JOURNAL OF CHEMOMETRICS
J. Chemometrics **13,** 331–341 (1999)

THREE-WAY DATA ANALYSIS OF POLLUTANT DEGRADATION PROFILES MONITORED USING LIQUID CHROMATOGRAPHY–DIODE ARRAY DETECTION

KIM JOHNSON,[1] ANNA DE JUAN[2] AND SARAH C. RUTAN[1]*
[1]*Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284-2006, USA*
[2]*Departament de Química Analítica, Universitat de Barcelona, E-08028 Barcelona, Spain*

SUMMARY

There are a wide variety of environmental contaminants that need to be better characterized. It is important to know the degradation pathways for these pollutants in order to better understand their effect on the environment. The hydrolysis of Glean® (chlorsulfuron), a sulfonylurea herbicide, is currently under study in this laboratory. Liquid chromatography coupled with diode array detection (LC–DAD) offers a means of studying these types of reactions. With this approach, three-way data are obtained—absorbance measurements as a function of chromatographic retention time, UV wavelength and reaction time. The direct trilinear decomposition (DTD) algorithm was applied to these data, presuming an internal trilinear structure in the data set. Upon inspection of the results, however, deviations from trilinearity were found that caused chemically meaningless results to be obtained using this approach. This difficulty was addressed by using the three-way multivariate curve resolution–alternating least squares approach (MCR–ALS). This technique allows for the optional use of the trilinearity constraint, and some additional constraints related to the features of the chromatographic and spectral profiles can be included. In the present work the application of this approach has been evaluated for resolving the overlapped responses that arise when measuring herbicide degradation reaction profiles. Copyright © 1999 John Wiley & Sons, Ltd.

KEY WORDS: three-way analysis; direct trilinear decomposition (DTD); multivariate curve resolution–alternating least squares (MCR–ALS); LC–DAD; pollutant degradation

INTRODUCTION

Over the past several years there has been increasing interest in the application of three-way data analysis methods.[1] These techniques allow chemists to obtain information from more complex experiments. Three-way approaches have increased resolving power relative to two-way data analysis techniques, allowing for qualitative and quantitative analyses.

In this work, three-way data analysis has been used to study an environmentally relevant reaction—the hydrolysis of Glean®, a low-application use sulfonylurea (SU) herbicide. The three-way data sets are formed using liquid chromatography–diode array detection (LC–DAD) matrices of the Glean samples obtained at different stages during the degradation process.

* Correspondence to: S. C. Rutan, Department of Chemistry, Virginia Commonwealth University, Richmond, VA 23284-2006, USA. E-mail: srutan@saturn.vcu.edu
Contract/grant sponsor: National Science Foundation; Contract/grant number: CHE-9709437.
Contract/grant sponsor: Department of Energy; Contract/grant number: DE-FG-88ER 13833.

CCC 0886–9383/99/030331–11 $17.50

Received 20 July 1998
Revised 5 November 1998
Accepted 11 February 1999

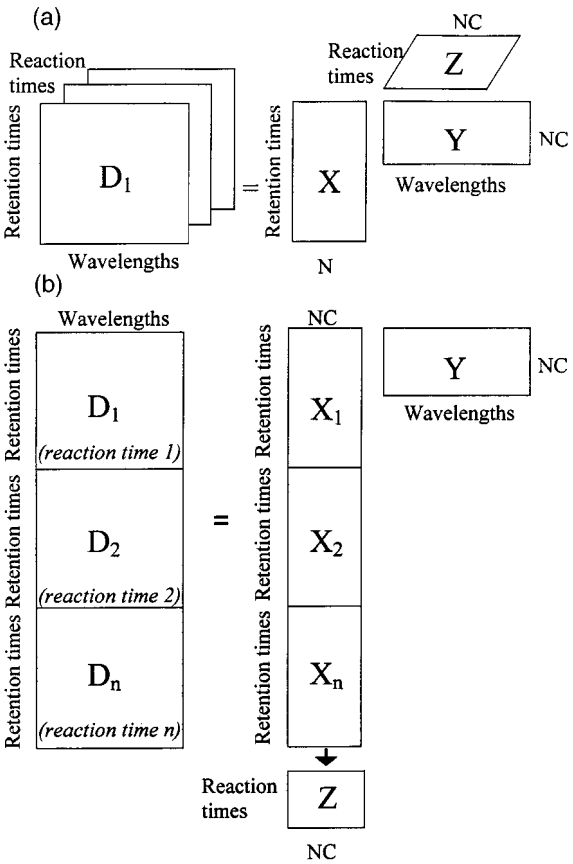

Figure 1. Decomposition of three-way data set according to (a) direct trilinear decomposition (DTD) and (b) multivariate curve resolution–alternating least squares (MCR–ALS). The D_i are the matrices that form the three-way data set, X and X_i are the matrices of elution profiles, Y is the matrix of spectra and Z is the matrix of kinetic profiles. NC is the number of components

Glean (2-chloro-*N*-[(4methoxy-6-methyl-1,3,5-triazin-2-yl)aminocarbonyl]benzenesulfonamide), or chlorsulfuron, is one of eight sulfonylurea herbicides currently under investigation in our laboratory. Glean is used to control weeds in cereal grain fields, such as wheat, barley and oats.[2] Unfortunately, tiny drift amounts of the SUs onto non-target crops have been shown to cause a drastic reduction in crop yields without affecting the plant growth or appearance.[3–5] The degradation of these herbicides is of importance in understanding how they affect these non-target crops. Hydrolysis is one of the most important degradation pathways of the SU herbicides.[6–15] Although many degradation studies have been done, most rely on measuring the disappearance of the SU herbicides and do not identify or quantify the hydrolysis products. For example, Dinelli *et al.* showed that capillary electrophoresis could be used to separate and monitor the metabolites of the SUs, but the compounds were not identified.[7] Reiser *et al.* have proposed some reaction mechanisms for the hydrolysis of Glean, and some of the intermediates and products have been identified by fast atom bombardment mass spectrometry.[8] Braschi *et al.* followed the hydrolysis of triasulfuron by LC with UV detection at 224 nm. The structures of the various metabolites were confirmed by synthesis and purification of the compounds, followed by mass spectral and infrared characterization.[10] However, no kinetic

measurements were undertaken in either of these studies. Degradation reactions of other pesticides have been characterized by measuring spectra at different pHs and reaction times using three-way analysis methods.[16]

In this work the LC–DAD data obtained during the hydrolysis reaction are analyzed using three-way data analysis methods. This approach allows the spectrochromatograms of the overlapped degradation products to be resolved. One of the long-term goals of our research is to develop general chemometric approaches for resolution of LC–DAD data of reactions. This new approach will help to eliminate the extensive LC method development typically required for characterization of complex reactions. In addition, these methods should be useful for studying commercial herbicide preparations or for following reactions *in situ*, i.e. reactions in soil or ground water.

LC–DAD provides *bilinear* two-way data (UV absorbance versus retention time and wavelength), i.e. each compound in the solution can be described as the outer product of one chromatographic profile and one spectrum. When LC–DAD matrices collected at different reaction times are coupled together, the resulting three-way data set theoretically presents a *trilinear* structure. A three-way data set D that is trilinear can be expressed according to the model[1]

$$\mathbf{D} = \mathbf{X} \otimes \mathbf{Y} \otimes \mathbf{Z} \tag{1}$$

where **X** is the matrix of chromatographic elution profiles, **Y** is the matrix of spectra and **Z** is the matrix of chemical degradation profiles. Chemically speaking, the existence of a trilinear structure would imply that the shapes of the spectrum and the elution profile of each chemical compound remain invariant in all the chromatographic runs recorded during the degradation process. While this can generally be assumed for the spectral profiles, there are normally more significant variations in the elution patterns between different chromatographic runs.

For strictly trilinear three-way data of low rank, direct trilinear decomposition (DTD) is usually considered to be the method of choice.[17] This method requires only knowledge of the number of independent contributions to the data set, which can include chemical species as well as background contributions. DTD always provides a unique solution, which coincides with the real one when the system is trilinear.[18] Unfortunately, LC–DAD data are seldom perfectly trilinear and DTD only yields chemically reasonable results in situations where the departures from the trilinear structure are very minor.[19]

Multivariate curve resolution using alternating least squares (MCR–ALS) can cope with non-trilinear three-way data sets because a different procedure is used for the data decomposition, as shown in Figure 1.[19,20] This procedure allows chromatograms of the individual species to vary from experiment to experiment, both in terms of shape and position, while the spectral profiles are assumed to be consistent. This relaxation of the trilinearity constraint implies that unique solutions cannot be guaranteed unless features such as selectivity are also present in the data set. However, application of constraints related to mathematical features (e.g. selectivity) or to properties of the chromatographic and spectral profiles (e.g. non-negativity, unimodality) decreases drastically the ambiguities in the solution and guides the iterative optimization process toward a chemically meaningful solution.

In this work, DTD and MCR–ALS have been applied to resolve the degradation profiles of the sulfonylurea herbicide Glean. Chemometric and chemical conclusions are inferred from the application of these two three-way methods of analysis.

EXPERIMENTAL

All experiments were done using a Hewlett Packard 1090 system LC equipped with a DAD. The column used was a 150 mm $\times$ 4·6 mm LUNA C-18 (Phenomenex) with 5 μm particles. The mobile

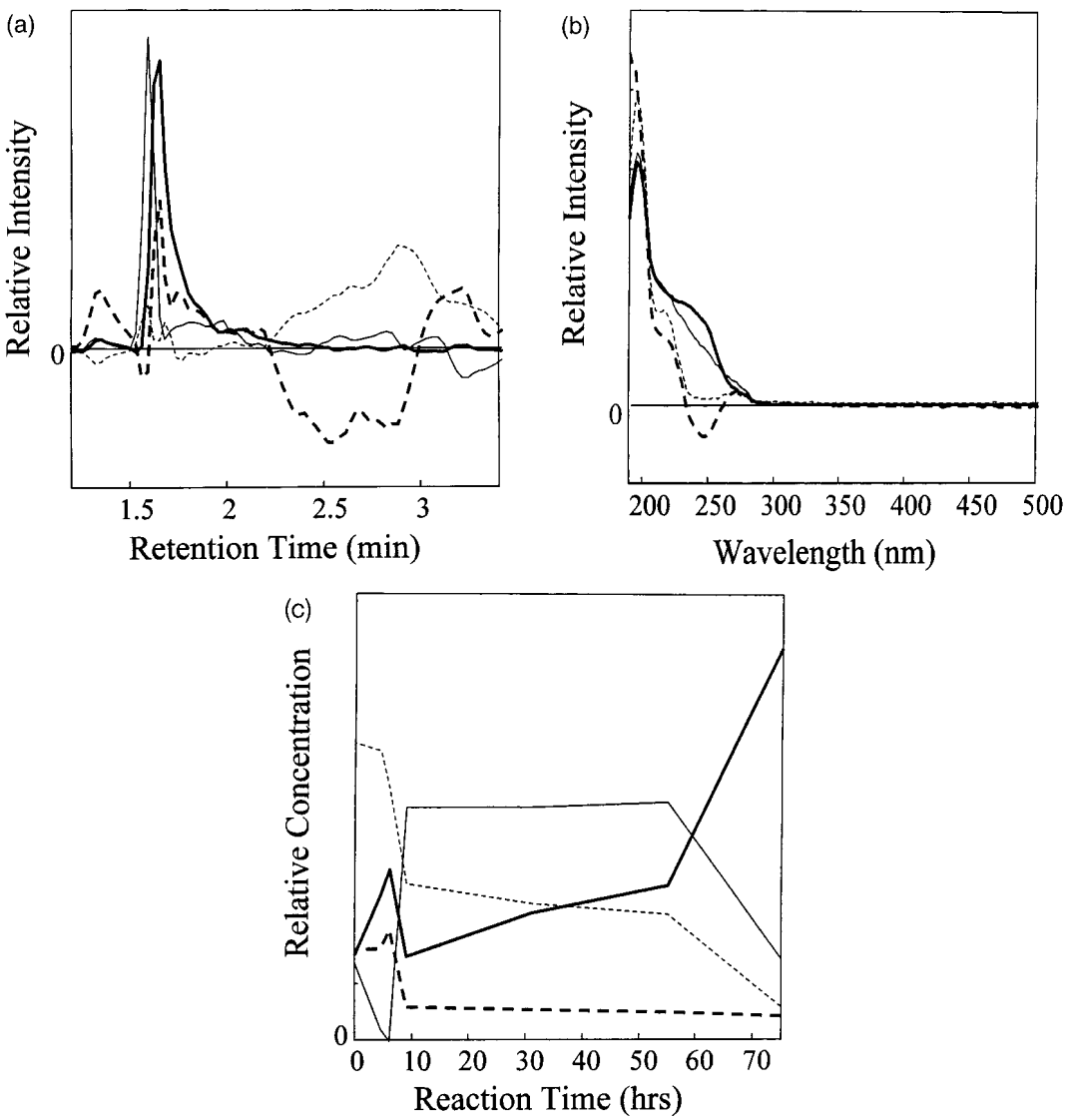

Figure 2. DTD results: (a) chromatographic profiles; (b) spectral profiles; (c) kinetic profiles (see Table 1 for component labels)

phase was 80% acetonitrile/water. Acetonitrile was HPLC grade from EM Science, and deionized water was filtered through a 0·4 μm filter before use. Injections were made by overfilling a 20 μl sample loop.

Data were collected using the HP 3-D Chem Station® software. Data were converted to ASCII format spectra using the LC3D.EXE program provided by Hewlett Packard as part of the user-contributed program package, or by using the dataout9.mac provided by R. Giuffre at Hewlett Packard.

The solutions for studying the hydrolysis of Glean were prepared by dissolving 25 μg ml^{-1} of the

herbicide in filtered, deionized water. An LC–DAD data matrix was obtained for the Glean solution prior to heating. The solution was placed in a closed sample vial and then in an oven heated to 55 °C. LC–DAD data were obtained after 4·5, 6, 9, 31, 55 and 75 h. The pH of these unbuffered solutions was approximately 6·5. Chromatographic data in the range from 1·2 to 3·4 min were obtained at 1·76 s intervals, and absorbance data were collected from 190 to 500 nm at 2 nm intervals.

Data analysis was carried out using programs written in Matlab, version 4·2.c.1 (Mathworks) that have been described previously.[19] Initial guesses for the MCR–ALS procedure were generated using DTD or evolving factor analysis (EFA).[21] EFA provides approximate profiles for components in systems where species appear and disappear in a sequential fashion, such as liquid chromatography.[22] MCR–ALS was then carried out on each individual data set, and finally on all data sets simultaneously.

Gas chromatography–mass spectrometry results were obtained using a Hewlett Packard GCD instrument. A linear temperature program from 70 to 230 °C at 20 °C min^{-1} was used with an initial time of 2·00 min. The inlet and detector temperatures were 250 and 280 °C respectively.

RESULTS AND DISCUSSION

The LC–DAD data matrices obtained at the different degradation times were organized to form a three-way data set of size $75 \times 157 \times 7$, where the first value refers to the number of retention time points in each chromatographic run, the second is the number of wavelength points in each spectrum, and the third is the number of chromatographic runs collected at different stages (times) of the kinetic degradation process.

Before performing the detailed analysis of the three-way data set, singular value decomposition (SVD) was carried out on each of the seven LC–DAD data matrices to get an initial guess of the rank of the data. The number of compounds ranged from three to four depending on the data matrix analyzed.

DTD analysis

Because of the simplicity of the method, DTD was first applied to resolve the three-way data set to get an initial idea as to the complexity of the data set. This method was run several times assuming the total rank of the system to be equal to three, four or five components. For each of these rank values, DTD was carried out by assigning the Z matrix to the chromatographic, spectral or kinetic profiles. The best results, in terms of chemically meaningful profiles and best data fit, were obtained by assuming four components, with the Z matrix assigned to the spectral direction (i.e. the spectral variations were the least informative). The lack of fit associated with these results was 23·8%, and the plots of the chromatographic, spectroscopic and kinetic profiles are shown in Figure 2. However, for these results, chemical inconsistencies were found in the shape of the profiles. Both the UV spectra and the chromatographic profiles showed negative regions, and the chromatographic profiles for single compounds had multiple peaks (Figures 2(a) and 2(b)). In addition, the kinetic profiles showed no trends consistent with degradation of the reactant or formation of intermediates or products (Figure 2(c)). These problems indicated that the data were not trilinear.

MCR–ALS analysis

Given the non-trilinear structure of the data set, MCR–ALS was selected as a more appropriate three-way analysis method for dealing with the Glean degradation data. The first step was to resolve each LC–DAD data matrix individually. Owing to the lack of prior spectral information and the higher resolution generally available in the chromatographic direction, MCR–ALS was run using

Table 1. Species present in mixtures (species correspondence)

Retention time (min):	1·7	1·7	2·9	2·4
Graphic representation:	Thick full line	Thin full line	Thick broken line	Thin broken line
Time (h)	(■)	(◆)	(●)	(▲)
0	present	present	absent	absent
4·5	present	present	present	present
6	present	present	present	present
9	present	present	present	present
31	absent	present	present	present
55	absent	present	present	present
75	absent	present	present	present

chromatographic initial estimates derived from EFA.[22] The results obtained provided an indication of which species were present in each mixture and also gave more chemically reasonable chromatographic profiles than the abstract EFA estimates. Some data matrices had to be resolved using four components, and the others gave similar results using either three or four components.

In view of the previous results, the initial estimates for the resolution of the three-way data set were built by linking the four-component chromatographic matrices (i.e. the X_i matrices) obtained as a result of each of the individual MCR–ALS analyses mentioned above. The columns in the X_i matrices were rearranged so that each column in the unfolded chromatographic matrix $[X_1; X_2; \ldots; X_n]$ represented the component spectra with the most similar shapes and chromatograms with similar retention times (chromatographic conditions were kept constant for all experimental runs).

Initially it was assumed that all four components were present in all seven data matrices. However, the results of the individual MCR–ALS analyses suggested that all components were not present in all matrices (some matrices could be reasonably resolved with less than four compounds). In addition, some components showed a minimal contribution and/or a dramatically shifted position in the chromatographic profiles obtained from the three-way resolution results. Thus the constraint of correspondence between species[21] (which defines the species present in each of the appended chromatographic runs) was applied, and certain components were assumed to be absent in some data matrices. Several logical variations were tried, and the best overall result gave chemically reasonable profiles and a lack of fit of 13·5%. This result corresponds to the distribution of components shown in Table 1. The related chromatographic, spectral and kinetic profiles obtained using these assumptions are shown in Figure 3. The residuals were decidedly non-random; there was no evidence of a trend in the residuals as a function of reaction time, therefore it can be assumed that the unmodeled components that show up in the residuals are not related to the reaction of interest. These components could likely come from some minor impurities present in the commercial pesticide, whose modelling would not provide any essential information about the reaction.

Only small variations were observed in the resolution results when MCR–ALS was run assuming species distributions slightly different from that shown in Table 1. The spectral shapes were similar, the chromatograms for the species initially assumed to be absent were small, and there were no significant changes in the lack of fit. These observations indicate that the resolution method and species distributions yielded valid results.

When looking at the chromatographic, spectral and kinetics profiles shown in Figure 3, the most striking feature is the presence of two chromatographic peaks practically identical in shape and eluting at virtually the same time (∼1·7 min). Both these peaks were present in the spectro-chromatogram of the starting material. The presence of more than one profile in the elution region of a main peak is often an artifact produced by the presence of heteroscedastic noise in the data set or by spectral distortions associated with finite DAD scan time problems.[23,24] A further study of the

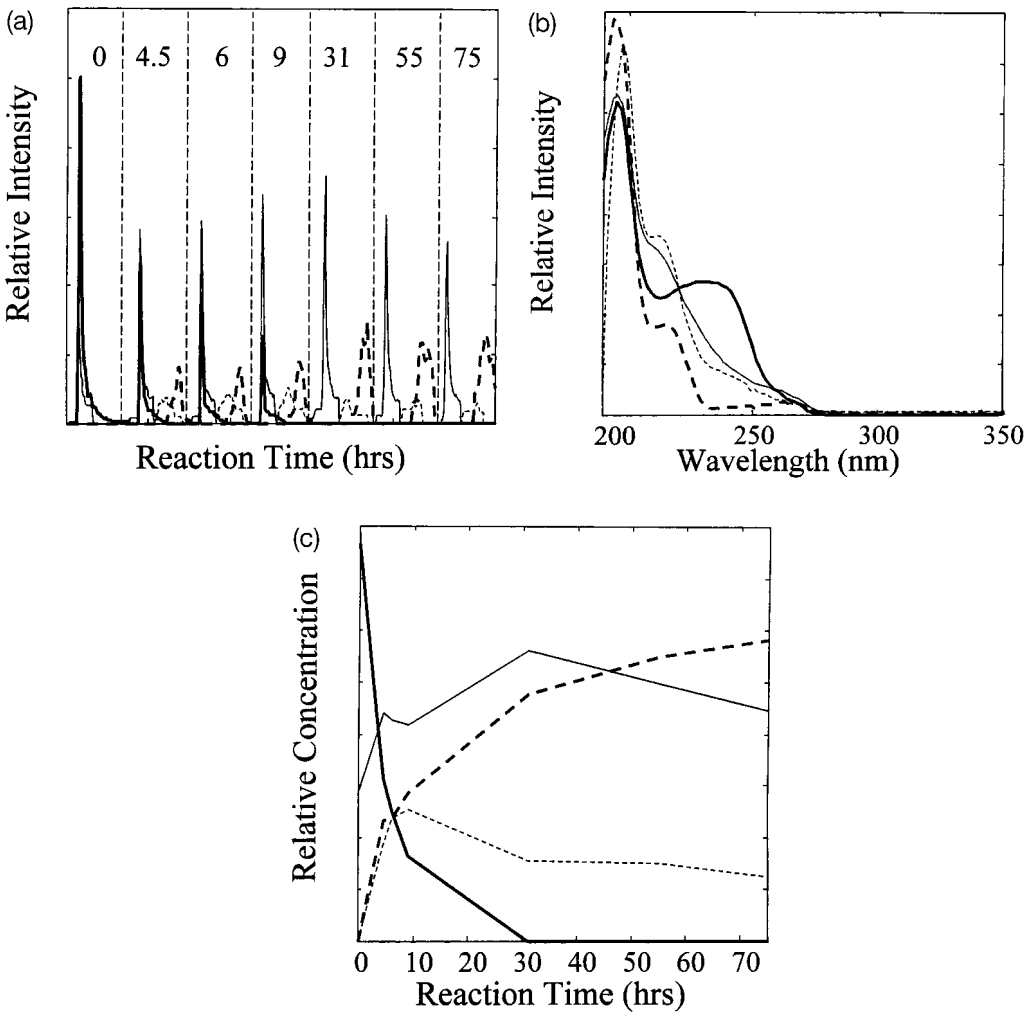

Figure 3. MCR–ALS results using chromatographic initial guesses: (a) chromatographic profiles; (b) spectral profiles; (c) kinetic profiles (see Table 1 for component labels)

degradation data set was carried out to determine if the two coincident chromatographic profiles could be really associated with the presence of two different chemical species.

The first and simplest check consisted of plotting the spectra collected at the peak maxima (~1·7 min) for each of the LC–DAD data matrices collected during the degradation process. The spectra shown in Figure 4 reveal a consistent and significant shape variation in the spectra obtained during the degradation process and confirmed that more than one chemical species was eluting in this region.

The overlap of these two chromatographic profiles suggested that it would be reasonable to use spectra instead of chromatographic profiles for the initial guesses. The spectral estimates for the two compounds eluting together were built by selecting the spectrum of the peak maximum at around 1·7 min in the first and last LC–DAD data matrices; the spectra for the other two compounds in the data set were built as the mean spectra from the elution region of these two last components in the first and last LC–DAD matrices (found at around 2·4 and 2·9 min respectively).

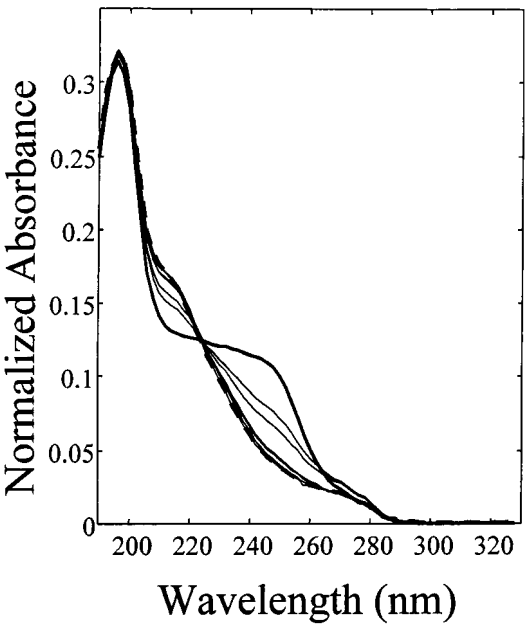

Figure 4. Normalized spectra obtained at maximum of main peak ($t_R \approx 1{\cdot}7$ min) for LC–DAD data matrices collected during degradation process. The spectrum recorded before the beginning of the degradation process is marked with a thick full line. The last spectrum, recorded at a reaction time equal to 75 h, is plotted with a thick broken line. The rest of the spectra are related to intermediate times during the reaction and evolve consistently with the reaction time

The MCR–ALS results obtained by using these spectral initial estimates agree quite well with those obtained using chromatographic initial estimates. These results can be seen in Figure 5, where the spectra obtained using the two approaches are overlaid. The similarity in the MCR–ALS profiles recovered from initial estimates based on chemical insight (the spectral estimates) and from estimates constructed by using data analysis (the chromatographic estimates obtained using EFA) supports the validity of these results.

Our original plan was to obtain mass spectra 'on-the-fly' subsequent to the LC–DAD analysis to aid in the identification of the reacting species. However, our particle beam LC–MS interface lacks the sensitivity to detect any of the species resolved from the LC–DAD data. We then turned to the mechanism proposed by Reiser *et al.*,[8] shown in Figure 6, and some GC–MS studies to further characterize this reaction. GC–MS cannot be used for degradation studies of the SU herbicides because these compounds are thermally labile and will decompose in the inlet or on-column. However, to help us get a better understanding of the Glean degradation, we injected one of our reaction mixtures onto the GC–MS, and three distinct peaks were observed. At a retention time of 7·0 min a mass spectrum consistent with the triazine amine structure (**3**) was obtained (strong peaks with *m/z* of 110 and 140). This compound was not present in the NIST98 database available on our data system. A peak eluting at 9·7 min was identified by the Hewlett Packard library search software as 4-chlorobenzenesulfonamide to 98% confidence. It is more likely, however, that this species is 2-chlorobenzenesulfonamide, the by-product (**2**) shown in Figure 6. Mass spectrometry is often unable to distinguish isomers, and the spectrum of 2-chlorobenzenesulfonamide was not present in the library. The third peak at 11 min showed a strong peak at *m/z* 149, often seen in mass spectra from

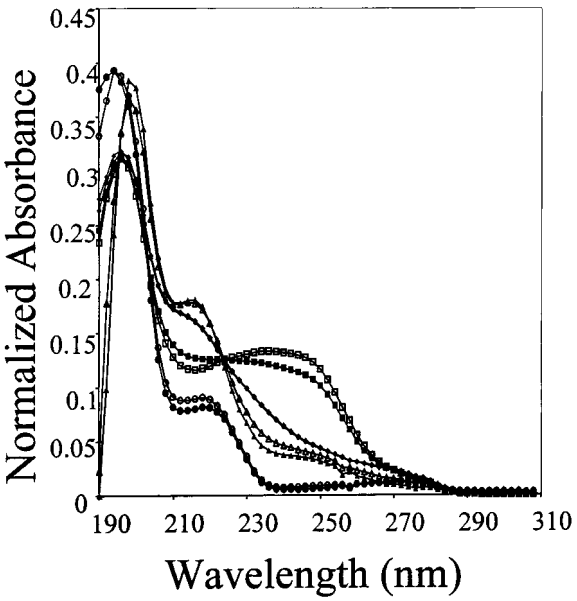

Figure 5. ALS spectra obtained using chromatographic initial estimates (lines with full symbols) and spectral initial estimates (lines with open symbols). The different symbols correspond to the different components in the degradation process, identified in Table 1

contamination of plasticizers used in plastic labware.

We were able to obtain a spectrochromatogram of 4-chlorobenzenesulfonamide for comparison with the spectra and chromatograms resolved by MCR–ALS, and these data showed reasonable agreement with the component eluting at 2·9 min. The triazine compound was not available through

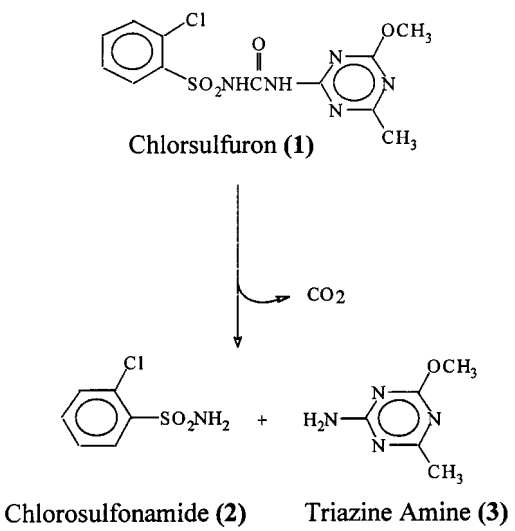

Figure 6. Proposed degradation pathway for Glean

any of our usual chemical suppliers. Our tentative conclusion is that the four resolved species shown in Figure 3 are Glean (thick full line eluting at 1·7 min), 2-chlorobenzenesulfonamide (thick broken line eluting at 2·9 min), 2-amino-4-methoxy-6-methyl-1,3,5-triazine (thin full line co-eluting with Glean at 1·7 min) and a phthalate or other contaminant species. These conclusions are consistent with literature reports that indicate this is the predominant pathway observed at pHs greater than 6.[10]

CONCLUSIONS

From the chemometric point of view, MCR–ALS has been shown to work satisfactorily in resolving a non-trilinear data set. The present example shows also the usefulness of the three-way data analysis approach for dealing with an interesting case of rank deficiency in the chromatographic data (i.e. the presence of two chemical compounds with highly similar elution profiles but different spectra). In such a situation, usually only one compound can be resolved from the two-way analysis of a single data matrix or, when two are modeled, the lack of fit associated with the results usually increases. By working with the three-way data set formed by coupling the rank-deficient matrix with others that have different concentration ratios of the co-eluting compounds, the two components can be successfully resolved. Based on the results reported above, we can design more in-depth studies to characterize these reactions under a variety of different reaction conditions. In addition, we are also interested in the development of general chemometric approaches for the resolution of three-way reaction data. We believe that reasonable progress toward this more general goal has been made.

ACKNOWLEDGEMENTS

The authors would like to acknowledge support for this research from the National Science Foundation, grant CHE-9709437, and the Department of Energy, grant DE-FG-88ER 13833. The authors would like to thank E. I. du Pont de Nemours & Co. for providing the Glean. They would also like to thank Wendy P. Wells for helpful discussions.

REFERENCES

1. K. S. Booksh and B. R. Kowalski, *Anal. Chem.* **66**, 782A–791A (1994).
2. E. M. Beyer, M. J. Duffy, J. V. Hay and D. D. Schlueter, in *Herbicides: Chemistry, Degradation and Mode of Action*, ed. by P. C. Kearney and D. D. Kaufman, pp. 117–189, Marcel Dekker, New York (1988).
3. J. S. Fletcher, T. G. Pfleeger, H. C. Ratsch and R. Hayes, *Environ. Toxicol. Chem.* **15**, 1189–1196 (1996).
4. A. S. Felsot, M. A. Bhatti and G. I. Mink, *J. Environ. Sci. Health B*, **31**, 831–845 (1996).
5. A. S. Felsot, M. A. Bhatti, G. I. Mink and G. Reisenauer, *Environ. Toxicol. Chem.* **15**, 452–459 (1996).
6. B. M. Berger and N. L. Wolfe, *Environ. Toxicol. Chem.* **15**, 1500–1507 (1996).
7. G. Dinelli, A. Vicari and A. Bonetti, *J. Chromatogr.* **700**, 195–200 (1995).
8. R. W. Reiser, A. C. Barefoot, R. F. Dietrich, A. J. Fogiel, W. R. Johnson and M. T. Scott, *J. Chromatogr.* **554**, 91–101 (1991).
9. G. Dinelli, A. Vicari, A. Bonetti and P. Catizone, *J. Agric. Food Chem.* **45**, 1940–1945 (1997).
10. I. Braschi, L. Calamai, M. A. Cremonini, P. Fusi, C. Gessa, O. Pantani and A. Pusino, *J. Agric. Food Chem.* **45**, 4495–4499 (1997).
11. L. D. Bray, N. E. Heard, M. C. Overman, J. D. Vargo, D. L. King, L. J. Lawrence and A. W. Phelps, *Pestic. Sci.* **51**, 56–64 (1997).
12. A. Vicari, R. L. Zimdahl, B. K. Cranmer and G. Dinelli, *Weed Sci.* **44**, 672–677 (1996).
13. C. Marucchini and R. Luigette, *Pestic. Sci.* **51**, 102–107 (1997).
14. J. Sabadie, *Weed Res.* **36**, 441–448 (1996).
15. J. P. Cambon and J. Bastide, *J. Agric. Food Chem.* **44**, 333–337 (1996).
16. E. Mallat, D. Barcelo and R. Tauler, *Chromatographia*, **46**, 342–350 (1997).
17. E. Sanchez and B. R. Kowalski, *J. Chemometrics*, **4**, 29–45 (1990).

18. J. B. Kruskal, In *Multiway Data Analysis*, ed. by R. Coppi, S. Bolasco, pp. 7–18, Elsevier, Amsterdam (1989).
19. A. de Juan, S. C. Rutan, R. Tauler and D. L. Massart, *Chemometrics Intell. Lab. Syst.* **40**, 19–32 (1998).
20. A. de Juan, A. Izquierdo-Ridorsa, R. Tauler, G. Fonrodona and E. Casassas, *Biophys. J.* **73**, 2937–2948 (1997).
21. R. Tauler, *Chemometrics Intell. Lab. Syst.* **30**, 133–146 (1995).
22. H. R. Keller and D. L. Massart, *Chemometrics Intell. Lab. Syst.* **12**, 209–224 (1992).
23. H. R. Keller and D. L. Massart, *Anal. Chim. Acta*, **263**, 21–28 (1992).
24. H. R. Keller, D. L. Massart, Y. Z. Liang and O. M. Kvalheim, *Anal. Chim. Acta*, **263**, 29–36 (1992).

