Monitoring of the acid catalysed esterification of ethanol by acetic acid using Raman spectroscopy†

Richmond J. Ampiah-Bonney and Anthony D. Walmsley*

Faculty of Science and the Environment, Department of Chemistry, University of Hull, Cottingham Road, Hull, UK HU6 7RX. E-mail: A.D.Walmsley@chem.hull.ac.uk; Fax: +44 1482 466416; Tel: +44 1482 465470

Received 6th September 1999, Accepted 19th October 1999

The acid catalysed esterification of ethanol by acetic acid has been monitored using Raman spectroscopy. It was found that noise and baseline shift due to fluorescence spectra and process dynamics obscured the essential chemical information in the data. Principal components analysis and regression analyses have been applied progressively to de-noise the data and to extract the Raman signal. It was found that the first principal component (PC1) contained the fluorescence, while the second principal component (PC2) contained the pure Raman spectra. It was then possible to reconstruct the original data using only the second principal component, and then use this reconstructed data to follow the progress of the reaction. To be sure that the data in PC2 represented the real chemical process, the kinetics of the reaction were calculated using selected wavelength profiles in PC2. It was shown that the reaction follows theoretical second order kinetics, which is expected of this esterification.

Introduction

The need for better process control of industrial batch processes is important, since batch processes are finding increasing applications in the chemical industry, pharmaceuticals, bio-technical industry and the brewery industry among many others.[1] Their advantage is their ability to produce high-value products within short manufacturing times. Moreover, the procedures in batch processes are relatively very simple; basically, the reactants are loaded into the reaction vessel, processed under controlled conditions and then the completed product is discharged.[2]

The variation in products from batch to batch needs to be minimised as much as possible. Many methods are used, ranging from basic 'knob-twisting' methods for controlling reaction conditions and feed rates to advanced programmable controllers. Batch reactions are complex in nature due to their finite duration, inconsistency in homogeneity and multi-component nature. Thus, even with extensive automation, control of a batch process is very challenging. Industry is rapidly moving away from post-production investigation due to its wastefulness both of resources and time.[3]

A more pro-active approach is to create a model that is capable of predicting chemical changes within the reaction vessel, with a view to controlling the process from a remote position. The initial stage of this work is the continuous monitoring of the process in order to acquire information on all stages and also to follow all ongoing chemical processes. This calls for the use of an *in situ* probe that conveys data to a spectrometer. Such data, when analysed with chemometric techniques, offer information vital to the modelling step.

Raman spectroscopy

Spectroscopic techniques are very useful in process analysis,[4] the most widely reported being vibrational spectroscopy.[5] When photons in a beam of electromagnetic radiation impinge inelastically on a molecule, the energy of the molecule changes by an amount ΔE_m characteristic of the molecule, the Raman effect.[6]

The use of Raman spectroscopy in industrial process control is relatively recent and rapidly increasing.[7–9] This is because Raman is more advantageous in many respects than many other conventional methods. Peaks in Raman spectra are abundant, well resolved and provide direct and clear chemical information, since they correspond to fundamental transitions.[10] This is an advantage over the more commonly used near-infrared (NIR) spectroscopy. Moreover, the light scattering nature of the Raman process allows simple, effective and stable fibre probe designs as compared to those needed for near- and mid-IR absorbance spectroscopy.[11]

The drawback with this technique is that, in addition to the Raman spectra, fluorescence spectra also arise, and this causes interferences. However, whereas Raman spectra always have multiple regions of uncorrelated spectral data, the same is not true for fluorescence spectra.[12] The presence of fluorescent reaction components and the occurrence of extra reflecting surfaces, like bubbles, enhance the occurrence of fluorescence spectra in Raman spectra. In fact, the fluorescence contribution can be so large as to render Raman data interpretation extremely difficult. Fluorescence can be minimised by a careful choice of the wavelength of the laser source for the Raman spectrometer, since the Raman scattering intensity decreases with laser wavelength, whereas fluorescence becomes troublesome at longer laser wavelengths.[7]

Principal components analysis (PCA)

The main goal of PCA is to reduce the size of a data set that has a large number of intercorrelated variables, and to retain as much of the information present in the original data as possible. When it is used for modelling, PCA is further applied to determine the minimum dimensionality needed to reproduce the original information within experimental measurement error.[12]

PCA reduces the spectral data (X) into principal component scores (T) and loadings ($\boldsymbol{P}$), according to the equation $\boldsymbol{X} = \boldsymbol{TP}^t + \boldsymbol{E}$, a linear transformation, where E is the X-residual matrix. The principal component scores are uncorrelated and are so

† Presented at R&D Topics, London, UK, April 12–14, 1999.

ordered that the first few retain most of the variation present in all of the original variables. Thus only a few of the transformed variables are needed in further procedures.[5] The correlation or the covariance matrix of the variables is decomposed into eigenvectors, each with an associated eigenvalue. The matrix of eigenvectors, called the loadings, contains information on how the variables relate to each other, while the scores give information on how the samples relate to each other. PCA is useful in visualising and analysing large data flows.[13]

In this work, PCA has been used to separate the true Raman spectra from the fluorescence spectra and other unwanted contributions in the data.

Kinetics

The reaction of interest, the sulfuric acid catalysed esterification of ethanol by acetic acid, is old and well established. The reaction is represented by the following scheme:

$$CH_3COOH + CH_3CH_2OH \underset{H^+}{\overset{H^+}{\leftrightarrow}} CH_3COOCH_2CH_3 + H_2O$$

The reaction goes through a five-step $A_{AC}2$ mechanism with known intermediates,[14] as follows:

$$RCOOH + H^+ \underset{k_{-1}}{\overset{k_1}{\leftrightarrow}} RC^+OOH_2 \quad (1)$$

$$RC^+OOH_2 + R'OH \underset{k_{-2}}{\overset{k_2}{\leftrightarrow}} RC(OH)_2OHR' \quad (2)$$

$$RC(OH)_2OHR' \underset{k_{-3}}{\overset{k_3}{\leftrightarrow}} RC(OH)(O^+H)_2OR' \quad (3)$$

$$RC(OH)(O^+H)_2OR' \underset{k_{-4}}{\overset{k_4}{\leftrightarrow}} RC^+(OH)OR' + H_2O \quad (4)$$

$$RC^+(OH)OR' \underset{k_{-5}}{\overset{k_5}{\leftrightarrow}} RCOOR' + H^+ \quad (5)$$

where R = CH_3 and R′ = CH_2CH_3. Step 4, which involves the production or consumption of a molecule of water, is the rate-determining step.

Generally, the reaction progress must show increasing product content and decreasing reactant content. For a second order reaction, the reciprocal of the amount of a reaction component must be directly proportional to the time. For this particular reaction therefore, the integrated form of its rate equation will be

$$\frac{1}{A_0 - B_0} \ln \frac{B_0 A}{A_0 B} = kt$$

where A_0 and B_0 are the initial concentrations of ethanol and acetic acid, A and B are the respective concentrations at time t and k is the overall rate constant, *i.e.*

$$k = \frac{k_1 k_2 k_3 k_4 k_5}{k_{-1} k_{-2} k_{-3} k_{-4} k_{-5}}$$

It follows that a plot of $1/(A_0 - B_0)$ *versus* t should be linear. The linearity of this plot is taken as proof of the order of the reaction.[15]

Experimental

Glacial acetic acid (AR grade) ethanol, and concentrated sulfuric acid (from Fisher Scientific, Loughborough, UK) are the starting materials. The reaction was performed in a l l reaction vessel set in a borosilicate glass water bath heated by a thermostated hotplate and fitted with a condenser and a glass stirrer driven by an electric powered motor. A model IMO/H0 Raman probe connected to an HL5-785-250 Kaiser optical system and a Hololab Series 5000 Raman spectrometer was fitted at one of the probe ports. The Raman probe uses a class 3B-focus laser. It was then connected to a computer with HOLOGRAM software that controls the spectrometer and probe as well as receiving and storing data loaded on it (Fig. 1).

Ethanol (300 ml) was poured into the reaction vessel, followed by 300 ml acetic acid. These volumes were used so that at least 60% of the length of the probe would be submerged in the liquid. The reactants were slowly brought to 78 °C while stirring. The reaction was initiated by the addition of 15 ml concentrated H_2SO_4 (catalyst). Data processing was done with the software MATLAB 5.2 (The Math Works Inc., Natick, MA, USA) and PLS_Toolbox 2.0 (Eigenvector Research, Inc., Manson, WA, USA). All data used for PCA were mean-centred. Smoothing of some of the spectral profiles was done using a 15-point Savitsky–Golay smoothing.

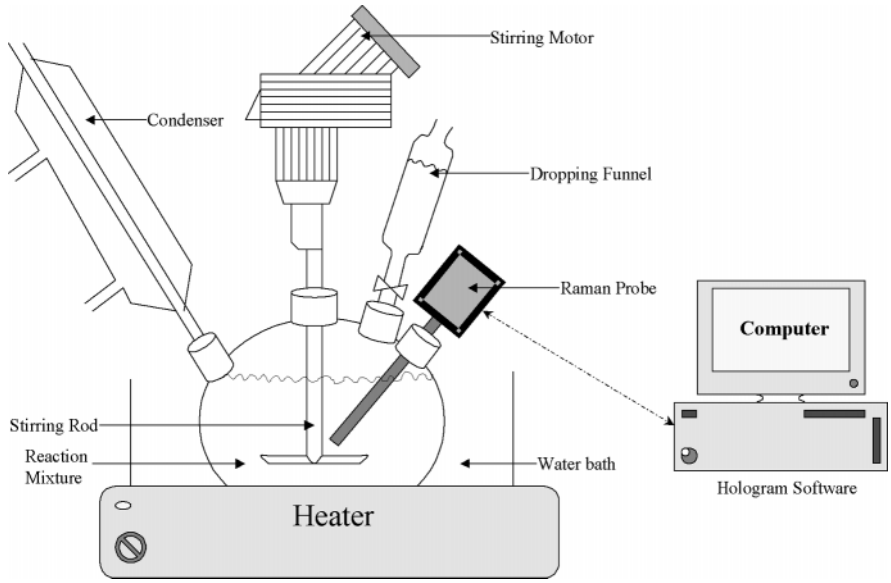

Fig. 1 Experimental set up showing the configuration of the Raman probe and other accessories in the 1 l vessel.

Results

Initially, the Raman spectra of the Raman-sensitive components of the reaction, *i.e.* ethyl acetate, ethanol and acetic acid, were obtained separately. These were used to generate simulated spectra, which act as an artificial representation of the reaction mixture. A plot of the three pure spectra and the simulated spectrum within a wavelength window of 400–1000 cm^{-1} is shown in Fig. 2.

Comparison of each of the pure spectra led to the assignment of the peaks in the reaction mixture to their respective compounds. Thus peaks at 427 and 683 cm^{-1} were identified as ethyl acetate, the 930 cm^{-1} peak was identified as ethanol and the 943 cm^{-1} peak was identified as acetic acid. One of the advantages of Raman spectroscopy is that water does not interfere with the normal Raman signal.

Data analysis

To remove the effect of fluorescence and other non-chemical contributions, PCA was performed on the mean-centred raw data. The result of PCA is shown in Table 1. We observe that the first PC accounts for as much as 99.34% of the variation in the data. This indicates that it contains a significant component of the data. The second principal component captures only 0.55% of the variation. The amount of variation captured thereafter is insignificant.

The scores and loadings of the first principal component (PC1) were recombined to regenerate the data based upon the variation in PC1. The resulting plot is shown in Fig. 3. The information here looks very much like the raw data, together with the huge baseline shift and the big noisy peaks. An observation of the scores on PC1 (shown in Fig. 4) shows that the information here is not that of the Raman spectra. The systematic rise and fall in Fig. 4 is a result of a noise effect that operates in cycles throughout the monitoring period. To ascertain the source of this noise, the profiles of wavenumbers 100 cm^{-1} and 1850 cm^{-1} (being the baseline values) in the mean-centred data were mapped out (Fig. 5). It is clear that they are the exact replicas of Fig. 4, the PC1 scores. Thus it is obvious that PC1 is only noise, specifically the baseline noise in the system. Since the fluorescence peaks form the major component, PC1 is largely the contribution from fluorescence. This implies that if PC1 were removed from the data set, the leftover would be the Raman data devoid of baseline shift and most of the noise. Fig. 6 shows the data remaining after PC1 has been removed. Most of the noise has been removed. Also, there is a common and well-defined baseline. The positions of the peaks are still maintained, although the contribution of acetic acid is still not captured.

With the expectation that the data have been de-noised, the profiles of ethanol and ethyl acetate are mapped out once more, using the data shown in Fig. 6. These profiles are shown in Fig. 7. What we see is a removal of the fluorescence data that were very high in intensity, but no effect on the undulations due to the process dynamics, temperature fluctuations, *etc*., as the undulation is still present.

Since PC1 contains the fluorescence spectra, it follows that PC2 will contain the next greatest variation in the data, *i.e.* the

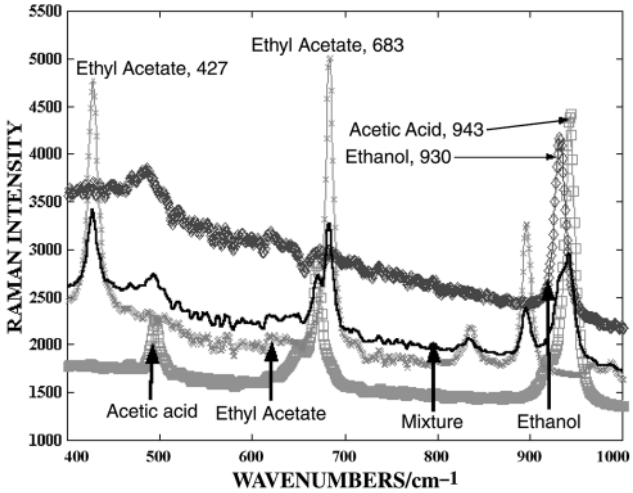

Fig. 2 Comparison of the pure Raman spectra of each reaction component and a synthetic mixture, in order to identify the sources of peaks in the reaction mixture.

Table 1 Percentage variance captured by the PCA model

Principal component number	Eigenvalue of covariance (*X*)	% Variance captured by this PC	Total % variance captured
1	1.61e + 007	99.34	99.34
2	8.84e + 004	0.55	99.88
3	1.40e + 004	0.09	99.97
4	7.27e + 002	0.00	99.97
5	3.01e + 002	0.00	99.98
6	1.62e + 002	0.00	99.98
7	1.28e + 002	0.00	99.98
8	9.99e + 001	0.00	99.98
9	9.04e + 001	0.00	99.98
10	8.03e + 001	0.00	99.98

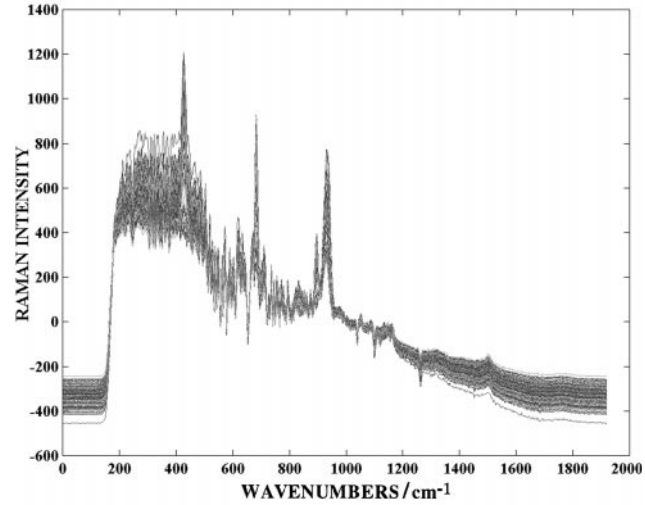

Fig. 3 Regenerated data based on variation in PC1.

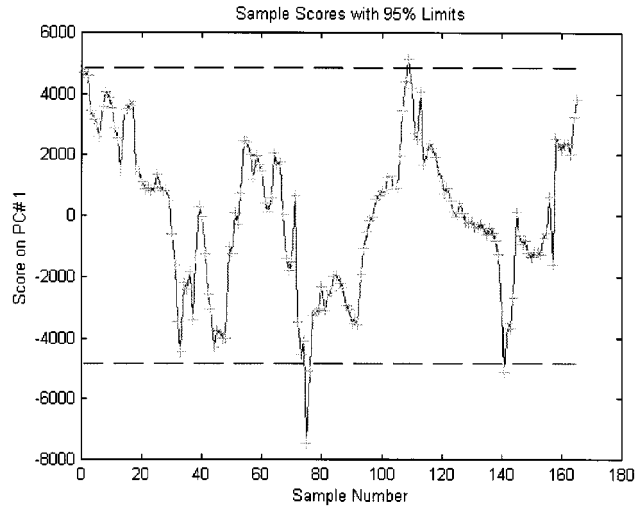

Fig. 4 Sample number *versus* scores on PC 1. This shows the progress of noise during the reaction time.

pure Raman spectra. Thus the data regeneration stage was repeated, based on the variation in the second principal component. Fig. 8 shows the regenerated Raman data using PC2. It is very obvious that all the noise is removed, as this is clearly what a set of Raman spectra should look like. In addition, the acetic acid peaks at 180, 493, 670 and 943 cm^{-1}, which were missing both in the original data and in PC1, are also captured. In comparison with Fig. 6, this is a much neater data set, with all the Raman data peaks standing out conspicuously. An observation of the scores for PC2 (Fig. 9) shows the expected progress of the reaction from the beginning to the tapering end. The reaction progress increases until around scan 40, when the rate reduces almost to a constant value. This confirms that PC2 contains largely pure Raman spectra.

Therefore, whereas PC1 captures the fluorescence data, PC2 captures only the Raman data. Again, the profiles of the peaks for ethanol, ethyl acetate and, this time, acetic acid are mapped out in Fig. 10.

Fig. 10 is actually the desired reaction progress, showing the true account of the chemical process that was being monitored. All the noise and the undulations are removed, and the result is a perfect picture of a reversible reaction. In comparison with Fig. 7, we observe the progressive removal first of the fluorescence noise and then the noise due to the system dynamics, leaving behind a true reaction profile.

Thus we see the effect of fluorescence spectra on Raman monitoring, and also the effectiveness of PC2 in selecting only

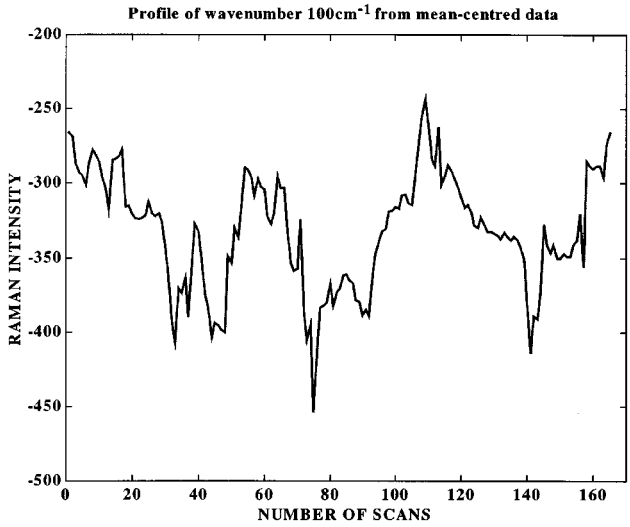

Fig. 5 Profile of wavenumber 100 cm^{-1} from the mean-centred data. This same profile is obtained for wavenumber 1850 cm^{-1} at the opposite end of the spectrum.

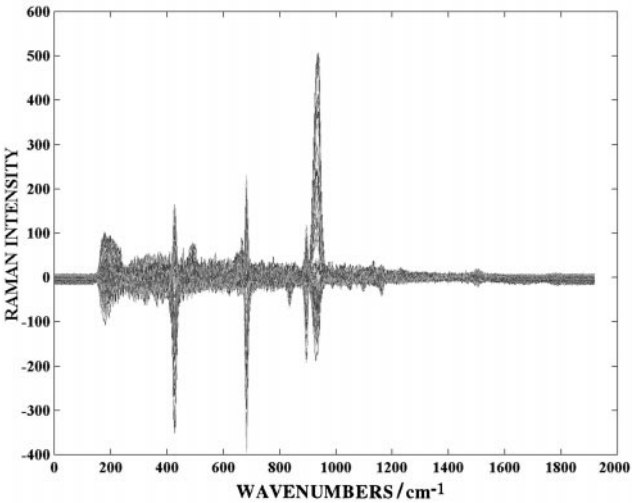

Fig. 6 PC1 results. Result of removing the regenerated Raman data obtained by the product of PC1 scores and loadings, from the mean-centred Raman data.

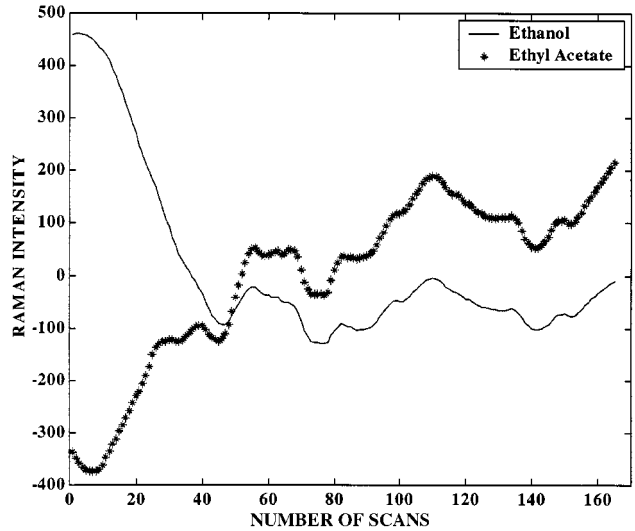

Fig. 7 Profiles of ethyl acetate and ethanol during the esterification, using the data remaining after removing PC1.

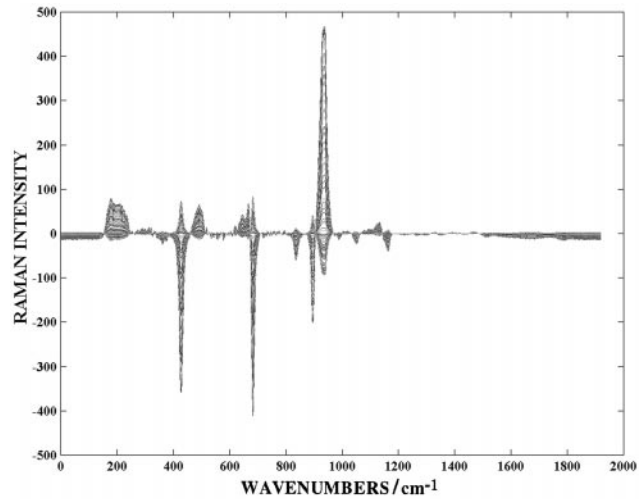

Fig. 8 PC2 results. Regenerated Raman spectra based on variation captured in the second PC.

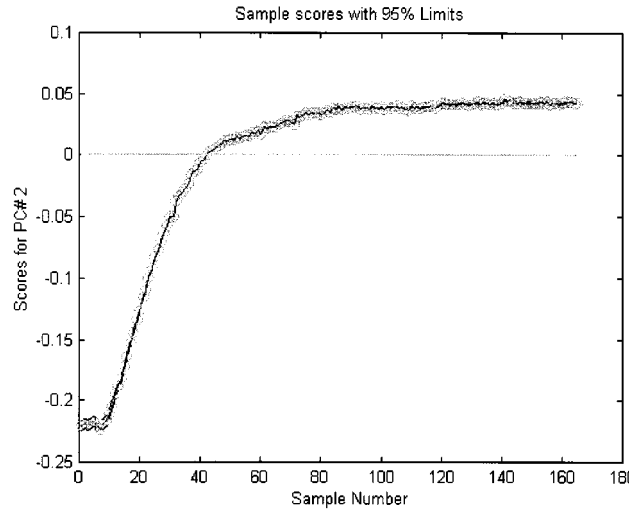

Fig. 9 Sample number *versus* scores for PC2. This shows the progress of the esterification reaction proper.

the most relevant data. It is useful to know that, whereas the data in Fig. 7 were smoothed (using Savitsky–Golay) for easier viewing, the data in Fig. 10 were not. This shows how accurately PC2 has captured the Raman spectra.

Kinetics

Until now we have observed the reconstituted data due to the variation captured in PC2 to be of the nature of Raman spectra and, based on our knowledge of the wavelengths at which the various reaction components occur, we have plotted their reaction profiles and are satisfied that they follow the expected trends. However, there is the need to validate these assertions theoretically, either by comparison with concentration data (which are unavailable in this work) or by testing the reconstituted data against a well-known chemical theory. In this work, the second option was chosen, and the PC2 data were tested for compatibility with second order reaction kinetics, as discussed earlier in the introduction section.

The kinetic plots have been generated for the region of highest reaction rate, *i.e.* scans 18–29. In Fig. 11, the kinetic plots are shown and, as expected, they are perfectly linear with R^2 values of 0.9578 for ethyl acetate (product) and 0.9891 for ethanol (reactant), thus satisfying the criterion for a second order reaction.

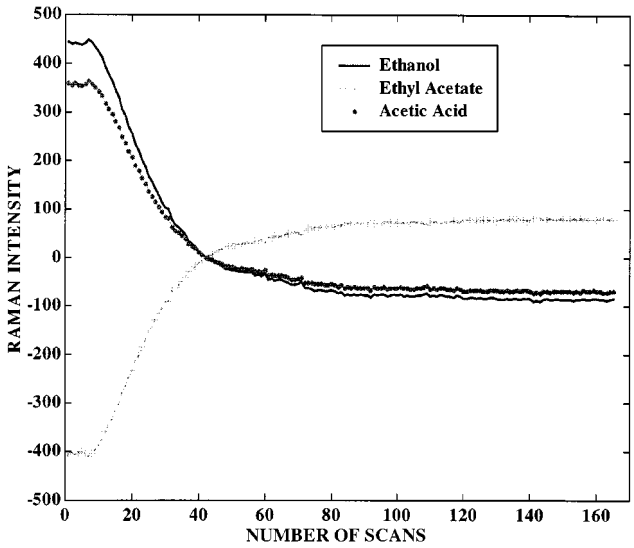

Fig. 10 Profiles of the peaks for ethanol, ethyl acetate and acetic acid during the esterification, using PC2.

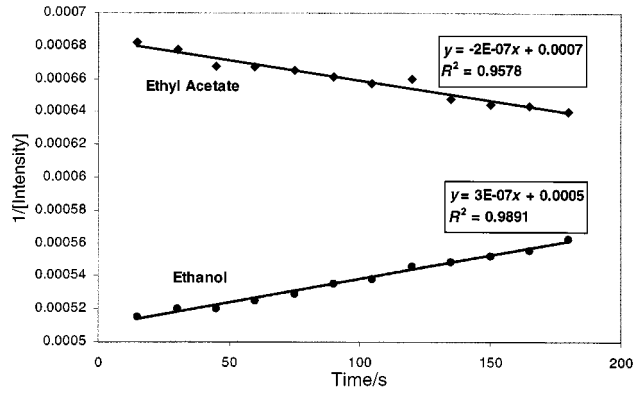

Fig. 11 Kinetic plots for ethyl acetate and ethanol during scans 18 to 29.

Conclusion

Raman spectroscopy has been efficiently used to collect detailed data on the esterification of ethanol and ethyl acetate. Principal components analysis has been applied on noisy Raman data to remove all the noise. Fluorescence spectra have been removed by PC1. The second PC contains all the pure Raman data, and clearly shows the progress of the reaction, totally devoid of noise and undulations due to the dynamics of the system. This technique can be applied to other Raman data sets. However, whether the Raman spectra will be described by PC1 or PC2 or another principal component depends on the degree of noise or fluorescence in the data. Thus, with a less noisy data set than that used in this work, PC1 may capture the pure Raman spectra, while in a noisier data set, the Raman data may be in PC3 for instance.

PCA has therefore been shown to be very useful in extracting information from raw data that are noise-ridden and difficult to explain. The fact that the data obtained from PC2 are actual Raman data has been confirmed by application of second order kinetic equations to the data.

Acknowledgements

We acknowledge the Centre for Process Analytics and Control Technologies (CPACT), Hull, UK and Clairet Scientific, Northampton, UK (for providing the Raman spectrometer and probe).

References

1 S. Wold, N. Kettaneh, H. Friden and A. Holmberg, *Chemom. Intell. Lab. Syst.*, 1998, **44**, 331.
2 K. S. Dahl, M. J. Piovoso and K. A. Kosanovich, *Chemom. Intell. Lab. Syst.*, 1999, **46**, 161.
3 A. C. Quinn, P. J. Gemperline, B. Baker, M. Zhu and D. S. Walker, *Chemom. Intell. Lab. Syst.*, 1999, **45**, 199.
4 C. E. Miller, *Chemom. Intell. Lab. Syst.*, 1995, **30**, 11.
5 W. Blaser, R. Bredeweg, R. Harner, M. LaPack, A. Leugers, D. Martin, R. Pell, J. Workman and L. Wright, *Anal. Chem.*, 1995, **67**, 47R.
6 N. B. Colthup, L. H. Daly and S. E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, London, 3rd edn, 1990, p. 60.
7 N. Ward, H. Edwards, A. Johnson, D. Fleming and P. Coates, *Appl. Spectrosc.*, 1996, **50**, 812.
8 J. B. Cooper, K. L. Wise, J. Groves and W. T. Welch, *Anal. Chem.*, 1995, **67**, 4096.
9 J. B. Cooper, K. L. Wise and B. J. Jensen, *Anal. Chem.*, 1997, **69**, 1973.
10 N. B. Colthup, L. H. Daly and S. E. Wiberley, *Introduction to Infrared and Raman Spectroscopy*, Academic Press, London, 3rd edn., 1990, p. 61.
11 J. B. Cooper, *Chemom. Intell. Lab. Syst.*, 1999, **46**, 231.
12 D. T. Andrews, L. Chen, P. D. Wentzell and D. C. Hamilton, *Chemom. Intell. Lab. Syst.*, 1996, **34**, 231.
13 J. E. Jackson, *A User's Guide to Principal Components*, Wiley, New York, 1991.
14 M. Edinborough, *Writing Organic Reaction Mechanisms*, Taylor and Francis, London, UK,1994, p. 140.
15 P. W. Atkins, *Physical Chemistry*, Oxford University Press, Oxford, 4th edn., 1990, pp. 786–787.

Paper 9/07200D