Multivariate Statistical Analysis of Mid-Infrared Spectra for the Online Monitoring of 2-Ethylhexyl Acrylate/Styrene Emulsion Copolymerization

Y. STAVROPOULOS,[1] O. KAMMONA,[1] E. G. CHATZI,[2] C. KIPARISSIDES[1,2]

[1] Department of Chemical Engineering and Chemical Process Engineering Research Institute, Aristotle University of Thessaloniki, P.O. Box 472, Thessaloniki, Greece 540 06

[2] Chemical Process Engineering Research Institute, P.O. Box 1517, Thessalonki, Greece 540 06

Received 17 August 2000; accepted 10 January 2001

ABSTRACT: This article describes the application of multivariate statistical process control techniques to a series of mid-infrared spectra collected online from a styrene/2-ethylhexyl acrylate emulsion copolymerization process. Principal component analysis of the mid-infrared spectral data indicated that *in situ* monitoring of the complex copolymerization process was feasible in the spectral region of interest. It was also observed that projection to latent structures or partial least squares (PLS) could be used for the effective indirect online prediction of individual monomer conversions and copolymer compositions over a substantial range of process operating conditions. A combination of the developed PLS methodology with a mid-infrared attenuated total reflection probe proved to be an effective tool for the efficient online characterization of polymer quality, thereby overcoming the lack of robust online conversion and composition measuring devices. © 2001 John Wiley & Sons, Inc. J Appl Polym Sci 82: 1776–1787, 2001

Key words: mid-infrared spectroscopy; emulsion copolymerization; principal component analysis; projection to latent structures; sensors; kinetics; latices

INTRODUCTION

Efficient polymer-quality control requires correct information on the state and behavior of the polymerization process; therefore, the availability of online measurements is essential. Despite ongoing developments in the field of hardware sensors for the online monitoring of polymerization processes, there is a remarkable lack of robust, highly sensitive and accurate sensing devices for the online monitoring of polymerization processes. The viscous nature of polymerization systems, the presence of several phases and species in polymerization media, and the complexity of polymer chain architecture are the main factors contributing to the limited number of applications of real-time analytical monitoring devices for the online characterization of polymer quality. In addition, extreme difficulties arise when we attempt to infer polymer quality from signals generated by various hardware sensors because of the complexity of the data produced.

Several remote fiber-optic probes suitable for collecting infrared (IR) spectra have been reported in the literature.[1] Near-infrared (NIR) spectroscopy received some early attention for the development of online analyzers because it does not require complicated hardware or sampling

Correspondence to: C. Kiparissides (cypress@alexandros.cperi.certh.gr).

Journal of Applied Polymer Science, Vol. 82, 1776–1787 (2001)

techniques. However, it suffers from nonspecificity of absorption peaks and possible nonlinearities associated with light scattering in heterogeneous systems. Research in the mid-infrared (MIR) spectral region has been sparse, although it contains a wealth of qualitative and quantitative information because MIR transmitting optical fibers, such as chalcogenide and metal halide optical fibers, have been developed only very recently and are not yet available in lengths longer than approximately a few meters. Finally, fiber-optic Raman sensors have been used because they require minimal alignment of samples with respect to the input laser beam or collection optics.

In an attempt to overcome some of the problems presented by the lack of appropriate online instrumentation, software sensors have been proposed to determine difficult-to-measure and unmeasurable process–quality variables. Multivariate statistical data analysis methods and, in some cases, projection to latent structures or partial least squares (PLS) techniques are considered state of the art. Most of the methods currently used either do not make any attempt to extract the underlying latent (information-carrying) variables from the data or at best use linear PLS approaches. However, there are technical limitations to these approaches, especially if the correlations being evaluated are nonlinear. The development of algorithms capable of accurately and reliably characterizing polymer-quality-associated data that can be integrated into the sensor–analyzer system itself is an important, and as yet unsatisfied, requirement.

A number of multivariate statistical projection methods have been presented in the literature, including principal component analysis (PCA) and PLS.[2–7] These methods are particularly suitable for analyzing large sets of correlated data. The information contained within the process and/or quality variables is summarized in terms of a reduced set of latent variables by the projection of the information down onto low-dimensional subspaces. PCA is used for explaining the variability in a single data block. It calculates latent vectors that are uncorrelated, called principal components (PCs), that describe the directions of greatest variability in the data set. Conceptually, PLS is similar to PCA except that it simultaneously reduces the dimensions of both process (**X**) and quality (**Y**) variable spaces to find the latent vectors for the **X** space that are most predictive of the **Y** space.[8] The combination of these projection methods with multivariate control charts underpins multivariate statistical process control (MSPC) methods.[8–10]

Several applications of multivariate statistical projection methods to the quantitative analysis of IR spectroscopic data have been reported in the literature.[11–16] A comparison of various approaches to IR multicomponent quantitative analysis, including K-matrix, multivariate least squares, PC regression, and PLS, was presented by Fuller et al.[11] Hazel et al.[12] described the application of multivariate analysis techniques, including PCA and PLS, to unreferenced MIR spectral data of hydrocarbon-contaminated wet soils. An evaluation of nonlinear model building techniques for the determination of glucose in biological matrices by NIR spectroscopy was presented by Ding et al.[13] Moreover, Chatzi et al.[14] showed that a novel and versatile attenuated total reflection (ATR) IR probe with the recently developed chalcogenide fibers as MIR transparent light guides was especially suitable for obtaining quality spectra in aqueous media and for monitoring characteristic monomer and polymer vibrations in styrene (STY)/2-ethylhexyl acrylate (2-EHA) emulsion polymerizations at the azeotropic composition.

This article describes the successful implementation of an MIR spectroscopic sensor in combination with PCA and PLS for the emulsion copolymerization process of STY/2-EHA. The dimensionality reduction aspects of PCA are exploited to develop MSPC schemes for monitoring the operating performance of the process. PLS is used to obtain online estimates of individual monomer conversions and copolymer composition.

EXPERIMENTAL

Several batch emulsion copolymerization experiments of 2-EHA and STY were carried out at two different initial monomer compositions (30 and 70 wt % STY), two polymerization temperatures (60 and 70°C), and three levels of emulsifier (0.0097, 0.0155, and 0.0176 mol/L) and initiator concentrations (0.0095, 0.013, and 0.015 mol/L). A preweighed amount of emulsifier dissolved in water was added into a 500-mL, water-jacketed glass reactor and was followed by the addition of a predetermined amount of STY and, subsequently, 2-EHA. The reactor was purged with nitrogen while the emulsion was being heated to the final polymerization temperature. When the temperature reached its prespecified level, the initiator

aqueous solution was added to the reaction mixture. The polymerization temperature was kept constant with the aid of a controlled-temperature recirculating bath.

An MIR optical-fiber probe[14] (Graseby Specac, Inc., Kent, UK) having a 19 mm head diameter and a 2-reflection 45° ZnSe crystal, was immersed into the reaction mixture immediately after monomer loading. The MIR probe was equipped with a set of 1.5-m chalcogenide optical fibers, which were especially suitable for use in the MIR region (i.e., 4500–900 cm^{-1}), and was connected to a Perkin Elmer 2000 Fourier transform infrared (FTIR) spectrophotometer purged with nitrogen and equipped with a wide-band mercury cadmium telluride detector. The frequencies of the spectrometer were calibrated to an accuracy better than 0.15 cm^{-1}, and the signal to noise (S/N) ratio was larger than 4000:1. ATR spectra were collected through the probe during polymerization from 4000 to 900 cm^{-1}. Two hundred scans recorded at 15-min intervals, starting from the time of the addition of the initiator ($t = 0$ min), were signal-averaged at a resolution of 4 cm^{-1}. The spectra were ratioed against a background recorded in the open air before the probe was immersed in the reactor. In addition, the total monomer conversion was measured offline at discrete time intervals by gravimetry and thermogravimetric analysis, and the copolymer composition was obtained through offline UV and IR measurements.[17] The standard deviations of the total monomer conversion and copolymer composition estimates were ± 0.03 and ± 0.02, respectively. The corresponding individual degrees of conversion for STY and 2-EHA were deduced on the basis of these direct experimental measurements.

MULTIVARIATE PROJECTION TECHNIQUES

The basic concepts and algorithms of PCA and PLS have been thoroughly presented in the chemometrics literature.[2–7] Their use in the multivariate monitoring of process operating performance has also been treated in recent articles.[8–10] Therefore, in this section we provide only a basic overview of the methods and their application to process and quality monitoring.

Consider a situation in which we have measurements on m process variables and k quality variables taken at n different times. These data can be arranged into an $(n \times m)$ process data matrix $\mathbf{X}$ and an $(n \times k)$ quality data matrix $\mathbf{Y}$. If we are interested in understanding and monitoring the variability only in the process variables ($\mathbf{X}$), we can perform PCA on $\mathbf{X}$. If we are more interested in studying and monitoring the variations in the process variables that are most influential on the quality and productivity variables ($\mathbf{Y}$), we should perform a PLS analysis with both $\mathbf{Y}$ and $\mathbf{X}$. Both these procedures are based on projecting the information in the high-dimensional data spaces ($\mathbf{X}$, $\mathbf{Y}$) down onto low-dimensional spaces defined by a small number of latent variables ($t_1, t_2, \ldots, t_A$). These new latent variables summarize all the important information contained in the original data sets.

In PCA, the objective is to compress the variance of $\mathbf{X}$ into a few uncorrelated latent variables (PCs). Accordingly, the mean-centered $\mathbf{X}$ matrix is decomposed as follows:

$$\mathbf{X} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^T + \mathbf{E} \tag{1}$$

The calculation of the PCs is based on the nonlinear iterative partial least squares (NIPALS) algorithm.[5] The latent score vectors $\mathbf{t}_a$ are computed sequentially from the input data for each new dimension ($a = 1, 2, \ldots, A$) such that the linear combination of the process variables ($x_j; j = 1, 2, \ldots, m$) defined by the latent variable $t_a = \mathbf{p}_a^T \mathbf{x}$ maximizes the variance of $\mathbf{X}$ that is explained at each dimension. The vector $\mathbf{p}_a$ is the loading vector whose elements, p_{ja}, express the contribution of each variable, x_j, toward defining the new PC, t_a. $\mathbf{E}$ is the matrix of residuals after the fitting of A PCs (e.g., the deviations of original data from the A-dimensional model). With highly correlated variables, the first few PCs (usually two or three) account for most of the variability present in the data. The remaining components are of limited significance and are typically attributed to noise inherent in the collected data. Hence, for monitoring it is often sufficient to consider only the first few dimensions ($a = 1, 2$) in eq. (1).

PLS is a multivariate regression method whose objective is to establish a predictive relationship between two data sets, $\mathbf{X}$ and $\mathbf{Y}$, and reduce problem dimensionality by compressing the covariance between $\mathbf{X}$ and $\mathbf{Y}$ into a low-dimensional subspace. The mean-centered $\mathbf{X}$ matrix is decomposed as shown in eq. (1). Similarly, the decomposition of the mean-centered $\mathbf{Y}$ matrix is represented as

$$\mathbf{Y} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{q}_a^T + \mathbf{F} \quad (2)$$

The calculation of the PLS components is based on the NIPALS PLS algorithm.[5] The latent score vectors $\mathbf{t}_a$ are computed sequentially for each PLS dimension ($a = 1, 2, \ldots, A$) such that the linear combination of the x's defined by the latent variable $t_a = \mathbf{w}_a^T\mathbf{x}$ and the linear combination of the y's defined by the latent variable $u_a = \mathbf{q}_a^T\mathbf{y}$ maximize the covariance between $\mathbf{X}$ and $\mathbf{Y}$ that is explained at each dimension. The vectors $\mathbf{w}_a$ and $\mathbf{q}_a$ are weight vectors whose elements, w_{ja} and q_{ka}, respectively, express the contribution of each variable, x_j and y_k, respectively, toward defining the new latent variables, t_a and u_a. $\mathbf{E}$ and $\mathbf{F}$ are the residual matrices of $\mathbf{X}$ and $\mathbf{Y}$, respectively, after the fitting of A latent variables.

The number of latent vectors (A) retained in the model is usually chosen by crossvalidation methods.[18] This involves holding back some of the observations, performing a PLS analysis, and computing the prediction error sum of squares (PRESS) for those observations left out. This is repeated until every observation has been left out once. The optimal order of the model (A) is taken as that minimizing the total PRESS. This procedure eliminates the overfitting that often occurs when models are fitted to correlated data and leads to increased prediction errors.

After a PCA or PLS model is built, process monitoring can be performed in the reduced latent variable space, instead of the original $\mathbf{X}$ block variables being used, by the plotting of the projected scores in the selected number of dimensions (e.g., individual scores plots, joint t_i–t_j plots, or plots of Hotelling's T-square statistic) and the residuals of the projection. Quadratic residual time evolution plots, such as a squared prediction error (SPE) plot, can be used for monitoring the deviations of original data from the MSPC model. The SPE of the $\mathbf{X}$ block is calculated as follows:

$$\mathrm{SPE}_{x,i} = \sum_{j=1}^{m} (x_{ij} - \hat{x}_{ij})^2 = \sum_{j=1}^{m} e_{ij}^2 \quad (3)$$

where $\hat{x}_{ij}$ is the value of x_{ij} predicted by the MSPC model. The state of the process at each time interval is statistically determined on the basis of appropriate control limits that define the range of acceptable values for the scores (t) and residuals (e). Two types of control limits are usually defined: a warning limit corresponding to a 5% significance level and an action limit corresponding to a 1% significance level. A shift in the score plane outside the control region indicates an increase in the magnitude of process variation that can still be explained by the MSPC model. However, when an increase in SPE beyond control limits is observed, a new event not included in the reference set has occurred, and the model is no longer valid.[9]

DATA ANALYSIS AND RESULTS

Data Set Characteristics

The available historical database consisted of a large number of online collected IR spectra (250 samples) and individual STY and 2-EHA conversion and copolymer composition values (204 samples) measured offline. The available data were representative of the process under investigation because they spanned every type of variability that had to be modeled. A mathematical model developed in the laboratory for the simulation of STY/2-EHA batch emulsion copolymerization[19] was employed to enrich the available database of offline measurements with quality data corresponding to intermediate time intervals.

A typical set of transmittance spectra from the historical database is presented in Figure 1. The spectral data used in the analysis consisted of IR transmittance measurements at 401 distinct wavelengths in the range of 1800–1400 cm^{-1}, which was shown to be adequate for the quantification of all types of structural units present in the system. Seven major spectral absorbances can be identified in Figure 1, with their maxima observed approximately at wavelengths of 1727, 1630, 1600, 1495, 1465, 1452, and 1407 cm^{-1}. In the 1550–1710-cm^{-1} region, the spectra of the emulsion have monomer and copolymer bands superimposed on the basic contours of the water spectrum.[14] However, manual subtraction of the water spectrum was not considered necessary on the basis of the performance of the statistical models developed.

The underlying peaks at 1630 and 1600 cm^{-1} correspond to the skeletal breathing vibration of the aromatic ring of the STY monomer and the corresponding structural unit in the copolymer, respectively. The peak at 1495 cm^{-1} is characteristic of the nature and position of the substituents

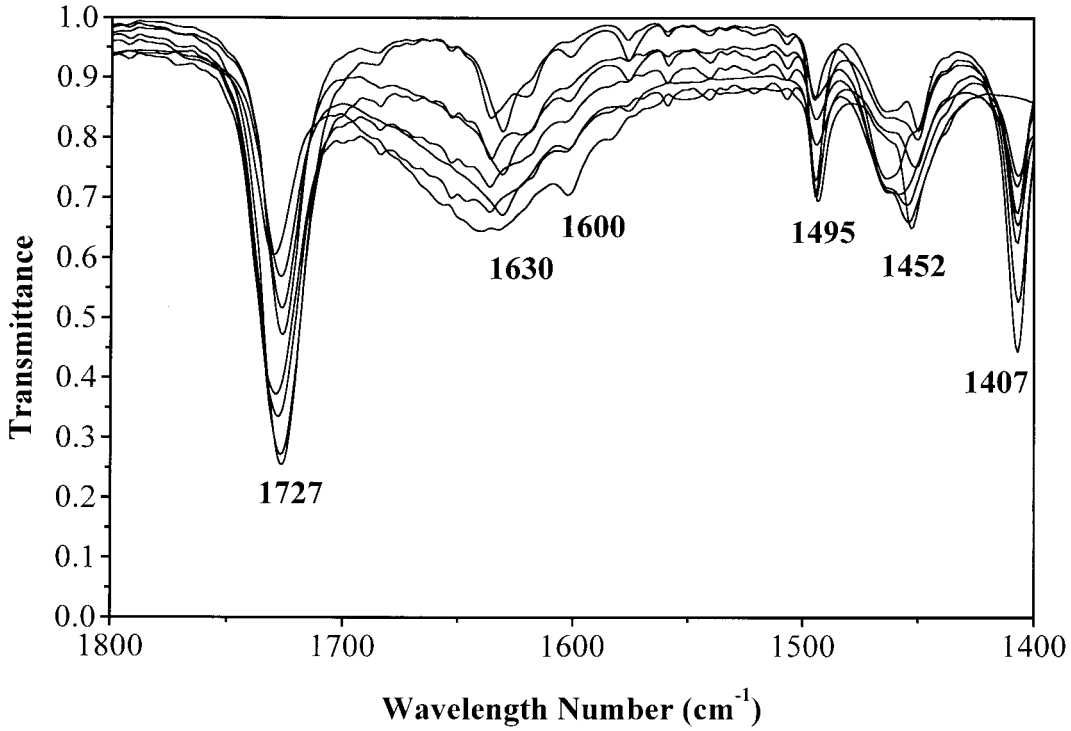

Figure 1 Indicative MIR spectra collected during STY/2-EHA emulsion copolymerization experiments.

of the STY aromatic ring, and its frequency is not very much affected by the polymerization reaction. The peaks at 1465 and 1452 cm^{-1} represent another skeletal ring breathing vibration of the STY monomer and its structural unit in the copolymer. Finally, the peaks at 1727 and 1407 cm^{-1} are characteristic of 2-EHA; more specifically, the position and shape of the 1727-cm^{-1} carbonyl stretching band is known to depend on copolymer composition,[17] whereas the 1407-cm^{-1} peak can be mainly attributed to the monomeric 2-EHA.

Figure 2 shows individual monomer conversion data for STY from a number of polymerization experiments carried out with initial STY weight fractions (W_{St}) of 0.3 and 0.7 under different experimental conditions (e.g., temperature, initiator concentration, and emulsifier concentration). W_{St} is defined as follows:

$$W_{St} = \left[\frac{\text{STY mass}}{\text{STY mass} + \text{2-EHA mass}}\right] \quad (4)$$

The STY conversion values for different experimental conditions follow a similar pattern, indicating that a global statistical model could be developed for the online prediction of STY monomer conversion. However, for a W_{St} value of 0.3 (see Fig. 3), the copolymer composition decreases with polymerization time, whereas it remains constant for a W_{St} value of 0.7 because of the azeotropic polymerization conditions.

To verify that all the samples of the historical database represent normal operation, we initially applied PCA and PLS to the whole data set. An investigation of the respective control charts (i.e., t and SPE plots) in the reduced latent plane did not reveal the presence of abnormal samples. A subset of the available historical data (84 samples for PCA and 67 samples for PLS, respectively) were selected to form a secondary data set upon which the developed models were validated. The selection was performed in such a way that the selected sample sets spanned the whole range of conversion. The remaining samples formed the reference data sets upon which the two models were built.

PCA Modeling and Validation

The **X** data matrix of reference spectra (166 $\times$ 401) was mean-centered at each spectral frequency prior to PCA modeling by the subtraction

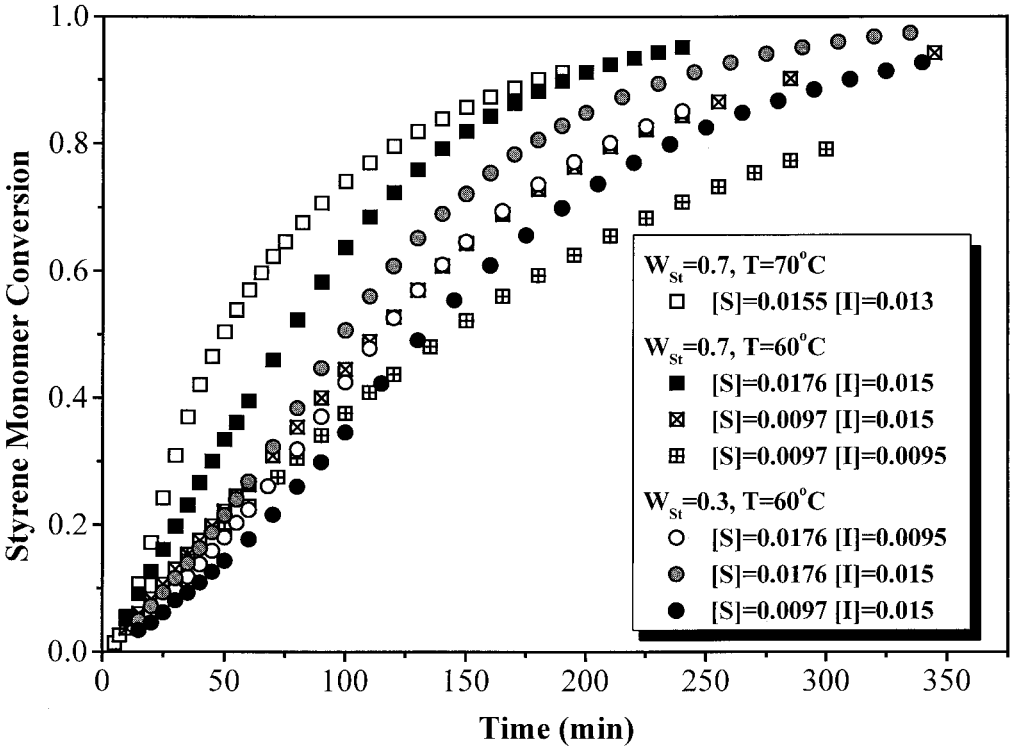

Figure 2 Time evolution of STY monomer conversion under different STY/2-EHA experimental conditions. [S] and [I] are the concentrations (mol/L) of the emulsifier and initiator, respectively.

of the average spectrum from each individual spectrum. This step is necessary for the effective application of the PCA and PLS techniques.[5]

The amount of spectral variability captured by each PC is listed in Table I for the first five latent dimensions. The first PC accounts for 80% of the variability in **X**, indicating that a strong linear correlation exists between the spectral frequencies. Two more PCs were considered significant on the basis of a visual inspection of their respective loadings, which account for 12 and 6.5% of the variance in **X**, respectively. The remaining components contribute very little variance to the model (i.e., $<1\%$) and are typically attributed to noise. This was verified by a number of criteria, such as Kaiser's rule[2,3] (i.e., the normalized eigenvalues of the retained PCs should be larger than the empirical limit of 1.0) and the second derivative of the imbedded error function proposed by Elbergali et al.[20] As shown in Table I, the second derivative exhibited a first maximum in the third PC, indicating that the amount of experimental noise mixed into the three-latent-factor scheme is minimal.

Figure 4 depicts the **X**-block loadings of the three PCs. These express the contribution of each individual variable (i.e., wavelength) toward defining each PC. By identifying those spectral regions that contribute the most to each PC, we can give a physical interpretation to the PCs; this information could be useful for selecting the dimensionality of the PCA model (i.e., identification of uninformative PCs), for attaining a better understanding of the underlying process phenomena and, consequently, for interpreting process faults or disturbances.

The first PC (see Fig. 4) is dominated by two positive loadings at 1727 and 1407 cm^{-1}, both ascribed to the monomeric 2-EHA, and by a strong negative contribution in the range of 1710–1500 cm^{-1}, which is mainly attributed to the water absorption. Because most of the spectral variability is accounted for by the first PC (i.e., 80%), the main source of disturbances in the copolymerization process can be related to the 2-EHA monomer.

Three areas dominate the second PC. The first one is in the area of the 1730-cm^{-1} band and is characterized by the strong contribution of a doublet attributed to the carbonyl stretching absorbances of the monomeric and polymeric 2-EHA at approximately 1720 and 1735 cm^{-1}, respectively.

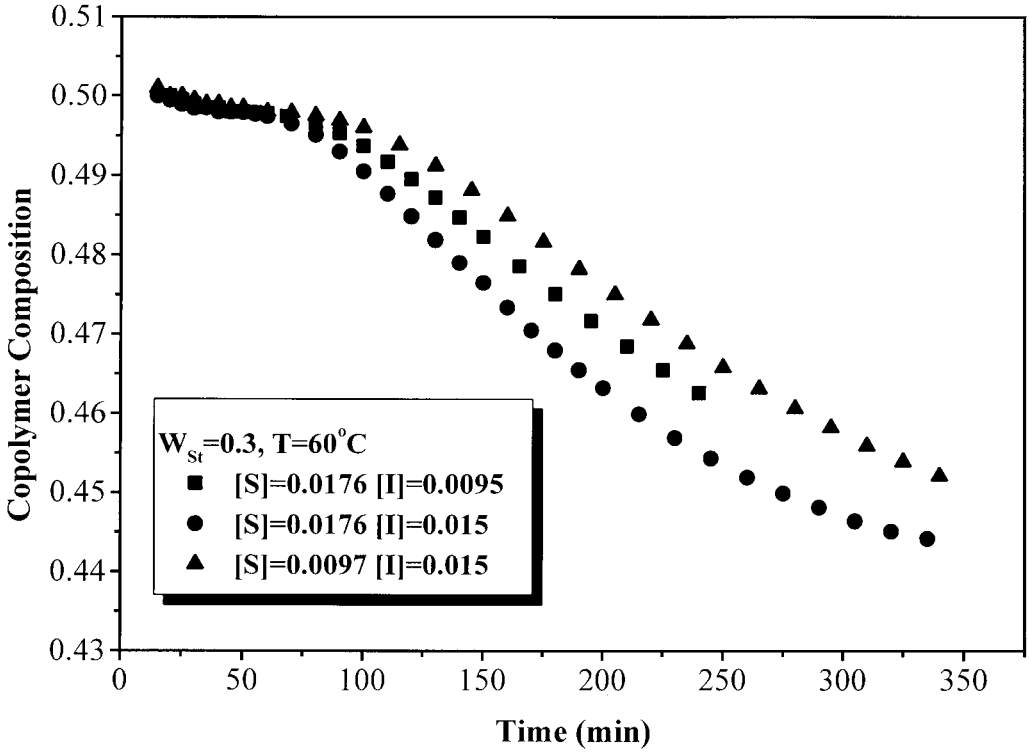

Figure 3 Time evolution of copolymer composition for different emulsifier ([S]) and initiator ([I]) concentrations (mol/L) and a W_{St} value of 0.3.

The second area consists of a positive peak at 1465 cm^{-1} and two negative peaks at 1495 and 1452 cm^{-1}, which correspond to the skeletal ring breathing vibration of the STY aromatic ring. Finally, the third area consists of a strong positive loading at 1407 cm^{-1} attributed to the monomeric 2-EHA. Hence, the loadings of the second PC show features characteristic of all species present in the reaction mixture and are, therefore, difficult to interpret. However, the scores in the second PC are grouped into two clusters according to the initial monomer composition, as described later (see Fig. 5). Consequently, on the basis of the previous results, the second PC is very likely to describe differences in the initial monomer mixture composition and in the copolymer composition.

Table I Sum of Squares (SSX) and Cumulative SSX Explained by the PCA Model and Calculated Values of the Second Derivative of the Imbedded Error [SD(IE)] Function[20]

Principal Component	SSX (%)	Cumulative SSX (%)	SD(IE)
1	80.1	80.1	—
2	12.1	92.2	−0.26
3	6.6	98.8	0.17
4	0.8	99.5	0.09
5	0.2	99.7	−0.04

Similar results arise for the third PC. The third PC is characterized by a strong positive loading at 1735 cm^{-1} attributed to the reacted 2-EHA and a strong negative contribution at 1407 cm^{-1}, which is characteristic of the unreacted 2-EHA monomer. In addition, there exists a strong positive contribution of the peaks at 1495 and 1452 cm^{-1}, which are characteristic of the STY structural unit in the copolymer. The scores in the third PC decrease with polymerization time. Further investigation of the individual variable contributions to the observed decrease indicated that it could be mainly attributed to the peaks at 1735 and 1452 cm^{-1}, and so the third PC is very likely to correspond to the extent of copolymerization.

Figure 5 illustrates the joint t_1–t_2 plot of the first two PCs of the PCA model. In this plot, each spectrum is plotted as a point in the low-dimensional PC space. Black circles represent reference

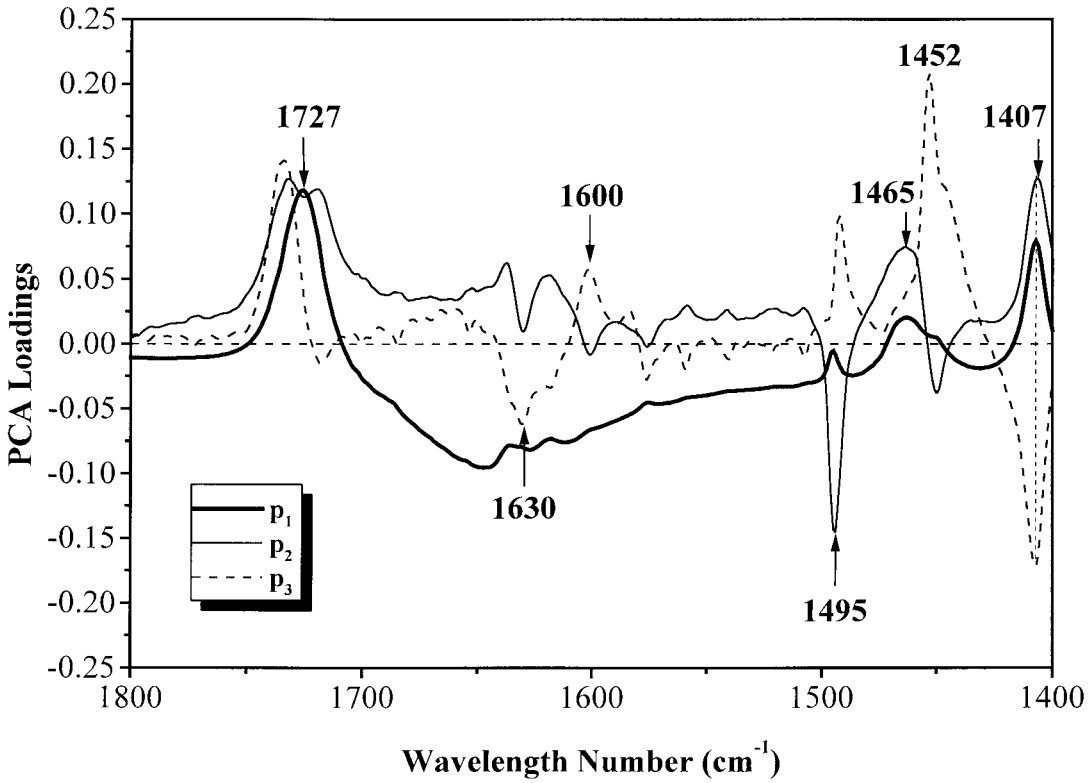

Figure 4 Loadings of the three PCs of the PCA model.

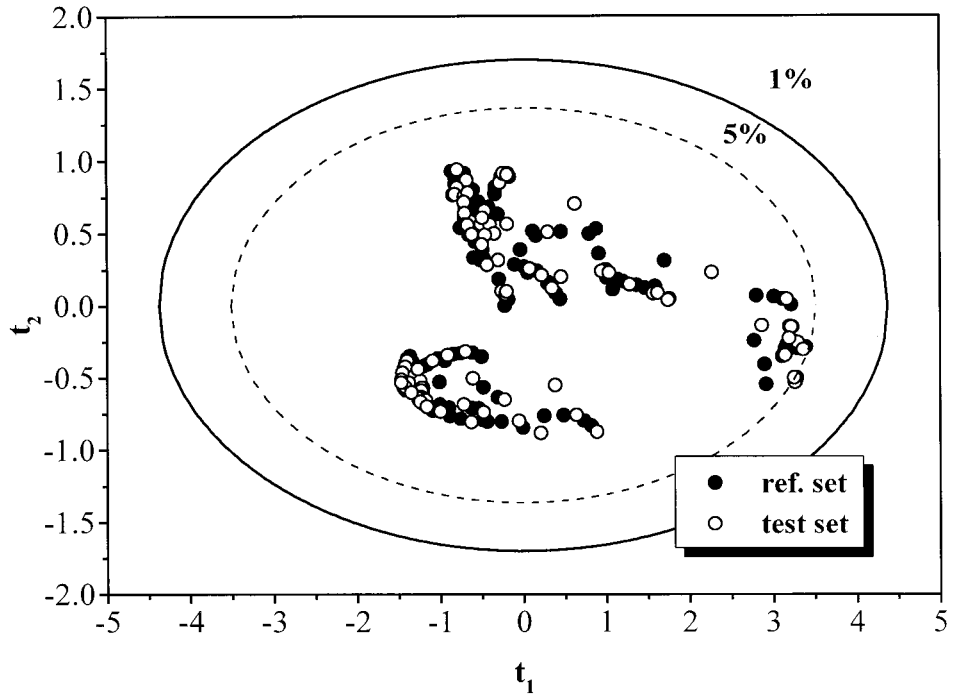

Figure 5 Joint t_1–t_2 plot of the first two PCs of the PCA model. The straight and dotted lines correspond to significance levels of 1 and 5%, respectively.

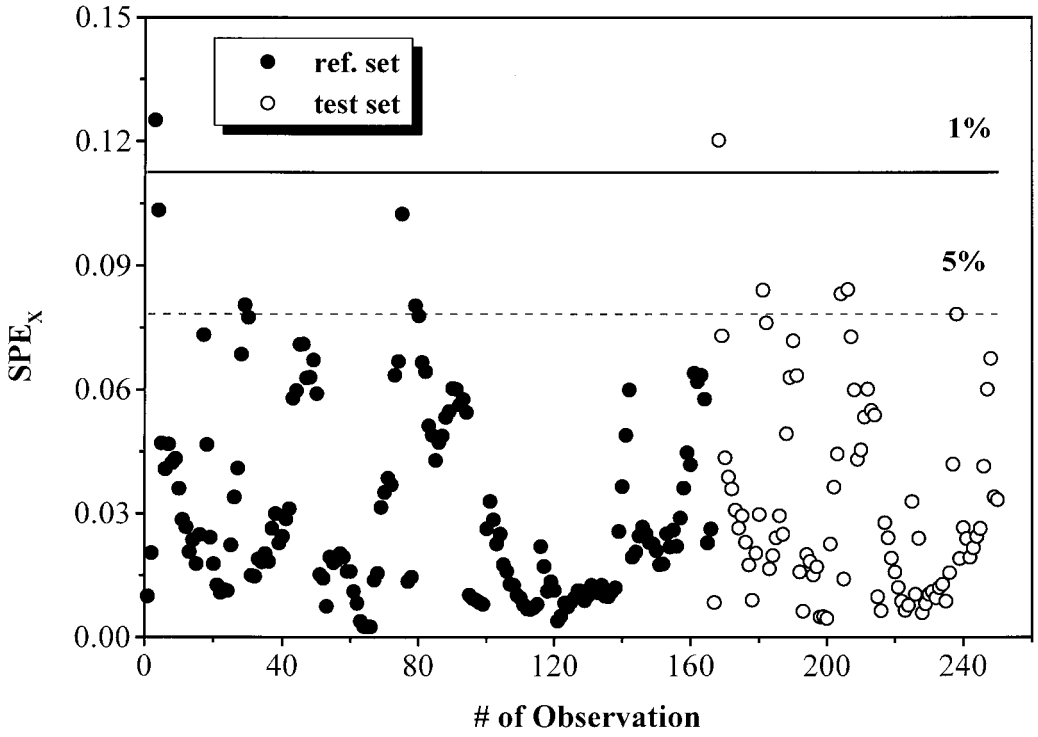

Figure 6 Quadratic **X**-block residuals (SPE_x) of the reference and test data based on a three-dimensional PCA model. The straight and dotted lines correspond to significance levels of 1 and 5%, respectively.

data, and white circles represent test data. The straight and dotted lines correspond to significance levels of 1 and 5%, respectively. These points are grouped according to W_{St} into two main clusters. More specifically, points that have positive t_2 values correspond to spectra with W_{St} equal to 0.7; points that have negative t_2 values correspond to spectra with W_{St} equal to 0.3. However, a third cluster of points with high positive t_1 values and small absolute t_2 values can be observed. These points refer to IR spectra characterized by strong water absorption, observed at the early stage of polymerization. As shown in Figure 6, these points have a high SPE, and some of them slightly exceed the 1% control limit. However, because the water contribution in the IR spectra was variable, these samples were considered to provide extreme but valuable information; therefore, they were not removed from the reference database.

From the aforementioned PCA analysis, one can conclude that a global three-dimensional PCA model can adequately describe the normal operation of the STY/2-EHA copolymerization process.

PLS Modeling and Validation

Subsequently, a PLS model was developed that was based on reference **X**-block spectra (137 $\times$ 401) and their respective **Y**-block quality values (137 $\times$ 3) for the prediction of individual STY and 2-EHA conversions and copolymer composition during polymerization. Before PLS analysis, both blocks were mean-centered. The three **Y**-block variables were further scaled to unit variance so

Table II Percentage Cumulative Sum of Squares Explained by the PLS Model and Calculated PRESS Based on Cross-Validation[18]

PLS Dimension	**X** Block	**Y** Block	PRESS
1	61.3	37.7	258.25
2	88.2	76.4	97.66
3	97.9	87.4	51.88
5	99.3	92.1	33.83
10	100.0	95.7	18.53
14	100.0	97.4	11.19
15	100.0	97.8	13.05

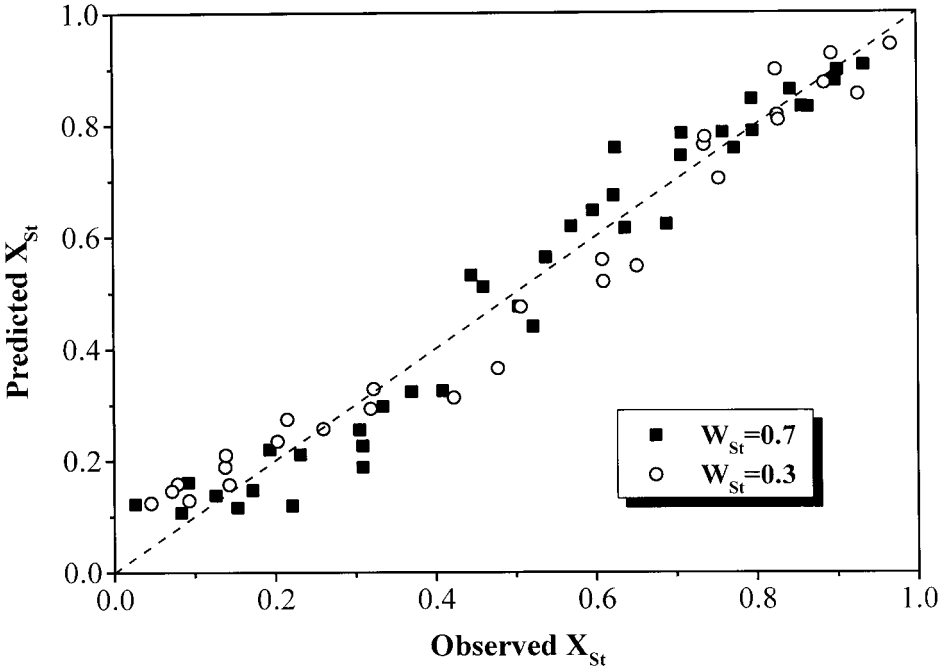

Figure 7 Predicted versus observed STY conversions (X_{St}) for test samples based on the PLS model.

that they were of equal importance to the projection.

The percentage of the cumulative **X**- and **Y**-block variability against the number of PLS components is listed in Table II. Most of the variability in **X** (98%) and **Y** (87.5%) is accounted for by the first three PLS components. On the basis of the crossvalidation technique,[18] the optimum number of latent variables to be retained in the model was identified. Crossvalidation indicated that a total of 14 latent variables should be used for optimal predictive ability (i.e., minimum PRESS), as shown in Table II. The crossvalidated root-mean-square error between predicted and observed **Y** values was 6% for STY conversion, 5.4% for 2-EHA conversion, and 1.3% for copolymer composition.

Figure 7 shows the predicted and observed values of STY conversion for each test sample. Black points represent STY conversions obtained at a W_{St} value of 0.7; white points correspond to a W_{St} value of 0.3. The predicted values of individual monomer conversions are in close agreement with the experimental measurements. Copolymer composition predictions obtained from the global PLS model and a local PLS model developed on the basis of a W_{St} value of 0.3 are presented in Figure 8. Although the local PLS model outperforms the global PLS model, the latter satisfactorily predicts (e.g., based on the experimental error of offline copolymer measurements) the copolymer composition, which means that the development of a different PLS model for each W_{St} value is not necessary. Very good agreement was also obtained between these results and the offline FTIR and UV composition measurements.[17]

CONCLUSIONS

MSPC techniques, including PCA and PLS, were applied to a series of MIR spectra collected online from an STY/2-EHA emulsion copolymerization process. PCA of the MIR spectral data indicated that *in situ* monitoring of the process under study was feasible in the spectral region of interest. A global three-dimensional PCA model adequately described the normal operation of the STY/2-EHA copolymerization process. In addition, PLS was used for the effective indirect online prediction of individual monomer conversions and copolymer composition over a substantial range of process operating conditions. A global PLS model satisfactorily predicted copolymer composition over a wide range of experimental conditions; therefore, the development of a different PLS model for each W_{St} value was not considered necessary. The

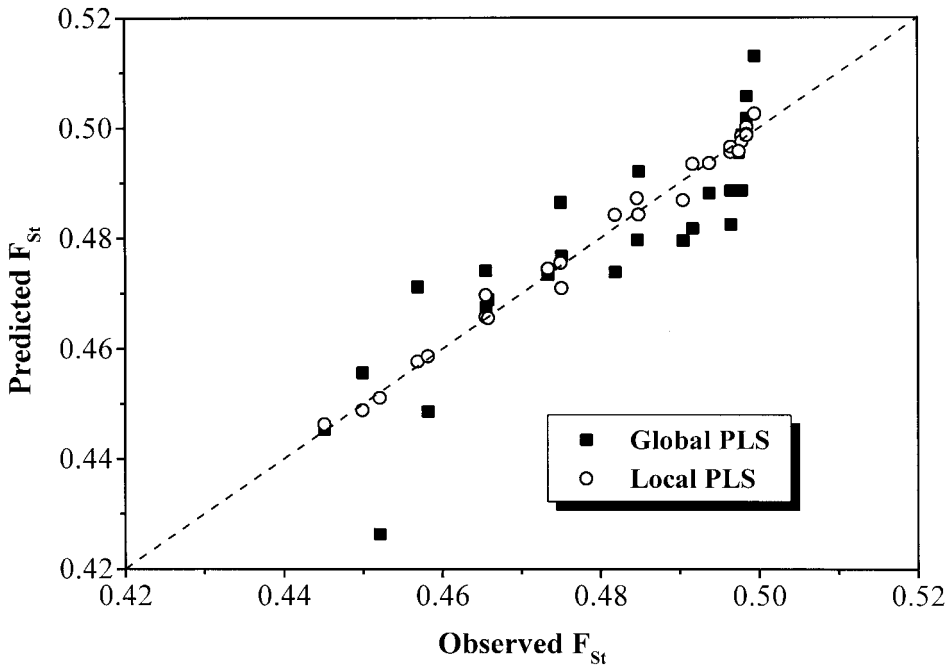

Figure 8 Predicted copolymer compositions versus observed copolymer compositions (F_{St}) for test samples corresponding to a W_{St} value of 0.3. Global and local PLS models were used; the local model was based only on data corresponding to a W_{St} value of 0.3.

overall performance of the PLS model leads us to the conclusion that the developed PLS methodology, in combination with the MIR ATR probe, has the potential to serve as an effective online polymer-quality monitoring tool. From the data presented, it is clear that the application of MSPC techniques is important for the efficient online characterization of polymer quality hindered at present by the lack of robust, highly sensitive and accurate sensing devices.

NOMENCLATURE

a	index for latent dimensions
I	index for observations (rows of **X** and **Y**)
j	index for process variables (columns of **X**)
k	index for quality variables (columns of **Y**)
n	number of observations
m	number of process variables
k	number of quality variables
A	number of latent dimensions retained in the MSPC model
x_j	**X** variable
y_k	**Y** variable
x_{ij}	value of **X** variable j for observation I
$\hat{x}_{ij}$	value of x_{ij} predicted by the PCA/PLS model
e_{ij}	residual of x_{ij} in PCA/PLS
t_a	latent variable of the **X** space
u_a	latent variable of the **Y** space
p_{ja}	loading for **X** variable j in latent dimension a calculated by PCA or PLS
q_{ka}	loading for **Y** variable k in PLS dimension a
w_{ja}	weight for **X** variable j in PLS dimension a
$\mathrm{SPE}_{x,i}$	squared prediction error calculated on **X** for observation i
PRESS	prediction error sum of squares
W_{St}	initial styrene weight fraction
F_{St}	molar fraction of the styrene units in the copolymer
[S]	initial concentration of emulsifier (mol/L)
[I]	initial concentration of initiator (mol/L)
T	polymerization temperature (°C)
STY	styrene
2-EHA	2-ethylhexyl acrylate
x	$(m \times 1)$ vector of process variables representing one observation
y	$(k \times 1)$ vector of quality variables representing one observation

$\mathbf{t}_a$	$(n \times 1)$ score vector for **X** in latent dimension a calculated by PCA or PLS
$\mathbf{p}_a$	$(m \times 1)$ loading vector for **X** in latent dimension a calculated by PCA or PLS
$\mathbf{q}_a$	$(k \times 1)$ loading vector for **Y** in PLS dimension a
$\mathbf{w}_a$	$(m \times 1)$ vector of weights for **X** variables in PLS dimension a
X	$(n \times m)$ matrix of process data
Y	$(n \times k)$ matrix of quality data
E	residual matrix for **X** in PCA or PLS
F	residual matrix for **Y** in PLS

REFERENCES

1. Kammona, O.; Chatzi, E. G.; Kiparissides, C. J Macromol Sci Rev Macromol Chem Phys 1999, 39, 57.
2. Jolliffe, L. T. Principal Component Analysis; Springer-Verlag: New York, 1986.
3. Jackson, J. E. A User's Guide to Principal Components; Wiley: New York, 1991.
4. Wold, S.; Esbensen, K.; Geladi, P. Chemom Intell Lab Syst 1987, 2, 37.
5. Geladi, P.; Kowalski, B. R. Anal Chim Acta 1986, 185, 1.
6. Hoskuldsson, A. J Chemom 1988, 2, 211.
7. Martens, H.; Naes, T. Multivariate Calibration; Wiley: New York, 1991.
8. Kresta, J. V.; MacGregor, J. F.; Marlin, T. E. Can J Chem Eng 1991, 69, 35.
9. MacGregor, J. F.; Kourti, T. Control Eng Practice 1995, 3, 403.
10. Skagerberg, B.; MacGregor, J. F.; Kiparissides, C. Chemometrics and Intelligent Laboratory Systems 1992, 14, 341.
11. Fuller, M. P.; Ritter, G. L.; Draper, C. S. Appl Spectrosc 1988, 42, 217.
12. Hazel, G.; Bucholtz, F.; Aggarwal, I.; Nau, G.; Ewing, K. Appl Spectrosc 1997, 51, 984.
13. Ding, Q.; Small, G. W.; Arnold, M. A. Appl Spectrosc 1999, 284, 333.
14. Chatzi, E. G.; Kammona, O.; Kiparissides, C. J Appl Polym Sci 1997, 63, 799.
15. Bhandare, P.; Mendelson, Y.; Peura, R. A.; Janatsch, G.; Kruse-Jarres, J. D.; Marbach, R.; Heise, H. M. Appl Spectrosc 1993, 47, 1214.
16. Shen, C.; Peacock, A. J.; Alamo, R. G.; Vickers, T. J.; Mandelkern, L.; Mann, C. K. Appl Spectrosc 1992, 46, 1226.
17. Chatzi, E. G.; Kammona, O.; Kentepozidou, A.; Kiparissides, C. Macromol Chem Phys 1997, 198, 2409.
18. Wold, S. Technometrics 1978, 20, 397.
19. Fratzikinakis, C.; Kammona, O.; Kiparissides, C. J Appl Polym Sci, to be submitted.
20. Elbergali, A.; Nygren, J.; Kubista, M. Appl Spectrosc 1999, 379, 143.