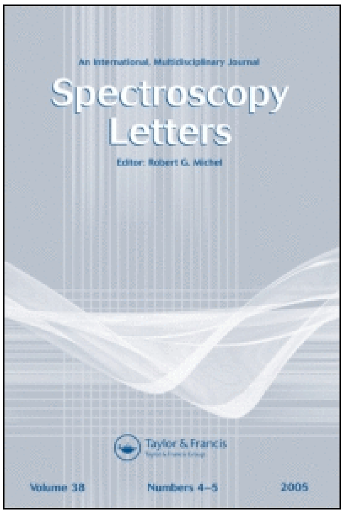

Spectroscopy Letters
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597299

Quantification of LDPE [Low Density Poly(Ethylene)], Lldpe [Linear Low Density Poly(Ethylene)], and Hdpe [High Density Poly(Ethylene)] In Polymer Film Mixtures “As Received” Using Multivariate Modeling with Data Augmentation (Data Fusion) and Infrared, Raman, and Near-Infrared Spectroscopy

Jerome J. Workman Jr[a]
[a] Kimberly-Clark Corporation, Analytical Science & Technology, Neenah, Wisconsin, USA

To cite this Article Workman Jr, Jerome J.(1999) 'Quantification of LDPE [Low Density Poly(Ethylene)], Lldpe [Linear Low Density Poly(Ethylene)], and Hdpe [High Density Poly(Ethylene)] In Polymer Film Mixtures “As Received” Using Multivariate Modeling with Data Augmentation (Data Fusion) and Infrared, Raman, and Near-Infrared Spectroscopy', Spectroscopy Letters, 32: 6, 1057 — 1071

To link to this Article: DOI: 10.1080/00387019909350050

URL: http://dx.doi.org/10.1080/00387019909350050

SPECTROSCOPY LETTERS, 32(6), 1057–1071 (1999)

QUANTIFICATION OF LDPE [LOW DENSITY POLY(ETHYLENE)], LLDPE [LINEAR LOW DENSITY POLY(ETHYLENE)], AND HDPE [HIGH DENSITY POLY(ETHYLENE)] IN POLYMER FILM MIXTURES "AS RECEIVED" USING MULTIVARIATE MODELING WITH DATA AUGMENTATION (DATA FUSION) AND INFRARED, RAMAN, AND NEAR-INFRARED SPECTROSCOPY

Keywords: polyethylene, data fusion, data augmentation, at-line, rapid, films

JEROME J. WORKMAN, JR
Kimberly-Clark Corporation, Analytical Science & Technology,
Neenah, Wisconsin 54956 USA

ABSTRACT

This work was completed to develop a routine method for *rapid or at-line* quantitative measurement of the blend composition of poly blends using NIR, IR, and Raman spectroscopy. The initial results indicate that spectroscopy can be used for predicting percent LDPE [low density poly(ethylene)], LLDPE [linear low density poly(ethylene)], and HDPE [high density poly(ethylene)] with accuracy values ranging from 1-5% absolute. The best results have been obtained by using mean-centered Raman spectra. Successful quantitative analysis was also achieved by combining NIR and Raman spectra using data augmentation (i.e., data fusion), autoscaling, and a variety of regression methods (chemometrics). Initial accuracy (as SEE, standard error of estimate for the regression) was 0.53-1.4 % absolute using Raman spectral data, and 1.5-4.5 % absolute for NIR/Raman augmented data. Infrared was found to have the least useful information content as compared to both NIR and Raman due to *in situ* sample thickness.

INTRODUCTION

Standard use of infrared spectroscopic techniques have not been shown to be useful for routine quantitative analysis for mixtures of LDPE, LLDPE, and HDPE in polymer films. This Technical Letter reports the development of a procedure enabling quantitative analysis of the film mixtures using standard spectroscopic measurements combined with chemometric methods. The results of multivariate regression modeling for IR (infrared, Figure 1), NIR (near infrared, Figure 2), Raman (Figure 3), and NIR + Raman (augmented spectra, Figures 4a and 4b) are shown in Tables 1 through 3; Figures 5a through 7b, and Figures 8a through 10b.

www.dekker.com

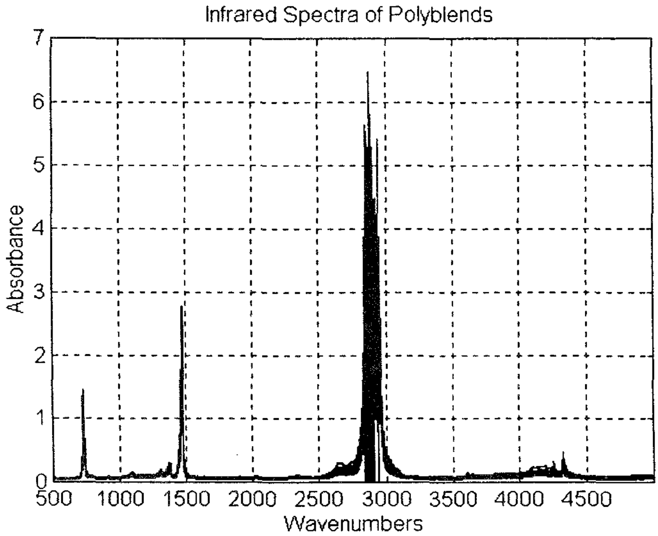

Figure 1. Infrared Spectra of Polyblend Films

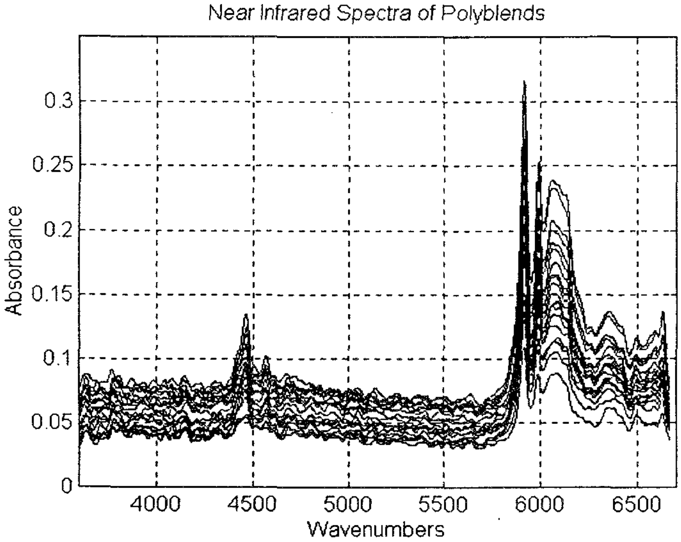

Figure 2. Near Infrared Spectra of Polyblend Films

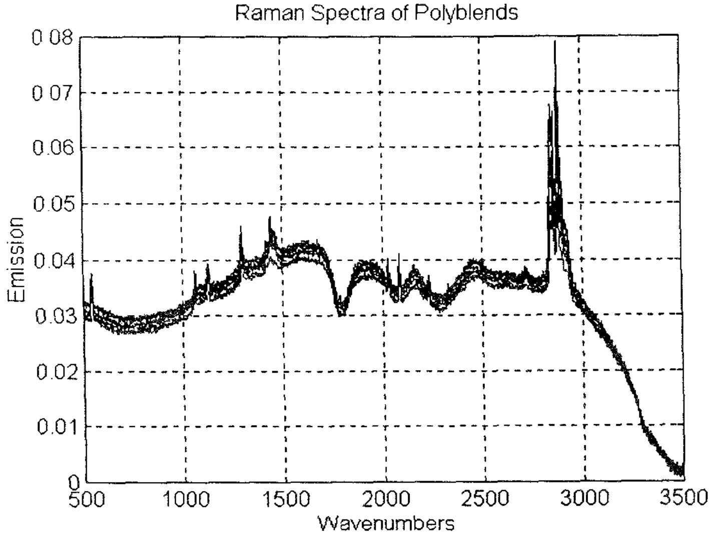

Figure 3. Raman Spectra of Polyblend Films

The learning (teaching) set composition is given graphically in Figure 11, and numerically in Table 4. A more complete learning set would be comprised of samples where intercorrelation (correlation coefficient) between analytes was minimized to ideally less than r = 0.2. Since this diversity in the sample set was not possible, the equations were tested using SEE (standard error of estimate) and RMSE (root mean square error for the regression) statistics to estimate the predictive performance under ideal conditions. Routine prediction of the film composition would involve measuring reference samples prior to measurement of the unknowns. The linear model would then be adjusted for slope and intercept to match the reference samples prior to prediction of the unknown(s).

EXPERIMENTAL

Film samples of 1.5 to 3.2 mils thickness were measured *as received* in transmission mode using a standard film holder for Raman, near infrared, and infrared measurements. The film samples were measured in duplicate by orienting the first sample with the MD (machine direction) of the film as horizontal, and then making a second measurement by rotating the film orientation by 90°; the second spectrum representing the CD (cross-wise direction as horizontal). The instruments used included a Bruker IFS 66 FT-NIR, a Bruker FRA 106 FT-Raman, and a Nicolet 510 FT-IR spectrophotometer; settings for recording the transmission spectra are shown in Table 5.

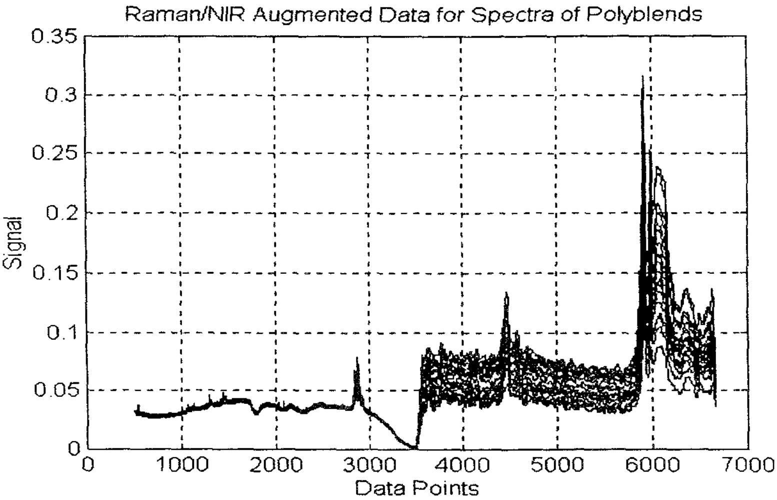

Figure 4a. Data Fusion between Raman and NIR

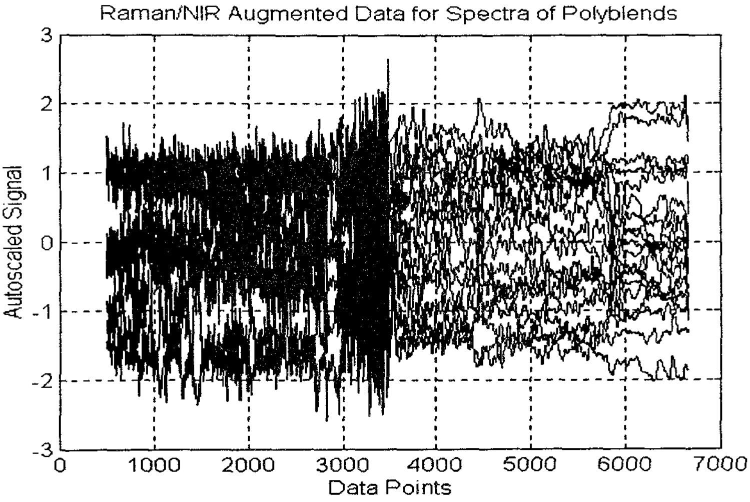

Figure 4b. Augmentation (Data Fusion) between Raman and NIR (Autoscaled)

Table 1. Results of LDPE Multivariate Modeling in order of best predictive model to worst.

Spectral Data Type	Math Preprocessing	Multivariate Method	Slope	Intercept	R2 [1]	SEE [2]	r [3]	RMSE [4]
Raman	Mean Centering	5 factor linear PLS	0.993	-64.549	0.993	1.376	0.997	65.015
NIR + Raman	Autoscaling + Augmentation	6 factor linear PLS	0.990	-64.372	0.990	1.622	0.995	65.020
NIR	Mean Centering	5 factor linear PLS	0.973	-63.237	0.973	2.694	0.986	65.057
NIR + Raman	Mean Centering + Augmentation	5 factor linear PLS	0.968	-62.924	0.968	2.916	0.984	65.068
NIR + Raman	Autoscaling + Augmentation	8 factor LWR3, local group = 6	0.883	8.312	0.914	4.490	0.956	4.941
IR	Mean Centering	5 factor linear PLS	0.884	-57.463	0.884	5.309	0.940	65.245

[1] R2 is the coefficient of determination

[2] SEE is the standard error of estimate for the regression

[3] r is the correlation coefficient for the regression

[4] RMSE is the root mean square error for the regression (uncorrected for bias)

Table 2. Results of LLDPE Multivariate Modeling in order of best predictive model to worst.

Spectral Data Type	Math Preprocessing	Multivariate Method	Slope	Intercept	R2 [1]	SEE [2]	r [3]	RMSE [4]
NIR + Raman	Autoscaling + Augmentation	6 factor linear PLS	0.981	-19.611	0.981	0.977	0.990	20.024
NIR + Raman	Mean Centering + Augmentation	5 factor linear PLS	0.978	-19.566	0.978	1.030	0.989	20.027
Raman	Mean Centering	5 factor linear PLS	0.974	-19.462	0.974	1.123	0.987	20.032
NIR	Mean Centering	5 factor linear PLS	0.973	-19.470	0.973	1.136	0.987	20.033
NIR + Raman	Autoscaling + Augmentation	8 factor LWR3, local group = 3	0.913	1.933	0.952	1.456	0.976	1.593
IR	Mean Centering	5 factor linear PLS	0.927	-18.549	0.927	1.834	0.963	20.090

[1] R2 is the coefficient of determination

[2] SEE is the standard error of estimate for the regression

[3] r is the correlation coefficient for the regression

[4] RMSE is the root mean square error for the regression (uncorrected for bias)

Table 3. Results of HDPE Multivariate Modeling in order of best predictive model to worst.

Spectral Data Type	Math Preprocessing	Multivariate Method	Slope	Intercept	R2 [1]	SEE [2]	r [3]	RMSE [4]
Raman	Mean Centering	5 factor linear PLS	0.999	-14.981	0.999	0.530	0.999	15.009
NIR + Raman	Autoscaling + Augmentation	6 factor linear PLS	0.995	-14.931	0.995	1.015	0.998	15.034
NIR + Raman	Mean Centering + Augmentation	5 factor linear PLS	0.977	-14.6502	0.977	2.264	0.988	15.174
NIR	Mean Centering	5 factor linear PLS	0.972	-14.584	0.972	2.462	0.986	15.206
IR	Mean Centering	5 factor linear PLS	0.885	-13.282	0.885	4.777	0.941	15.836
NIR + Raman	Autoscaling + Augmentation	8 factor LWR3, local group = 3	0.842	2.156	0.874	4.799	0.935	5.355

[1] R2 is the coefficient of determination

[2] SEE is the standard error of estimate for the regression

[3] r is the correlation coefficient for the regression

[4] RMSE is the root mean square error for the regression (uncorrected for bias)

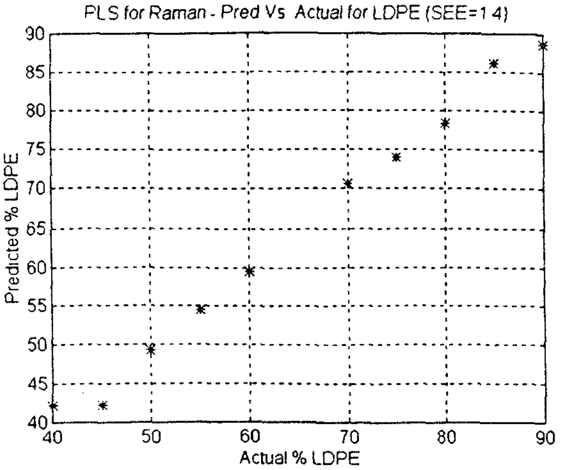

Figure 5a. Predicted Vs. Actual for % LDPE

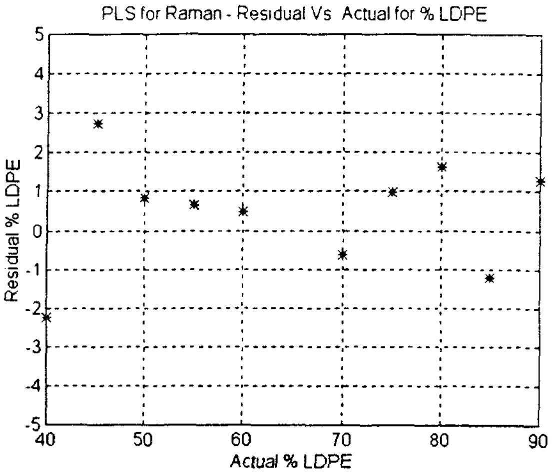

Figure 5b. Residual Vs. Actual for % LDPE

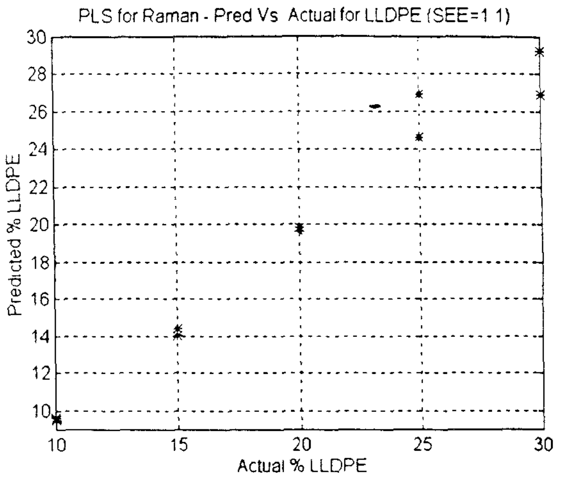

Figure 6a. Predicted Vs. Actual for % LLDPE

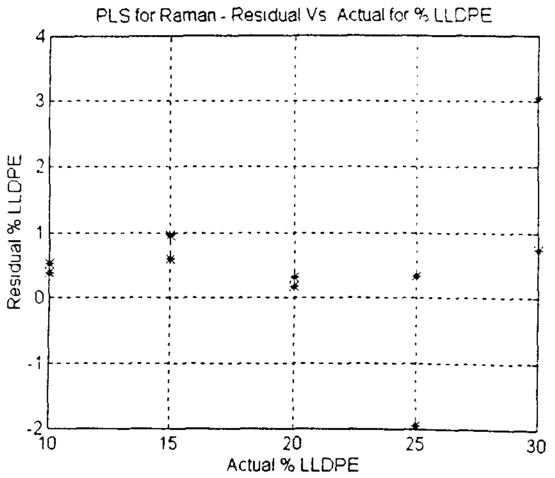

Figure 6b. Residual Vs. Actual for % LLDPE

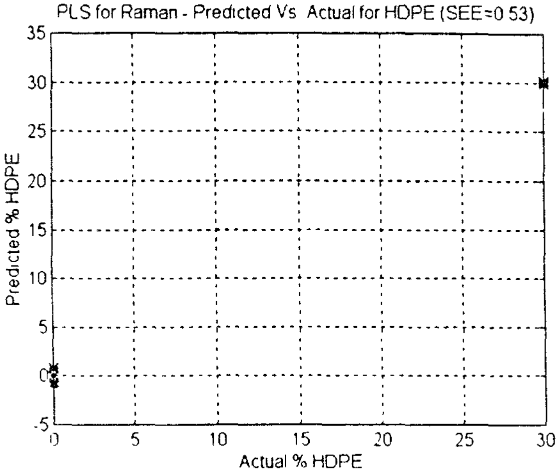

Figure 7a. Predicted Vs. Actual for % HDPE

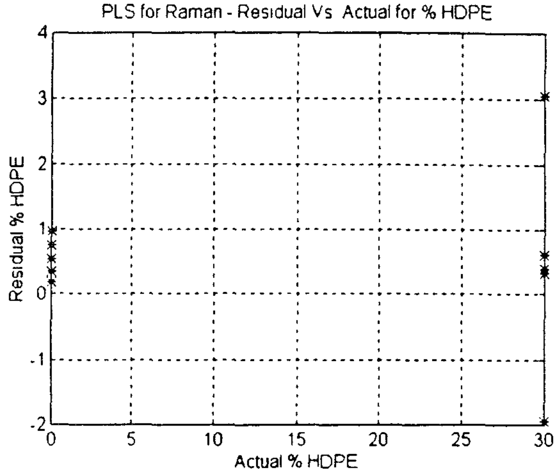

Figure 7b. Residual Vs. Actual for % HDPE

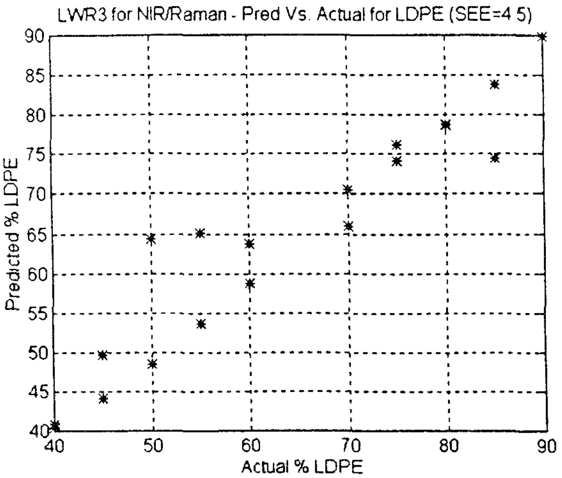

Figure 8a. Predicted Vs. Actual for % LDPE

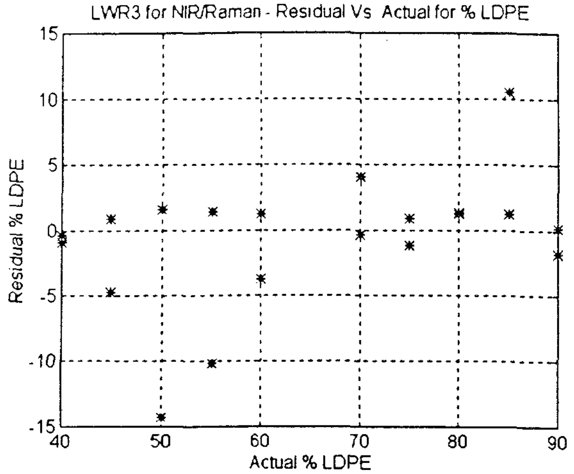

Figure 8b. Residual Vs. Actual for % LDPE

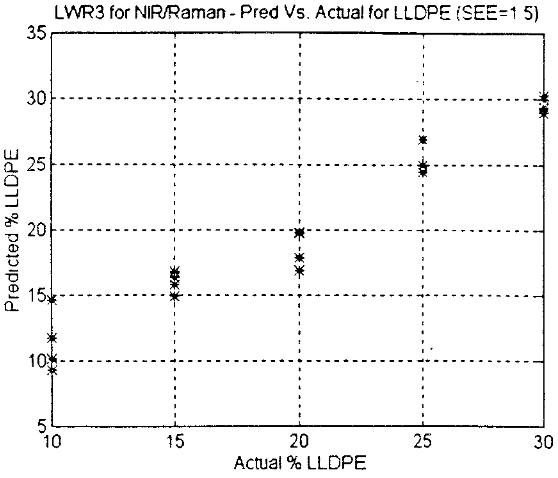

Figure 9a. Predicted Vs. Actual for % LLDPE

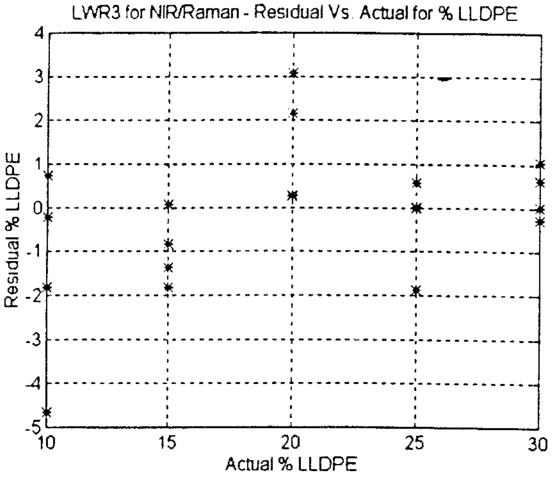

Figure 9b. Residual Vs. Actual for % LLDPE

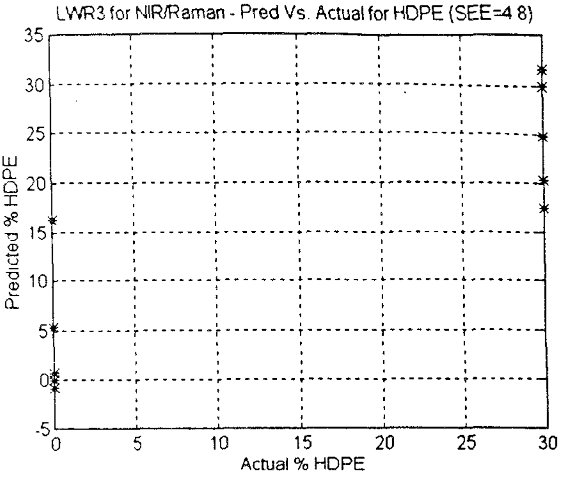

Figure 10a. Predicted Vs. Actual for % HDPE

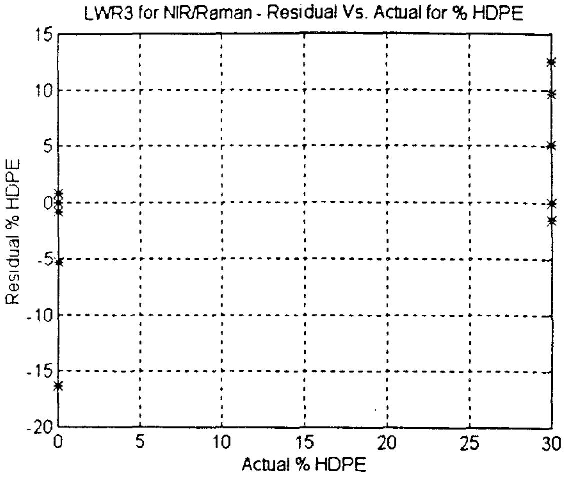

Figure 10b. Residual Vs. Actual for % HDPE

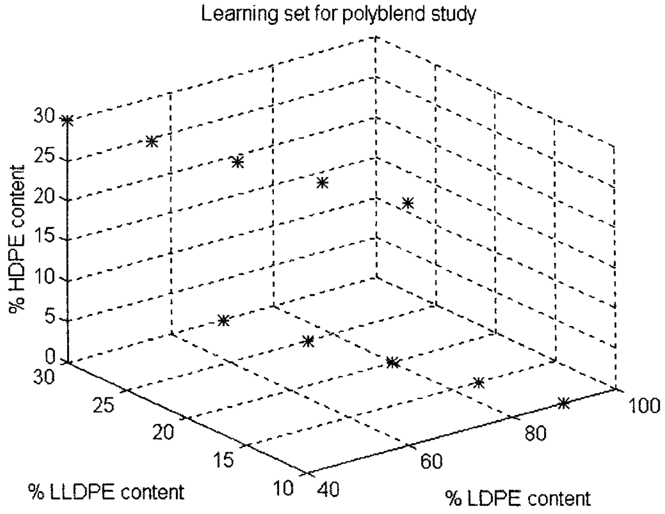

Figure 11. Three-dimensional illustration of percent film composition for the learning set samples (Each point represent a single film sample).

Table 4. Learning Set composition (percent component)

Sample Number	% LDPE	%LLDPE	%HDPE
1	90	10	0
2	85	15	0
3	80	20	0
4	75	25	0
5	70	30	0
6	60	10	30
7	55	15	30
8	50	20	30
9	45	25	30
10	40	30	30

Table 5. Instrument and data recording settings for polyblend film measurements

Spectrum Type	Frequency Range (cm-1)	Resolution (cm-1)	No. Co-added scans	No. of Data Points	Digital Precision	Background
IR (infrared)	5000 - 500	4	32	4096	32-bit	air
NIR (near infrared)	6667-3571	16	32	1301	32-bit	air
Raman	3500-500	4	500	1501	32-bit	N/A

The duplicate sample spectra for IR, NIR, and Raman measurements were converted to Grams/32 (*.spc) format, exported to ASCII XY (*.prn) format and imported into MATLAB™ 4.2 for chemometric modeling [3]. Modeling steps were performed using the Chemometrics Expert Toolbox (CPAC algorithms) [4] and the PLS_Toolbox (Eigenvector Research) [5].

RESULTS AND DISCUSSION

Regression methods such as MLR (multiple linear regression), PCR (linear principal components regression), PLSR (linear partial least squares regression), and LWR3 (locally weighted regression using PLS) [1,2] were tested on the learning set using a variety of mathematical preprocessing steps prior to calculation of the regression equations. A description of the mathematical preprocessing methods follows. *Autoscaling* is a mathematical step performed on a matrix **A** of spectra with rows as samples and columns as data channels, which produces a matrix **aA** that has columns with a mean of 0 and a variance of 1. *Mean Centering* is a function performed on a matrix **A** of spectra which calculates the mean value (in absorbance, transmission, or emission) for the column values (data channels) and subtracts this value from each of the data channels. In effect this is subtracting the mean spectrum from all of the other spectra in the learning set (see Figures 12a and 12b for illustration). *Data augmentation* is a procedure whereby spectra from different measurement techniques (i.e., NIR and Raman) are connected, fused or augmented (see Figures 4a and 4b). The resulting set of spectra or vectors can be mathematically preprocessed using mean centering or autoscaling, and the regression equation computed following this preprocessing step. Note: chromatograms can also be combined with spectra to increase the information content of the vectors to predict the sample chemistry. Data augmentation is used to increase information content in an analytical situation where a single measurement technique lacks this information content.

The final multivariate models were compared by selecting the model with the lowest SEE values. For the quantitative measurement of each of the components the Raman spectra, or Raman augmented with near infrared spectra, provided the best predictive equations for LDPE, LLDPE, and HDPE in polymer film blends. The SEE (standard error of estimate for the regression) and RMSE (root mean square error for the regression) are used to estimate the prediction error under ideal conditions for each spectroscopic and mathematics combination. The SEE Calculation of these parameters is given by

$$\mathrm{SEE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i + b_0)^2}{n-1}}$$

where y_i is the actual concentration of each sample (i); $\hat{y}_i$ is the predicted concentration of each sample (i); b_0 is the bias between the sum of the actual concentration values (all y_i) and the predicted concentration values (all $\hat{y}_i$); and n is the number of samples (i).

The RMSE (root mean square error) is the uncorrected (for bias) SEE value, given below. Note that all the symbols from the equation are the same as those used for the SEE equation above.

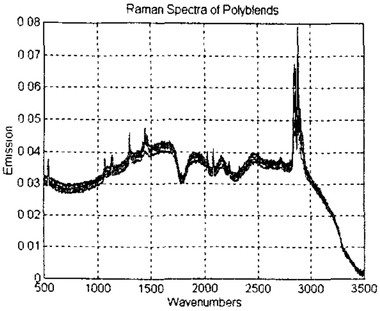

Figure 12a. Raman Emission Spectra of Polyblend Films

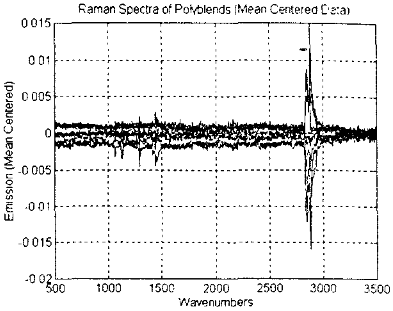

Figure 12b. Raman Emission Spectra of Polyblend Films (Mean Centered)

Table 6. Raman Spectra Data Channels (in cm-1) of Importance for Prediction of Composition using Step-Wise Multiple Linear Regression

	Data Channels	(cm-1) Selected (in	order of importance)
Data Processing	LDPE (R2=0.992)	LLDPE (R2=0.986)	HDPE (R2=0.980)
Mean Centered	3374	3324	3454
	3254	1062	3412
	3392	3466	2744
None	LDPE (R2=0.937)	LLDPE (R2=0.906)	HDPE (R2=0.909)
	3450	3500	3480
	3484	3464	3476
	3474	2880	3478

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-1}}$$

For LDPE, Raman and NIR + Raman provided SEE values of 1.376 and 1.622 percent LDPE, respectively (Table 1). For LLDPE, NIR + Raman resulted in SEE values of 0.977 and 0.989 percent LLDPE, depending upon the mathematical preprocessing steps used (Table 2). In the case of HDPE prediction, SEE results of 0.53 and 1.015 percent HDPE were obtained using Raman and NIR + Raman, respectively (Table 3).

As a test to determine which frequencies in the Raman spectra exhibited the greatest variance correlating with a change in composition Table 6 was computed. This table represents the data channels (frequencies) with the highest change in emission which is correlated to change in composition numbers for each of the LDPE, LLDPE, and HDPE. A comparison is made between the learning set spectra with no math preprocessing and for the same data, mean centered prior to computing the step-wise multiple linear regression. Note that the 3450 to 3484 wavenumber region seems to be quite significant for the prediction of constituent concentrations. This is unusual since there are no known Raman bands of particular importance relative to chemical structure in this region. The thickness or optical properties (i.e., reflectivity, scattering, etc.) was the probable cause in the selection of this spectral region.

REFERENCES

1. H. Martens, and T. Naes, Multivariate Calibration, John Wiley & Sons, Inc., New York (1989).

2. H. Mark, and J. Workman, Statistics in Spectroscopy, Academic Press, Boston (1991)

3. MATLAB™ Software, The Mathworks, Inc., 24 Prime Park Way, Nattick, Mass. 01760

4. Chemometrics m-scripts from the University of Washington, Laboratory for Chemometrics, Chemistry Dept., University of Washington, Seattle (1994).

5. PLS_Toolbox from Eigenvector Research, Inc., 830 Wapato Lake Road, Manson, WA 98831

Date Received: January 15, 1999
Date Accepted: August 20, 1999