



Cite this: *Analyst*, 2026, **151**, 2024

## Raman spectroscopic monitoring of multiproduct chemical reaction kinetics; the case of ester hydrolysis

Muhammed Kashif,\* Mark E. Keating and Hugh J. Byrne 

Raman Spectroscopy (RS) is a powerful technique for the identification of molecules based on the characteristic fingerprint spectra of their vibrational modes. Although challenging, real time spectroscopic monitoring of reactions and processes has great potential value in multiple fields, including process analysis, bio reactors, cell therapies and *in vitro* metabolomics. Refined chemometrics methodologies are required to datamine the kinetic evolution of multivariate spectral mixtures to establish the constituent reactants and products, as well as the characteristic rates of the reaction. To explore the capabilities and challenges, RS was used to study the chemical kinetics of propyl acetate hydrolysis in an aqueous environment at room temperature, *in situ*, as a model reaction. The continuous conversion of propyl acetate to 1-propanol and acetic acid was monitored periodically over 250 min using RS with a 532 nm laser source. Simulated admixture solutions, mimicking the reaction from pure reactants to pure products conversion, were also recorded for comparison. Problem based nonlinear least squares (NLS) fitting was applied to both the actual reaction and simulated solution data sets using pure components spectra of both the reactants, propyl acetate, water and products, 1-propanol and acetic acid, in order to visualise and confirm the trends and kinetics of the reaction components. Multivariate Curve Resolution-Alternating Least Squares analysis (MCR-ALS) with kinetic constraints was applied to further resolve the concentration and spectral profiles and to quantify the rates of the reaction. It is demonstrated that MCR-ALS could not accurately resolve the evolving reaction species with respect to concentration, due to rank deficiency. To enhance the analysis, a data augmentation approach was used, seeding the measured datasets with the spectra of the pure components to bias the initial singular value decomposition and spectral unmixing process, resulting in an improved resolution of the systematic variation of concentration dependent data to monitor the kinetic evolution of the reaction mixture. The required seeding weights were optimized by visualizing the sum residual error (SUMR) in least squares fitting of the actual components with the identified pure components by MCR-ALS. Minimum SUMR values for admixtures were found at a seeding weight of 10 000X, while 100X was found to be optimum for the actual reaction. This proof of concept can further pave the way for better analysis and understanding of cascade reactions, and ultimately, potentially of metabolomic pathways.

Received 3rd October 2025,  
Accepted 21st February 2026

DOI: 10.1039/d5an01055a

rsc.li/analyst

## 1 Introduction

Raman spectroscopy holds exceptional potential for the analysis of materials, due to its ability to provide detailed molecular fingerprints based on vibrational modes.<sup>1</sup> Unlike other spectroscopic techniques, Raman can perform non-destructive analysis of solids, liquids, and gases without extensive sample preparation, making it highly versatile across diverse fields.<sup>2</sup> Its sensitivity to molecular structure, crystallinity, phase tran-

sitions, and chemical composition enables precise differentiation between components of complex mixtures.<sup>3,4</sup> Raman's ability to deliver rapid and label-free identification continues to place it at the forefront of advanced materials characterization.<sup>5</sup>

Raman spectroscopy has emerged as a highly versatile tool for the study of reaction kinetics in pharmaceutical processes, disease diagnostics, cellular functions, and real-time monitoring of chemical and biological species.<sup>6,7</sup> In pharmaceutical applications, Raman spectroscopy enables precise tracking of drug crystallization, polymorphic transitions, and degradation pathways under process-relevant conditions, providing crucial insights for quality control and formulation stability.<sup>8</sup> Raman

Physical to Life Sciences Research Hub, FOCAS, Technological University Dublin, City Campus, Dublin D08 CKP1, Ireland. E-mail: Muhammed.Kashif@TUDublin.ie



spectroscopy excels in capturing onsite chemical changes, such as oxidation, hydrolysis, and catalytic processes, with high temporal resolution,<sup>9</sup> making it invaluable for understanding complex reaction mechanisms in both chemical and biological contexts.<sup>10,11</sup> In process analytical technology (PAT), Raman spectroscopy enables continuous analysis of critical quality attributes (CQAs) and process parameters (CPPs).<sup>12,13</sup> In bioreactor monitoring, it is widely applied for tracking metabolites, nutrients, and biomass in real time.<sup>14–16</sup> Its integration with chemometrics and multivariate analysis allows simultaneous quantification of multiple analytes, supporting efficient upstream and downstream process control.<sup>15,17</sup> Within cellular environments, Raman spectroscopy offers non-invasive, label-free observation of intracellular dynamics, enabling real-time studies of metabolic pathways, drug uptake, and organelle-specific reactions without disrupting biological integrity.<sup>18</sup>

Despite the widely accepted promise of Raman spectroscopy, accurately datamining the spectroscopic signatures and quantifying the kinetic parameters of the processes of interest remains a challenge. In this context, the technique of multivariate curve resolution, alternating least squares (MCR-ALS) fitting can be employed to extract independent spectral profiles from a complex mixture,<sup>19</sup> and the analysis can be constrained using kinetic rate equations to describe the evolution of the component concentrations over time, in a hard-soft modelling approach.<sup>20</sup> Such an approach has been employed to analyse the *in vitro* uptake of chemotherapeutic drugs in biological cells,<sup>21–23</sup> as well as stem cell differentiation.<sup>24</sup>

MCR-ALS exhibits notable limitations when applied to complex reaction systems, however, particularly those involving multiple reactants and products. The case of complete and incomplete kinetic processes, in which the starting or end point points are respectively missing from the experimental analysis, was considered by de Juan *et al.*<sup>25</sup> The study also assessed the influence of drift of the spectral baseline or of an absorbing interference in the kinetic process, as monitored by UV-Visible absorption spectroscopy. For the analysis of complex cascade reactions, faithful resolution of the spectral components of parallel or competing processes has been recognised as a challenge in a number of studies, and the issue has been addressed in part by the use of complementary techniques, such as chromatography,<sup>26</sup> or spectral analysis of similar compounds,<sup>27</sup> which are analysed in a multiset structure.<sup>28</sup>

A promising solution to address the challenges of resolving complex reaction systems is the recently demonstrated seeded-

MCR-ALS approach.<sup>29</sup> In this method, the data matrix is deliberately augmented with known spectral profiles, either full spectra of pure components or selected features as “seeds”, prior to curve resolution. This strategic seeding significantly biases the multivariate optimization toward chemically meaningful solutions, effectively guiding the algorithm to unmix overlapping spectra in complex mixtures. By integrating prior spectral knowledge directly into the MCR-ALS workflow, seeded-MCR-ALS offers a robust means to disentangle multiple reactants and products even in multiproduct reactions thus mitigating rotational ambiguity and improving interpretability of concentration change profiles in complex kinetic studies.

In this research work, the propyl acetate acid hydrolysis reaction (Fig. 1) was studied to demonstrate the proof of concept, and for better understanding, eleven simulated admixtures of reactants and products were prepared to mimic the conversion of 100% reactants to 100% products. The contributions of the respective constituent component spectra to the evolution or the reaction data were initially illustrated using a problem based nonlinear least squared NLS fitting approach. The limitations of MCR-ALS in resolving simultaneously evolving components are demonstrated, and then the seeded approach was utilized for better identification of pure reaction components in the singular value decomposition (SVD) stage of the MCR-ALS algorithm. Least squares fitting was applied to optimize the seed weight by observing the sum residual error (SUMR) of the MCR-ALS results.

Fig. 1 shows the reaction scheme of propyl acetate hydrolysis, in which the OR group of an ester is replaced by an OH from water, resulting in the corresponding carboxylic acid (acetic acid) and alcohol (propan-1-ol). The reaction is promoted by H<sub>2</sub>SO<sub>4</sub> and for methyl ester typically takes place at 20–80 °C. In this study hydrolysis of propyl ester was monitored at room temperature ~20 °C using Raman spectroscopy.

Fig. 2 illustrates the simulated Raman spectral evolution and the corresponding reaction kinetics for the acid-catalyzed hydrolysis of propyl acetate. In Fig. 2(a), the simulation depicts the progressive changes in Raman spectra of the pure components over the course of the reaction, modeled using 100 acquisition points from the initial state to reaction completion, thereby capturing the continuous spectral transformation associated with reactant consumption and product formation. Fig. 2(b) presents the reaction kinetics obtained from supervised MCR-ALS analysis, clearly demonstrating a systematic decrease in the reactant contributions accompanied by a corresponding increase in the product profiles, consistent with the expected reaction progression.

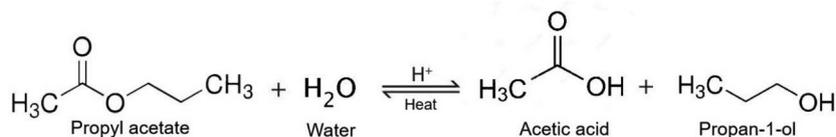


Fig. 1 Generic reaction scheme of propyl acetate hydrolysis.



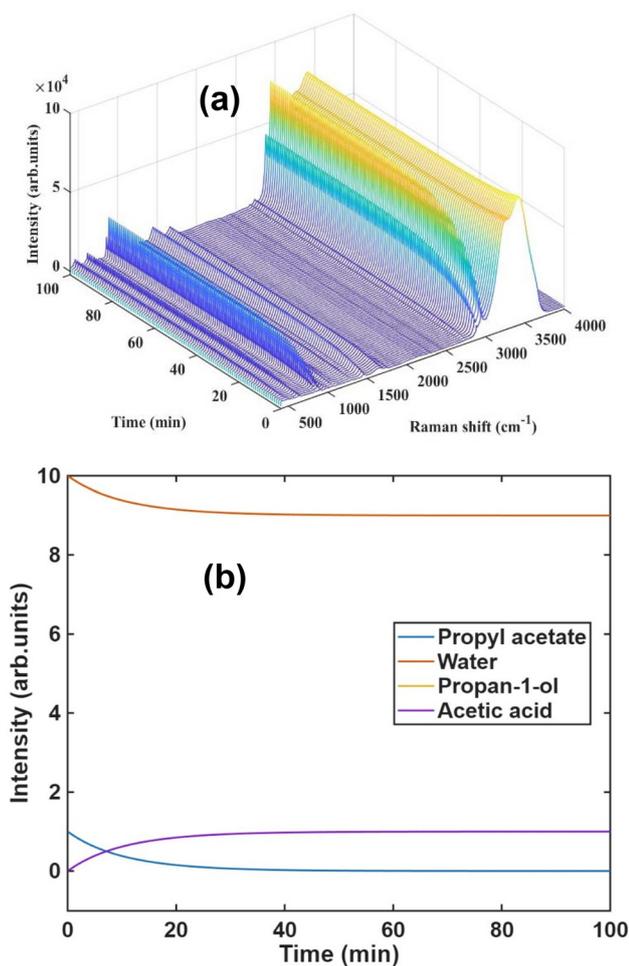


Fig. 2 The simulation for the Raman spectral changes of pure components as the reaction proceeds considering 100 points of acquisition to the completion of reaction (a). The reaction kinetics generated by supervised MCR-ALS from simulated data, illustrating the decrease in reactants and increase in the products accordingly (b). "Note that the curves representing the evolution of the propan-1-ol and acetic acid reactants are overlaid."

## 2 Methodology

### 2.1 Materials

Chemicals used for the study, propyl acetate (99%) Sigma-Aldrich, 1-propanol ( $\geq 99\%$ ) Fisher Scientific, acetic acid ( $\geq 99.7\%$ ) Fisher Scientific and Sulfuric acid (95%) was sourced from Thermo Scientific (Ireland). Deionized water was used for preparing solutions.

Propyl acetate and water were mixed with the 1:10 by volume, taking 5 mL of propyl acetate and 50 mL of water, 1 mL of concentrated  $\text{H}_2\text{SO}_4$  was used to catalyze the reaction.<sup>30</sup> Reaction was allowed to continue for 250 min at room temperature around 20 °C under gentle stirring at 500 rpm. Progress of reaction was monitored periodically with the interval of 5–10 min using Raman spectroscopy. Eleven simulated admixture solutions were made at room temperature for the

comparison according to the volume ratios shown in Table 1S. The "admixture 0" contains 0% products and 100% reactants while "admixture 10" contains 100% products and 0% reactants. The 10% change of reactants concentration to the products by volume has been considered in each admixture from 0–10<sup>th</sup> admixture.

### 2.2 Spectroscopic analysis

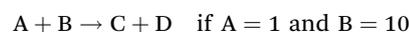
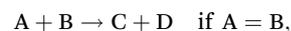
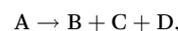
Raman spectroscopic analysis was performed using an Instruments S.A. Labram Raman confocal spectroscopic microscope, with a 532 nm semiconductor diode laser as source and a 1800 grooves per mm grating, producing a spectral dispersion of  $1.65 \text{ cm}^{-1}$  per pixel. Detection of the scattered light was performed using a Peltier cooled, 16-bit dynamic range CCD detector with  $1024 \times 256$  pixels. The laser power at the sample was measured as  $\sim 20 \text{ mW}$ , and the beam was focussed through the liquid, such that the focal depth was immersed in the sample. The scattered Raman signal was integrated for 10 seconds with 3 repetitions and measured over a spectral range of 200 to  $3600 \text{ cm}^{-1}$  with respect to the source frequency. The software used for the Raman spectral acquisitions was LabSpec 5.

Data was recorded by placing the beaker containing the reaction mixture under the  $10\times$  lens. Raman acquisitions are made keeping the focus inside the reaction mixture. The same conditions were used to record simulated solutions data. Once acquired, each spectrum was baseline corrected using the asymmetric least squares (ALS) method, noise subtracted and lightly filtered using a third order linear model to improve.

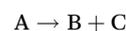
**2.2.1 Data processing.** Data preprocessing was performed using MATLAB R2024a statistical software. Smoothing of data is done by Savitsky Golay smoothing using polynomial of 5 over 9 points. Asymmetric least squares followed by rubber band background correction was used to correct the background, keeping all the data in single matrix. MCR-ALS analysis was conducted using MCR-ALS Graphical User Interface 2.0 in MATLAB.

**2.2.2 Data analysis.** Problem based NLS fitting was applied to both the actual reaction and simulated solutions data sets using pure components spectra of both the reactants, propyl acetate, water and products, 1-propanol and acetic acid. MCR-ALS was applied to further resolve the concentration and spectral profiles of the kinetics of the reaction. Unsupervised MCR-ALS analysis was conducted by calculating the number of components through SVD. Three different approaches considering 4, 3 and 2 components were studied. The supervised MCR-ALS analysis was conducted using nonnegative least squares (NNLS) and kinetic constraints. Following 5 models were applied for kinetic constraints.

4 components:



3 components:



2 components:



Spectral profiles (sopt) and concentration profiles (copt) were analysed to resolve the kinetics of the reaction by MCR-ALS.

**2.2.3 Data seeding.** Seeded data sets were generated by following the methodologies of Keating and Byrne.<sup>29</sup> The reaction data set was augmented by adding all four reaction components to the same matrix. Similarly, the simulated admixtures data set was also augmented with the addition of the components. Fig. 3 shows the schematic illustration of the data seeding process.

MCR-ALS normally uploads a reaction data set ( $R_{\text{data}}$ ) in the form of an  $(n \times m)$  matrix of  $n$  spectra of spectral range  $m$  data-points. In the data seeding process,  $R_{\text{data}}$  was augmented by adding the spectra of each of the four reaction components,  $C_A$ ,  $C_B$ ,  $C_C$ ,  $C_D$ , to the same matrix, in the form of  $(1 \times m)$  matrices, resulting in a matrix of dimensions  $(n + 4 \times m)$ . In order to explore the effect of biasing the algorithm in favour of the components spectra the value of seeding weight “ $X$ ” was varied from 1–10 000 and sum residual error (SUMR) of the model was monitored continuously to get minimal SUMR values and best fit of the components in the data. This seeding protocol was applied to both the experimental and simulated reaction spectral datasets. Fig. 4 illustrates the

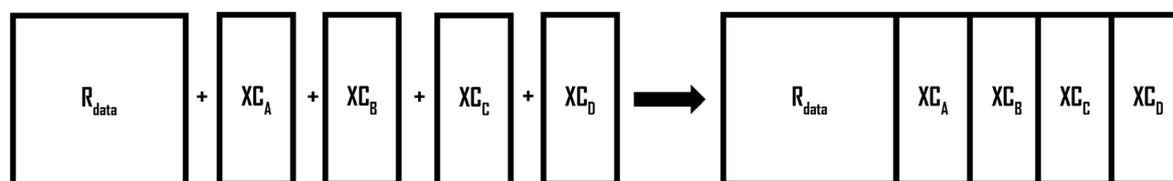


Fig. 3 Schematic illustration of the process of seeding the reaction data set. The  $(n \times m)$  reaction data set ( $R_{\text{data}}$ ) is augmented with the  $(1 \times m)$  spectra of the reaction components ( $C_A$ ,  $C_B$ ,  $C_C$  and  $C_D$ ), weighted by a factor  $X$ .

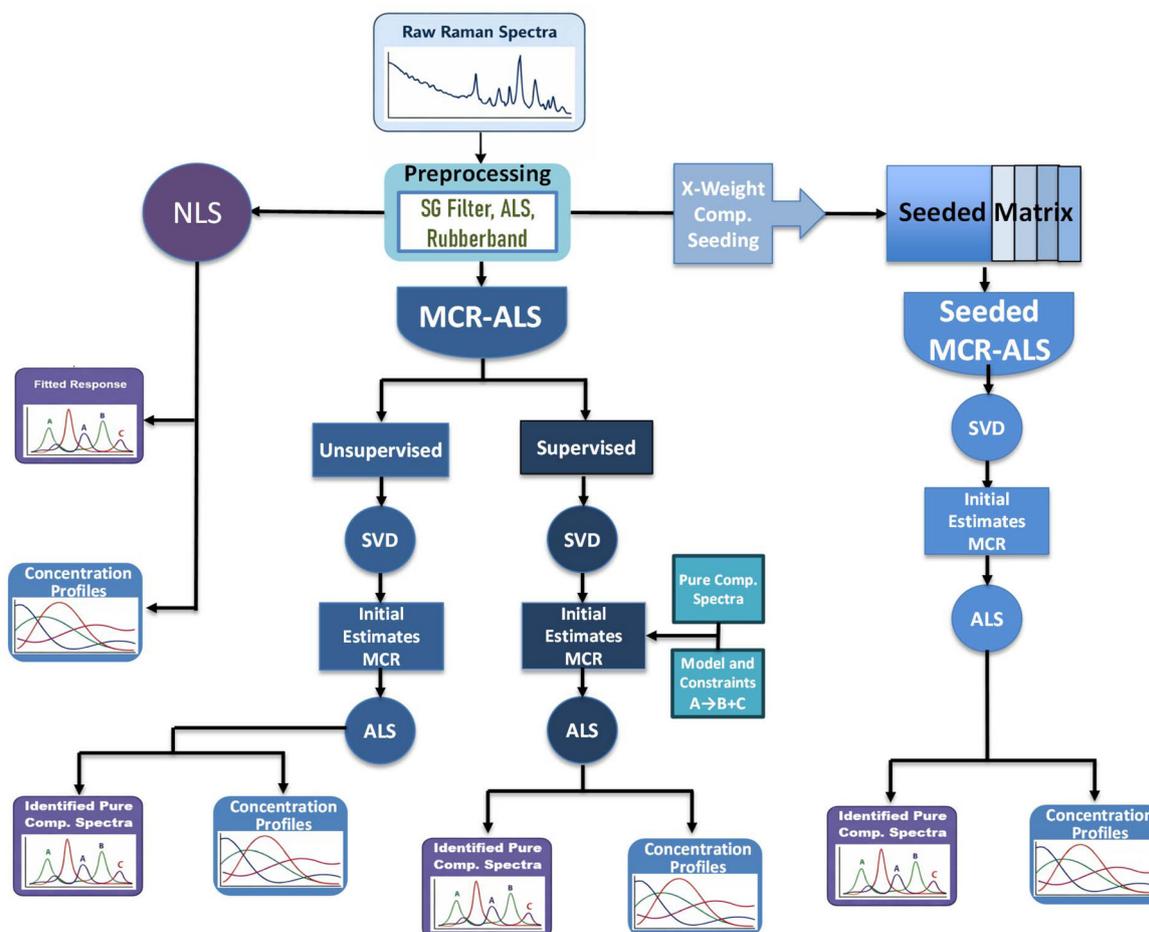


Fig. 4 Flowchart diagram for the chemometric analysis.



overall workflow of the chemometric analysis performed in this study.

### 3 Results

Fig. 5(a) shows the spectra of identical volumes of pure components of the reaction of propyl acetate with water to produce 1-propanol and acetic acid, while Fig. 5(b) shows the observed evolution of the spectral profile of the reaction mixture over a period of 250 min. It is notable that the spectrum of the starting mixture of 5 mL propyl acetate in 50 mL water is dominated by that of propyl acetate, and, for example, the most prominent features of water at  $\sim 3300\text{ cm}^{-1}$  are hardly visible. In contrast, as the reaction progresses beyond 150 min, the features of water in this region begin to dominate.

Multiple changes in spectral features are observed as the reaction proceeds; the most robust changes include the evolution of two distinct peaks in the region of  $800\text{--}900\text{ cm}^{-1}$  linked to products formation, while a continuous decrease of features in the region below  $800\text{ cm}^{-1}$  and above  $900\text{ cm}^{-1}$  to  $3100\text{ cm}^{-1}$  occurs. A similar behavior was observed in the Raman spectral acquisition of the 11 simulated admixtures of the reaction components. Fig. 6(a) shows the spectral data for the simulated admixtures of the gradually increasing product ratios as compared to the reactants in the mixtures.

#### 3.1 Water suppression

A similar unusual behavior of water is observed in both the actual reaction and the simulated admixtures of the components, further verified by studying the 10 mL of propyl acetate in 5 mL–500 mL of water dilutions. Figure 6(b) shows the Raman spectral acquisitions of the multiple dilutions of propyl acetate in water. It is observed that the signature peaks for the  $\text{-OH}$  group of water around  $\sim 3300\text{ cm}^{-1}$  do not appear up to the 1:10 dilution ratio. The features were observed at 1:20 dilution ratio and above, which confirms that certain

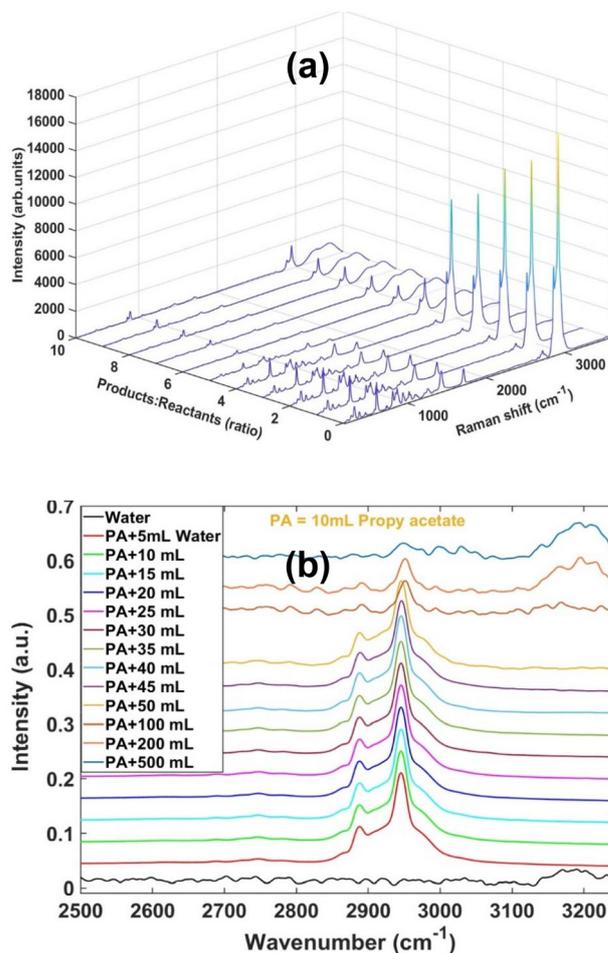


Fig. 6 Spectral evolution of the simulated admixture solutions of the reaction components with increasing ratio of the products to reactants from 0–100% (a). Raman spectra of water dilutions of the 10 mL propyl acetate in 5 mL–500 mL of water (b).

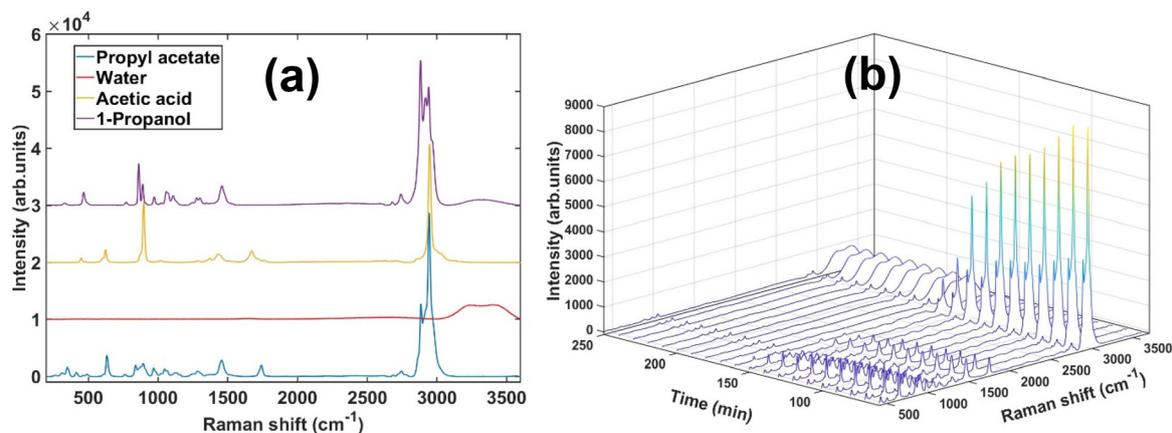


Fig. 5 (a) Preprocessed Raman spectra of the reactants and products of the hydrolysis of propyl acetate (offset for visual clarity). (b) Spectral evolution of the reaction mixture over a period of 250 min.



spectroscopic features may be suppressed in the presence of other molecules at some ratios of the mixtures. In terms of the spectroscopic evolution, the water spectrum emerges as product in the reaction instead of the reactant and may be seen as increasing component in the reaction kinetics plots.

### 3.2 Problem based NLS

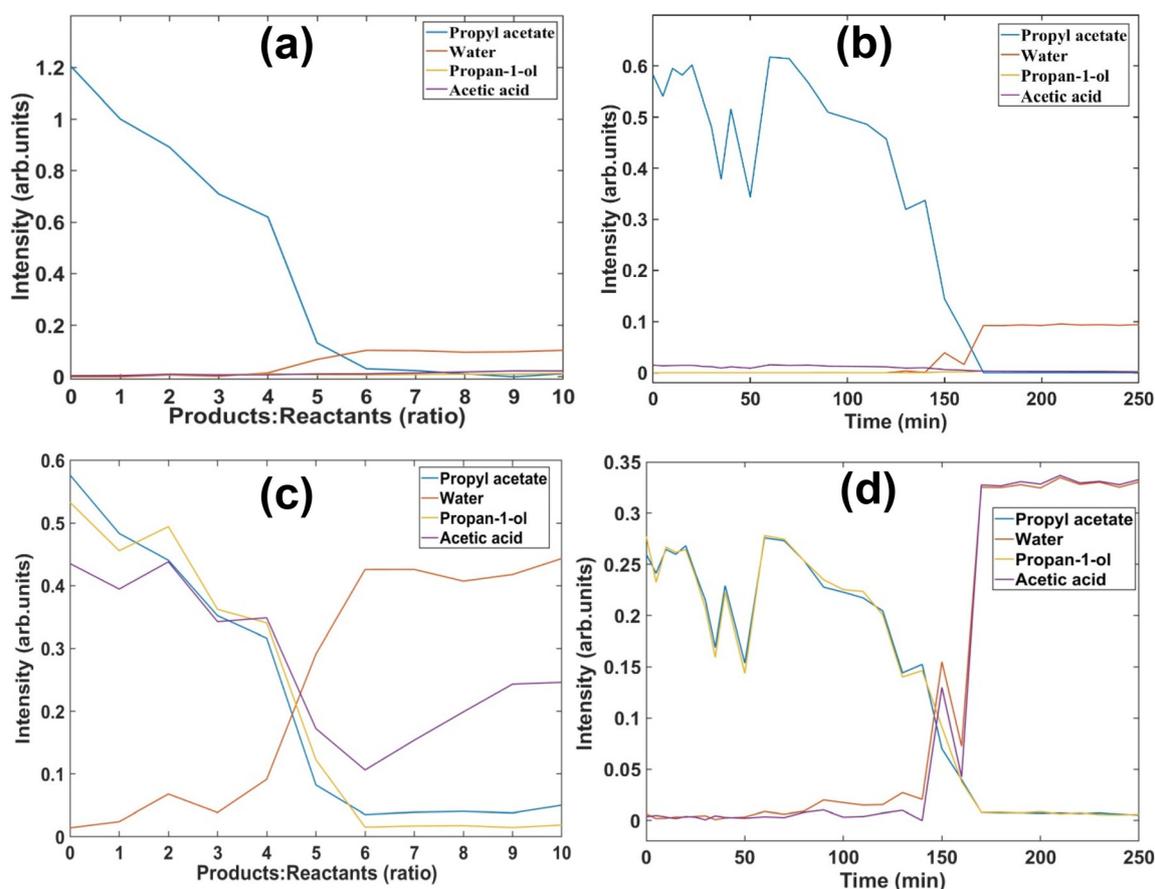
Problem-based nonlinear least squares (NLS) fitting was employed to investigate the evolution of the simulated and experimental reaction mixtures, based on the observation that the spectra of the physical admixtures of the components did not correspond to the proportionate linear superposition of the spectra of the pure compounds. Fig. 7(a) and (b) show the concentration changes for the 4 components for the simulated admixtures and actual reaction respectively. The time evolution of concentrations does not seem to be accurate, because the fitted spectral responses do not exactly match the spectra recorded experimentally, even for the simulated admixtures. This happens because the solution driven by the problem based NLS is based on the pure components, and the program

struggles to identify the exact features in the actual data leading to misfitting. Fig. 1S–11S in the SI show the spectral fitting of the simulated admixtures, while Fig. 12S–41S show the spectral fitting of the reaction data.

These findings confirm that the application of NLS fitting to elucidating kinetic parameters in systems in which spectral data are complex and overlapping, does not offer precise insights into the dynamic behaviour of the multi-component processes like ester hydrolysis reactions, in which more than one component react to give two or more components *e.g.*  $A + B \rightarrow C + D$ . These algorithms are quite good for the simple process in which change is considered as a whole *e.g.*  $A \rightarrow B$ .<sup>31,32</sup>

### 3.3 MCR-ALS (unsupervised)

SVD using the MCR-ALS returned two components with notably different values for both the simulated admixtures data and the reaction data. Selection of 4 components in SVD and calculations for the identification of pure components fails to accurately identify the actual components, not only in



**Fig. 7** Kinetic evolution of problem based NLS fitting (a) the simulated admixtures as a function of reactant product ratio (mimicking the progression over the reaction timescale), (b) the actual reaction as a function of time. Evolution of the components calculated by Unsupervised MCR-ALS (c) Simulated admixtures data as a function of reactant product ratio (mimicking the progression over the reaction timescale) (d) Reaction data as a function of time.



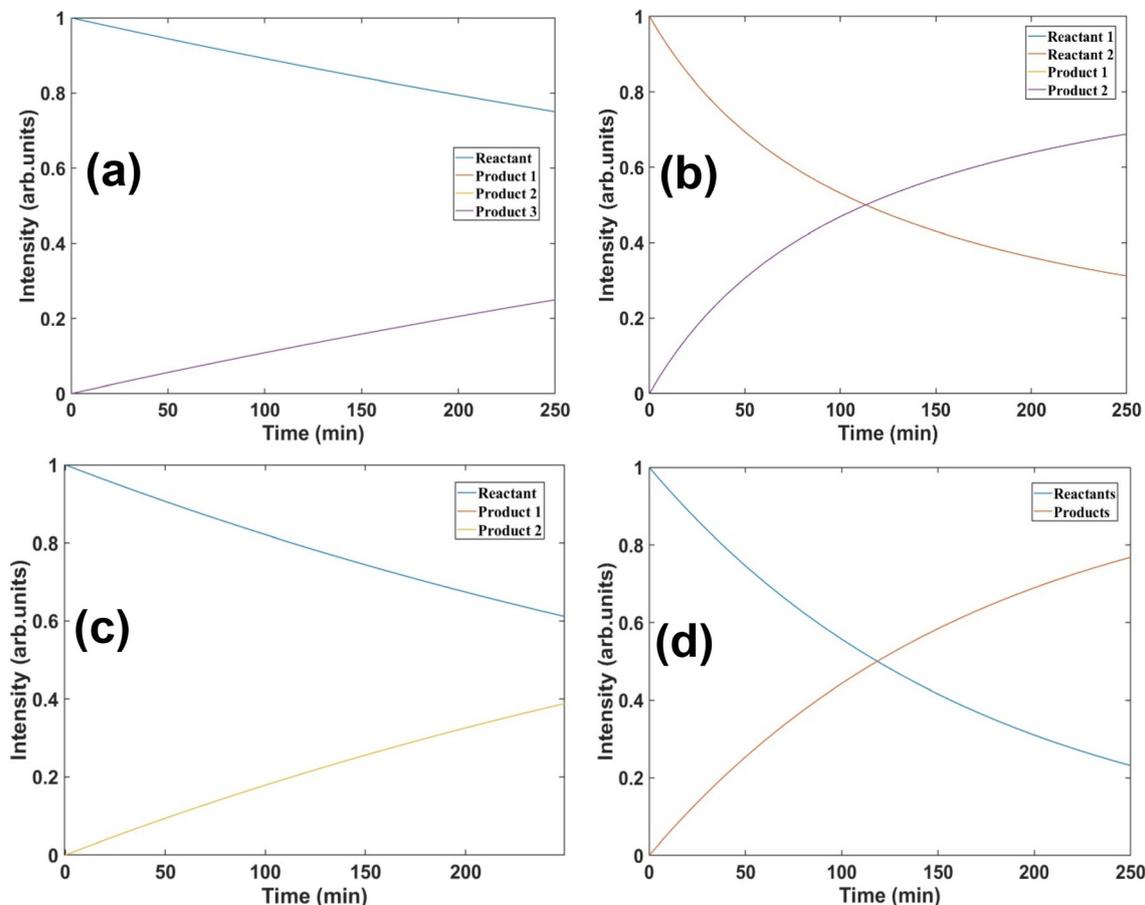
the reaction data but also in the simulated admixtures data. Fig. 7(c) shows the time evolution of the reaction components generated by unsupervised MCR-ALS in simulated admixture data, and Fig. 7(d) shows the time evolution of reaction components in the reaction data. The time evolution reveals inaccurate changes in the reactants consumption and products formation due to ambiguous identification of components, resulting in the poor kinetic profiling. Nevertheless, the analysis does reproduce the suppression of the water signal at the early stages, and its re-emergence as the reaction progresses, as described in Section 3.1. The selected components by the algorithm are shown in Fig. 42S and 43S for the reaction data and admixtures data respectively.

### 3.4 MCR-ALS (supervised)

Supervised MCR-ALS was performed by manually selecting 4, 3 and 2 components, in turn, in the component selection window of the MCR-ALS GUI, instead of SVD calculations, and uploading the pure spectra of all four (water, propyl acetate, propanol and acetic acid), three (propyl acetate/water solution, propanol and acetic acid) and two (propyl acetate/water solu-

tion and propanol/acetic acid mixture) components, respectively. NNLS and kinetic constraints with four different models were applied, in an attempt to force the analysis. Fig. 8(a) shows the supervised MCR-ALS time evolution for reaction data for 4 components  $A \rightarrow B + C + D$ , Fig. 8(b) shows the time evolution for reaction data for 4 components of kinetic model  $A + B \rightarrow C + D$ , Fig. 8(c) shows the time evolution for reaction data for 3 components using kinetic model  $A \rightarrow B + C$  and Fig. 8(d) shows the time evolution for reaction data for 2 components using kinetic model  $A \rightarrow B$ .

MCR-ALS can be constrained by kinetic models to quantify the rate of evolution. Although, the reaction of Fig. 1 should be represented by a model of the form  $A + B \rightarrow C + D$ , although, as A and B, and B and C might be expected to evolve simultaneously, they may be equally represented by a more simplistic model of the form  $A \rightarrow B$ . However, the observed suppression of the Raman signal of water in the presence of propyl acetate, and its re-emergence as the propyl acetate is consumed as the reaction progresses means that the Raman spectroscopic representation of the reaction may be better described by a model of the form  $A \rightarrow B + C + D$ , or  $A \rightarrow B + C$ . All four models were



**Fig. 8** Time evolution of components in reaction data using supervised MCR-ALS (a) 4 components  $A \rightarrow B + C + D$  (b) 4 components  $A + B \rightarrow C + D$  (c) 3 components  $A \rightarrow B + C$  (d) 2 components  $A \rightarrow B$ . In all cases, MCR-ALS effectively only resolves two components, as simultaneously evolving reactants and/or products are not distinguished.



therefore used as kinetic constraints of the MCR-ALS analysis of both the simulated and experimental reaction data sets. All the models depict time evolution of just as two components in the data and do not exhibit the correct spectral variation for the actual 4 components.

Fig. 44S shows the MCR-ALS time evolution for the 4 components for the model  $A + B \rightarrow C + D$  if  $A = 1$  and  $B = 10$ . This also does not correctly represent the correct time evolution profile for the 4 components due to inappropriate selection of the components. While Fig. 45S–47S show the component selection by the algorithm for three different models considering 4 components in the data. Fig. 48S(a) shows the supervised MCR-ALS time evolution for the simulated admixtures data for 4 components  $A \rightarrow B + C + D$ . Fig. 48S(b) shows the time evolution for the simulated admixtures data for 4 components of kinetic model  $A + B \rightarrow C + D$ , Fig. 44S(c) shows the time evolution for the simulated admixtures data for 3 components using kinetic model  $A \rightarrow B + C$  and Fig. 44S(d) shows the time evolution for the simulated admixtures data for 2 components

using kinetic model  $A \rightarrow B$ . Fig. 49S shows the time evolution for the simulated admixtures data for 4 components using kinetic model  $A + B \rightarrow C + D$  if  $A = 1$ ,  $B = 10$ . While the Fig. 50S–52S illustrates the component selection for different models of 4 components for the admixtures data. Indeed, all the models show a time evolution characteristic of either only two components or incorrect, and do not follow the correct spectral variations for the actual components due to incorrect selection of components. In all cases, MCR-ALS effectively only resolves two components, as simultaneously evolving reactants and/or products are not distinguished. It suggests that, while MCR-ALS is a powerful technique for the simple processes in which spectral changes are considered as a whole, from one point to other,<sup>33–35</sup> it has limitations when applied to multi-component processes.

### 3.5 Seeded MCR-ALS

Fig. 9(a) shows the concentration profile results for seeded MCR-ALS analysis of the simulated admixtures data at a com-

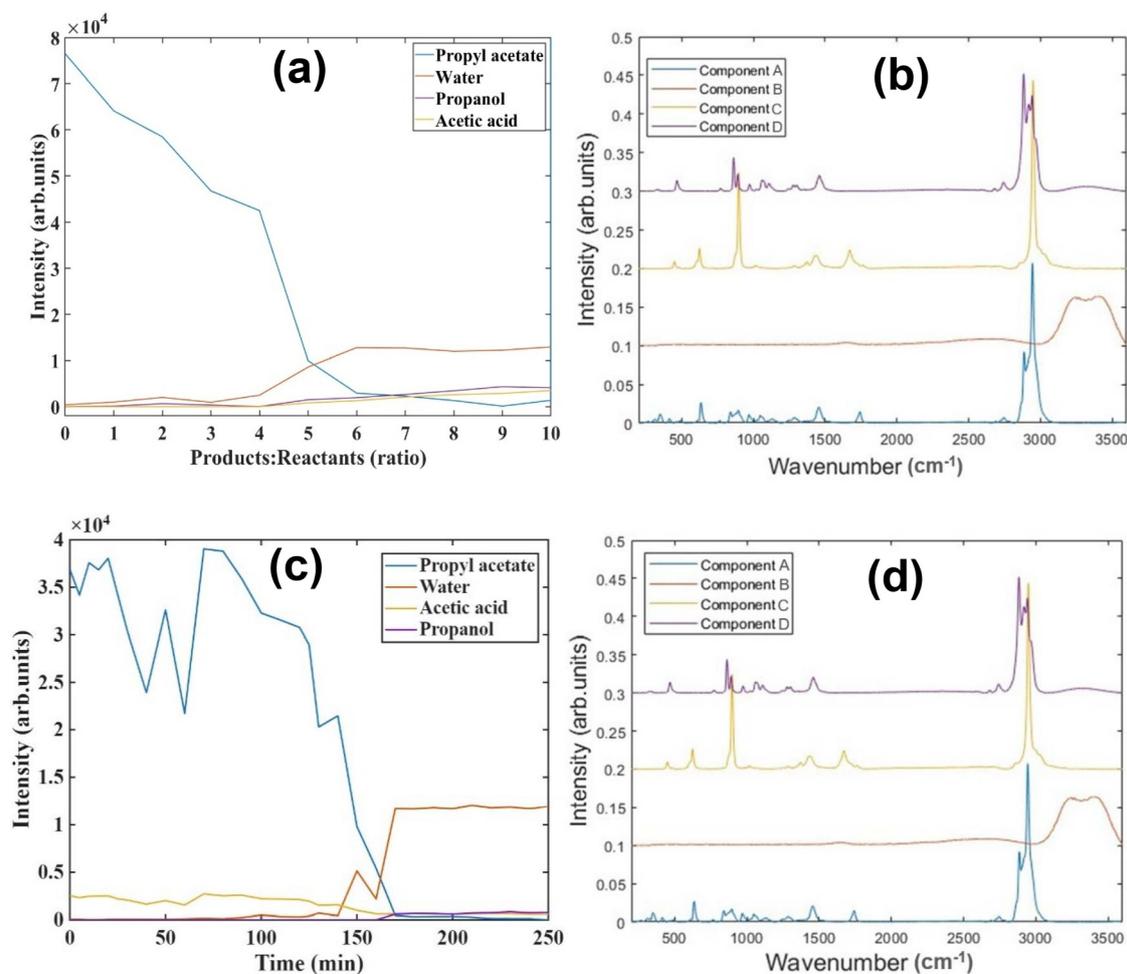


Fig. 9 Changing components profile with time for simulated admixtures data by seeded MCR-ALS (a), identified components spectra in the data by 10 000X weighted seeding of the components (b), changing components profile with time for reaction data by seeded MCR-ALS (c), identified reaction components in the data by 100X weighted seeding of the reaction components (d).



ponent seeding weight of 10 000X, and Fig. 9(b) shows the components identified by the seeded MCR-ALS. The components spectra at this seeding weight are nearly identical to the original spectra of the pure components, shown in Fig. 5(a). As the components in the simulated admixtures change from pure reactants to the pure products, the spectral changes are translated to concentration profiles by MCR-ALS accordingly. Similarly Fig. 9(c) shows the concentration profile for seeded MCR-ALS analysis of the reaction data at 100X components seeding and Fig. 9(d) shows the components identified by the seeded MCR-ALS. The components spectra are again quite similar to the original spectra of the reaction components, shown in Fig. 5(a). This suggests that the augmentation of data with component seeds at specific weights enhances the accurate identification of the components in the data and results in correct concentration changes over time.

### 3.6 Optimization of seeding

Accurate identification of the components in the data is crucial for correct kinetic profiling of the components in the MCR-ALS.<sup>36</sup> In this regard, the seeding weighting “X” was systematically varied to optimize the process. The identified components are fitted with the original spectra of the reaction components and the least squares fitting error values were monitored to ensure the correct identification of the components.<sup>37</sup> Minimal or close to zero error values ensure the accurate component identification in the data by algorithm.

#### 3.6.1 Seeding optimization for simulated admixtures.

Fig. 10 shows the seeding weight optimization of the components for simulated admixtures plotted *versus* SUMR values. Table 1 shows the change in SUMR values of each component as the seeding weight is increased. Further

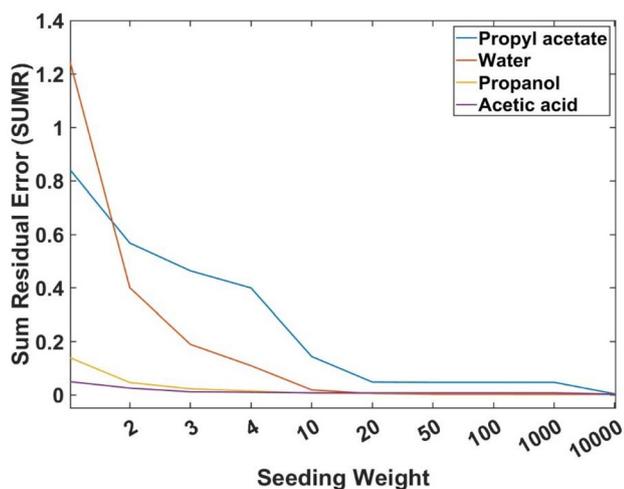


Fig. 10 Seeding weight optimization for simulated admixtures, seeding weight vs. SUMR.

Table 1 SUMR values for participating all 4 components in simulated admixtures at different seeding weights from 1–10 000

Seeding Weight (X)	SUMR component A (propyl acetate)	SUMR component B (water)	SUMR component C (propanol)	SUMR component D (acetic acid)
1	0.8456	1.2582	0.1408	0.0502
2	0.5681	0.4008	0.0466	0.0257
3	0.4642	0.1895	0.0231	0.0119
4	0.3999	0.1095	0.0148	0.0101
10	0.1441	0.0191	0.0068	0.0082
20	0.0484	0.0053	0.0059	0.0078
50	0.0474	0.0023	0.0059	0.0078
100	0.0474	0.0021	0.0059	0.0078
1000	0.0474	0.0021	0.0059	0.0078
10 000	0.0043	0.0021	0.0030	0.0034

increase of seed weights above 10 000X does not further reduce the SUMR values. The minimal SUMR values for all four components were obtained at 10 000X seeding weights, which are close to zero, 0.0043, 0.0021, 0.0030 and 0.0034 respectively. At this seeding weight, MCR-ALS is capable of accurately identifying the participating 4 components in the simulated admixtures.

Fig. 11(a)–(d) shows the comparison of original components spectra of the components fitted with the identified spectra of the components along with the residual spectra, multiplied by 10. The flat lines of the residual spectra confirm that the original spectra and the fitted spectra (identified by the MCR-ALS in data) are effectively identical, providing the proof of concept that seeding with an appropriately weighted seed, the algorithm can correctly identify the components that results in generation of accurate concentration profile/kinetics of the process.

**3.6.2 Seeding optimization for reaction.** Fig. 53S shows the seeding weight optimization of the components for propyl acetate hydrolysis reaction plotted *versus* SUMR values. Table 2 shows the change in SUMR values of each component as the seeding weight is increased up to 10 000X. SUMR values are minimal at 100X seed weights. The further increase in seed weights does not reduce the SUMR values. The SUMR values for all four components obtained at 100X seeding weights are 0.0474, 0.0021, 0.0058 and 0.0078 respectively. At this seeding weight, MCR-ALS is capable of accurately identifying the participating 4 components in the simulated admixtures.

Fig. 54S(a)–(d) shows the comparison of original components spectra of the reaction components fitted with the identified spectra of the components along with the residual spectra multiplied by 10. The straight lines of residual spectra confirm that the original spectra and the fitted spectra (identified by the MCR-ALS in data) are quite similar. Depicting the proof of concept that seeding with a proper weight augment the algorithm to correctly identify the components that results in generation of accurate concentration profile/kinetics of the process.



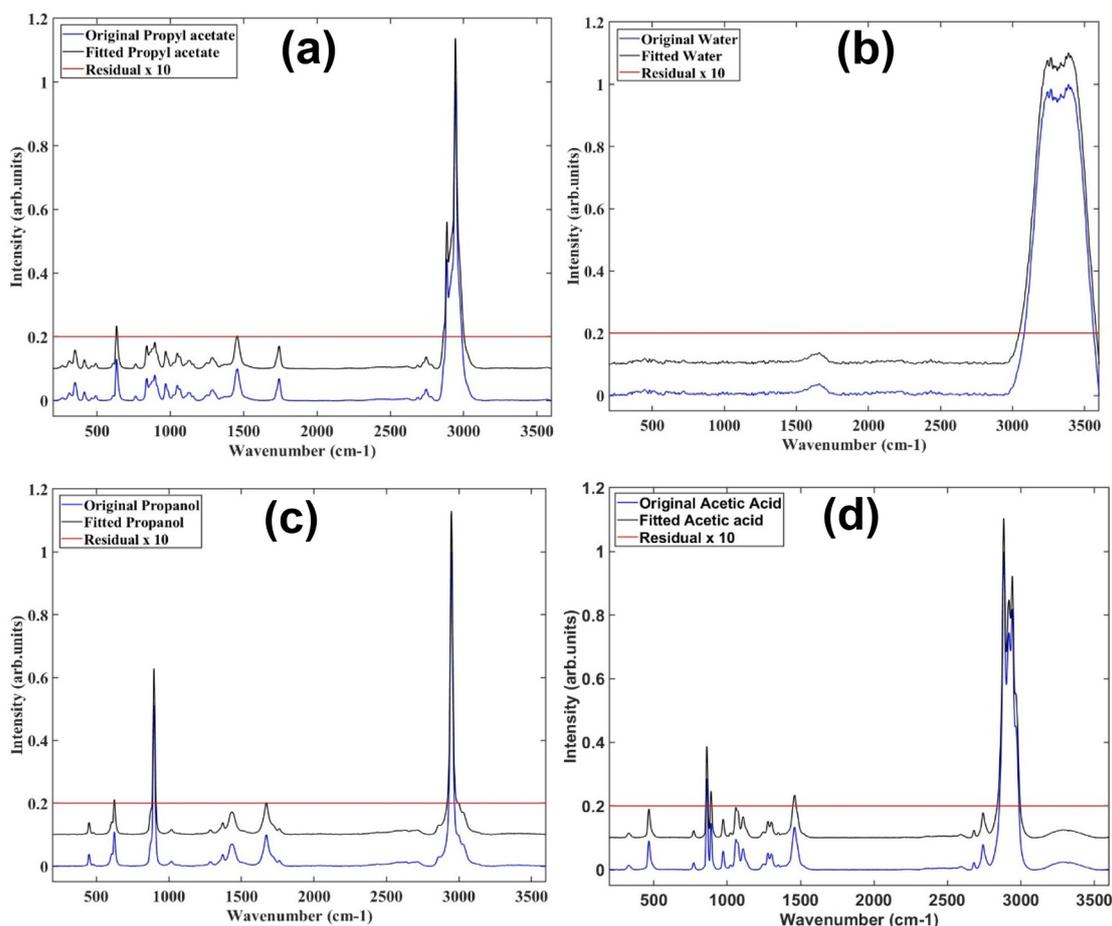


Fig. 11 Original components spectra vs. identified components spectra by seeded MCR-ALS along with 10X residual response spectra for simulated admixtures, offset for clarity (a) component A: propyl acetate (b) component B: water (c) component C: 1-propanol (d) component D: acetic acid.

Table 2 SUMR values for participating all 4 components in in hydrolysis reaction at different seeding weights from 1–10 000

Seeding weight (X)	SUMR component A (propyl acetate)	SUMR component B (water)	SUMR component C (propanol)	SUMR component D (acetic acid)
1	1.4247	1.9852	0.0620	0.1407
2	1.0224	0.6699	0.0521	0.0799
3	0.8674	0.3198	0.0470	0.0516
4	0.7752	0.1849	0.0437	0.0316
10	0.3386	0.0319	0.0205	0.0093
20	0.0511	0.0085	0.0059	0.0079
50	0.0475	0.0027	0.0058	0.0078
100	0.0474	0.0021	0.0058	0.0078
1000	0.0474	0.0021	0.0059	0.0078
10 000	0.0474	0.0021	0.0059	0.0078

## 4 Discussion

The primary objective of the current study was to explore the limitations of MCR-ALS and to evaluate the effect of seeding on the data analysis of kinetically varying mixtures, using the hydrolysis of propyl acetate reaction as a model example. The

study was initially performed at temperatures of 20 °C, 40 °C and 80 °C. At higher temperatures, the reaction proceeds more rapidly, but the risk of product loss through evaporation also increases. To avoid such losses, a closed container with a condenser system is required, which makes it difficult to record the Raman spectra *in situ*. A temperature of 20 °C was therefore chosen such that the Raman analysis could be performed *in situ*.

The initial analysis of the starting reactants indicated that the Raman spectrum of the mixture was not a weighted (1 : 10) linear combination of the spectra of propan-1-ol and water on their own, and analysis of the reaction mixture as it progressed gave the surprising result that the signatures of water emerged as though it were a reaction product. In order to investigate this phenomenon further, admixtures of systematically varied proportions of reactants and products were prepared to simulate the evolution of the reaction mixtures. No heat was applied to promote the reaction, however, and so the composition of these admixtures is precisely known. The spectra of the simulated mixtures were seen to faithfully reproduce the evolution of those of the reaction mixture, and therefore no further analysis was deemed necessary. Note, it is assumed



that the tetrahedral intermediate formed in ester hydrolysis is very short lived, low in concentration,<sup>38</sup> and therefore does not contribute significantly to the spectroscopic signatures, measured over seconds. Time resolved resonance and ultrafast Raman spectroscopies are used for reaction intermediate detection.<sup>39</sup>

The observed behavior can be attributed to the change in polarizability due to changes in intermolecular water interactions in the presence of a solute.<sup>40</sup> Studies have consistently demonstrated that the presence of solutes can alter the local hydrogen-bond network of water, resulting in shifts, intensity changes, and band-shape modifications in the O–H stretching and bending regions of the vibrational spectrum.<sup>41</sup> While propyl acetate is nonpolar and sparingly soluble in water, both the propanol and acetic acid products are capable of hydrogen bonding and thus do not significantly perturb the local clustering of the bulk water, resulting in the reappearance of water signals.<sup>42</sup>

Although the NLS approach can resolve the evolution of the reaction products, MCR-ALS fails to do so, essentially because of the issue of rank deficiency, as previously highlighted by de Juan *et al.*<sup>20</sup> The rank of a matrix describing  $R$  reactions involving  $S$  species is the minimum of  $(R,S)$ . In the case of the reaction  $A > B + C$ , for example, the rank is two, which is insufficient for the SVD process to identify that there are three independent species, and the analysis rather returns a representation of the evolution of their mixture. Such rank deficiency has been overcome by, for example, augmenting the data by independent pathways, essentially an additional parallel reaction to increase the rank of the matrix.<sup>43,44</sup> Examples include the parallel use of chromatographic and spectroscopic data,<sup>26,45</sup> and augmentation of the dataset of a parallel reaction pathway with that of a model compound reaction.<sup>27</sup>

As an alternative approach, the process of seeding the multivariate analysis with weighted spectra of the known constituent components can be employed. The process of seeding multivariate analysis was first described in the context of Principal Component Analysis (PCA),<sup>29,46</sup> which is one of the methodologies commonly used in the initial SVD stages of MCR-ALS. In the case of rank deficiency, SVD will not distinguish two simultaneously evolving components. However, the process of seeding biases the variance of the data set towards one component such that they are more easily distinguishable by PCA. Note that, in the process, the data is mean-centred, and thus the seed will bias the mean, and the mean-centred “score” of each datapoint in the comparative analysis. In Fig. 10 the effect of systematically varying the seeding weight on the resolved component fitting error has been shown for each of the components, although the same weighting was employed for each, at all points in the analysis. It can be seen that the propyl acetate requires the highest seeding weight, while acetic acid requires the lowest for a similar improvement in accuracy of resolution of the component. While further examination is required, including independently varied weightings, which of the species requires

higher or lower seeding weight may depend upon the complexity of the spectrum, and/or the similarity of the spectra of simultaneously evolving components.

The approach of seeding the MCR-ALS resulted in the successful resolutions of the reaction components, and their kinetic evolution, promising potential in a range of applications, potentially extending to systems biology and metabolic flux analysis (fluxomics),<sup>47–49</sup> although application of the MCR-ALS technique to evolving spectra of the *in vitro* intracellular<sup>50</sup> or extracellular<sup>51,52</sup> environment may pose additional challenges.

## 5 Conclusion

This study demonstrates the potential and limitations of Raman spectroscopy and chemometrics for real-time monitoring of chemical changes in a multiproduct reaction. Using propyl acetate hydrolysis as a model system, it is demonstrated that Raman spectroscopy can monitor the progress of the reaction, *in situ*, with applications in process analytical technology for pharmaceutical and bioreactors, for example. However, the ability to specifically identify the profiles of individual, simultaneously evolving components and their kinetic evolution using multivariate chemometric techniques is limited by rank deficiency. It is demonstrated that biasing the variance at the initial stages of singular value decomposition by seeding the data set with the known spectra of the constituent components enables an accurate determination of their relative concentrations at the different stages of the reaction, and the reaction kinetics. Such “seeding” approaches may considerably impact the potential of such label free spectroscopic techniques in process monitoring. While both problem-based nonlinear squares fitting and MCR-ALS provided insight into the spectral and concentration trends of reactants and products, neither method alone was sufficient for precise resolution of evolving species. By employing a seeded MCR-ALS approach with optimized seeding weights, the analysis achieved improved spectral unmixing and more reliable concentration profiles, reducing residual errors and enhancing kinetic interpretation. These findings highlight the importance of data augmentation strategies in overcoming the challenges of spectroscopic monitoring of complex reaction mixtures. The proposed methodology offers a promising pathway for advancing the application of Raman spectroscopy in metabolic studies, cascade reactions, and broader process analytical contexts.

## Author contributions

Hugh Byrne: supervision; funding acquisition; conceptualization; methodology; writing – review and editing. Muhammad Kashif: investigation; data curation; writing original draft; conceptualization; methodology; formal analysis. Mark Keating:



conceptualization; formal analysis; writing – review and editing.

## Conflicts of interest

Authors declare no conflict of interests.

## Data availability

Data for this article are available at Zenodo.org at <https://doi.org/10.5281/zenodo.17252250>.

Supplementary information is available. See DOI: <https://doi.org/10.1039/d5an01055a>.

## Acknowledgements

MK is supported by Research Ireland Frontiers for the Future Award 20/FFPP/8517. MEK is supported by Research Ireland Frontiers for Partnership Award 21/FFP-A/9091.

## References

- 1 E. Smith and G. Dent, *Modern Raman spectroscopy : a practical approach*, John Wiley & Sons Ltd, 2019.
- 2 R. S. Das and Y. K. Agrawal, *Vib. Spectrosc.*, 2011, **57**, 163–176.
- 3 B. A. Nogueira and C. Castiglioni, *Crystals*, 2024, **14**, 251.
- 4 D. Cui, L. Kong, Y. Wang, Y. Zhu and C. Zhang, *Environ. Sci. Ecotechnol.*, 2022, **11**, 100187.
- 5 G. Cutshaw, S. Uthaman, N. Hassan, S. Kothadiya, X. Wen and R. Bardhan, *Chem. Rev.*, 2023, **123**, 8297–8346.
- 6 M. Chen, D. Liu, L. Qiao, P. Zhou, J. Feng, K. W. Ng, Q. Liu, S. Wang and H. Pan, *Chem. Eng. J.*, 2023, **461**, 141939.
- 7 M. H. G. Costa, I. Carrondo, I. A. Isidro and M. Serra, *Biotechnol. Adv.*, 2024, **77**, 108472.
- 8 A. Silge, K. Weber, D. Cialla-May, L. Müller-Böttcher, D. Fischer and J. Popp, *TrAC, Trends Anal. Chem.*, 2022, **153**, 116623.
- 9 C. Hess, *Chem. Soc. Rev.*, 2021, **50**, 3519–3564.
- 10 S. Nakashima, T. Ogura and T. Kitagawa, *Biochim. Biophys. Acta, Bioenerg.*, 2015, **1847**, 86–97.
- 11 A. Yao-Lin, H. J. Ze, X. T. Wang, Y. Zhang, H. Zhang, Q. N. Zheng, J. C. Dong, J. H. Tian and J. F. Li, *Curr. Opin. Electrochem.*, 2023, **42**, 101381.
- 12 K. A. Esmonde-White, M. Cuellar, C. Uerpmann, B. Lenain and I. R. Lewis, *Anal. Bioanal. Chem.*, 2017, **409**, 637–649.
- 13 Y. Tang, X. Wang, G. Zhou, S. Guo, Z. Li, Y. Hu and W. Li, *J. Anal. Test.*, 2024, **9**, 136–152.
- 14 N. Yang, PhD Thesis, Université Paris-Saclay, 2024.
- 15 Y. K. Lin, H. Y. Leong, T. C. Ling, D. Q. Lin and S. J. Yao, *Chin. J. Chem. Eng.*, 2021, **30**, 204–211.
- 16 S. Dzurendova, P. M. Olsen, D. Byrtusová, V. Tafintseva, V. Shapaval, S. J. Horn, A. Kohler, M. Sotkowski, I. Marova and B. Zimmermann, *Microb. Cell Fact.*, 2023, **22**, 1–14.
- 17 A. S. Rathore, M. Pathak, R. Jain and G. P. S. Jadaun, *AAPS J.*, 2016, **18**, 793–800.
- 18 N. Patil, Z. Mirveis and H. J. Byrne, *Spectrochim. Acta, Part A*, 2025, **340**, 126363.
- 19 A. de Juan and R. Tauler, *Anal. Chim. Acta*, 2021, **1145**, 59–78.
- 20 A. De Juan, M. Maeder, M. Martínez and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2000, **54**, 123–141.
- 21 D. Perez-Guaita, G. Quintas, Z. Farhane, R. Tauler and H. J. Byrne, *Talanta*, 2020, **208**, 120386.
- 22 D. Pérez-Guaita, G. Quintás, Z. Farhane, R. Tauler and H. J. Byrne, *Cells*, 2022, **11**, 1555.
- 23 V. Notarstefano, A. Belloni, P. Mariani, G. Orilisi, G. Orsini, E. Giorgini and H. J. Byrne, *Analyst*, 2023, **148**, 4365–4372.
- 24 F. Ravera, E. Efeoglu and H. J. Byrne, *Analyst*, 2025, **150**, 3445–3456.
- 25 A. De Juan, M. Maeder, M. Martínez and R. Tauler, *Chemom. Intell. Lab. Syst.*, 2000, **54**, 123–141.
- 26 A. Jayaraman, S. Mas, R. Tauler and A. de Juan, *J. Chromatogr. B:Anal. Technol. Biomed. Life Sci.*, 2012, **910**, 138–148.
- 27 N. Mouton, O. Devos, M. Sliwa, A. de Juan and C. Ruckebusch, *Anal. Chim. Acta*, 2013, **788**, 8–16.
- 28 G. M. Escandar and A. C. Olivieri, *J. Chromatogr. A*, 2019, **1587**, 2–13.
- 29 M. E. Keating and H. J. Byrne, *Spectrochim. Acta, Part A*, 2025, **340**, 126369.
- 30 K. Yates, R. A. McClelland, M. L. Bender, R. D. Ginger, J. P. Uink, C. A. Bunton, E. D. Hughes and C. K. Ingold, *J. Am. Chem. Soc.*, 1966, **75**, 35.
- 31 R. Luce, P. Hildebrandt, U. Kuhlmann and J. Liesen, *Appl. Spectrosc.*, 2016, **70**, 1464–1475.
- 32 Y. Sun, W. Wu and J. Xin, *Signal, Image and Video Process.*, 2017, **11**, 651–658.
- 33 A. Stella, F. Bonnier, A. Tfayli, F. Yvergnaux, H. J. Byrne, I. Chourpa, E. Munnier and C. Tauber, *J. Biophotonics*, 2020, **13**, e202000136.
- 34 M. C. Valeriano, A. M. Neto, A. C. F. Batista and M. B. Mamián-López, *Spectrochim. Acta, Part A*, 2024, **318**, 124498.
- 35 J. P. Smith, F. C. Smith, J. Ottaway, A. E. Krull-Davatzes, B. M. Simonson, B. P. Glass and K. S. Booksh, *Appl. Spectrosc.*, 2017, **71**, 1816–1833.
- 36 H. Motegi, Y. Tsuboi, A. Saga, T. Kagami, M. Inoue, H. Toki, O. Minowa, T. Noda and J. Kikuchi, *Sci. Rep.*, 2015, **5**, 15710.
- 37 D. Y. Lin, L. J. Wei and Z. Ying, *Biometrics*, 2002, **58**, 1–12.
- 38 R. A. McClelland and L. J. Santry, *Acc. Chem. Res.*, 1983, **16**, 394–399.
- 39 S. Umopathy, S. K. Sahoo and A. W. Parker, *Appl. Spectrosc.*, 2011, **65**(10), 1087–1115.
- 40 N. da Silva, L. A. Ferreira, A. I. Belgovskiy, P. P. Madeira, J. A. Teixeira, E. K. Mann, J. Adin Mann Jr, W. V. Meyer,



- A. E. Smart, V. Y. Chernyak, V. N. Uversky and B. Y. Zaslavsky, *J. Mol. Liq.*, 2021, **335**, 116288.
- 41 A. Jumabaev, S. J. Koyambo-Konzapa, H. Hushvaktov, A. Absanov, B. Khudaykulov, U. Holikulov, Z. Ernazarov, N. Issaoui, O. M. Al-Dossary and M. Nsangou, *J. Mol. Model.*, 2024, **30**(10), 349.
- 42 F. Hammami, N. Issaoui and S. Nasr, *Comput. Theor. Chem.*, 2021, **1199**, 113218.
- 43 A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou and R. Tauler, *Chemom. Intell. Lab. Syst.*, 1997, **38**, 183–196.
- 44 J. Saurina, S. Hernández-Cassou, R. Tauler and A. Izquierdo-Ridorsa, *J. Chemom.*, 1998, **12**, 183–203.
- 45 R. Tauler, *Tech. Instrum. Anal. Chem.*, 2000, **21**, 689–736.
- 46 B. Gardner, J. Haskell, P. Matousek and N. Stone, *Analyst*, 2023, **149**, 205–211.
- 47 M. Cascante and S. Marin, *Essays Biochem.*, 2008, **45**, 67–81.
- 48 G. Winter and J. O. Krömer, *Environ. Microbiol.*, 2013, **15**, 1901–1916.
- 49 S. Cortassa, V. Caceres, L. N. Bell, B. O'Rourke, N. Paolucci and M. A. Aon, *Biophys. J.*, 2015, **108**, 163–172.
- 50 N. Patil, Z. Mirveis and H. J. Byrne, *Spectrochim. Acta, Part A*, 2026, **348**, 127156.
- 51 N. Patil, Z. Mirveis and H. J. Byrne, *Spectrochim. Acta, Part A*, 2025, **340**, 126363.
- 52 Z. Mirveis, N. Patil and H. J. Byrne, *Spectrochim. Acta, Part A*, 2025, **340**, 126308.

