Analysis of three- and four-way data using multivariate curve resolution-alternating least squares with global multi-way kinetic fitting

Ernst Bezemer [1], Sarah C. Rutan *

Department of Chemistry, Virginia Commonwealth University, PO Box 842006, Richmond, VA 23284-2006, USA

Abstract

This paper demonstrates a novel implementation of an alternating least squares (ALS) algorithm for resolving three- and four-way data. Computer-simulated multi-way data are studied as well as the multi-way data obtained from typical kinetic experiments observed using liquid chromatography with diode array detection (LC-DAD) and UV–visible spectroscopy. Each data set is analyzed using this new multi-way ALS algorithm, not only providing estimates of the spectral profiles (and retention profiles in the case of LC-DAD measurements) for each of the components involved, but also simultaneously estimating the rate constants for the reaction steps at different experimental conditions using a global kinetic analysis. However, when the reaction conditions do not require that all the rate constants are identical for each experiment, as is the case when the reactions are observed at different temperatures, the data analysis still benefits from the common information present in the data, such as spectral and retention profiles, as well as a common reaction mechanism.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Kinetic constraint; Hydrolisis; Nonlinear regression; Liquid chromatography; Diode array detection

1. Introduction

The determination of reaction rate constants can be accomplished by fitting of experimentally determined concentration profiles of reactants, intermediates and products as a function of time. The analysis becomes more complicated when the responses of the reactants and the products overlap and one cannot directly determine their concentration profiles over time. One can apply many curve resolution methods to these overlapped species in order to mathematically resolve the individual concentration profiles [1–3] and subsequently fit these with any of the available kinetic fitting programs, for example, Gepasi [4–6] and WinNonlin® [7].

Many groups have integrated these kinetic fitting routines into their curve resolution algorithms [8–14] reducing this previous multi-step process, into a single data analysis procedure. The results of these analyses provide both the individual component response profiles (such as spectra) and the rate constants for the reaction.

Many researchers have used the analytical solutions for the differential equations to estimate rate constants of spectroscopic data during a multivariate curve resolution method for a consecutive first-order model [15] as well as first- and second-order kinetic models [12,16]. These methods lack the flexibility of being able to analyze more complex kinetic models for which the analytical solutions of the differential equations cannot be obtained. Moreover, studying different kinetic models (other than the three mentioned above) requires modifications to the program code, rather than a change in the input parameters. Using the analytical solutions for the differential equations increases the calculation speed; however, when modern desktop computers are used to solve the differential equations numerically, only a few seconds are added to the analysis.

Gemperline et al. have implemented kinetic fitting routines based on chemical representations of the component reactions. This is accomplished by parsing the reaction mechanism as entered by the user. The resulting differential equations are solved, and this approach allows for the analysis of kinetic models of any complexity [17]. Moreover, this method also allows for the global estimation of rate constants of multiple

---

reaction kinetic experiments simultaneously. However, the method implemented by Gemperline et al. is not based on a curve resolution algorithm but on the Levenberg–Marquardt (LM) optimization of the projection of the concentration profiles into the null space spanned by the data [17]. The approach described here utilizes an ALS algorithm that allows for the flexible implementation of various constraints. Moreover, Gemperline's approach has only been applied to spectroscopic measurements and, thus, is limited to three-way data (wavelength, reaction time and reaction conditions).

The chemometric method described in this paper is applicable to the analysis of multi-way experiments with two-way kinetic data forming four-way data sets (the measured response as a function retention time, wavelength, reaction time and initial reaction conditions). The extended kinetic fitting algorithm described in this paper is implemented within the ALS algorithm, which allows for the application of multiple constraints to help resolve chromatographically or spectrally overlapped components and to limit the results to a chemically reasonable solution [3,18]. The underlying assumption is that the individual component responses (such as spectra or chromatograms) do not change with variations in the reaction conditions. This is typically the case for different initial concentrations of the reactants but may also be the case for reactions studied at different temperatures.

This new chemometric algorithm has been characterized by fitting computer-simulated data in order to compare the results obtained with the parameters used to generate the data. Secondly, the performance has been investigated by analyzing data from typical kinetic experiments. In the first experiment, the hydrolysis of the dye molecule 4-nitrophenyl-pivalate (2,2-dimethylpropionate) (4NPP) was studied using a thermostated cuvette holder in a diode-array UV–vis spectrometer. The final products from this reaction are 4-nitrophenol and pivalate [19]. This experiment resulted in a two-way data set of absorbance measurements as a function of wavelength and reaction time. In order to get an insight into the thermodynamic parameters of this reaction, the spectroscopic experiment was repeated at several different temperatures resulting in a three-way data set of absorbance measurements as a function of wavelength, reaction time and reaction temperature.

In a second experiment, the decomposition of an herbicide (Ally™) was followed by injecting aliquots of the reaction mixture into an LC-DAD instrument at regular time intervals while the reaction mixture was kept at a constant temperature, as reported in an earlier publication [20]. The resulting three-way data sets consist of absorbance measurements as a function of wavelength, retention time and reaction time. However, when multiple reaction mixtures are held at different temperatures, this results in a combined four-way data set of absorbance measurements as a function of wavelength, retention time, reaction time and reaction temperature.

These two experiments serve as examples of data types that can be analyzed using this new approach.

2. Theory

2.1. Multivariate curve resolution–alternating least squares

Two-way LC-DAD data are frequently resolved using multivariate curve resolution methods, where the model is given as [21]

$$\mathbf{D} = \mathbf{R}\mathbf{S}^T + \mathbf{E} \tag{1}$$

where $\mathbf{D}$ (R × S) is a spectrochromatographic data matrix with R retention times and S wavelengths, $\mathbf{R}$ (R × N) contains the chromatographic profiles for N species present in the mixture, $\mathbf{S}$ (S × N) contains the spectral profiles and $\mathbf{E}$ (R × S) is the error matrix. The iterative procedure to minimize the error matrix ($\mathbf{E}$) is based on alternation between Eqs. (2) and (3) until a minimal improvement is achieved or a preset number of iterations is accomplished.

$$\mathbf{S} = \mathbf{R}^{\dagger}\mathbf{D} \tag{2}$$

$$\mathbf{R} = \mathbf{D}(\mathbf{S}^T)^{\dagger} \tag{3}$$

where † indicates the pseudo-inverse of the matrix. The ALS algorithm requires an initial estimate of either $\mathbf{S}$ or $\mathbf{R}$, which can be obtained using, for example, SIMPLISMA or evolving factor analysis (EFA), respectively [22,23].

The most general model used for multi-way analysis in this work can be expressed by

$$\mathbf{D}_k = \mathbf{R}_k\mathbf{Q}_k\mathbf{S}_k^T + \mathbf{E}_k \tag{4}$$

where $\mathbf{D}_k$ is the spectrochromatogram (R × S) for the kth time point in a kinetic experiment, $\mathbf{R}_k$ (R × N) contains the pure component chromatograms for the kth reaction time point, $\mathbf{S}_k$ (S × N) contains the pure component spectra for the kth time point, $\mathbf{Q}_k$ is an (N × N) diagonal matrix constructed from the $k_{1k}\ldots k_{Nk}$ elements in the N × K kinetic profile matrix, $\mathbf{K}$ and $\mathbf{E}_k$ is the error matrix for the kth time point. In the discussion that follows, we follow closely the nomenclature suggestions given by Smilde et al.[24]. Briefly, italic lower case are scalar values, upper case values represent array dimensions, bold lower case represent vectors, bold upper case represent matrices, and underlined bold upper case is reserved for tensors. Tensors, i.e., $\underline{\mathbf{D}}$, may be matricized, where the R × S × K tensor, can be represented as $\mathbf{D}_{RK \times S}$, indicating the data are represented in RK rows and S columns.

The ALS algorithm must be initialized, and there are two possible cases for the initial estimates. If the initial guess is given as an R × N × K tensor $\underline{\mathbf{Z}}$, with the R × N pure component chromatograms for each of the K time points obtained from two-way ALS fits of each $\mathbf{D}_k$ matrix, then the $\underline{\mathbf{R}}$ and $\mathbf{K}$ arrays are obtained from the initial guess $\underline{\mathbf{Z}}$ by normalization where

$$k_{nk} = \|\mathbf{z}_{nk}\| \tag{5}$$
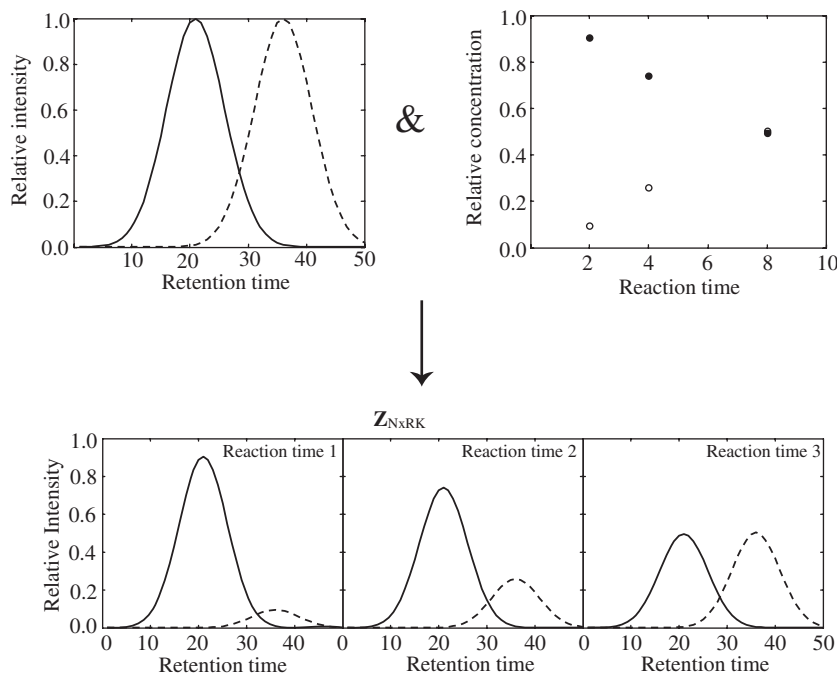
and

$$\mathbf{r}_{nk} = \mathbf{z}_{nk}/k_{nk} \tag{6}$$

Fig. 1. A graphical depiction of the $\underline{\mathbf{Z}}$ tensor, matricized as $\mathbf{Z}_{N \times RK}$ for use in step (6) of Table 1. Reactant (—); product (- - - -).

The structure of the matricized $\underline{\mathbf{Z}}$ array, $\mathbf{Z}_{N \times RK}$, is depicted in Fig. 1. The $\underline{\mathbf{S}}$ tensor is determined as follows

$$\mathbf{S}_k = \mathbf{D}_k^T \left(\mathbf{Q}_k \mathbf{R}_k^T \right)^{\dagger}. \tag{7}$$

Alternatively, the initial guess may be given in terms of the $S \times N$ pure component spectra for the k different time points, $\underline{\mathbf{Y}}$, an $S \times N \times K$ tensor, and the initial values for the $\underline{\mathbf{R}}$, $\underline{\mathbf{S}}$ and $\underline{\mathbf{K}}$ arrays are found in an analogous fashion to that shown in Eqs. (5)–(7).

The iterative calculations of the algorithm are carried out as shown in Table 1, depending on the structure of the model, where the iterations are started with step (1) in the case of a spectral initial guess ($\underline{\mathbf{Y}}$), or with step (6) in the case of a chromatographic initial guess ($\underline{\mathbf{Z}}$). If there is no trilinear or bilinear structure at all, then K independent two-way ALS analyses are carried out. However, this algorithm allows for the presence of components with different structures, for example,

a background component lacking all structure while another component might be trilinear. The component with the lowest level structure determines which column of Table 1 is used. Components that have a higher order structure are constrained according to a method described earlier [25]. A graphical depiction of one of the steps in the algorithm, step (6), for trilinear data is shown in Fig. 2.

In the present example, trilinearity implies that each species spectral profile is the same for each retention measurement and reaction time, each species chromatographic profile is the same for each wavelength and reaction time and each species kinetic profile is the same for each retention time and wavelength. In chromatographic measurements, however, retention profiles may change in shape depending on the concentration, giving nonbilinear retention behavior. A graphical depiction of the $\mathbf{Z}_{N \times RK}$ matrix for a nonbilinear $\underline{\mathbf{R}}$ is shown in Fig. 3.

2.2. Constraints

During the iterations, after a variable is recalculated, the appropriate constraints may be applied, such as non-negativity and unimodality, and these have been described earlier by several investigators [18,21]. The profiles adjusted by the constraints are subsequently used for the next step in the calculation. While this approach does not give true least-squares solutions, it is a commonly used approach that has been used to speed up calculations, and it allows for constraints to be applied on a component-by-component basis [26–33].

The kinetic model is implemented in a similar fashion to that described earlier [14], in that the kinetic profiles calculated during the ALS iterations are fit to the differential equations, and the optimal reaction parameters are used to simulate new

Table 1
Equations for MCR-ALS algorithm

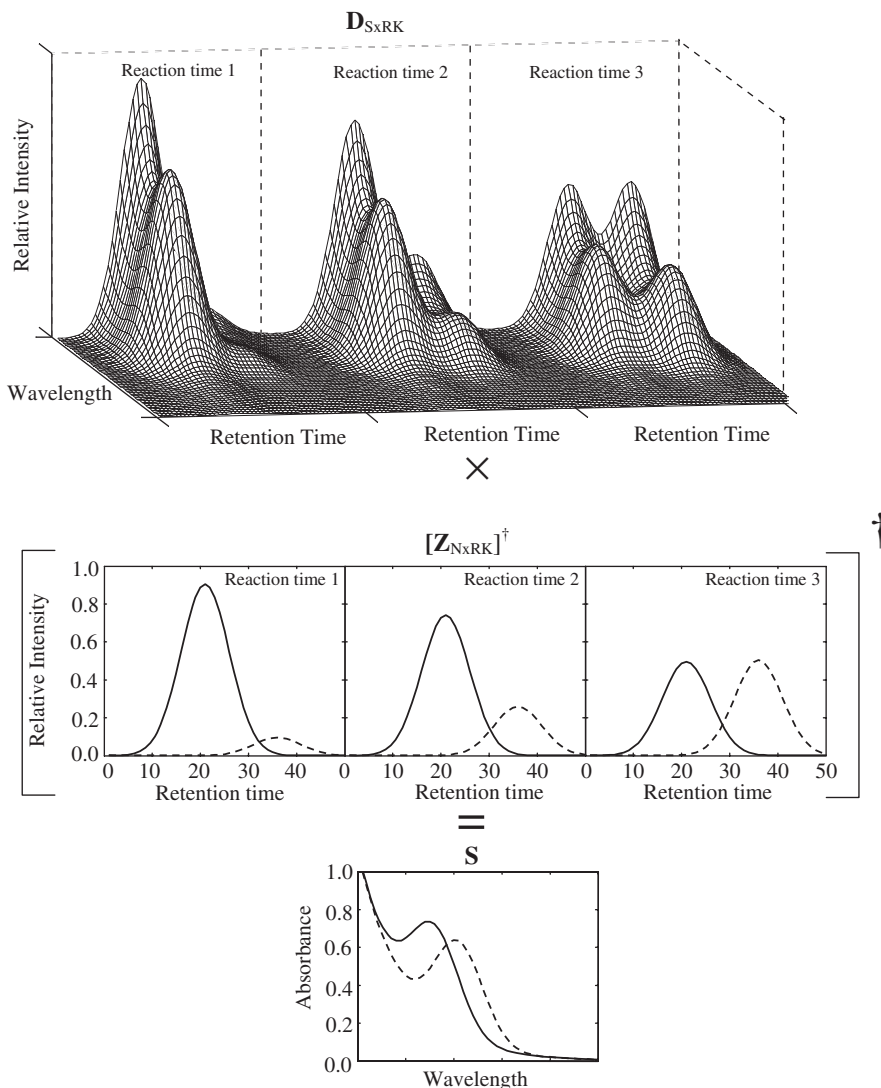| Bilinear $\mathbf{S}$	Trilinear	Bilinear $\mathbf{R}$	Step
	$\mathbf{R} = \mathbf{D}_{R \times SK} \mathbf{Y}_{N \times SK}^{\dagger}$		(1)
		$\mathbf{Y}_{SK \times N} = \mathbf{D}_{SK \times R} (\mathbf{R}^T)^{\dagger}$	(2)
		$k_{nk} = |\mathbf{y}_{nk}|$	(3)
		$\mathbf{S}_{nk} = \mathbf{y}_{nk}/k_{nk}$	(4)
$\mathbf{Z}_k = \mathbf{Q}_k \mathbf{R}_k^T$	$\mathbf{Z}_k = \mathbf{Q}_k \mathbf{R}^T$		(5)
$\mathbf{S} = \mathbf{D}_{S \times RK} \mathbf{Z}_{N \times RK}^{\dagger}$			(6)
$\mathbf{Z}_{RK \times N} = \mathbf{D}_{RK \times S} (\mathbf{S}^T)^{\dagger}$			(7)
$k_{nk} = \|\mathbf{z}_{nk}\|$			(8)
$\mathbf{r}_{nk} = \mathbf{z}_{nk}/k_{nk}$			(9)
	$\mathbf{Y}_k = \mathbf{Q}_k \mathbf{S}^T$	$\mathbf{Y}_k = \mathbf{Q}_k \mathbf{S}_k^T$	(10)
	$\mathbf{X}_n = \mathbf{r}_n \mathbf{s}_n^T$		(11)
	$\mathbf{K} = (\mathbf{D}_{K \times RS} \mathbf{X}_{N \times RS}^{\dagger})^T$		(12)

Fig. 2. A graphical depiction of step of the ALS algorithm for trilinear data. Reactant (—); product (- - - -).

kinetic profiles, which are subsequently used in the next iteration. The description of the implementation of the reaction mechanism as well as the individual kinetic fitting is described in an earlier paper [14]. In that work, simplex optimization was used to find the optimal kinetic parameters for the reaction, whereas, in this work, the method is extended to include global or local searches for the optimal rate constants, as well as the incorporation of the Gauss–Newton (GN) and LM optimization methods that are implemented in the optimization toolbox for Matlab by Mathworks [34].

2.3. Two-way kinetic model fitting

When different reaction conditions or initial concentrations are used, a UV–visible kinetic experiment will result in two-way kinetic data. Moreover, if each reaction condition results in a three-way data set (e.g., LC-DAD) the overall data would be four-way data. Note that these data are not quadrilinear. Unlike spectral profiles and retention profiles, the different kinetic profiles (at different reaction conditions) have a non-linear relationship with each other.
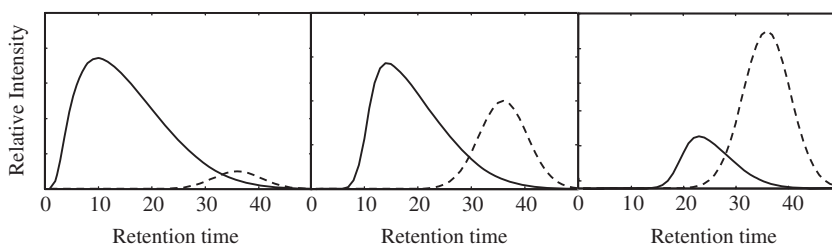

Fig. 3. A graphical depiction of the $\mathbf{Z}_{NxRK}$ matrix for nonbilinear $\underline{\mathbf{R}}$. Reactant (—); product (- - - -).
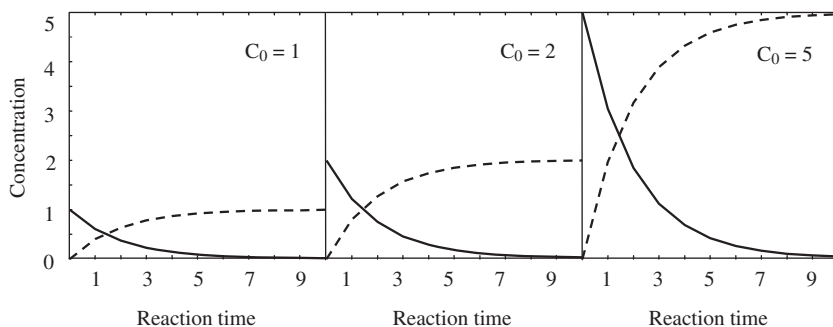
Fig. 4. An example of augmented kinetic profiles for first-order kinetics, starting with 1, 2 and 5 as the initial concentration of reactant. Reactant (—); product (- - - -).

In order to facilitate the analysis of four-way data, the data for the different kinetic experiments are unfolded as shown in Fig. 4, thus simplifying the four-way analysis to a three-way ALS as described above. This method maintains the bilinear or trilinear structure of the data and takes advantage of the common properties of the data (identical spectral and/or retention profiles). During the iterations of the ALS algorithm, whether using Eqs. (2) and (3) for two-way analysis or the equations shown in Table 1 for three-way analysis, the intermediate kinetic profiles contain all the profiles for each experiment unfolded as depicted in Fig. 4.

The algorithm that performs the kinetic fitting allows for the use of a single kinetic model to be applied to all the different reaction condition resulting in a global fit of all the data. This method is useful when duplicates are analyzed simultaneously or when different initial reagent concentrations are used, as these experiments can be described by a single kinetic model. However, when experiments are performed that have similar models but different rate constants (e.g., at different reaction temperatures), the algorithm allows for a local fit of the kinetic profiles, resulting in a set of rate constants for each reaction kinetic experiment. Subsequently, these rate constants may be examined for their adherence to an activation model such as the Arrhenius equation.

3. Experimental

Most chemicals were purchased from Aldrich and used without further purification, while the Ally™ herbicide was donated by DuPont, as acknowledged previously [20].

The hydrolysis of 4NPP was characterized at a fixed pH as follows. A pH 9.00 buffer was prepared by titrating a 0.1 M potassium borate solution with hydrochloric acid at each temperature. A 30.0 mg quantity of 4NPP was dissolved in 5.00 mL of methanol. A 10.0 μL aliquot of this solution was added to 2.50 ml of the buffer in a stoppered cuvette for the

UV–visible experiments. The UV–visible spectrophotometer was a Hewlett-Packard (Palo Alto, CA) 8453 with a Hewlett-Packard 89090A temperature-controlled cell compartment set at 20, 37 or 50 °C. Spectra were collected from 200 to 500 nm at 1-nm intervals. The reaction was monitored approximately every minute at the beginning of the reaction, and the interval was increased as the reaction progressed.

The hydrolysis of methyl-2-[[(4-methoxy-6-methyl-1,3,5-triazin-2-yl) aminocarbonyl] aminosulfonyl] benzoate (Ally™) has been studied previously [20]. The procedure used was to dissolve 10.0 mg of Ally™ into 0.500 ml of acetonitrile. This solution was then diluted to 100.0 mL using a 0.0050 M aspartic acid buffer titrated to pH 2.00. Aliquots of this Ally™ solution were stored in closed vials and kept in waterbaths held at 25, 35, 45 and 53 °C. At each time point, a sample was injected through a 20-μl sample loop into the LC-DAD instrument. The LC-DAD instrument was a Hewlett-Packard 1090 LC with diode array detector equipped with a Phenomenex LUNA C_{18} (15×4.6 mm) column with 5-μm silica particles, running an isocratic mobile phase consisting of 50% acetonitrile and 50% of the pH 2.00 aspartic acid buffer at 0.2 mL/min. Spectra (200–500 nm) of the eluent were obtained at approximately 1-s intervals.

3.1. Generation of simulated data

The synthetic data were generated using the enzyme kinetic model shown in Fig. 5.
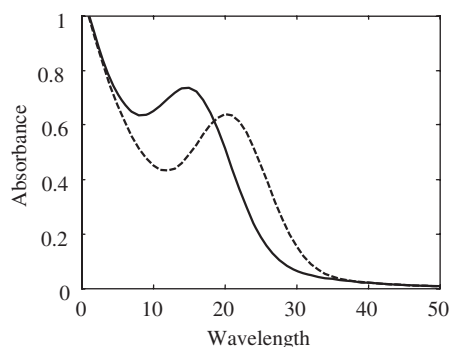
Fig. 6. Simulated spectral profiles used for the synthetic data sets. Reactant (—); product (- - - -).

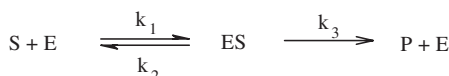$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} P + E$$

Fig. 5. Enzyme kinetic model. Conversion by an enzyme (E) of a substrate (S) forming the enzyme–substrate complex (ES) intermediate and the product (P).
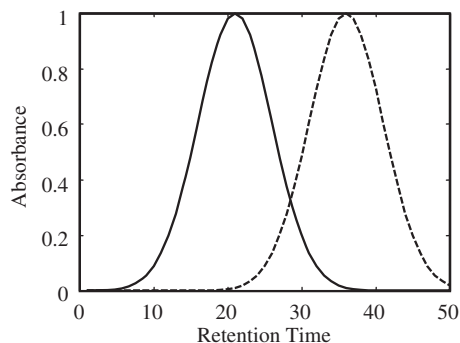
Fig. 7. Simulated retention profiles used for the synthetic data sets. Reactant (—); product (- - - -).

The three micro-rate constants for this model are related to the Michaelis constant and the v_{max} by Eqs. (8) and (9) [35].

$$K_m = \frac{k_2 + k_3}{k_1} \tag{8}$$

$$v_{max} = k_3 [E]_T \tag{9}$$

where $[E]_T$ is the total enzyme concentration.

Enzyme parameters are typically reported in the literature as v_{max} and K_m. In the present work, instead of fitting the data to the classical Michaelis–Menten model, which assumes steady state conditions, we fit the micro-rate constant model directly. Although the k_1 and k_2 are not well defined, the K_m values calculated from Eq. (8) are expected to be reasonably accurate.

In the simulations carried out in this work, we followed the enzymatic substrate conversion at different substrate levels using hypothetical UV–vis experiments as well as LC-DAD experiments. The data were created using the spectral and retention profiles shown in Figs. 6 and 7, respectively. The micro-rate constants used for the simulations were set to values of 0.58 μM^{-1} s^{-1}, 4.0 s^{-1} and 0.29 s^{-1}, for k_1, k_2 and k_3, respectively. These rate constants result in $K_M = 7.5$ μM and $v_{max}/[E]_T = 17.6$ nmol P/nmol E/min, which are similar to the enzyme parameters found for the O-demethylation of dextromethorphan by CYP2D6 in the presence of antibody [36]. The enzyme concentration was 0.03 μM for all simulated experiments, while the initial substrate concentrations were 5, 10, 20, 50, 100, 200 and 500 μM.

The total time was calculated by adding 5000–100 times the initial substrate concentration. The total time was divided into 10 quadratically separated time points according to Eq. (10) resulting in the time points shown in Table 2.

$$\left(\text{time point} \times \frac{\sqrt{\text{total time}}}{10} \right)^2 . \tag{10}$$

Normally distributed random numbers were added corresponding to S/N = 100 at the peak maximum to simulate homoscedastic measurement noise.

3.2. Data analysis algorithm

The three-way ALS algorithm was written in the MATLAB (versions 5.3–7.0) programming environment from Mathworks [34] while the kinetic fitting routine as described in an earlier paper [14] was modified to fit multi-dimensional kinetic profiles. The program was executed on various x86 computers running Windows XP.

4. Results and discussion

4.1. Computer simulated data

In typical enzyme experiments, a mixture of enzyme and substrate is incubated for a predetermined period of time and a single aliquot is analyzed to determine the concentration of product formed during the reaction. The initial rate of reaction estimated from this data point is plotted vs. the initial substrate concentration [35]. In a previous paper we investigated the influence of using a single substrate concentration but following the concentrations of the substrate and product over time rather than only at a single incubation time [37]. This additional time-dependent information aided in the determination of the micro-rate constants. However, under certain reaction conditions some sets of micro-rate constants are indistinguishable from others within the error of the experiments. By adding multiple substrate concentration experiments to our data, it may now be possible to relieve this ambiguity, while at the same time reducing the number of experiments (different substrate concentration levels) that need to be performed from many to a few. The multiple incubation times used were calculated according to Eq. (10) and were dependent on the initial substrate concentration, as shown in Table 1. The selected time points were chosen empirically, using a quadratic function to sample the reaction less frequently towards the end of the reaction. Other investigators have determined optimal sampling schedules for kinetic experiments based on analysis of the information function for some kinetic models that have analytical solutions for the set of differential

Table 2
List of reaction time points (seconds) for simulated enzyme reaction data at each substrate level[a]

Time point	Substrate concentration (μM)						
	5	10	20	50	100	200	500
1	55	60	70	100	150	250	550
2	220	240	280	400	600	1000	2200
3	495	540	630	900	1350	2250	4950
4	880	960	1120	1600	2400	4000	8800
5	1375	1500	1750	2500	3750	6250	13,750
6	1980	2160	2520	3600	5400	9000	19,800
7	2695	2940	3430	4900	7350	12,250	26,950
8	3520	3840	4480	6400	9600	16,000	35,200
9	4455	4860	5670	8100	12,150	20,250	44,550
10	5500	6000	7000	10,000	15,000	25,000	55,000

[a] Values for the time points were calculated using Eq. (10).

Table 3
Summary of analysis of enzyme kinetic data from simulated three-way UV–visible spectroscopy experiments

Optimization method	Initial concentration (μM)	Global fit	ALS fit error (%)	Rate constants[a]			Michaelis constant (μM)[b]
				k_1 (μM^{-1} s^{-1})	k_2 (s^{-1})	k_3 (s^{-1})	
Simplex	All	Yes	2.08	0.21	0.85	0.29	5.43
Levenberg Marquardt	All	Yes	2.08	0.43	2.10	0.29	5.56
Simplex	5	No	2.09	0.19	0.73	0.23	5.05
	10			0.15	0.79	0.29	7.20
	20			0.16	0.78	0.29	6.69
	50			0.19	0.68	0.28	5.05
	100			0.16	0.81	0.29	6.88
	200			0.22	0.44	0.29	3.32
	500			0.38	0.0002	0.28	0.74
Levenberg Marquardt	5	No	2.10	0.47	2.13	0.23	5.02
	10			0.36	2.36	0.29	7.36
	20			0.43	2.20	0.27	5.74
	50			0.47	2.14	0.28	5.15
	100			0.37	2.29	0.29	6.97
	200			0.65	1.85	0.29	3.29
	500			4.00	0.0001	0.28	0.07

[a] The true values are 0.58 μM^{-1} s^{-1}, 4.0 s^{-1} and 0.29 s^{-1} for k_1, k_2 and k_3, respectively.
[b] The Michaelis constant calculated from the true values using Eq. (8) is 7.40 μM.

equations to be solved [38–40]. Our design follows the general pattern of sampling found to be optimal for related models [38].

4.1.1. Generation of initial estimates for the analysis of the simulated data

The initial estimates were created by first performing EFA [23] on each data matrix, assuming the presence of two components. These coarse initial estimates were refined by using two-way ALS on each data matrix using just the non-negativity constraint. In the case of LC-DAD (four-way data) these two-way ALS results were further refined by performing three-way ALS using the non-negativity and trilinearity constraints. The initial guesses for the rate constants were 0.1, 5 and 0.5 for k_1, k_2 and k_3, respectively.

4.1.2. Simulated enzyme kinetics using UV–visible spectroscopy resulting in three-way data

The three-way data set (wavelength, reaction time and initial concentration) obtained from this simulation was analyzed using global analysis (one set of rate constants for all kinetic experiments) or with a local kinetic fit (a set of rate constants for each kinetic experiment). The latter method still takes advantage of the fact that the spectral profiles and reaction kinetic model are identical for all experiments. Furthermore, the ALS algorithm was modified to be able to use the different optimization algorithms that are available in the optimization toolbox for Matlab.

Table 3 shows the result of the three-way analysis; the results from the Gauss Newton optimization were virtually identical to those obtained for the Levenberg–Marquardt (LM)

Table 4
Summary of analysis of enzyme kinetic data from simulated four-way LC-DAD experiments

Optimization method	Initial concentration (μM)	Global fit	ALS fit error (%)	Rate constants[a]			Michaelis constant (μM)[b]
				k_1 (μM^{-1} s^{-1})	k_2 (s^{-1})	k_3 (s^{-1})	
Simplex	All	Yes	6.03	0.44	2.86	0.29	7.16
Levenberg Marquardt	All	Yes	6.03	0.59	3.99	0.29	7.25
	5	No	6.03	0.17	0.93	0.28	7.12
	10			0.17	0.95	0.28	7.24
	20			0.17	0.96	0.29	7.35
	50			0.18	0.97	0.29	7.00
	100			0.17	1.01	0.29	7.65
	200			0.17	1.02	0.29	7.71
	500			0.19	0.88	0.29	6.16
Levenberg Marquardt	5	No	6.04	5.51	5.51	0.51	15.4
	10			4.78	4.78	0.47	15.4
	20			3.86	3.86	0.42	15.3
	50			3.30	3.30	0.29	7.18
	100			3.33	3.33	0.29	7.54
	200			3.34	3.34	0.29	7.56
	500			3.19	3.19	0.29	6.00

[a] The true values are 0.58 μM^{-1} s^{-1}, 4.0 s^{-1} and 0.29 s^{-1} for k_1, k_2 and k_3 respectively.
[b] The Michaelis constant calculated from the true values using Eq. (8) is 7.40 μM.

optimization; therefore, only the results from the latter algorithm are shown here. The results from using the default optimization method available in the base Matlab package, based on simplex optimization, are also shown in Table 3.

In all cases, the spectral and kinetic profiles estimated by the algorithm were similar to the profiles used to generate the data, within the simulated noise level. However, the response surface for this model and parameter set has a very shallow minimum [37], leading to ambiguity in the kinetic model parameters. This results in different combinations of k_1 and k_2 that do not change the overall ALS fit error (calculated as described previously [3]) significantly for any of the analyses, as shown in Table 3. Although the estimated values of the rate constants were different from those that were used to simulate the data, it is clear that using a global analysis removes the ambiguity of some of the individual kinetic experiments, specifically the experiments with high initial substrate levels. A two-level–two-factor design for the initial guess showed that the resulting k_1 and k_2 values were highly dependent on the starting values, whereas the k_3 and the K_M value calculated according to Eq. (8) were independent of the starting values employed. It should be noted that the latter two parameters are of most interest in typical biochemical enzyme studies.

4.1.3. Simulated enzyme kinetics using an LC-DAD resulting in four-way data

In this case four-way data were simulated and analyzed, with the independent variables consisting of wavelength, retention time, reaction time and initial substrate concentration. The results of the four-way ALS resolution using multi-way kinetic fitting are summarized in Table 4. The addition of the extra data dimension significantly improves the accuracy of the kinetic fitting results. Specifically, the k_3 value is more accurately predicted by the global fits and in many of the local fits. In addition, although the estimated values of the k_1 and k_2 rate constants were different in most cases from those that were used to simulate the data, the Michaelis constant calculated from the micro-rate constants using Eq. (8) is more accurately estimated in many instances. The most accurate parameter estimates are obtained from the global LM analysis,
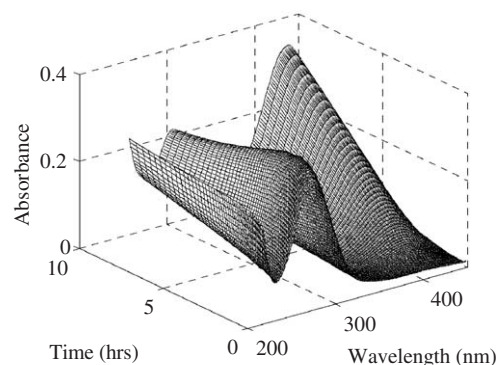

Fig. 8. UV–visible spectra of 0.107 mM 4NPP in pH 9.00 buffer reacting at 20 °C.

clearly indicating the advantage of using a global search over a local analysis.

4.1.4. Simulated traditional enzyme kinetic experiments using an LC-DAD resulting in three-way data

In traditional enzyme kinetic experiments, the enzyme and substrate are incubated for a specific time. For this type of data set, the reaction time dimension is missing, resulting in three-way data with independent variables of wavelength, retention time and initial substrate concentration.

One can imagine that the incubation time would make significant impact on the results of the analysis as during short incubations hardly any substrate is converted, while, on the other hand, during a near infinitely long incubation, hardly any substrate is left. Either case would result in data with very limited information about the rates of the reaction. The same 10 time points were used as described earlier, and the subsequent analysis was done for each row of Table 2. The fit parameters for the single time point simulations at multiple substrate concentrations are shown in Table 5. The ALS fit errors showed negligible differences between the LM and simplex kinetic fitting methods, but the resulting parameters were different indicating the presence of local minima in the response surface. Once again, the LM optimization method performed identically to the GN search method (results not shown), and these methods showed

Table 5
Summary of ALS fit results of LC-DAD enzymatic data using only one kinetic time point using multiple initial substrate concentrations

Time point	ALS fit error (%) using simplex	Simplex rate constants[a]				ALS fit error (%) using LM	LM rate constants[b]			
		k_1 ($\mu M^{-1} s^{-1}$)	k_2 (s^{-1})	k_3 (s^{-1})	K_M (μM)		k_1 ($\mu M^{-1} s^{-1}$)	k_2 (s^{-1})	k_3 (s^{-1})	K_M (μM)
1	5.02	0.42	0.0018	15.6	37.1	5.02	1.20	25.1	17.3	35.3
2	5.24	0.25	1.47	0.33	7.20	5.24	0.73	4.95	0.33	7.23
3	5.47	0.23	1.21	0.29	6.52	5.47	0.89	5.47	0.29	6.47
4	5.84	0.18	1.03	0.30	7.39	5.84	0.69	4.82	0.30	7.42
5	6.28	0.20	1.06	0.30	6.80	6.28	0.57	3.60	0.30	6.84
6	6.62	0.19	1.13	0.29	7.47	6.62	0.66	4.56	0.29	7.35
7	7.02	0.18	1.00	0.29	7.17	7.02	0.64	4.23	0.29	7.06
8	6.94	0.28	1.65	0.29	6.93	6.94	0.61	3.92	0.29	6.90
9	6.43	0.22	1.22	0.29	6.86	6.43	0.75	4.77	0.29	6.75
10	5.61	0.27	1.72	0.29	7.44	5.61	0.90	6.40	0.29	7.43

[a] The true values are 0.58 $\mu M^{-1} s^{-1}$, 4.0 s^{-1} and 0.29 s^{-1} for k_1, k_2 and k_3 respectively.
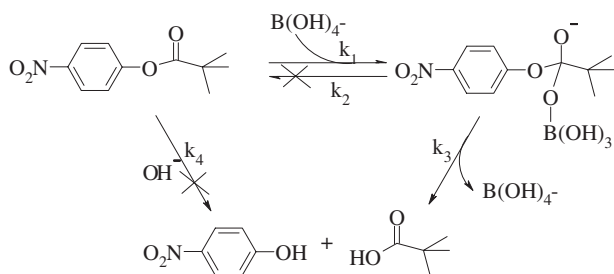[b] The Michaelis constant calculated from the true values using Eq. (8) is 7.40 μM.

Fig. 9. Proposed reaction scheme for hydrolysis of 4NPP.

Fig. 11. Arrhenius plot for observed rate constants for k_1 (—) and k_3 (- - - -) for NPP hydrolysis mechanism shown in Fig. 9.

significant improvement over the simplex method in regards to the accuracy of the estimation of the micro-rate constants. Using the earlier and the later time points resulted in the poorest estimates for the rate constants as predicted by the limited kinetic information present in those data. The intermediate time points resulted in similar rate constants for each fitting method, with once again the LM algorithm resulting in sets of rate constants close to the true value. The estimated K_M values for the simplex and LM fits were similar and were close to the true value, except in the case of the fit using the earliest time point.

The ALS fit error, however, is the best for the simulations based on the earlier and later time points. The curve resolution was performed in all cases assuming the presence of two components, while at the earliest and latest time points, only one component is present in the data. The substrate is the only species present at short incubation times, while the product is the only species present at the longest incubation times. The second component in these cases described the noise in the data and therefore resulted in a better overall fit error due to overfitting.

4.2. Chemical kinetic data

In order to investigate the real-world performance of the extended kinetic fitting algorithm, the base catalyzed hydrolysis of 4NPP was followed by UV–visible spectroscopy and the acid-catalyzed hydrolysis of Ally™ was followed using LC-DAD experiments.

4.2.1. UV–Vis spectroscopy of hydrolysis of 4NPP

The dye molecule 4NPP was chosen because both the reactant and product would have a significant UV–visible
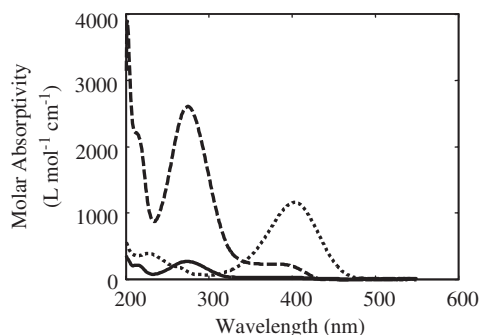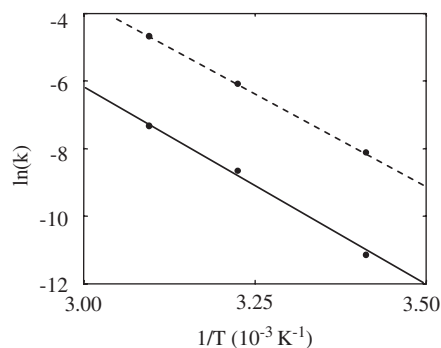
absorption spectrum due to the nitrophenyl moiety while the sterically hindered pivalate group results in a slow hydrolysis reaction for easy observability. The hydrolysis of 4NPP was followed using UV–visible spectrophotometer at 20, 37 and 50 °C, resulting in a three-way data set of absorbance measurements as a function of wavelength, reaction time and temperature.

4.2.1.1. Data pretreatment. The collected data were transferred to Matlab, and a baseline offset was added so that the absorbance at 550 nm for each reaction time point was equal to zero. The different experiments at each temperature were augmented together into a single three-way data set.

4.2.1.2. Initial estimates. Singular value decomposition indicated the presence of three components. EFA was used to obtain an initial estimate for the concentration profiles, and the normalized EFA profiles were used as the initial guess for the three-way ALS analysis.
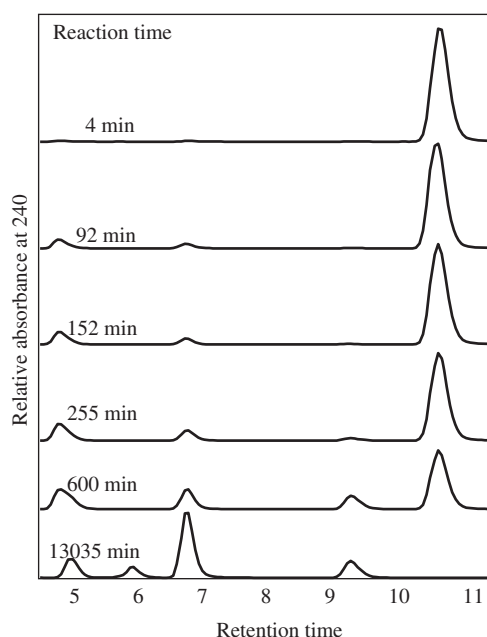

Fig. 10. ALS resolved spectral profiles for the hydrolysis of NPP. Reactant (—); (- - - -) intermediate; product (·······).

Fig. 12. Chromatograms (measured at 240 nm) of Ally in pH 2.00 buffer at 25 °C at different reaction times.

4.2.1.3. Choice of kinetic model. From the outset, it was clear that the observed kinetic profiles did not correspond to a simple first-order kinetic model. A maximum absorption can be seen after the start in some experiments, as shown in Fig. 8. The literature indicates that in the base-catalyzed hydrolysis of 4NPP in phosphate buffer, there is a phosphate-bound intermediate produced by nucleophilic attack of the phosphate on the carbonyl carbon [19]. The borate ion is a similar nucleophile which may also attack the carbonyl carbon forming an analogous borate-bound intermediate as postulated by Fig. 9. When the data was fit using this model, k_2 and k_4 resulted in very small values ($<10^{-8}$) indicating that the direct hydrolysis by water/hydroxide as well as the reverse (equilibrium) reaction were negligible. Therefore, the model used during the three-way ALS analysis was a simplified to a consecutive first-order reaction model (k_1 and k_3 only).

4.2.1.4. Three-way ALS of the 4NPP hydrolysis studied using UV–visible spectroscopy. Three-way ALS was performed using only non-negativity and the kinetic model as constraints. The resolved spectra are shown in Fig. 10. The spectrum of the reactant and the intermediate are very similar in shape, as would be expected for the mechanism shown in Fig. 9 due to the presence of a common chromophore.

An Arrhenius plot of the estimated rate constants using this method is shown in Fig. 11 and indicates the linear relationship between the reciprocal temperature and the natural log of the rate constants. The activation energy for second step (k_3, 90 ± 3 kJ/mol) is slightly less than that of the first step in the reaction (k_3, 101 ± 7 kJ/mol), while the individual rates are orders of magnitude lower, demonstrating that the second step is the rate-determining step (errors were determined from the statistics of the linear fit).

4.2.2. LC-DAD investigation of the hydrolysis of Ally^{TM}
The acid-catalyzed hydrolysis of Ally™ was followed using an LC-DAD instrument as described previously [20]. These experiments resulted in a four-way data set (absorbance measurements as a function of retention time, wavelength,
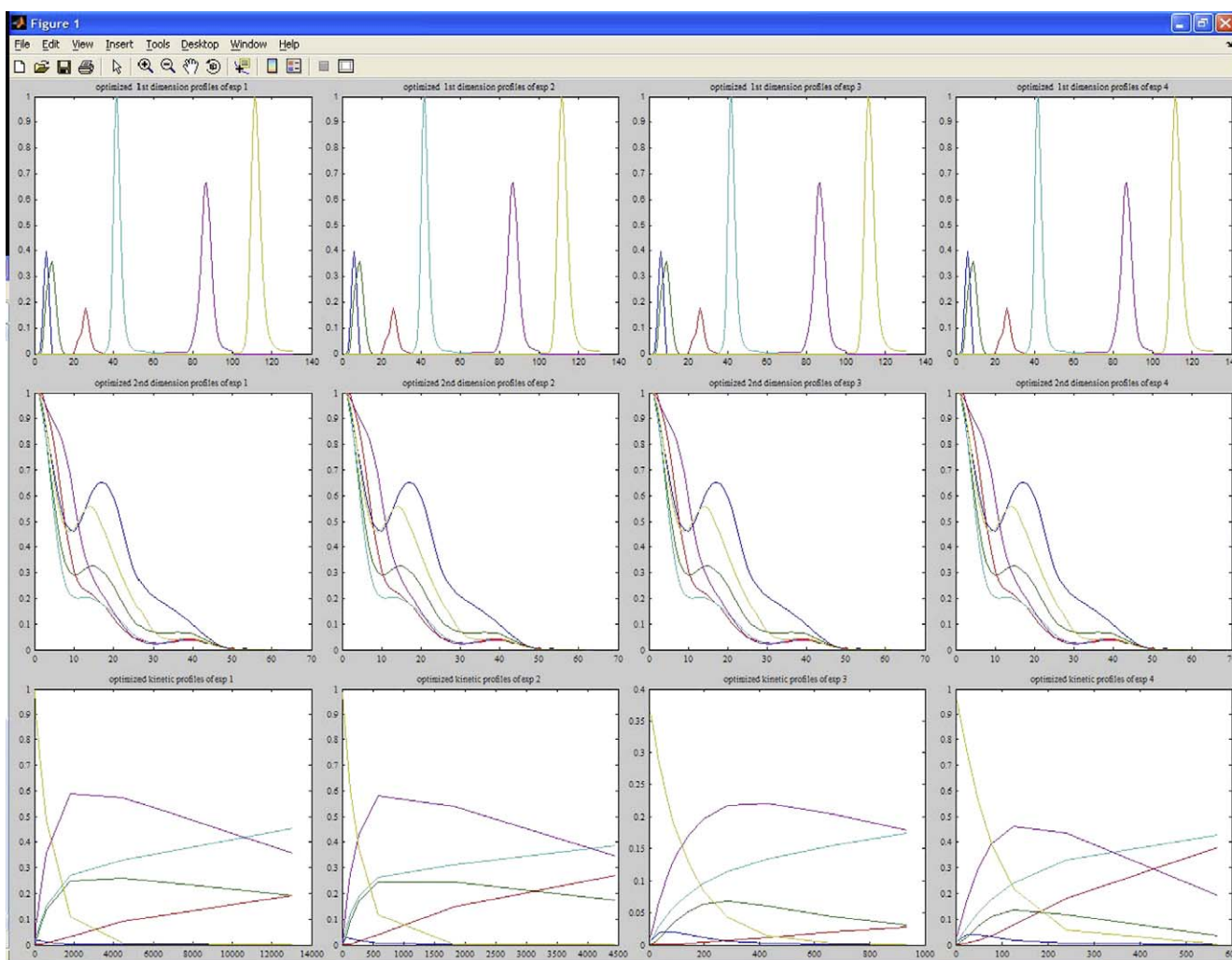

Fig. 13. Sample screen shot from the four-way ALS program. Each of the four columns corresponds to an experiment at a different temperature, 25, 35, 45 and 53 °C, with the rows representing the resolved chromatograms, spectra and kinetic profiles, respectively. Note the longer time scales for the experiments carried out at lower temperatures (lower left panel).
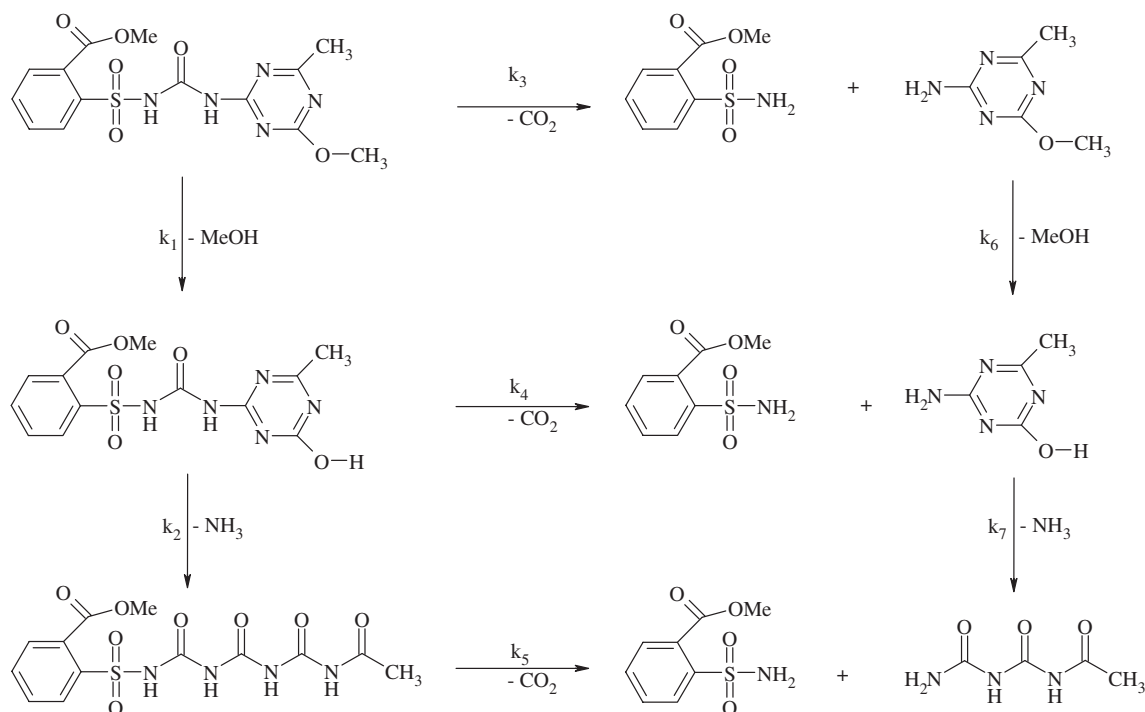
Fig. 14. Reaction scheme for the hydrolysis of Ally. Reproduced with permission from Ref. [20]. Copyright 2001 American Chemical Society.

reaction time and temperature) that was analyzed using the present algorithm.

4.2.2.1. Pretreatment of Ally[TM] data. The LC-DAD data were aligned by interpolating the DAD scan times at an interval of 1.5 s so that all spectra within the different chromatograms were taken at the same retention time, and the chromatograms were cropped to exclude the system peak and the baseline after the last peak eluted, as described in the previous publication [20]. An example of part of this data set is shown in Fig. 12. The three-way data sets from the experiments at each temperature were subsequently augmented to form the four-way data set, consisting of absorbance values as a function of wavelength, retention time, reaction time and temperature.

Table 6
Rate constants for the hydrolysis Ally hydrolysis using four-way ALS analysis

Rate constant[a]	Temperature				E_a (kJ/mol)[b]	E_a (kJ/mol)[c]
	25 °C	35 °C	45 °C	53 °C		
k_1 (10^{-4} min^{-1})	8.49	25.6	54.0	87.7	67.2±5.1	70±14
k_2 (10^{-4} min^{-1})	0.44	1.70	1.57	19.6	95±33	116±31
k_3 (10^{-4} min^{-1})	3.49	10.9	22.5	32.1	64.3±7.4	70±13
k_4 (10^{-4} min^{-1})	0.13	<0.001	3.49	8.30	123±8	247±58
k_5 (10^{-4} min^{-1})	0.54	1.24	<0.001	<0.001	—[d]	—[d]
k_6 (10^{-4} min^{-1})	170	296	243	497	25±11	25±11
k_7 (10^{-4} min^{-1})	0.61	1.29	32.6	81.8	154±30	—[d]

[a] Mechanism shown in Fig. 14.
[b] Calculated from this work.
[c] Calculated from Ref. [20].
[d] Inadequate agreement with Arrhenius model to determine an activation energy.

4.2.2.2. Initial estimates for four-way Ally[TM] data analysis. The initial estimates were obtained as described in the previous paper [20]. The chromatogram was divided into peak envelopes, followed by EFA to create the initial estimates for the concentration profiles of these peak envelopes. Two-way ALS using only non-negativity was used to refine these coarse concentration estimates from EFA.

The results of all these two-way ALS analyses were merged together to form the starting points for concentration profiles for three-way ALS analysis. The results from the three-way ALS, using non-negativity, trilinearity and selectivity of the retention profiles, formed the starting point for the four-way analysis. This method of analyzing data of increasing complexity, starting with the combined results from analysis of less complex data was shown to be successful in previous studies [20].

4.2.2.3. Resolving four-way Ally[TM] hydrolysis data. The three-way data sets at the different reaction temperatures were augmented to form one four-way data set, while the starting point for the concentration profiles was the product of the augmentation of the results of the individual three-way analysis of each data set at that temperature. These starting estimates were then refined by performing ALS on this four-way data set using non-negativity on all components, trilinearity and selectivity on the retention profiles, but without the kinetic fitting. These results were used to construct the starting point for the final four-way ALS resolution.

Fig. 13 shows the graphical output of the program, with each column containing the optimized profiles at each reaction condition. In this case, all the retention and spectral profiles for each component were found to be the same for every

temperature; therefore, the trilinearity constraint was employed. The retention time shifts from experiment to experiment were only minor due to careful experimental procedures. It should be noted, however, that the program is able to handle severe chromatographic shift or spectral changes by lowering the trilinearity requirement for some or all the components.

Earlier research [20] had established that the acid hydrolysis pathway of Ally™ follows the scheme shown in Fig. 14, in which the herbicide is degraded by cleavage of the sulfonamide bond, or by demethylation of the triazine ring. The activation energies for each of the elementary reaction steps shown in Fig. 14 were estimated using the Arrhenius model and are summarized in Table 6. These activation energies are compared to those based on the previous three-way analysis. It can be seen that the agreement between the values is generally good, but the precision of the activation energies calculated using the four-way method is improved by a factor of two, demonstrating the reliability of the present approach.

5. Conclusion

In this paper we have demonstrated a four-way alternating least-squares multivariate curve resolution algorithm that can take advantage of multi-way reaction kinetic data that have identical reaction kinetic models but not necessarily identical reaction rate constants or initial reaction conditions.

From the analysis of simulated data, we conclude that the use of four-way data results in better estimation of the rate constants compared to the use of three-way data by reducing the ambiguity associated with the indistinguishability of different kinetic models.

Analysis of experimental data obtained by monitoring the hydrolysis of 4NPP and Ally™ demonstrated that this algorithm can find rate constants at different temperatures simultaneously. The higher order advantage of similar spectral and retention profiles as well as the identical reaction mechanism assisted in this analysis.

The algorithm presented here may be used for many other multi-way kinetic data as well. For example, data from the enzyme conversion of different concentration levels of drug candidates by liver enzymes followed by LC-MS analysis could be treated using this approach. This method would result in faster determination of the enzyme parameters and thus direct further research towards those pharmaceuticals with more favorable metabolic profiles [41].

Acknowledgements

References

[1] R. Bro, C.A. Andersson, H.A.L. Kiers, J. Chemom. 13 (1999) 295–309.
[2] R. Bro, J. Chemom. 10 (1996) 47–61.
[3] E.C. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst. 60 (2002) 239–251.
[4] P. Mendes, Comput. Appl. Biosci. 9 (1993) 563–571.
[5] P. Mendes, Trends Biochem. Sci. 22 (1997) 361–363.
[6] P. Mendes. Gepasi. (2000). http://gepasi.dbs.aber.ac.uk/softw/gepasi.html.
[7] WinNonlin. (2005). Mountain View, CA, Pharsight.
[8] S. Bijlsma, H.F.M. Boelens, H.C.J. Hoefsloot, A.K. Smilde, J. Chemom. 16 (2002) 28–40.
[9] H. Haario, V.-M. Taavitsainen, Chemom. Intell. Lab. Syst. 44 (1998) 77–98.
[10] E. Furusjo, L. Danielson, Anal. Chim. Acta 373 (1998) 83–94.
[11] K. Héberger, A.P. Borosy, J. Chemom. 13 (1999) 473–489.
[12] T.J. Thurston, R.G. Brereton, Analyst 127 (2002) 659–668.
[13] A. de Juan, M. Maeder, M. Martínez, R. Tauler, Chemom. Intell. Lab. Syst. 54 (2000) 123–141.
[14] E. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst. 59 (2001) 19–31.
[15] S. Bijlsma, D.J. Louwerse, A.K. Smilde, J. Chemom. 13 (1999) 311–329.
[16] S. Bijlsma, H.F.M. Boelens, A.K. Smilde, Appl. Spectrosc. 55 (2001) 77–83.
[17] P.J. Gemperline, G. Puxty, M. Maeder, D. Walker, F. Tarczynski, M. Bosserman, Anal. Chem. 76 (2004) 2575–2582.
[18] A. de Juan, Y. Vander Heyden, R. Tauler, D.L. Massart, Anal. Chim. Acta 346 (1997) 307–318.
[19] C.J. O'Connor, R.G. Wallace, Aust. J. Chem. 37 (1984) 2559–2569.
[20] E.C. Bezemer, S.C. Rutan, Anal. Chem. 73 (2001) 4403–4409.
[21] R. Tauler, D. Barcelo, TRAC 12 (1993) 319–327.
[22] W. Windig, Chemom. Intell. Lab. Syst. 36 (1997) 3–16.
[23] M. Maeder, A. Zilian, Chemom. Intell. Lab. Syst. 3 (1988) 205–213.
[24] A. Smilde, R. Bro, P. Geladi, Multi-way Analysis: Applications in the Chemical Sciences, Wiley, New York, 2004.
[25] E. Bezemer, S.C. Rutan, Chemom. Intell. Lab. Syst. 60 (2002) 239–251.
[26] A. de Juan, Y.V. Heyden, R. Tauler, D.L. Massart, Anal. Chim. Acta 346 (1997) 307.
[27] A. de Juan, M. Maeder, M. Martinez, R. Tauler, Anal. Chim. Acta 442 (2001) 337–350.
[28] A. de Juan, M. Maeder, M. Martinez, R. Tauler, Chemom. Intell. Lab. Syst. 54 (2000) 123–141.
[29] A. de Juan, R. Tauler, J. Chemom. 15 (2001) 749–772.
[30] J. Diewok, A. de Juan, M. Maeder, R. Tauler, B. Lendl, Anal. Chem. 75 (2003) 641–647.
[31] J. Saurina, S. Hernandez-Cassou, R. Tauler, A. Izquierdo–Ridorsa, J. Chemom. 12 (1998) 183–203.
[32] E.J. Karjalainen, U.P. Karjalainen, Anal. Chim. Acta 250 (1991) 169–179.
[33] E. Furusjo, L.G. Danielsson, E. Konberg, M. Rentsch-Jonas, B. Skaggerberg, Anal. Chem. 70 (1998) 1726–1734.
[34] Mathworks. MATLAB. (1999). Natick, MA.
[35] L. Stryer, Biochemistry, W. H. Freeman, New York, 1988.
[36] Q. Mei, C. Tang, Y. Lin, T.H. Rushmore, M. Shou, Drug Metab. Dispos. 30 (2005) 701–708.
[37] E. Bezemer, S.C. Rutan, Anal. Chim. Acta 490 (2003) 17–29.
[38] H.F.M. Boelens, D. Iron, J.A. Westerhuis, G. Rothenberg, Chem. Eur. J. 9 (2003) 3876–3881.
[39] D. Iron, H.F.M. Boelens, J.A. Westerhuis, G. Rothenberg, Anal. Chem. 75 (2003) 6701–6707.
[40] J.A. Westerhuis, H.F.M. Boelens, D. Iron, G. Rothenberg, Anal. Chem. 76 (2004) 3171–3178.
[41] J.H. Lin, A.Y.H. Lu, Pharmacol. Rev. 49 (1997) 403–449.