Development of Calibration Models for Quality Control in the Production of Ethylene/Propylene Copolymers by FTIR Spectroscopy, Multivariate Statistical Tools, and Artificial Neural Networks

Emilio Marengo,[1] Valentina Longo,[1] Elisa Robotti,[1] Marco Bobba,[1] Fabio Gosetti,[1] Orfeo Zerbinati,[1] Silvana Di Martino[2]

[1]Department of Environmental and Life Sciences, University of Eastern Piedmont, 15100 Alessandria, Italy
[2]Polimeri Europa, Ferrara Plant, Piazzale G. Donegani 12, 44100 Ferrara, Italy

ABSTRACT: Principal component regression (PCR), partial least squares (PLS), StepWise ordinary least squares regression (OLS), and back-propagation artificial neural network (BP-ANN) are applied here for the determination of the propylene concentration of a set of 83 production samples of ethylene–propylene copolymers from their infrared spectra. The set of available samples was split into (a) a training set, for models calculation; (b) a test set, for selecting the correct number of latent variables in PCR and PLS and the end point of the training phase of BP-ANN; (c) a production set, for evaluating the predictive ability of the models. The predictive ability of the models is thus evaluated by genuine predictions. The model obtained by StepWise OLS turned out to be the best one, both in fitting and prediction. The study of the breakdown number of samples to be included in the training set showed that at least 52 experiments are necessary to build a reliable and predictive calibration model. It can be concluded that FTIR spectroscopy and OLS can be properly employed for monitoring the synthesis or the final product of ethylene–propylene copolymers, by predicting the concentration of propylene directly along the process line. © 2008 Wiley Periodicals, Inc. J Appl Polym Sci 109: 3975–3982, 2008

Key words: FTIR spectroscopy; back-propagation artificial neural network; principal component regression; partial least squares; ethylene–propylene copolymers; chemometrics

INTRODUCTION

Ethylene–propylene elastomers[1,2] constitute a family of amorphous products whose properties are due to the high flexibility of the macromolecular chain.

Ethylene–propylene rubbers and elastomers (called EPDM and EPM) continue to be one of the most widely used and fastest growing synthetic rubbers, having both specialty and general-purpose applications.

Ethylene–propylene rubbers are valuable for their excellent resistance to heat, oxidation, ozone, and weather aging because of their stable, saturated polymer backbone structure. Properly black pigmented and non-black compounds are color stable. As non-polar elastomers, they have good electrical resistivity as well as resistance to polar solvents, such as water, acids, alkalies, phosphate esters, and many ketones and alcohols. Amorphous or low crystalline grades have excellent low temperature flexibility with glass transition points of about −60°C.

The variation of composition and the varied distribution of the monomeric units imply different applicative characteristics: from here the necessity of gaining structural information related to the properties of the polymer under investigation.

EPM suppliers use different test methods to characterize polymers. The most widespread method to determine the molecular composition of EPM, and the only one recognized by the American Society of Testing of Materials (ASTM), is the infrared method ASTM D 3900.[3]

In this article, a procedure is proposed consisting in the calibration of infrared spectroscopic data using reference data obtained by ^{13}C NMR analysis.[4]

The final target of this work is the development of multivariate regression models relating the infrared spectrum of a selected type of polymer to propylene concentration in EPM copolymers.

In other words, the use of a different analytical approach usually employed in the NIR region has been evaluated for the determination of the monomer concentration using the entire infrared MIR range, without limiting the calculation to the ratio between infrared bands (classic approach of ASTM D 3900).
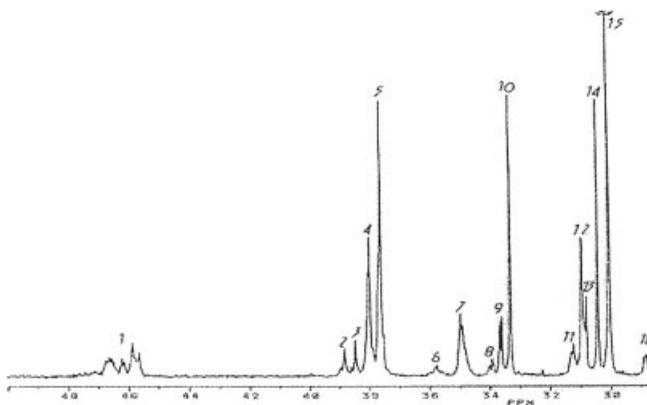
Figure 1 A typical spectrum ^{13}C NMR of a generic EPM.

Different "hard" regression methods were employed to calculate the propylene concentration: principal component regression (PCR),[5] partial least squares (PLS),[6–11] and StepWise regression (SWR) in forward search.[12,13]

Finally, the supervised back-propagation neural networks (BP-ANN)[14–16] were applied and the results were compared to those obtained by the previous methods. The possible application of ANNs was tested here to verify the presence of nonlinear effects that cannot be modeled through the other linear regression tools investigated.

This multivariate approach has already been successfully applied by our research group to other industrial problems.[17,18]

THEORY

Quantitative methods for monomer determination of EPM

Infrared spectroscopy

IR test method ASTM D 3900[3] covers the determination of the proportion of ethylene and propylene in copolymers and terpolymers over a range from 35 to 65% w/w of ethylene.

This test exploits the ratio between the absorbance of methyl groups from propylene units at 8.65 μm (1156 cm^{-1}) versus the absorbance of methylene sequences from ethylene units at 13.85 μm (722 cm^{-1}). A series of known EPM polymers (whose concentration is determined through ^{13}C NMR spectroscopy) is used to build a calibration curve of $A_{8.65}/A_{13.85}$ versus % w/w of ethylene:

$$\text{Ethylene concentration}(\%) = a - b \ \ln(A_{8.65}/A_{13.85})$$

where a and b are the regression coefficients to be calculated.

The % propylene is obtained as complement to 100.

The % ethylene necessary to build the calibration curve is obtained by NMR spectroscopic determinations carried out on a set of samples covering the overall range of concentration.

^{13}C NMR spectroscopy

^{13}C nuclear magnetic resonance spectrometry was then used to determine the molecular composition of a set of EPM standards: the NMR method[4] is the same used in ASTM D 3900 to determine the monomer concentration of the reference samples.

A typical spectrum ^{13}C NMR of an EPM is reported in Figure 1: the spectrum was recorded by a Bruker-Avance 300 NMR spectrometer (7.2 T). The polymer was dissolved in 1,1,2,2,-tetrachloroethane-d_2 ($C_2D_2Cl_4$). Chemical shifts are given relative to hexametyldisilane (0.037 ppm with respect to TMS). A delay time of 10 s between the 90° ^{13}C pulses was used; scans: 6600; spectral range: 12,000 Hz, temperature: 120°C.

The number written next to each spectral peak is related to the allocation of the chemical shift reported in Table I: primary, secondary, and tertiary carbons are labeled, respectively, P, S, and T; in addition, the two Greek letters in the subscripts define the position of the carbon relative to the nearest tertiary carbons in the chain. The method followed to determine the monomer concentration is based on the combination of the signal areas of primary and secondary carbons.

The % of ethylene and propylene are then computed as

$$\text{Mol}\,\%\,\text{ethylene} = \frac{(N_0 - N_1)}{(N_0 + N_1)} \times 100$$

$$\text{Mol}\,\%\,\text{propylene} = \frac{(2N_1)}{(N_0 + N_1)} \times 100$$

where

$N_0 = $ total number of methylenes
$\quad = (S_{\alpha\alpha} + S_{\alpha\beta} + 3S_{\beta\beta} + 2S_{\beta\gamma} + 5S_{\gamma\gamma} + 3S_{\gamma\beta} + S_{\delta\delta})$
$N_1 = $ total number of methyls $= (P_{\beta\beta} + P_{\beta\gamma} + P_{\gamma\gamma})$

PLS and PCR

PCR and PLS regression are multivariate statistical projection methods. Model components are extracted in such a way that the first principal component (PC) contains the largest amount of information, followed by the second PC, etc.

The optimal number of PCs modeling useful information but avoiding overfitting is determined with the help of the residual variances in prediction.

TABLE I
Chemical Shift of the ^{13}C NMR Peaks of Figure 1

Peak	Carbon type	Chemical shift	Sequence	Integration limits
1	$S_{\alpha\alpha}$	48.1–45.3	$S_{\alpha\alpha}$	48.50–44.50
2	$S_{\alpha\gamma}$	38.8	$r - S_{\alpha\gamma}$	
3	$S_{\alpha\delta}$	38.4	$r - S_{\alpha\delta}$	40.00–36.50
4	$S_{\alpha\gamma}$	37.96	$m - S_{\alpha\gamma} + m - S_{\alpha\gamma}$	
5	$S_{\alpha\delta}$	37.58	$m - S_{\alpha\delta} + m - S_{\alpha\delta}$	
6	$S_{\alpha\beta}$	35.7	$r - S_{\alpha\beta}$	36.20–34.30
7	$S_{\alpha\beta}$	34.9	$m - S_{\alpha\beta} + m - S_{\alpha\beta}$	
8	$T_{\gamma\gamma}$	33.9	$T_{\gamma\gamma}$	
9	$T_{\gamma\delta}$	33.6	$T_{\gamma\delta}$	34.29–32.80
10	$T_{\delta\delta}$	33.3	$T_{\delta\delta}$	
11	$T_{\beta\gamma}$	31.2	$T_{\beta\gamma} (m) + T_{\beta\gamma} (r)$	
12	$T_{\beta\delta}$	30.9	$T_{\beta\delta} (m)$	31.91–30.61
13	$S_{\gamma\gamma}$	30.8	$S_{\gamma\gamma} + T_{\beta\gamma} (r)$[a]	
14	$S_{\gamma\delta}$	30.4	$S_{\gamma\delta}$	30.61–30.23
15	$S_{\delta\delta}$	30.0	$S_{\delta\delta}$	30.23–29.32
16	$T_{\beta\beta}$	28.8–28.5	$T_{\beta\beta} (mm) + T_{\beta\beta} (mr+rr)$	29.15–28.22
17	$S_{\beta\gamma}$	27.85	$S_{\beta\gamma}$	28.22–27.63
18	$S_{\beta\delta}$	27.45–26.30	$S_{\beta\delta}$	27.63–26.63
19	$S_{\beta\beta}$	24.9	$S_{\beta\beta}$	25.60–23.95
20	$P_{\beta\beta}$	22.0–21.3	$P_{\beta\beta} (mm)$	
21	$P_{\beta\gamma}$	21.3–20.6	$P_{\beta\beta} (mr) + P_{\beta\gamma} (m) + P_{\beta\delta}(m)$	22.50–19.00
22	$P_{\gamma\gamma}$	20.6–19.5	$P_{\beta\beta}(rr) + P_{\beta\gamma} (r) + P_{\beta\delta} (r) + P_{\gamma\gamma}$	

[a] $S_{\gamma\gamma}$ is obtained by the formula $[(S_{\beta\delta} - P_{\gamma\delta})/2]$.

In PCR, a PC analysis is first performed on X, then a regression model is built relating the PCs on X to the Y variable. In PCR, therefore, the definition of X components is determined prior to the regression analysis, and the Y variables not playing a role at this stage.

PLS is a method where information in the original X-data is projected into a small number of underlying ("latent") variables called PLS components. The Y-data are actively used in estimating the "latent" variables to ensure that the first components are the most relevant ones for predicting the Y variables. Interpretation of the relationship between X-data and Y-data is then simplified as this relationship in focused on the smallest possible number of components.

StepWise ordinary least squares regression

SWR is the more exploited method to select a small number of variables from the original data. Here, the forward selection (FS) approach was used: it starts with a model where no variables are included and gradually adds a variable at a time until a determined criterion of arrest of the procedure is satisfied. The variable being included in the model in each step is the one providing the greatest value of the iterative F-Fisher ratio; in other words, the jth variable is included in the model with p variables already included if

$$F_j^+ = \max_j \left[\frac{RSS_p - RSS_{p+j}}{S^2_{p+j}} \right] > F_{in}$$

where, S^2_{p+j} is the variance calculated for the model containing p variables plus variable j; $RSS_p = \sum_p (y_p - \hat{y}_p)^2$ is the residual sum of squares of the model with p variables. RSS_{p+j} is the residual sum of squares of the model with p variables plus variable j.

The F value calculated is compared to a reference value (F_{in}), usually set at values ranging from 1 to 4 : 1 corresponding to a more permissive selection, including in the final model a larger number of variables, while 4 corresponds to a more severe selection.

Artificial neural network

Artificial neural networks are mathematical algorithms that can be used to solve complex problems by simulating the function of the human brain. They are mainly dedicated to modeling the behavior of complex systems, where they usually provide better results than ordinary least squares (OLS), especially when nonlinear relationships are present.

Supervised BP-ANN build models, classify patterns, and make predictions according to patterns of input/output they have learned.

The back-propagation network is the most popular ANN used for calibration; it consists of

- An input layer, where each neuron is associated to an experimental variable (in this case the wavelengths of the infrared spectrum);
- One or more hidden layers with a variable number of neurons;
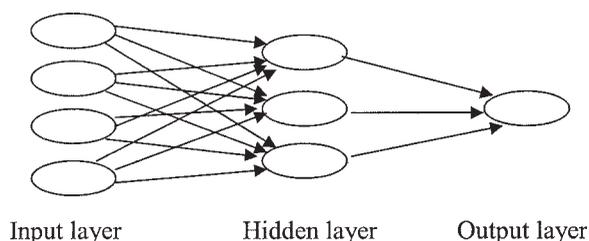
Figure 2 Scheme of a typical BP-ANN.

- An output layer, where each neuron is associated to a response (in this case the propylene concentration).

In the feed forward back-propagation training algorithm, the signal moves from the input layer toward the output layer (Fig. 2). In this process, each neuron uploads all the neurons of the following layer, transferring a portion of the value (input) it has accumulated. The portion of signal that is transferred is regulated by a transfer function.

The choice of the network architecture is very important because it determines the ability of the network to predict unknown responses; it consists in selecting the number of hidden layers, the number of neurons in each hidden layer, the connection pattern of each neuron and the specific transfer functions to be applied between the different neuronal layers (here, logistic transfer function).

Since a high number of original variables contrasts with the use of ANNs (each variable corresponds to an input neuron), it was necessary to reduce the size of the spectral dataset. This was achieved here through a smoothing procedure with step 10. This transformation reduces the number of wavelengths to 236.

The back propagation algorithm attempts to minimize the difference (or error) between the desired and actual output according to an iterative procedure. For each iteration, the initial weights involved in the network are adjusted by the algorithm, so that the error is decreased along a descending direction. Two parameters, called learning rate (set here at 0.3) and momentum (set here at 0.3), are used for controlling the size of weight adjustment along the descending direction and for dampening oscillations of the iterations.

The secret of building successful neural networks is to know when to stop the training phase. In fact if the net is trained for a too short time it will not learn the data patterns while if the net is trained for too long, it will learn the noise and memorize the data by heart (overfitting). To solve the problem of overfitting, the validation procedure adopted is fundamental. Here, the dataset was split in three sets:

- Training set (48 samples), used for training the network (model building);

- Test set (20 samples), used for selecting the end of the training phase;
- Production set (15 samples), used for testing the real predictive ability of the network.

The samples assigned to the test set were selected from the training set so that they represent the entire experimental domain: low, medium, and high concentration of propylene.

Index for estimating fitting and prediction ability

The fitting ability of artificial neural models and regression models was evaluated by the coefficient of multiple determination, R^2, calculated as

$$R^2 = 1 - \frac{\sum\limits_{i=1,n} (\hat{y}_i - y_i)^2}{\sum\limits_{i=1,n} (y_i - \bar{y})^2}$$

where

n = number of samples of the training set
$\hat{y}_i$ = predicted value of the response of the ith experiment
y_i = experimental value of the ith experiment
$\bar{y}$ = average response of n experiments.

This expression can be calculated using either the experiments of the training set, R^2(tr), leading to the classical coefficient of multiple determination that gives information on the model fitting ability, or the experiments of the production set, R^2(pro), which is very effective for evaluating the predictive ability of the model.

The root mean square error (RMSE) between the measured and predicted values is estimated as

$$\text{RMSE} = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{n}}$$

This parameter again can be calculated using both the training (RMSEF, RMSE of fitting) or the production (RMSEP, RMSE of prediction) set of experiments, to achieve information about fitting and predictive ability, respectively.

EXPERIMENTAL

Dataset

The copolymers examined in this work are statistic or random copolymers, i.e., polymeric chains in which the monomeric units are randomly distributed; these polymers are completely saturated.

The copolymers dataset used for regression models, constituted by 83 samples obtained directly
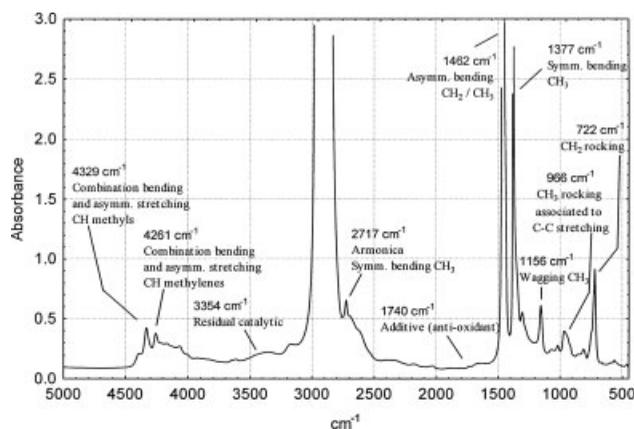
Figure 3 Example of spectrum IR of a generic copolymer ethylene–propylene.

from the production plant, was divided into two groups:

Sixty-eight samples in the training set (for the construction of the model);
Eight subgroups with 10 samples in the test set every time (systematic iterative procedure);
Fifteen samples in the production set (to calculate the predictive ability of the model).
The copolymer samples correspond to the products obtained from the production in ''Polimeri Europa'' company (Ferrara plant, Italy).

The 83 samples uniformly cover a range of propylene concentrations from 29 to 53% w/w.

IR and NMR spectroscopic determination of monomer concentrations

For each sample the infrared spectrum in absorbance[19,20] was recorded from 5000 to 450 cm^{-1} (2360 wavelengths).

IR spectra were recorded by a Perkin–Elmer FT-Spectrum One System: 16 scans, resolution of 4

TABLE II
% Cumulative Explained Variance in Y of Each Principal Component in PCR and PLS Model

Principal components	Y − % cumulative variance	
	PCR	PLS
PC1	0.13	67.91
PC2	48.70	79.28
PC3	50.86	95.90
PC4	51.63	96.98
PC5	91.36	97.91
PC6	93.94	98.76
PC7	94.75	99.15
PC8	95.92	99.34
PC9	96.35	99.56
PC10	96.49	99.68

TABLE III
Summary of R^2 and RMSE for the Concentration of Propylene with PCR and PLS Models

	R^2(tr)	R^2(pro)	RMSEF	RMSEP
PCR	0.94	0.96	1.67	1.61
PLS	0.98	0.96	0.81	1.43

cm^{-1}, using a deuterated triglycine sulfate detector. The absorbance infrared spectrum of a generic ethylene–propylene copolymer is shown in Figure 3. The monomer concentration of the standard polymers was determined through a ^{13}C NMR[4] method based on the combination of spectral peaks. The NMR determinations were carried out by a Bruker Avance-300 spectrometer, collecting 6600 scans at 120°C with an angle of impulse of 90° using deuterated tetrachloride–ethane ($C_2D_2Cl_4$) as solvent.

RESULTS AND DISCUSSION

PCR and PLS

PCR and PLS were applied to the final data matrix of size 83 × 2360 (83 being the samples and 2360 being the spectroscopic variables).
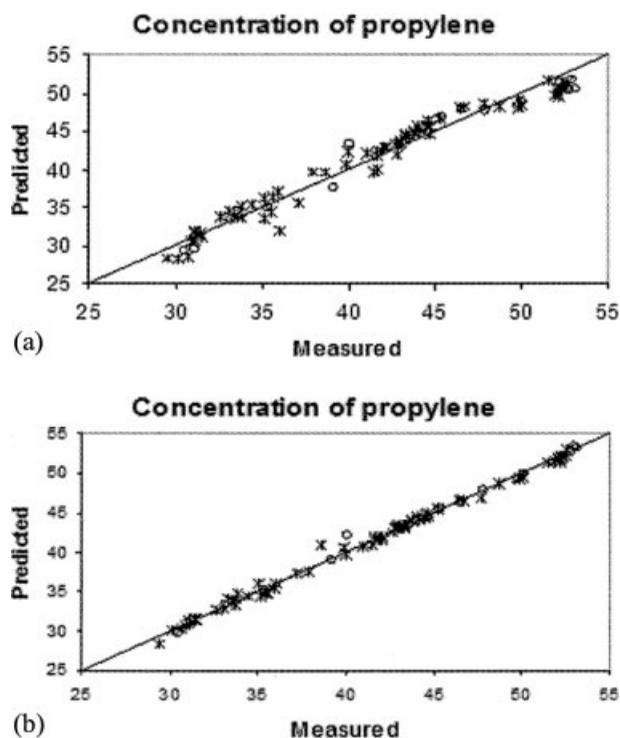
Figure 4 PCR (a) and PLS (b) results: predicted versus experimental concentration of propylene for training set (*) and for production set (○). The straight lines represent the target perfect accordance between predicted and experimental responses. PLS shows better results both in fitting and prediction, since both training and production set points lay almost along the straight line.

Table II reports the % of cumulative explained variance for PCR and PLS, respectively.

The first 7 PCs were retained as significant for PCR and the first 6 PCs for PLS. They explain, respectively, 94.75% and 98.76% of total variance.

Table III reports the R^2 and RMSE values obtained, both in fitting and prediction for the regression procedures adopted. It can be noticed that PCR and PLS models perform satisfactorily both in fitting and prediction even if PLS guarantees the best results. Figure 4 represents the predicted versus experimental values for the samples of the training set (represented as asterisks) and for those of the production set (represented as circles) for PCR [Fig. 4(a)] and PLS models [Fig. 4(b)], respectively. Figure 4 shows that PLS [Fig. 4(b)] performs better than PCR, both in fitting and prediction: for what regards PLS in fact the samples lay almost along a line, while more significant deviations from linearity can be noticed for PCR.

PLS thus performs better both in fitting and in prediction, when compared to PCR.

Figure 5 reports the regression coefficients (expressing the relationship between variation in the predictors and in the response) in the PCR (- - - line) and PLS models (— line). The coefficients show the relative importance of the X variables in the model calculated by only the relevant PCs. A positive coefficient shows a positive correlation with the response (increasing the value of that variable, the response increases), and a negative coefficient shows a negative correlation (increasing the value of that variable, the response decreases). Predictors with a coefficient around 0 are negligible.

The trends of the regression coefficients are very similar in PCR and PLS.

Figure 5 shows that almost the same spectral regions exhibit large coefficients on both PCR and PLS models, as expected. An increase of the spectral regions at 722 and 1500 cm^{-1} corresponds to a decrease of propylene concentration, while an

Figure 6 Stepwise OLS results (F_{in} = 4): predicted versus experimental concentration of propylene for production set (○). The straight line represents the target perfect accordance between predicted and experimental responses. For what regards prediction, OLS performs better than PCR and PLS, since the objects of the production set lay better along the straight line.

increase of the bands at 1371 cm^{-1} corresponds to an increase of propylene concentration.

It is important to point out that PLS and PCR show that more spectral regions are important for the determination of the propylene concentration of copolymers respect to the method ASTM D 3900.

Stepwise OLS regression

A SWR was then carried out on the original variables and compared to the results obtained by PCR and PLS. The StepWise procedure was applied with a FS. The analysis gave effected with different F_{in} (from 3 to 10) but all gave the same results. Here, the results of the test with F_{in} = 4 are shown.

Figure 5 Plot of regression coefficients in PCR and PLS.

TABLE IV
Summary of R^2 and RMSE for the Concentration of Propylene of BP-ANNs with Different Number of Neurons of Hidden Layer

236 neurons input layer; 1 neuron output layer; logistic transfer function	R^2(tr)	R^2(pro)	RMSEF	RMSEP
2 neurons; hidden layer	0.96	0.92	1.42	2.21
3 neurons; hidden layer	0.99	0.94	0.62	1.80
4 neurons; hidden layer	0.99	0.98	0.47	0.99
5 neurons; hidden layer	0.99	0.95	0.51	1.74
6 neurons; hidden layer	0.99	0.96	0.54	1.52
7 neurons; hidden layer	0.99	0.97	0.48	1.31
8 neurons; hidden layer	0.99	0.96	0.53	1.51
9 neurons; hidden layer	0.99	0.94	0.68	1.82
10 neurons; hidden layer	0.99	0.97	0.57	1.39

The procedure applied selected an optimal subset of only six original variables at $\sim$ 709, 941, 960, 1384, 2561, and 2830 cm^{-1}.

The experimental versus predicted concentration of propylene for the production set of the StepWise model with F_{in} = 4 are shown in Figure 6.

The coefficients of multiple determination for the training set and the production set with SWR are, respectively, 0.99 and 0.99 while RMSEF = 0.66 and RMSEP = 0.81.

In this case, SWR performed significantly better than PCR and PLS, in fact improvements are observed both in fitting and predictive ability.

Back-propagation network

Artificial neural networks[21–28] were applied to the dataset of copolymers to verify the possible existence of nonlinear effects that cannot be accounted for by linear methods as PCR, PLS, and StepWise OLS.

Different architectures were tried here changing number of neurons in the hidden layer. Table IV reports the R^2 and RMSE values obtained, both in fitting and prediction of BP-ANNs with different number of neurons in the hidden layer. It can be noticed that the best results were obtained selecting one hidden layer containing four neurons.

The experimental versus predicted concentration of propylene of the production set for the best network is shown in Figure 7.

For what regards the R^2 values, the BP-ANN model performs better, both in fitting and prediction with respect to PCR and PLS models.

TABLE V
Summary of R^2 and RMSE for the Concentration of Propylene with all the Regression Procedures Adopted

	R^2 (tr)	R^2(pro)	RMSEF	RMSEP
PCR	0.94	0.96	1.67	1.61
PLS	0.98	0.96	0.81	1.43
StepWise OLS	0.99	0.99	0.66	0.81
BP-ANN	0.99	0.98	0.47	0.99

The BP-ANN with respect to Stepwise OLS does not succeed to improve the error committed neither in fitting (RMSEF = 0.47) nor in prediction (RMSEP = 0.99).

Table V reports the R^2 and RMSE values obtained both in fitting and prediction for all multivariate statistical tools adopted.

The results obtained prove that StepWise OLS is the best method for this specific application and provides very good results in particular for what regards its predictive ability (it guarantees lower RMSEs lower).

Breakdown number of samples for model calculation

Finally, the stability of the best calibration model so far obtained as a function of the number of samples of the training set was investigated. The "breakdown number of samples" means how many samples in the training set are necessary for attaining a calibration model that is able to explain the information in the data and predict correctly the value of the response (propylene concentration) of the samples in the production set. This was evaluated by gradually reducing the number of samples in the training set; in particular, groups of eight samples were eliminated every time. The samples removed for each cycle were selected in such a way that all the experimental domains were represented in the training set (low, medium, and high propylene concentration).

The trend of the coefficient of multiple determination [Fig. 8(a)] and RMSE [Fig. 8(b)] of the training and the production sets as a function of the number of groups of samples eliminated from the training set are shown in Figure 8.

Figure 8(a) shows that the model predictive ability (R^2 (pro)) decreases progressively, as expected, with the reduction of the number of samples present in the training set. In particular, a model built with 52 samples loses the ability to generalize and to predict satisfactorily (high RMSEP, Fig. 8(b)] when compared to a model with 68 samples in the training set. This means that a higher number of samples is necessary to be able to predict the propylene concentration of unknown samples with Step Wise OLS.

Figure 7 BP-ANN results: predicted versus experimental concentration of propylene for production set (○). The straight line represents the target perfect accordance between predicted and experimental responses. BP-ANN shows better results for the production set if compared to PCR and PLS models, since the points are better distributed along the straight line. The results are however comparable to OLS model.

Figure 8 Trend of R^2 for training set and production set (a) and trend of RMSE for training and production set (b).

CONCLUSIONS

The final target of this work was the development of multivariate regression models relating the infrared spectrum of ethylene/propylene copolymers to the propylene concentration in the macromolecule.

It can be asserted that it is possible to take advantage of FT-IR spectroscopy for the determination of the concentration of propylene in copolymers.

Usually calibration methods on copolymers involve the use of NIR spectroscopy, but in this particular case, very good results were obtained using the complete MIR spectral range.

The model obtained with StepWise OLS, Table V, regression turned out to be the best one is in fitting is in prediction (R^2(pro) = 0.99 and RMSEP = 0.81), provided sufficient samples are available to calculate a stable and reliable model.

The ANN, trained through the back-propagation algorithm, does not provide better results compared to StepWise OLS. Moreover, ANNs require relatively long calculation times.

References

1. El-Sabbagh, S. H. Polym Test 2003, 22, 93.
2. Ciampelli, F.; Bucci, G.; Simonazzi, T.; Santambrogio, A.; La Chimica e l'industria 1962, 44, 489.
3. ASTM D 3900–95. Determination of ethylene units in EPM (ethylene–propylene copolymers) and EPDM (ethylene-propyl-ene-diene terpolymers). American Society for testing and materials, Pennsylvania; p 629.
4. Di Martino, S.; Kelchtermans, M. J Appl Polym Sci 1995, 56, 1781.
5. Davies, A. M. C. Spectrosc Eur 1995, 7, 36.
6. Bastien, P.; Vinzi, V.; Tenenhaus, M. Comput Statist Data Anal 2005, 48, 17.
7. Luinge, H. J.; Van der Maas, J. H.; Visser, T. Chemom Intell Lab Syst 1995, 28, 129.
8. Geladi, P.; Kowalski, B. Anal Chim Acta 1986, 185, 1.
9. Wold, H. O. A. In Encyclopaedia of the Statistical Sciences, Vol. 6; Johnson, N. L., Kotz, S., Eds.; Wiley: New York, 1988; pp 581–891.
10. Luinge, H. J.; van der Maas, J. H.; Visser, T. Chemom Intell Lab Syst 1995, 28, 129.
11. Marengo, E.; Todeschini, R. Chemom Intell Lab Syst 1992, 12, 117.
12. Burkholder, T. J.; Lieber, R. L. J Biomech 1996, 29, 235.
13. Steyerberg, E. W.; Eijkemans, M. J. C.; Habbema, J.; Dik, F. J Clin Epidemiol 1999, 52, 935.
14. Mazzetti, A. Reti Neurali Artificiali, Apogeo, Italy, 1991.
15. Hopfield, J. Neural Networks and Physical System; MIT Press: Cambridge, MA, 1982.
16. Zupan, J.; Gasteiger, J. Neural Network for Chemist: An Intro-duction; VCH: Weinheim, 1993.
17. Marengo, E.; Soave, C.; Gennaro, M. C.; Robotti, E.; Bobba, M.; Lenti, M. Ann di Chim (Rome) 2004, 94, 219.
18. Marengo, E.; Bobba, M.; Robotti, E.; Lenti, M. Anal Chim Acta 2004, 511, 313.
19. Celina, M.; Ottesen, D. K.; Gillen, K. T.; Clough, R. L. Polym Degrad Stab 1997, 58, 15.
20. Wilhelm, P. Micron 1996, 27, 341.
21. Wang, Y. Chin Sci Abstracts Ser B 1995, 14, 16.
22. Robb, E. W.; Munk, M. E. Mikrochim Acta (Wien) 1990, 1, 131.
23. Tanabe, K.; Tamura, T.; Uesaka, H. Appl Spectrosc 1992, 46, 807.
24. Smits, J.; Melssen, W. J.; Buydens, L. M. C.; Kateman, G. Che-mom Intell Lab Syst 1994, 22, 165.
25. Garson, G. D. AIExpert 1991, 6, 45.
26. Cirovic, A. D. Trends Anal Chem 1997, 16, 148.
27. Aleksander, I. Introduction to Neural Computing; Thompson International Press: London, 1995.
28. Smith, M. Neural Networks for Statistical Modelling; Van Nos-trand Reinhold: New York, 1993.