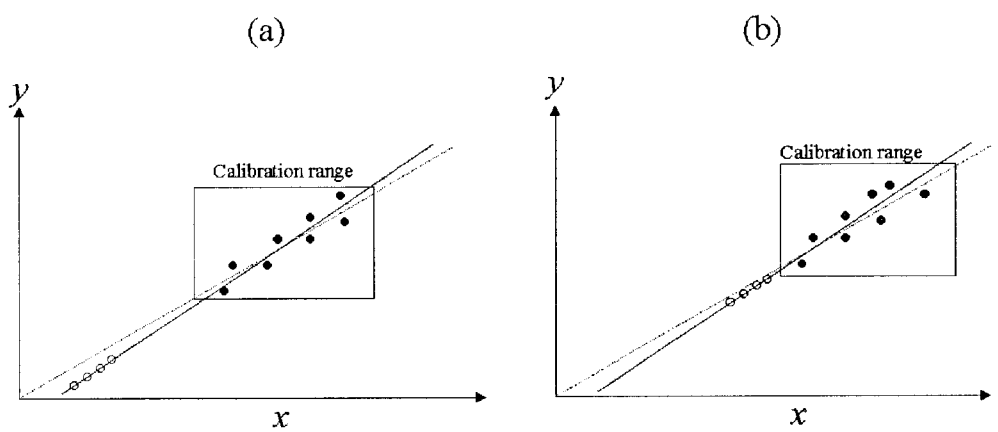

Figure 12 Illustration of extrapolation cases. The dashed line represents the real line, "ideal model," and the solid line represents the model used for the prediction of the open circles.

TABLE II
New Synthetic Samples

Samples	VA(g)	BA(g)	Latex(g) NR2	Water(g)
1[a,b]	0	0	10.002	0
2[a,b]	0	0	0.117	5.075
3[a,b]	0	0	0.295	5.014
4[a,b]	0	0	0.555	4.995
5[a,b]	0	0	1.013	5.072
6[a,b]	0	0	3.000	5.039
7[a,b]	0	0	5.000	5.038
8[a,b]	0	0	9.009	5.112
9[b]	1.080	0.064	4.055	5.011
10[b]	1.005	0.035	6.029	5.045
11[b]	1.012	0.028	8.142	5.068
12[a,b]	1.018	0.016	9.996	5.007
13[a]	0.066	1.005	1.001	5.038
14[a]	0.347	2.517	1.022	5.061
15[a]	0.501	2.004	3.016	5.013
16[a]	2.003	0.510	7.010	5.028
17[a]	0.507	2.107	8.007	5.016
18[a,b]	0.057	0	10.027	5.032
19[a,b]	0.158	0	5.022	5.059
20[a,b]	0.549	0	10.014	0
21[a]	1.526	0	10.037	0

[a] Samples used on the model for vinyl acetate.
[b] Samples used on the model for butyl acrylate.

calibration models for polymerization monitoring. In particular, they point to possible difficulties with extrapolation beyond the monomer concentration range used in the calibration step.

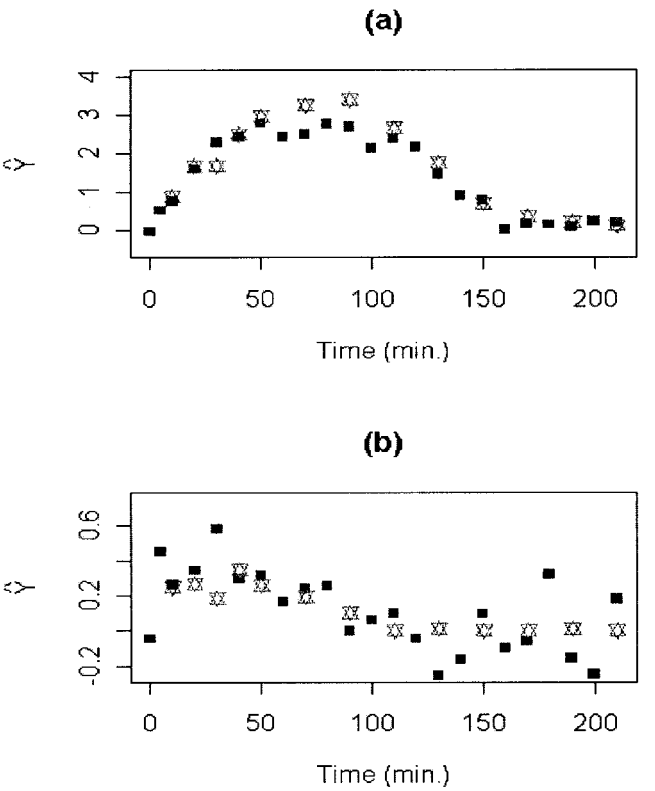

Figure 14 Predicted ($\hat{Y}$) monomer concentration during polymerization NR1 for (a) vinyl acetate and (b) butyl acrylate. Stars correspond to the GC quantification and squares to the PLS model predictions.

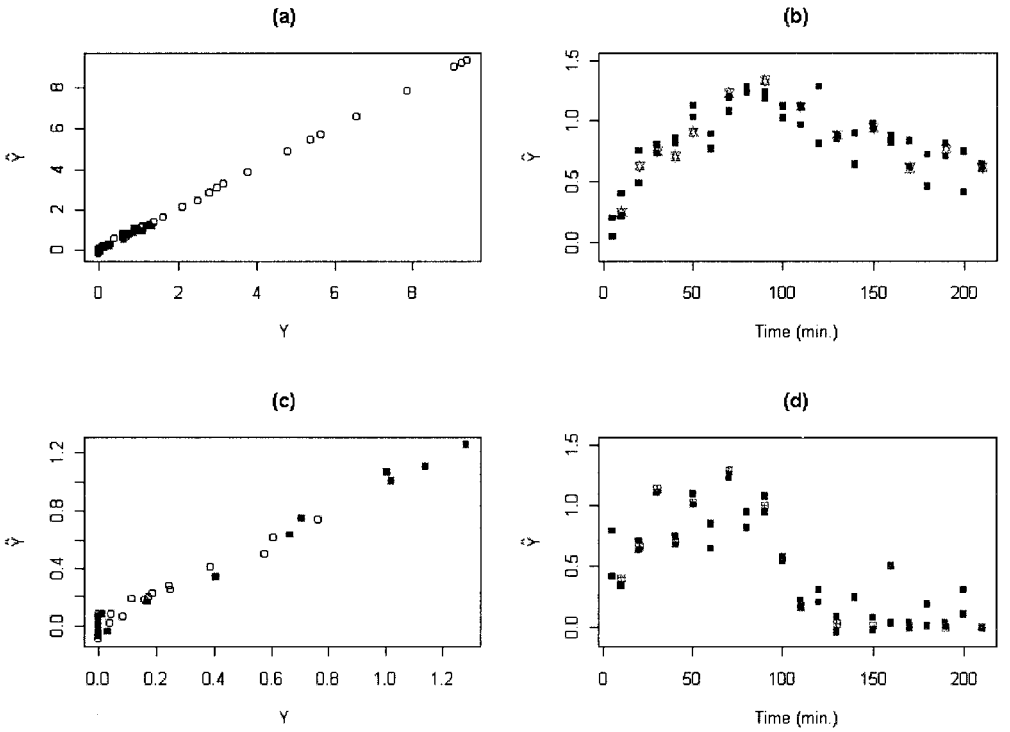

Figure 13 Predicted ($\hat{Y}$) versus actual monomer concentration in (a) and (c) for the new model fit. In (b) and (d), predicted ($\hat{Y}$) monomer concentration during polymerization NR2 for vinyl acetate and butyl acrylate, respectively, with duplicate analysis of each sample. Stars correspond to the GC quantification and squares to the PLS model predictions.

The best results were obtained employing the model based entirely on spectra collected from samples taken from the same reaction. This model has little practical importance in polymerization monitoring, but shows that, even under a spectroscopic setup arranged for a fast spectral acquisition, with reduced signal-to-noise ratio, it is possible to perform good monomer quantification.

The last model was fitted with spectra collected from synthetic samples and from samples collected during a reaction (referred to as NR2). This model was used to estimate monomer concentrations of reaction NR1, resulting in satisfactory predictions. These results are interesting because the concentration of butyl acrylate is very small during the reaction, less than 0.5% w/w, and because the spectra used in the model fitting were collected with a larger number of scans than those collected during reaction NR1. These conditions mimic a realistic performance of a calibration model for process monitoring, where the data used for the model fitting are independent of the process to be monitored and are collected by a spectroscopic setup with a lower noise level than may be present during the process.

A PROFIX fellowship of CNPq-Brazil for P.H.H. Araujo and financial support from CNPq and FAPESP (grant number 01/13,017–1) are gratefully appreciated. The authors also thank Prof. Frank Quina for revising the manuscript.

APPENDIX

Polymer emulsion reactions

The semicontinuous polymerization NR1 was performed by charging the reactor with an initial mass composed of water (420 g), emulsifier [sodium lauryl sulfate (SLS), 2.00g] and pH buffer (calcium carbonate, 0.50 g), which was kept in a nitrogen atmosphere for 1 h. Addition of monomer and initiator (sodium persulfate) to the reactor by feed stream 1, composed of 55.00 g of vinyl acetate plus 55.00 g of butyl acrylate, and feed stream 2, with 30.00 g of water plus 1.00 g of sodium persulfate, was then begun. Feed stream 1 was added at 1.3454 mL/min and feed stream 2 at 0.3375 mL/min. After 90 min, the charge of both feed streams had been completely added to the reactor. The average temperature during the 210 min of reaction was 59.9°C (standard deviation equal to 0.75°C)

The semicontinuous polymerization NR2 was also performed by charging the reactor with a initial mass composed of water (420 g), emulsifier (SLS, 2.00 g), and pH buffer (calcium carbonate, 0.50 g), which was kept in a nitrogen atmosphere for 1 h. Monomer and initiator (sodium persulfate) were then added to the reactor by feed stream 1, composed of 16.50 g of vinyl acetate plus 93.50 g of butyl acrylate, and feed stream 2 with 30.00 g of water plus 1.00 g of sodium persulfate. Feed stream 1 was added at 1.3454 mL/min and feed stream 2 at 0.3375 mL/min. After 90 min, the charge of both feed streams had been added to the reactor. The average temperature during the 210 min of reaction was 59.3°C (standard deviation equal to 0.96°C).

References

1. Kiparissides, C. Chem Eng Sci 1996, 51, 1637.
2. Dimitratos, J.; Eliçabe, G.; Georgakis, C. AIChE J 1994, 40 1993.
3. Laserna, J. J. Modern Techniques in Raman Spectroscopy; John Wiley & Sons: New York, 1996.
4. Hendra, P.; Jones, C.; Warnes, G. Fourier Transform Raman Spectroscopy, Instrumentation and Chemical Application; Ellis Horwood: Chichester, England 1991; p. 127.
5. Wang, C.; Vickers, T. J.; Mann, C. K. Appl Spectrosc 1993, 47, 928.
6. Al-Khanbashi, A.; Hansen, M. G.; Wachter, E. A. Appl Spectrosc 1996, 50, 1089.
7. Özpozan, T.; Schrader, B.; Keller, S. Spectrochim Acta 1997, 53, 1.
8. Bauer, C.; Amram, B.; Agnely, M.; Charmot, D.; Sawatzki, J.; Dupuy, N.; Huvenne, J.-P. Appl Spectrosc 2000, 54, 528.
9. Charmot, D.; Amram, B.; Drochon, B.; Agnely, M.; Armitage, P.; Pere, E. Patent EP1163504, 2000.
10. van den Brink, M. On-Line Monitoring of Polymerization Reactions by Raman Spectroscopy: Application to Control of Emulsion Copolymerizations and Copolymerizations Kinetics; Ph. D. Thesis, Eindhoven University of Technology, Eindhoven, The Netherlands, 2000.
11. Van den Brink, M.; Pepers, M.; van Herk, A. M.; German, A. L. Polym React Eng 2001, 9, 101.
12. Reis, M. M.; Araújo, P. H. H.; Sayer, C.; Giudici, R. Macromolecular Symposia 2004, 206, 165.
13. Jackson, J. E. A User's Guide to Principal Component; John Wiley & Sons: New York, 1991.
14. Geladi, P.; Kowalski, B. R. Anal Chim Acta 1986, 185, 1.
15. de Jong, S. Chemom Intell Lab Syst 1993, 18, 251.
16. Lin-Vien, D.; Colthup, N. B.; Fateley, W. G.; Grasselli, J. G. The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules; Academic Press: San Diego, CA, 1991.
17. Ellis, G.; Claybourn, M.; Richards, S. E. Spectrochim Acta 1990, 46A, 227.