ELSEVIER

Analytica Chimica Acta 321 (1996) 75–95

ANALYTICA CHIMICA ACTA

A comparative study of several chemometric methods applied to the treatment of two-way kinetic-spectral data for mixture resolution

Yu-Long Xie [1], Juan José Baeza-Baeza, Guillermo Ramis-Ramos *

Department of Analytical Chemistry, University of Valencia, E-46100 Burjassot, Valencia, Spain

Received 16 May 1995; revised 31 July 1995; accepted 31 July 1995

Abstract

A comparative study was conducted to investigate the performance of several chemometric methods applied to the treatment of two-way kinetic-spectral data with the aim of resolving mixtures. The methods involved are non-linear least squares regression performed using the Powell algorithm, the linear and extended Kalman filter, and partial least squares regression. Both simulated and experimental data were processed. The coupling reaction of diazotized sulfanilamide with arylamines to give azodyes was monitored spectrophotometrically. Binary mixtures of the substrates with different values of the rate constant ratio and with varied degrees of spectral overlap were resolved. The effects of several influence factors have been studied using numeric simulation. The advantages and the limitations of each method have been evaluated.

Keywords: Chemometrics; Kinetic methods; Spectrophotometry; Mixture resolution

1. Introduction

During the last decade, and in comparison to equilibrium methods, kinetic methods of analysis have become increasingly popular owing to their simplicity, precision, short analysis time, reduced susceptibility to interferences and easy automation [1]. The area of multicomponent determinations is one of the most active in the context of kinetic analysis [1,2]. Mixtures of reacting analytes have been resolved in the past by means of the method of proportional equations [3,4], however, only a small fraction of the data collected was used, which led to a poor precision. With the introduction of the computer in the chemical laboratory, kinetic methods of analysis have become more powerful. Most often, kinetic systems follow non-linear relationships and thus, nonlinear regression of the initial concentrations of the analytes and the rate constants should be applied [5–7]. In the least squares methods, the sum of the squares of the differences between the measured signal and the reconstructed signal is optimized. In order to assure convergence and to achieve accurate results, good initial estimates of the parameters of the model are necessary.

The Kalman filter, a nonlinear regression algorithm, is also widely used in the kinetic determination of mixtures [8]. In the Kalman filter, the mean

* Corresponding author.

[1] On leave from the Department of Chemistry, Xiangtan University, Xiangtan, China.

SSDI 0003-2670(95)00532-3

square error of the estimates of the parameters (which are called the state variables) is optimized. The least squares and Kalman filter methods are declared to provide similar results under certain conditions [8]. The Kalman filter is a recursive algorithm, and initial estimates of the state variables and also initial guesses for the variance of these estimates and the measurement variance should be furnished. The Kalman filter has various versions, i.e., the linear [8], extended [8], adaptive [9] and robust [10] versions that are designed to address troublesome analysis in different areas. The original form of the Kalman filter, the linear Kalman filter requires the model to be linear with respect to the state variables. The linear Kalman filter has been often used for kinetic determinations, in which first or pseudo first order reactions are assumed, and the rate constants of the reactions are well-known and supposed to be invariant from run to run [11–15]. The latter assumption is a serious drawback since the rate constants are functions of many experimental factors (e.g., temperature, pH, ionic strength and so on). Thus, the results of the linear Kalman filter would be affected drastically by the value of the rate constants provided to the algorithm [11].

If the above conditions are not met, then the non-linear form of the Kalman filter, the extended Kalman filter, may be used [8,16–23]. Since the Kalman filter was developed to be used with linear models, then non-linear models should be linearized. Usually a Taylor's series expansion is used, where an inherent error caused by truncation of high order terms are introduced and it renders the extended Kalman filter statistically non-optimal. In addition, the extended Kalman filter is more prone to the influence of the initial guesses of the parameters than the linear counterpart [17,18,20]. For the extended Kalman filter method adopted by most researchers and in this study, the rate constants are assumed to be time-invariant. The only difference between this version of the extended Kalman filter and its linear counterpart is that the rate constants are regarded as parameters to be estimated and are adjusted in the process of filter. However, the commonly used extended Kalman filter tolerates a certain variation of the rate constants [22,23]. The compensation of time-variant rate constants caused by the temperature variation was studied by Corcoran and Rutan [17,18]. For multicomponent kinetic analysis, more applications of the linear Kalman filter than that of the extended Kalman filter are found in the literature [1].

The least squares and Kalman filter algorithms are model based, and can be regarded as the so-called hard-modelling methods, in which accurate model information is indispensable to obtain good results. Instead, in soft-modelling methods, no model should be previously assumed, and an empirical model is derived from the data themselves. Multiple linear regression, principal component regression or partial least squares regression (PLS) are used to build up the model from a number of mixtures known as the calibration samples. These procedures have been used extensively in equilibrium methods of analysis, but they have been scarcely applied to kinetic analysis [24–28]. However, a rapid popularization of soft modelling methods in the kinetic multicomponent determinations is to be expected, owing to their efficiency in managing with the problems derived from the incomplete models, imprecise data sets and interactions among the components, which are frequently found in kinetics. Also it is no necessary to supply initial guesses of the parameters of the model, as required by the nonlinear regression methods.

In most of the reported procedures kinetic curves monitored on a single wavelength have been used, and relatively few applications have taken advantage of the multi-wavelength data sets generated by an array detector [22,23,28]. Simulated and actual experiments have demonstrated that the extended Kalman filter with the use of multiple wavelength detection can tolerate more serious spectral overlap and smaller kinetic differences between the analytes than the single wavelength approach [22,23]. The three way PLS method of Wold et al. [29] has been extensively used for the calibration of chromatographic systems [30–33]. Recently, it has been used to treat three way kinetic-spectral data arrays [28]. The three-way PLS based on three way kinetic-spectral data arrays has been claimed to provide better results than the two way kinetic PLS method, particularly with mixtures having both a low rate constant ratio and small spectral differences [28].

In this work, the performance of the Powell algorithm (nonlinear least squares regression), the linear and the extended Kalman filters, and the PLS method when applied to two-way kinetic-spectral data were

comparatively studied using both simulated and experimental data. Numeric simulation was implemented to investigate the effects of different factors of the models. The potential of these techniques for the kinetic simultaneous determination of mixtures of drugs was evaluated by using the coupling reaction of diazotized sulfanilamide with the *o*-, *m*- and *p*-aminobenzoic acids. Binary mixtures of these compounds with different ratios of the rate constants, and various degrees of the spectral overlap of the colored products were resolved. The advantages and limitations of each method were discussed from the viewpoint of practical application.

2. Theoretical

2.1. Nomenclature

Lowercase bold characters are used for column vectors, uppercase bold characters for two-way matrices, and underlined italic uppercase bold characters for three-way matrices. The transpose of a matrix or a vector is represented by the superscript T. Unless otherwise stated, both lowercase and uppercase plain characters are used for scalars, lowercase and uppercase plain characters are also used as running indices and to indicate the number of dimensions of the vectors or matrices, respectively.

2.2. Kinetic model

Consider a system of L active substrates, reacting with a common reagent, R, to form L similar but not identical products, P_l, following the pseudo-first order reactions:

$$C_l + R \rightarrow P_l \qquad (1)$$

where C_l denotes the l-th component and P_l is the corresponding product. We will assume that only the products absorb within the monitored wavelength region, and that the absorbances are additive and follow the Lambert-Beer law. If the spectral scan in the whole wavelength range at each time point can be regarded as instantaneous, then:

$$A_{i,j} = \Sigma S_{i,l} c_l \left(1 - \exp\left(-k_l t_j\right)\right) + B \qquad (2)$$

where $A_{i,j}$ is the absorbance of the mixture at the wavelength λ_i (i = 1,...,I) and the time t_j (j = 1,...,J), $S_{i,l}$ is the molar absorptivity of the P_l product at the λ_i wavelength, c_l is the initial concentration of C_l (l = 1,...,L), k_l is the corresponding rate constant, and B is the background absorbance which is assumed to be constant. Thus, Eq. 2 contains the complete time–spectral kinetic information of the mixture. Next, the way the different algorithms have been implemented to evaluate the initial concentrations is briefly explained.

2.3. Powell algorithm

The Powell algorithm was implemented following the indications giving by Rao [34]. The one-dimensional searching was performed with the quadratic interpolation method. Initial guesses of the parameters to be estimated together with the initial searching directions and the step size to be used were provided. The coordinate axes were used as the initial directions. The parameters to be estimated were the concentrations of the analytes, the rate constants and the background absorbance.

2.4. Linear and extended Kalman filter

The equations of the Kalman filter algorithm are listed in Table 1, and a brief explanation with emphasis on the way the multichannel computation was implemented, is given next.

The Kalman filter algorithm is based on two equations, one is the system dynamic equation (Eq. 3) and the other one is the measurement model equation (Eq. 4). In the linear filter, the system state vector $\mathbf{x}(k)$ consists of the initial concentrations of the analytes, and sometimes the background absorbance is also included. In the extended filter, the rate constants of the reactions have also been included as state variables. If the rate constants are known accurately, only the concentrations of the analytes should be estimated, then Eq. 2 will be linear with respect to the state variables. On the other hand, if the rate constants are not so certain, and they should be regarded as state variables to be estimated from the measurements, then Eq. 2 will be nonlinear with respect to the state variables. In most applications of the extended Kalman filter and also in this

Table 1
Equations of the Kalman filter algorithm

System dynamic equation	$\mathbf{x}(k) = \mathbf{F}(k,k-1)\mathbf{x}(k-1) + \mathbf{w}(k)$	(3)
Measurement model equation		
multiple channel linear:	$\mathbf{z}(k) = \mathbf{H}^T(k)\mathbf{x}(k) + \mathbf{v}(k)$	(4a)
single channel linear:	$\mathbf{z}(k) = \mathbf{h}^T(k)\mathbf{x}(k) + \mathbf{v}(k)$	(4b)
multiple channel nonlinear:	$\mathbf{z}(k) = \mathbf{F}(\mathbf{x}(k))\mathbf{v}(k)$	(4c)
single channel nonlinear:	$\mathbf{z}(k) = \mathbf{f}(\mathbf{x}(k)) + \mathbf{v}(k)$	(4d)
State estimate extrapolation	$\mathbf{x}(k/k-1) = \mathbf{x}(k-1/k-1)$	(5)
Error covariance extrapolation	$\mathbf{P}(k/k-1) = \mathbf{P}(k-1/k-1)$	(6)
State estimate update		
multiple channel:	$\mathbf{x}(k/k) = \mathbf{x}(k/k-1) + \mathbf{K}(k)\mathbf{g}(k)$	(7a)
single channel:	$\mathbf{x}(k/k) = \mathbf{x}(k/k-1) + \mathbf{k}(k)\mathbf{g}(k)$	(7b)
Error covariance update		
multiple channel:	$\mathbf{P}(k/k) = [\mathbf{I} - \mathbf{K}(k)\mathbf{H}^T(k)]\mathbf{P}(k/k-1)$	(8a)
single channel:	$\mathbf{P}(k/k) = [\mathbf{I} - \mathbf{k}(k)\mathbf{h}^T(k)]\mathbf{P}(k/k-1)$	(8b)
Kalman gain		
multiple channel:	$\mathbf{K}(k) = \mathbf{P}(k/k-1)\mathbf{H}(k)[\mathbf{H}^T(k)\mathbf{P}(k/k-1)\mathbf{H}(k) + \mathbf{R}(k)]^{-1}$	(9a)
single channel:	$\mathbf{k}(k) = \mathbf{P}(k/k-1)\mathbf{h}(k)[\mathbf{h}^T(k)\mathbf{P}(k/k-1)\mathbf{h}(k) + \mathbf{R}(k)]^{-1}$	(9b)
Innovation		
multiple channel linear:	$\mathbf{g}(k) = \mathbf{z}(k) - \mathbf{H}^T(k)\mathbf{x}(k/k-1)$	(10a)
single channel linear:	$\mathbf{g}(k) = \mathbf{z}(k) - \mathbf{h}^T(k)\mathbf{x}(k/k-1)$	(10b)
multiple channel nonlinear:	$\mathbf{g}(k) = \mathbf{z}(k) - \mathbf{F}(\mathbf{x}(k))$	(10c)
single channel nonlinear:	$\mathbf{g}(k) = \mathbf{z}(k) - \mathbf{f}(\mathbf{x}(k))$	(10d)

work, the rate constants are treated as time-invariant, so the state transition matrix **F** is an identity matrix as that in the linear filter. In the extended Kalman filter, if the rate constants are assumed to be time-dependent, then a state transition matrix describing such dependency is required [8,17,18]. If the rate constants are adjusted from time to time in the filtering process, then more inaccuracy in the values of the rate constants may be tolerated. The state error vector **w(k)** terms were usually taken as zeros since most of the dynamic models used in analytical chemistry are deterministic. The matrix **H(k)** (or vector **h(k)**) in Eq. 4a (or in Eq. 4b) is the linear measurement function matrix (or vector), while in Eq. 4c and 4d it denotes the nonlinear measurement function.

Both the linear and extended Kalman filter share the same presentation of the algorithm (Eqs. 5 to 10), but when the extended filter is adopted, the measurement function **H** (or **h**) in the recursive algorithm should be calculated from the nonlinear measurement model of Eq. 4c (or Eq. 4d) by using the Taylor's series expansion. This is applied to the current estimate of the state variables with the truncation after the linear term. When multiple wavelength spectral–kinetic data are used in the filter, the inversion operation in the computation of the Kalman gain is required. The results of the Kalman filter depend on the initial estimates of the state variables, as it also occurs with the Powell algorithm. Additionally, the Kalman filter results rely on the initial guesses for the variance of the corresponding estimate values, and on the variance of the measurement noise.

2.5. PLS

Implementation of two-way PLS involves to build up a calibration model based on the so-called calibration data matrix, say **X**, obtained by recording the absorbance of N mixtures of known composition at I different wavelengths. The PLS model for matrix form data is:

$$\mathbf{X} = \Sigma \mathbf{t}_j \mathbf{p}_j^T + \mathbf{E} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (11)$$

$$\mathbf{Y} = \Sigma \mathbf{u}_j \mathbf{q}_j^T + \mathbf{F} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}$$

$$\mathbf{U} = \mathbf{T}\mathbf{B} + \mathbf{H}$$

and the prediction equation for **Y** is:

$$\mathbf{Y} = \mathbf{TBQ}^{\mathrm{T}} \quad (12)$$

where $\mathbf{T} = \{\mathbf{t}_1,...,\mathbf{t}_J\}$ is the score matrix, and $\mathbf{P} = \{\mathbf{p}_1, ...,\mathbf{p}_J\}$ is the loading matrix for **X**, and where **U** and **Q** for **Y** have the same meaning as **T** and **P** for **X**. The subscript J represents the number of principal components retained in the model.

In a kinetic determination, when a diode array detector is used to record the spectra of the reacting mixtures, the data obtained for the calibration mixtures constructs a three-way data array. In this case, Eq. 2 can be extended to:

$$A_{n,i,j} = \Sigma S_{i,l} C_{n,l} \left\{1 - \exp\left(-k_l t_j\right)\right\} + B \quad (13a)$$

$$= \Sigma S_{i,l} C_{n,l} K_{j,l} + B$$

that in matrix-tensor form is:

$$\underline{A} = \Sigma \mathbf{s}_l \otimes \mathbf{c}_l \otimes \mathbf{k}_l \quad (13b)$$

where $\underline{A}$ is a $N \times I \times J$ three way response data array and $A_{n,i,j}$ is its typical element, $\mathbf{c}_l$ is the vector of the concentration of the l-th analyte, $\mathbf{s}_l$ is the molar absorptivity vector of the P_l product and $\mathbf{k}_l$ is a vector containing the kinetic information for the l-th component and whose generic element is $\{1 - \exp(-k_l t_j)\}$. The symbol $\otimes$ represents the tensor product [29].

The generalization of PLS to $\underline{X}$ and $\underline{Y}$ three-way data arrays has been described by Wold et al. [29]. The three way PLS leads to the following model:

$$\underline{X} = \mathbf{T}. \otimes \underline{P}^{\mathrm{T}} + \underline{E} \quad (14)$$

$$\underline{Y} = \mathbf{U}. \otimes \underline{Q}^{\mathrm{T}} + \underline{F}$$

$$\mathbf{U} = \mathbf{TB} + \mathbf{H}$$

To estimate the parameters of the three way model, Wold et al. [29] suggested to unfold the three way data array in the direction which leaves the first mode intact, which in our case is the concentration mode. Thus, the model parameters can be estimated on the basis of the unfold data matrices.

The NIPALS algorithm [35] is used to decompose the data array, and cross validation (leave-one-out procedure) is adopted to determine the number of latent variables to be retained in the calibration model.

3. Experimental

3.1. Apparatus

Kinetic measurements were done in a Hewlett-Packard HP 8452A photodiode array spectrophotometer provided with a 1-cm quartz cell. The pH values were adjusted with a Crison MicroPH 2001 pH-meter. The spectrophotometer and data transfer were controlled by an IBM 486 compatible microcomputer, and calculations were also conducted on a 486 type computer. All the computation programs were written in MATLAB (Math Works, Sherborn, MA).

3.2. Reagents, solutions and procedures

Analytical reagent grade *o*-, *m*- and *p*-amino benzoic acids (ABA, Merck, Darmstadt, Germany), sodium dodecyl sulphate (SDS, Fluka, Buchs, Switzerland), sulfanilamide (Sigma, St. Louis, MO), sulfamic acid (Fluka), sodium nitrite (Fluka), and citric acid monohydrate (Panreac, Barcelona, Spain) were used. Distilled demineralized water (Barnstead, Sybron, Taunton, MA) was used throughout. All other reagents were analytical grade. The series of pH buffers of 0.25 mol l^{-1} citric acid were prepared by adjusting the pH potentiometrically with a sodium hydroxide solution. Stock solutions of *o*-, *m*- and *p*-ABA were prepared by solving 17.5, 17.5 and 18.5 mg, respectively, in 1 ml ethanol and then diluting with water to 50 ml. A 4×10^{-2} mol l^{-1} sulfanilamide stock solution was prepared in 0.3 mol l^{-1} HCl. A 0.2 mol l^{-1} $NaNO_2$, 0.5 mol l^{-1} sulfamic acid and 20% SDS solutions were made with water. To prepare the 1×10^{-2} mol l^{-1} diazonium ion solution, 12.5 ml of the sulfanilamide solution was introduced into a 50 ml volumetric flask, 15 ml $NaNO_2$ was added, the mixture was allowed to react for 10 min, 15 ml sulfamic acid was added to destroy the excess nitrite, and after 15 min the volume was completed up to the mark with water. The diazonium ion solution was renewed daily.

Mixtures were made by introducing the adequate volumes of the ABA stock solutions, 20 ml buffer and 2.5 ml 20% SDS into a 25 ml volumetric flask, and adding water up to the mark. A 2.25 ml volume

of the mixture was transferred into a dry 1-cm quartz cell, and then 0.25 ml diazonium ion solution was injected into the cell with a 500 μl regulable piston pipette, immediately after pressing the start button to acquire the data. Mixing was done by bubbling the solution in the cell four times with the piston pipette. The data collection was delayed 10 s thus to avoid the mixing period of the solution. The data were acquired within the 10–910 s range with 30 s intervals between the successive wavelength scans. The spectra were recorded within a given wavelength range (the spectral resolution of the HP8452A spectrophotometer is 2 nm). The blank was prepared in the absence of the ABAs and measured in the same way. The blank absorbance was always substracted from the sample absorbance before the data processing. The first time point was removed, thus a total of 30 time points were used in the data processing.

Table 2
Synthetic data sets for the investigation of the effect of the number of wavelengths

S for both species:	2000 mol l^{-1}
Width of peaks:	15 nm
Position of peaks:	565 and 585 nm for species 1 and 2, respectively
Wavelength range:	550 to 600 nm with 1 nm interval
Number of wavelength points:	51 (from 1 to 51 were used)
Rate constants:	5×10^{-3} and 3.5×10^{-3} s^{-1} for species 1 and 2, respectively
Time range:	10 to 910 s with 30 s interval
Number of time points:	31 (all the points were always used)
Standard deviation of noise:	0.1% of the maximum absorbance value
Background absorbance:	0.005

Datasets	δk(%) [a]	δt(s) [b]	Mixture composition($\times 10^{-4}$ mol. l^{-1})		
			No.	Species 1	Species 2
A	0.0	0	1	1.0	1.0
B	1.0	0	2	1.0	3.0
C	5.0	0	3	1.0	5.0
D	10.0	0	4	3.0	1.0
1	0.0	0	5	3.0	3.0
2	0.0	4	6	3.0	5.0
3	1.0	4	7	5.0	1.0
4	5.0	4	8	5.0	3.0
5	10.0	4	9	5.0	5.0
6	20.0	4	10	2.0	2.0
7	30.0	4	11	2.0	4.0
8	40.0	4	12	4.0	2.0
9	50.0	4	13	4.0	4.0

Initial values of the parameters for the Kalman filter and Powell algorithm [c]

Linear Kalman filter:	x_0 = (0.00 0.00 0.00)	$R = 10^{-6}$ $\sigma^2 = 5 \times 10^{-4}$
Extended Kalman filter:	x_0 = (0.00 0.00 0.005 0.0035 0.00)	$R = 10^{2}$ $\sigma^2 = 5 \times 10^{-4}$
Powell algorithm:	x_0 = (0.00 0.00 0.005 0.0035 0.00)	
	Step size = (10^{-4} 10^{-4} 10^{-6} 10^{-6} 10^{-3})	
	PLS calibration samples:	
	Mixture Nos. 1, 3, 5, 7 and 9	

[a] Normally distributed random numbers multiplied by δk (a percentage of the nominal value of k) were added to the nominal values of the rate constants.
[b] The timing imprecision for each mixture in the data set is an evenly distributed random number within 0 and 1 multiplied by δt.
[c] x_0 is the vector of the initial state variables. The first two elements are the concentrations of both species, and the last one is the background absorbance. For the extended Kalman filter and Powell algorithm, the other two elements are the rate constants. R is the variance of the absorbance noise, and σ^2 is the variance of the estimates.

3.3. Generation of simulated data

The spectra of mixtures of two components were simulated by using gaussian shaped peaks. The peak positions of the two components were fixed at 565 nm and 585 nm, respectively, and the width (standard deviation of the peaks) was fixed to 15 nm, or as otherwise specified. Data were generated within the 550–600 nm range with 1 nm interval between them. The nominal values of the first order rate constants used were 5×10^{-3} and 3.5×10^{-3} s^{-1}, and they were allowed to change a finite amount along the time scale in a random way. The amount of variation was an arbitrary value, δk, multiplied by a random number which followed a normal distribution. Thus, the rate constant changed along the kinetic process for the same mixture in a random manner. The time range monitored was from 10 to 910 s in 30 s intervals. In order to simulate the situation in which the start of the data acquisition period would not accurately coincide with the start of the reaction, a timing imprecision was introduced. For this purpose, an arbitrary time quantity, δt (in seconds), was multiplied by an evenly distributed random number within the 0–1 range, and added to the time values of the mixtures. In this way, different mixtures of the same data set had a different time shift. For each data set, a total of 13 mixtures were simulated. The perturbed rate constants, the timing imprecision together with the standard spectra and the composition of the mixtures were used to generate the response data according to the kinetic model

Table 3
Synthetic data sets for the investigation of the effect of the ratio of the kinetic rate constants and the spectral overlap [a]

S for both species:	2000 mol^{-1} l						
Width of peaks:	10 nm						
Position of peaks:	varied						
Wavelength range:	550 to 600 nm with 1 nm interval						
Number of wavelength points:	51 (11 points from 570 to 580 nm were used)						
Rate constants:	3.5×10^{-3} s^{-1} for species 2						
Time range:	10 to 910 s with 30 s interval						
Number of time points:	31 (all the points were always used)						
Standard deviation of noise:	0.1% of maximum absorbance value						
Background absorbance:	0.005						
Variation of rate constants:	10%						
Timing imprecision (δt):	2 s						
Composition of mixtures:	the same as in Table 2						
The degree of spectral overlapping							
Experiment No.	1	2	3	4	5	6	
Peak of species 1	565	567	569	571	573	575	
Peak of species 2	585	583	581	579	577	575	
Degree of overlap	2.0	1.6	1.2	0.8	0.4	0.0	
The difference of kinetic rate constants							
Experiment No.	1	2	3	4	5	6	7
Rate constant of species 1(k_1)	5.00	4.75	4.50	4.25	4.00	3.75	3.50
Ratio of rate constants (k_1/k_2)	1.43	1.36	1.29	1.21	1.14	1.07	1.00
Initial values of the parameters for the Kalman filter and Powell algorithm							
Linear Kalman filter:	$x_0 = (0.00\ 0.00\ 0.00)$			$R = 10^{-6}$		$\sigma^2 = 5 \times 10^{-4}$	
Extended Kalman filter:	$x_0 = (0.00\ 0.00\ k_1\ k_2\ 0.00)$			$R = 10^{2}$		$\sigma^2 = 5 \times 10^{-4}$	
Powell algorithm:	$x_0 = (0.00\ 0.00\ k_1\ k_2\ 0.00)$						
	Step size = $(10^{-4}\ 10^{-4}\ 10^{-6}\ 10^{-6}\ 10^{-3})$						
	Calibration samples of PLS:						
	Mixtures No. 1, 3, 5, 7 and 9						

[a] The meaning of the symbols is given in Table 2.

of Eq. 13. A constant was added as background absorbance and zero-mean random numbers with a gaussian distribution were added to the absorbance values to simulate the experimental noise. The standard deviation of the added noise was a certain percentage of the maximum absorbance of each mixture.

Here, we did not try to distinguish if the variation of a rate constant is produced within-run or between-run, but the rate constants changed always from time to time and from mixture to mixture, thus to simulate the perturbations introduced in the real experiment by small variations of the experimental and instrumental conditions during data collection. For each data set, a $13 \times 51 \times 31$ three way data array of simulated data consisting of concentration, wavelength modes and time was obtained. The parameters and the composition of the mixtures used in the simulation are listed in Table 2. These data sets were used to investigate the effect of the number of wavelengths on the performance of the different algorithms.

In order to study the effects of the difference of the kinetic rate constants and the spectral overlap, other synthetic data sets were used. Different rate constant ratios were obtained by decreasing the value of the rate constant of species 1 while maintaining a fixed rate constant for species 2. A total of 7 different rate constant ratios were considered with a maximum of 1.43 ($5 \times 10^{-3}/3.5 \times 10^{-3}$) and a minimum of 1 ($3.5 \times 10^{-3}/3.5 \times 10^{-3}$). The spectral overlap was evaluated as the ratio of the distance between the two peaks and the width of the peaks. The degree of spectral overlap was changed by shortening the distance between the peaks while maintaining a constant peak width. A total of 6 situations were considered in which the maximum ratio was 2.0 and the minimum one was 0. Therefore, a total of 42 data sets with varied rate differences and spectral overlaps were synthesized. In each data set the 13 mixtures listed in Table 2 were used. Again, each data set gave a $13 \times 51 \times 31$ three way data array. The parameters used in this simulation experiment are presented in Table 3.

As the Powell algorithm for nonlinear regression and the Kalman filter need initial guesses of the parameters to be estimated, and the Kalman filter also needs guesses of the variance of these initial estimates and the variance of the experimental noise, another relatively 'clean' data set without perturbation of the rate constants and timing imprecision was synthesis to investigate the effects of the initial

Table 4
Synthetic data set for the investigation of the effect of the initial estimates of the rate constants [a]

S for both species:	2000 mol^{-1} l		
Width of peaks:	15 nm		
Wavelength range:	550 to 600 nm with 1 nm interval		
Number of wavelength points:	51		
Number of wavelengths used:	1 (575 nm), 5 (573–577 nm) and 21 (565–585 nm)		
Rate constants:	5.0×10^{-3} and 3.5×10^{-3} s^{-1} for species 1 and 2, respectively		
Time range:	10 to 910 s with 10 s interval		
Number of time points:	91 (all the points were always used)		
Rate constants used in evaluation:	$(1.00 \pm 0.25) \times k$		
Standard deviation of noise:	0.5% of maximum absorbance value		
Background absorbance:	0.005		
Variation of rate constants:	0%		
Timing imprecision (δt):	0 s		
Composition of mixture:	1.000×10^{-4} mol l^{-1} for both species		
Initial values of the parameters for the Kalman filter and Powell algorithm			
Linear Kalman filter:	$\mathbf{x}_0 = (0.00\ 0.00\ 0.00)$	$R = 10^{-6}$	$\sigma^2 = 5 \times 10^{-4}$
Extended Kalman filter:	$\mathbf{x}_0 = [0\ 0\ (1 \pm 0.25) \times k_1\ (1 \pm 0.25) \times k_2\ 0]$	$R = 10^{-2}$	$\sigma^2 = 5 \times 10^{-4}$
Powell algorithm:	$\mathbf{x}_0 = [0\ 0\ (1 \pm 0.25) \times k_1\ (1 \pm 0.25) \times k_2\ 0]$		
	Step size = $(10^{-4}\ 10^{-4}\ 10^{-6}\ 10^{-6}\ 10^{-3})$		

[a] The meaning of the symbols is given in Table 2.

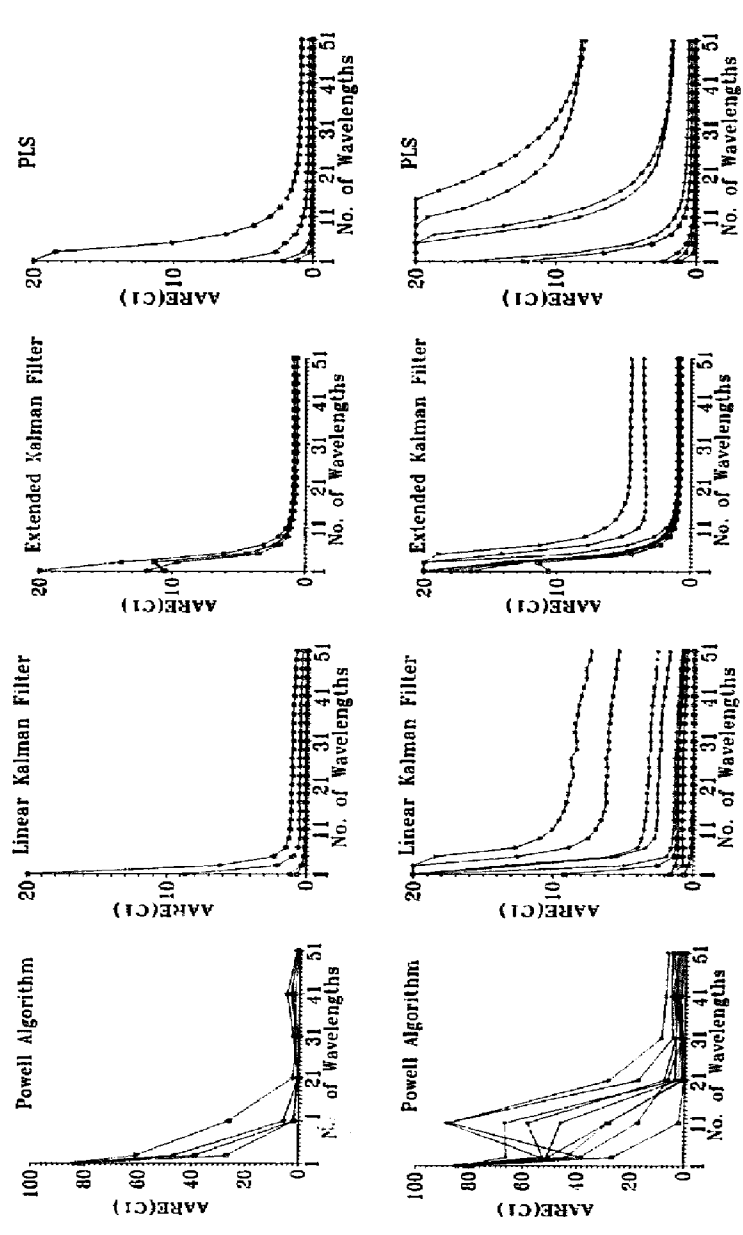


Fig. 1. Influence of the number of wavelengths when $\delta t = 0$ and $\delta k \leq 10\%$ (upper part), and when $\delta t = 4$ s and $\delta k \leq 50\%$ (lower part). From left to right are the results of Powell, linear Kalman filter, extended Kalman filter and PLS method, successively. The error in the vertical axis is the value of *AARE*(%) calculated from Eq. 15. Attention should be paid to the different scale for the Powell algorithm. In the upper part, the four lines were obtained with increasing values of δk (datasets A, B, C and D in Table 2). In the lower part, the nine lines were obtained with increasing values of δk and at two values of δt (from dataset 1 to dataset 9 in Table 2).

guesses. The parameters used to synthesize the data are listed in Table 4, where only one mixture was generated and the sampling interval for time was shortened from 30 s to 10 s to guarantee the convergence of the Kalman filter. This data set was treated by changing the initial rate constants up to a ±25% of the nominal values in 2.5% intervals. A total of 441 different combinations of initial rate constants (21 × 21) were evaluated with three different number of wavelengths (i.e., 1, 5 and 21 wavelengths).

3.4. Data processing

For the linear and extended Kalman filters and the Powell algorithm, the values of the initial estimates of the concentration were always set to zero in order to avoid bias. The initial estimates of the rate constants were always taken as the nominal values used in the simulation, except in the simulation experiment addressed to study the effect of the initial values of the rate constants. Other initial values used are given in Tables 2–4.

The average absolute relative error used for the evaluation of the results was:

$$AARE_l = \frac{\sum_{n=1}^{N} \left| \frac{\hat{c}_{nl} - \hat{c}_{nl}}{\hat{c}_{nl}} \right| \times 100}{N} \qquad (15)$$

where $\hat{c}_{nl}$ is the estimate for the true concentration of the l-th component in the n-th mixture, c_{nl}, and N is the number of mixtures.

PLS is an indirect calibration procedure that makes use of a set of calibration mixtures instead of using the standard spectra and the rate constants of the components in the calibration process. Therefore, the comparison between PLS and the other algorithms will not be straightforward. We have always used the same five mixtures in each data set as the calibration mixtures (No. 1, 3, 5, 7 and 9 in Table 2), to build up the calibration model, and to predict the concentration of the analytes in the other eight mixtures. The average error defined in Eq. 15 and obtained with PLS over these eight mixtures was compared to the error obtained also with Eq. 15, but using the thirteen mixtures with the other three methods.

4. Results and discussion

4.1. Numeric simulation

4.1.1. Influence of the number of wavelengths at several δt and δk values

First, simulation studies were performed for the kinetic determination of mixtures in which the products were assumed to absorb with a given degree of spectral overlap (i.e., 1.33), and also with a fixed ratio of the first-order rate constants (i.e., 1.43), and the number of wavelengths was varied. Two groups of data sets were generated (Table 2). In one group, the variation of the imprecision of the rate constants (δk) was in the range of 0% to 10% for four data sets, but no timing imprecision was assumed therein. With another group of data sets, a more serious variation of the rate constants was performed and a maximum random timing imprecision of $\delta t = 4$ s was considered in the data synthesis. The number of wavelengths incorporated in the data processing was increased from only a single wavelength to the whole spectrum (51 wavelength points) with an increment of 2 at a time. However, for the Powell algorithm, a larger increment of the number of selected wavelengths was used, thus to reduce the very large overall computation time. The wavelength located in the centre of the spectrum (the 26th wavelength point, 575 nm) was used for the single wavelength computation, and it was also fixed as the centre of the wavelength range when more wavelengths were adopted. The *AARE* values against the number of wavelengths and for species 1 are shown in Fig. 1. The corresponding results for species 2 were entirely similar and are not shown. Owing to the relatively large errors given by the Powell algorithm, different *AARE* scales were used in the presentation of the results (see Fig. 1).

From Fig. 1, it can be seen that if only a single wavelength was used, it was difficult to obtain acceptable estimates for the concentrations. For the linear Kalman filter and PLS, the acceptable results seemed to be obtained for the data sets with variations of the rate constants less than 5% in the single wavelength situation. It is beyond expectation that the extended Kalman filter gave worse results than its linear counterpart with some of the data sets. Even for the 'cleanest' data set (data set A and data

set 1), it could not provide accurate estimates. The possible reason is that we tried to use constant initial guesses for the covariance (variance) of the state estimates and measurement noise, but different mixtures in the same simulation conditions were affected in a slightly different way by these initial guesses. In other words, the filter parameters may not be optimal in some cases. However from the viewpoint of the practical application, we still adopted fixed initial guesses for the computation of all the mixtures in the data sets. In addition and in comparison to the other algorithms, it seemed much more difficult for the Powell algorithm to provide acceptable results when only a single or a few wavelengths were used in the computation.

In all cases, the estimates converged quickly to a certain error level as the number of wavelengths used in the computation increased. Thus, the utilization of multiple wavelengths in the evaluation of the results is obviously superior to the single wavelength situation for all the methods. Of course, this superiority is obtained at the expense of more data and more time for computation which is not a problem for the off-line treatment of the kinetic data. To reach a given value of the error, the number of wavelengths required was more or less the same both for the linear and extended Kalman filter, but the Powell algorithm needed more wavelengths than the Kalman filter. A large number of wavelengths was also required by PLS, which is reasonable since no model is assumed in PLS and the calibration is performed from the data by themselves. Comparing the data sets with a rate constant variation smaller than 10%, it can be observed in Fig. 1 that the introduction of a timing imprecision did not alter the results significantly, which hints that the major factor affecting the precision was the variation of the rate constants rather than the timing imprecision. From the lower part of Fig. 1, it can also be seen that, if sufficient wavelengths were used, higher than 30% variation in the rate constants could be tolerated, however, the number of the sufficient wavelengths were different for the different procedures. In general, more wavelengths were needed with PLS than with the other algorithms.

It seems that the linear Kalman filter can tolerate some variation of the rate constants when many wavelengths are used. From Fig. 1, one cannot see significant differences between the results of the linear and extended Kalman filters. One possible reason is the zero mean feature of the noise used in the simulation, and the signal averaging effect resulting from the many wavelengths incorporated. When only a single wavelength was used, the spectral difference was not utilized and the error would critically depend on the error of the rate constants. Instead, in the multiple wavelength situation, the influence of the error of the rate constants decreased, as the number of wavelengths used increased. A larger number of wavelengths means to add more information to the data set, and therefore, to improve the quality of the data set. Another intrinsic factor may probably be that there is no significant difference between the linear and extended Kalman filter used, because in both cases, the rate constants which can be adjusted pointwise in the extended Kalman filter were assumed to be time-invariant. However, our experience and that of other researchers [19] has proved that the estimated accuracy for the concentrations and for the rate constants in the extended Kalman filter are not always the same, but sometimes, the concentrations can be estimated much more accurately than the rate constants. In such a case, the variation of the rate constants may affect the linear and extended Kalman filter in a similar way. The next point that should be sticked out is that the extended Kalman filter was largely affected by the initial guesses of the variance of the state variables and of the measurement noise, which should be chosen carefully for the system investigated, and that could cause problems in the practical application.

The performance of the Powell algorithm deserves more attention. For the mixtures in the same data set, and when the same initial parameters were used, the Powell algorithm very often succeeded in some cases and absolutely failed in others. Therefore, it is imperative to choose suitable initial values for each mixture in order to achieve good results, but the problem is the lack of heuristic rules to do that. It should be pointed out that, qualitatively, the Powell algorithm needed much more computation time (about 50 times in average) than the other algorithms, and also there was no direct proportion between the computation time and the number of wavelengths incorporated.

Finally, the slightly slow convergence of the re-

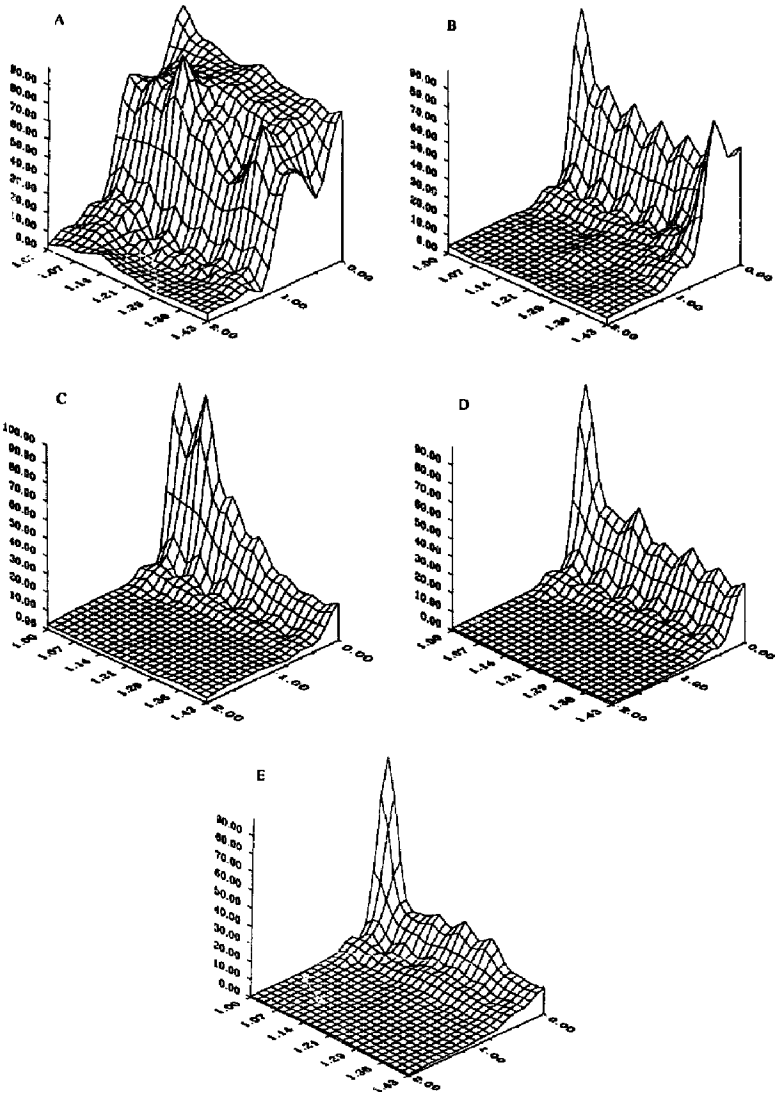


Fig. 2. The error (*AARE*) as a function of the rate constants (from 1.00 to 1.43) and the spectral overlap (from 2.00 to 0.00). Species 1 (A) and 2 (B) for the Powell algorithm, and species 1 for the linear (C) and extended (D) Kalman filters and for PLS (E).

sults of PLS to a low error level did not mean that PLS is inferior to the hard modelling methods. For the sake of simplicity, only five mixtures were used to construct the calibration model for all the cases. When 'unclean' data sets are dealt with, larger calibration sets should be used, and better results are to be expected.

4.1.2. Influence of the kinetic rate constant ratio and spectral overlap

The degree of spectral overlap and the ratio of the rate constants were varied. A total of 42 data sets were generated according to the simulation conditions described in the Experimental section and that are listed in Table 3. The number of wavelengths

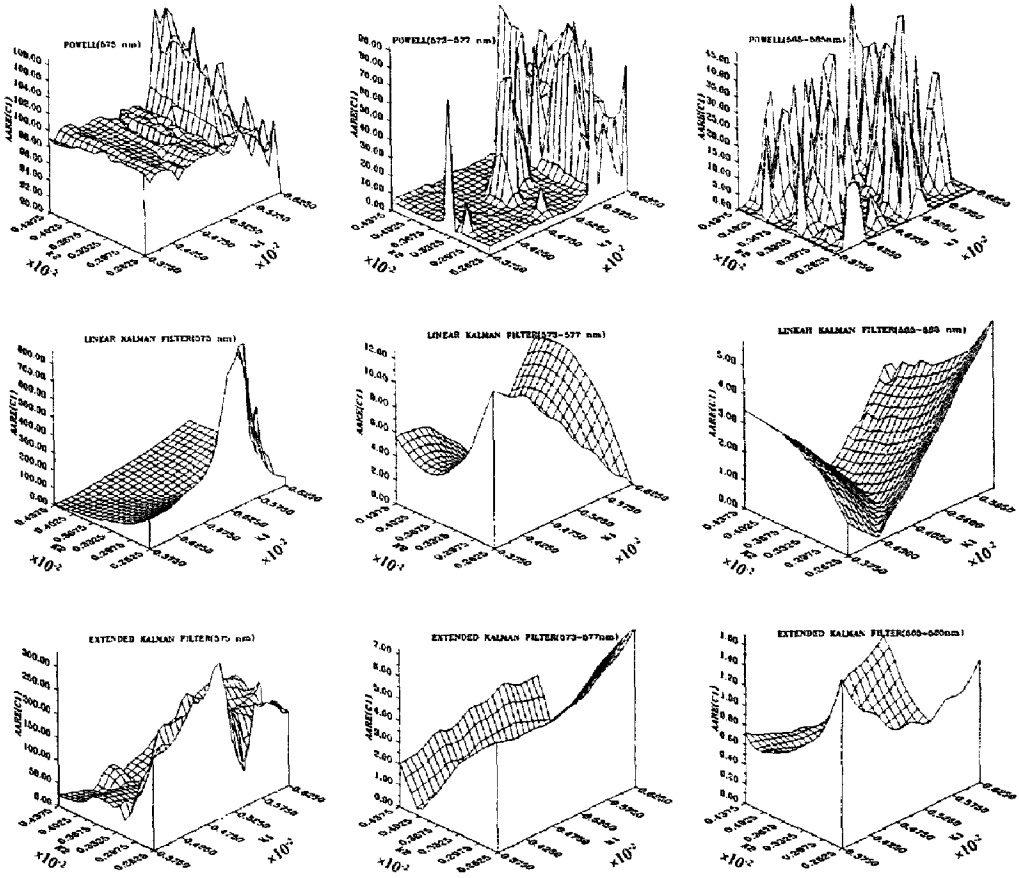


Fig. 3. Influence of the accuracy of the initial guesses of the rate constants. For the Powell algorithm (upper part), linear Kalman (middle part) and extended Kalman filter (lower part). From left to right, results obtained with 1, 5 and 21 wavelengths. Attention should be paid to the changes of scale of the *AARE* axis.

used in the computation was 11, and the nominal values of the rate constants were used as the initial guesses in the Powell algorithm and the Kalman filter. The *AARE* of the estimated concentrations obtained using the different algorithms are shown in Fig. 2A–E. For the convenience of the graphic presentation, the *AARE* values were normalized to a maximum of 100%. As expected, it is observed that the error increases as both the spectral overlap increases (its numerical estimation decreases from 2.00 to 0.00) and the ratio of the rate constants decreases (from 1.43 to 1.00). The results were very good except with the Powell algorithm, and with those cases where both a serious spectral overlap and a very small ratio of the rate constants existed. Since the two components showed very similar error behaviour for all the algorithms, except for the Powell algorithm, only the results of one component are shown in Fig. 2C–E. It was reasonable to have a similar error behaviour for both components, because their spectra have the same shape, and the rate constants and the variations of concentrations in the mixtures were also similar. The results with the Powell algorithm might not reflect the real situation, since the Powell algorithm was not so stable and very often missed the optimum in this study.

In Fig. 2A–E, it is also observed that the accuracy of the estimates was affected much more by the spectral overlap than by the kinetic difference. In the extreme situation in which the rate constants were identical, most of the methods still supplied reliable results, so long as there was a sufficiently large spectral difference. It should be considered that mixtures with very small kinetic differences were used, and that the variation of the rate constant ratio (from 1.43 to 1.00) would only slightly change the shape of the kinetic curves, and therefore, it could not bring much variation in the resultant kinetic–spectral data. Relatively, the spectral variation derived from the change of spectral overlap (from 2.0 to 0.0) was larger. Therefore, the results would be more sensitive to the spectral variation than to the kinetic difference. These results coincided with the results obtained by other authors [22,23].

Table 5
AARE values (species 1/species 2, in percentage) for the data set of Table 4 when the nominal values of k_1 and k_2 were provided as the initial guesses

No. of wavelengths	Powell	Linear Kalman	Extended Kalman
1	99/98	0.40/0.77	0.29/2.80
5	5.0/6.5	0.61/0.24	1.7/0.67
21	0.23/0.14	0.077/0.095	0.17/0.39

4.1.3. Influence of the initial guesses on the nonlinear regression techniques

The effects of the initial estimates were also studied by using simulated data. Usually, no a priori knowledge about the concentrations of the analytes in the mixtures is available, so it would be reasonable to set them to zero in order not to introduce bias. Initial estimates of the rate constants should be also provided to the Powell algorithm and the extended Kalman filter. With the linear Kalman filter, the value of the rate constants is also required to construct the measurement function. In practice, approximate values of the rate constants can be estimated. We are here interested in investigating the effect of the inaccuracy of the initial rate constants provided to these algorithms. A simulated data set in the conditions given in Table 4 was generated for this purpose. The initial rate constants provided to the algorithm were changed in the range of $\pm 25\%$ of the nominal values. In each mixture, the concentrations of both components were estimated using 1, 5 and 21 wavelengths.

For simplicity, the *AARE* values of only one component are shown in Fig. 3. Similar results were obtained for the other component. The *AARE* values obtained with the nominal values of k_1 and k_2 are shown in Table 5 and the *AARE* values given with erroneous values of the rate constants are presented in Fig. 3. The results of Fig. 3, upper part, indicate that in the single wavelength situation the Powell algorithm converged without finding the optimum. With the same initial parameters, however, the results were improved when more wavelengths were adopted, although in many cases, the optimum was still missed. The estimate errors decreased remarkably as more wavelengths were adopted for the Kalman filter. This was reasonable, since more measurement information was included in the computation and used to improve the estimates. Owing to the possibility of adjusting erroneous initial rate con-

stants, it could be expected that the results given by the extended Kalman filter should be better than those given by the linear Kalman filter, however, the differences were small. Once again, the results shown in Fig. 3 stressed the advantage of the use of multiple wavelengths.

When the initial values of k_1 and k_2 provided to the algorithms were modified, the linear and extended Kalman filters give rise to narrow elongated valleys of the *AARE* value, which include the position of the k_1 and k_2 nominal values (Fig. 3). This was attributed to the mutual cancellation of the influence of the systematic errors of the two constants. This suggests that at least in some cases the system has not the necessary information to distinguish between k_1 and k_2. Therefore, the system has not either the information required to adjust independently k_1 and k_2 to the nominal values. Consequently, it is not surprising that the extended Kalman filter had not the capability of improving the results given by the linear Kalman filter.

The effects of varying the initial guesses of the estimates of the variance of the state variables, and of the measurement variance (required only by the Kalman filter) were also investigated. We have found that, qualitatively, it is the relative rather than the absolute magnitude of the state covariance and the measurement variance that influence the estimates. Also, they have more influence on the extended Kalman filter than on the linear Kalman filter. For the linear Kalman filter, when these values were varied in a relatively large range, they did not degrade the results significantly, but that was not the case for the extended Kalman filter. We also found that the initial guesses could be taken in a wider range as the number of wavelenghts used in the evaluation increased. However, since the choice of initial values depends on the system investigated, i.e., on the ratio of the rate constants and spectral overlap, relative concentration of compenents and level of noise, the choice of optimal initial values is not a simple matter.

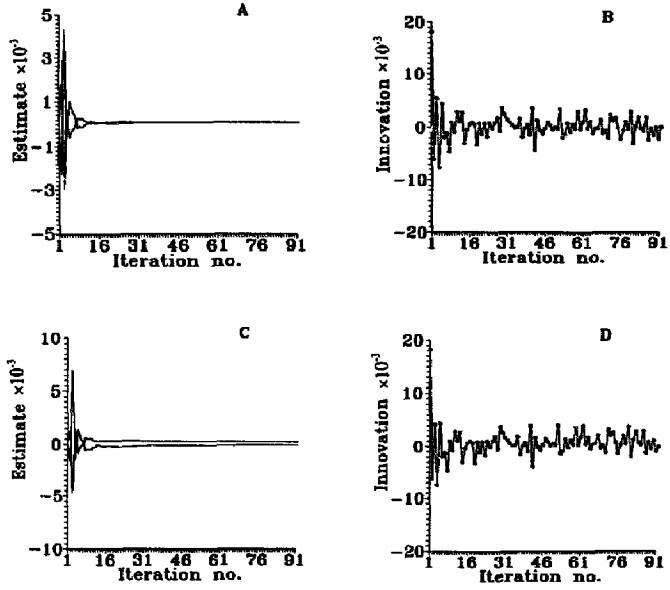


Fig. 4. Sequences of the state estimates (left part) and the corresponding innovation sequences (right part) for the linear Kalman filter when accurate (upper part) and erroneous (lower part) initial guesses of k_1 and k_2 were provided to the algorithm.

In the Powell algorithm, another factor studied was the initial step size for the one dimensional searching. In the literature there is no criterion of how to define the searching step size when a new searching direction should be used in the computation cycling. Most often, if the initial searching step size corresponding to the rate constants was too large, the algorithm diverged. We found that it is advisable to use a relatively small initial step for the directions referring to the rate constants, since the initial values of the rate constants might be sufficiently accurate.

The Powell algorithm, indeed, did not give any means of detecting model errors. However, from our experience and that of others [19], the Kalman filter algorithm cannot provide more advantage at this point. We have found that sometimes the white noise characterized innovation sequence and the strongly smoothed state sequence might correspond to both very accurate and very absurd estimates. Some examples which were extracted from the study of the effects of the initial rate constants (data sets obtained in the conditions of Table 4) are given in Fig. 4. Fig. 4A shows a state sequence evolved with the linear filter, and Fig. 4B is the corresponding innovation sequence. The number of wavelengths used for the evaluation was 1 and the initial rate constants were the nominal values used in the data synthesis (which refers to the central point in the left plot of Fig. 3, middle part). In this case, the concentrations were estimated accurately (0.996×10^{-4} for species 1 and 1.0077×10^{-4} for species 2). The filter feature reflected by Fig. 4A and B coincided with the fact that the estimates were highly accurate. However, when erroneous initial rate constants were used (3.75×10^{-3} and 2.625×10^{-3} s^{-1} for species 1 and 2, respectively, which corresponds to the bottom corner point in the plots of Fig. 3), the inaccuracy of the estimates of the concentration (168% for species 1 and −173% for species 2) were not reflected by the innovation plot and state sequence plot (Fig. 4C and D). The innovation sequence in Fig. 4C was also characterized by the white noise and the state sequences were also smoothed. Furthermore, all the final variances of the estimated concentrations in these two situations were of about 10^{-12}. Therefore, it was hard to say which results should be accepted if we did not have the concentrations in advance.

The same occurred with the extended Kalman filter. In the investigation of the initial guess of the variance of experimental noise, two extended filters were carried out. In both cases, the same initial guesses were used, but in one of the filters the initial variance of the experimental noise was assumed to be 10^{-2}, and in the other filter it was given the value 10^{-4}. Random innovation sequences were obtained, and in both cases, the state sequences were smooth, but the accuracy of the estimates differed largely. The estimated concentrations and rate constants for these two experiments are given in Table 6. It is deduced that the characteristics of the innovation and the values of the error variance of the Kalman filter cannot be directly used as criteria to estimate the accuracy of the evaluated state variables [11], since these characteristics were related with an exact model.

4.2. Treatment of experimental data

4.2.1. Optimization of the reaction conditions

Since only the basic form of the ABAs (the non-ionic free amines) are sufficiently activated to couple with diazonium ions, the reaction rate is largely affected by pH [36]. Thus, at pH < 3.5, the

Table 6
Effect of the initial guess of the absorbance variance, R, on the performance of the extended Kalman filter [a]

R	State variable	c_1	c_2	k_1	k_2
		($\times 10^{-4}$ mol l^{-1})		($\times 10^{-3}$ s^{-1})	
–	Nominal value	1.0000	1.0000	5.0000	3.5000
10^{-2}	Estimated value	1.0029	1.0280	5.0042	3.4854
10^{-4}	Estimated value	1.7680	0.2889	4.4483	2.9515

[a] Data set described in Table 4. The initial guesses for the state variables were $\mathbf{x}_0 = (0.00\ 0.00\ 0.005\ 0.0035\ 0.00)$. The initial variance guess for the estimates was $\sigma^2 = 5 \times 10^{-4}$.

Table 7
First-order rate constants for *o*-, *m*- and *p*-ABA at different pH values in a 2% SDS medium

Substrates	λ_{max} (nm)	pH	$k \pm s_k$ ($s^{-1} \times 10^{-3}$) [a]	k ($s^{-1} \times 10^{-3}$) [b]
m-ABA	360	3.90	8.16 ± 0.40	9.24
p-ABA	368	3.90	1.67 ± 0.11	1.45
o-ABA	370	3.80	4.48 ± 0.08	4.67
m-ABA	360	3.80	7.14 ± 0.07	8.28

[a] Obtained from three solutions with different substrate concentrations by the Powell algorithm.
[b] Mean value refined by the extended Kalman filter.

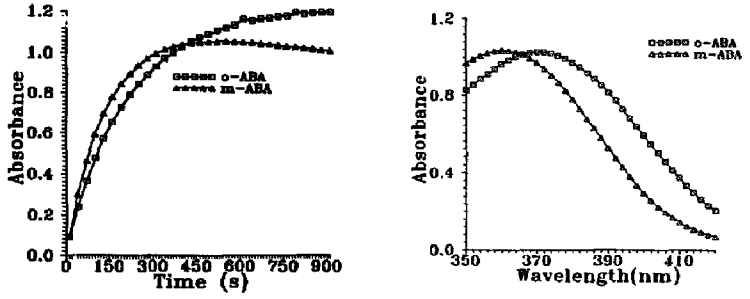


Fig. 5. Kinetic curves of the coupling reaction (left) and azodye spectra (right) for 5.06×10^{-4} mol l^{-1} *o*-ABA and *m*-ABA.

Table 8
Composition of the binary mixtures

Experiment No.	*m*-ABA/*p*-ABA ($\times 10^{-5}$ mol l^{-1})		Experiment No.	*o*-ABA/*m*-ABA ($\times 10^{-5}$ mol l^{-1})	
1	1.839	1.945	14	2.299	2.299
2	3.219	1.945	15	2.299	3.679
3	4.598	1.945	16	2.299	5.058
4	1.839	3.403	17	3.679	2.299
5	3.219	3.403	18	3.679	3.679
6	4.598	3.403	19	3.679	5.058
7	1.839	4.861	20	5.058	2.299
8	3.219	4.861	21	5.058	3.679
9	4.598	4.861	22	5.058	5.058
10	2.759	3.889	23	3.219	2.759
11	3.679	2.917	24	4.139	4.599
12	2.299	3.403	25	4.139	2.759
13	3.219	2.431	26	3.219	4.599

Initial values of the parameters for the Kalman filter and Powell algorithm [a]

Linear Kalman filter:	$x_0 = (0.00\ 0.00\ 0.00)$	$R = 10^{-6}$	$\sigma^2 = 5 \times 10^{-4}$
Extended Kalman filter:	$x_0 = (0.00\ 0.00\ k_1\ k_2\ 0.00)$	$R = 10^2$	$\sigma^2 = 5 \times 10^{-4}$
Powell algorithm:	$x_0 = (0.00\ 0.00\ k_1\ k_2\ 0.00)$		
	Step size $= (10^{-4}\ 10^{-4}\ 10^{-6}\ 10^{-6}\ 10^{-3})$		

Calibration samples for PLS

Mixtures No. 1, 3, 5, 7 and 9 for *m*- ABA/*p*-ABA and mixtures No. 14, 16, 18, 20 and 22 for *o*-ABA/*m*-ABA.

[a] The meaning of the symbols is given in Table 2. For the extended Kalman filter, the initial guesses of k_1 and k_2 are given in the last column of Table 7.

reactions were too slow, and at pH higher than 5.5, some of the analytes coupled in less than 1 min, which was not suitable for the manual mixing procedure used. Also, at pH values higher than 5.5, the absorbance of the reagent blank was large. This has been shown to be due to hydrolysis of the diazonium ion to yield a phenol which couples with the excess reagent [37]. Also, at higher pH values, the azo dyes were unstable, and the absorbance decreased rapidly after reaching a maximum. It was observed that the addition of SDS alleviated this problem and, therefore, a final concentration of 2% SDS was used.

To evaluate the rate constants, three solutions of each substrate at increasing concentrations were prepared, the procedure given above was applied, and the rate constants were calculated from the corresponding kinetic curves by the Powell algorithm and the extended Kalman filter, which were implemented to treat the data given by a single component. The results of the Powell algorithm were used as the initial values to be refined by the extended Kalman filter. The calculated rate constants were used in the computation afterwards. As expected, the results given in Table 7 showed a large variation of the rate constants with pH. The kinetic curves and the spectra of the azo dyes obtained with two *o*-ABA and *m*-ABA solutions are shown in Fig. 5. It can be observed that the *m*-ABA azodye was not stable.

4.2.2. *Resolution of binary mixtures*

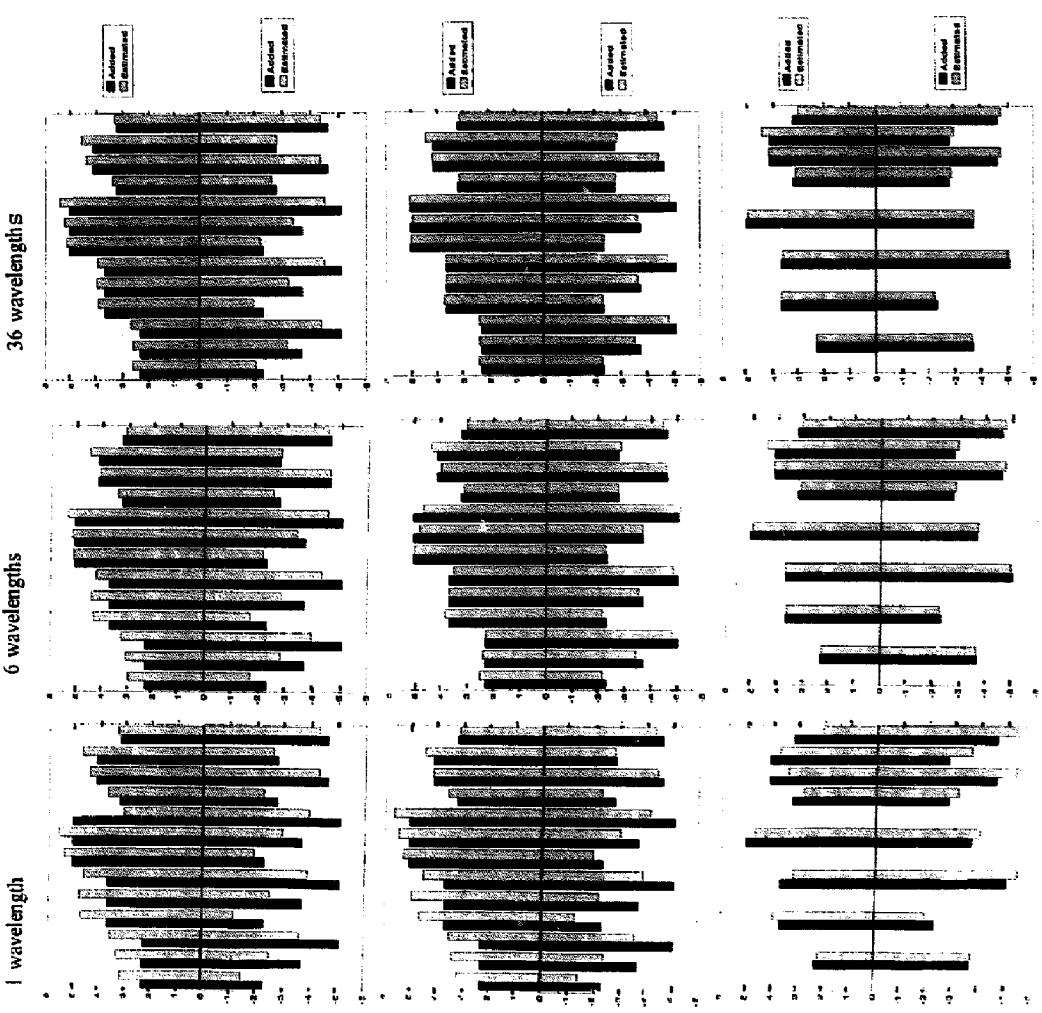
The spectral–kinetic data of two series of binary mixtures were obtained in the conditions given in Table 7. The data were treated by the four methods, and the results were evaluated using three different number of wavelengths (1, 6 and 41 for *m*-ABA/*p*-ABA mixtures, and 1, 6 and 36 for *o*-ABA/*m*-ABA mixtures). The initial parameters provided to the Powell and the Kalman filter algorithms are listed in Table 8. The nominal compositions of the mixtures are also given in Table 8. The *AARE* values of both components for the two series of binary mixtures and for the different methods are listed in Table 9. For the Powell algorithm, the estimated error was extremely large for some of the mixtures. For PLS, since five mixtures were used as the calibration samples, only the results of the remain eight mixtures were used in the computation of *AARE*. Except with the Powell algorithm, the *m*-ABA/*p*-ABA mixtures gave satisfactory results with all the methods, even when only a single wavelength was used. The

Table 9
Errors (*AARE*%, species 1/species 2) of the estimated concentrations of the binary mixtures using different algorithms

	m-ABA/*p*-ABA		
No. of wavelengths used	1	6	41
Powell	– [a]	56/20	61/7.8
Linear Kalman filter	2.7/2.9	2.9/5.4	3.4/4.6
Extended Kalman filter	8.3/4.6	4.0/3.8	3.5/3.6
PLS	2.6/1.8	2.6/1.7	2.8/1.9
	o-ABA/*m*-ABA		
No. of wavelengths used	1	6	36
Powell	– [a]	– [a]	– [a]
Linear Kalman filter	23/23	14/13	8.0/9.1
Extended Kalman filter	21/25	4.1/3.3	2.4/3.3
PLS	13/17	2.8/2.8	2.5/2.7

[a]Extremely large absurd values.

Fig. 6. Added (full bars) and estimated (grey bars) concentrations for the *o*-ABA and *m*-ABA mixtures with the linear (upper part) and extended Kalman filter (middle part) and with PLS (lower part). In each plot, the upper half are the concentrations of species 1 and the lower half shows the concentrations of species 2. In each figure, from left to right are the results with 1, 6 and 36 wavelengths, and in each plot, the bars from left to right are the thirteen mixtures in the order given in Table 8. Attention should be paid to the absence of the bars referring to the calibration mixtures for PLS.


36 wavelengths
6 wavelengths
1 wavelength
Added
Estimated

incorporation of more wavelengths did not improve the results, because of the relative large difference of the rate constants for the components in this system.

This was not the case for the *o*-ABA/*m*-ABA mixtures. The single wavelength evaluation for the *o*-ABA/*m*-ABA mixtures gave much higher error than that in the *m*-ABA/*p*-ABA mixtures, and the increase of the number of wavelengths improved the results remarkably. The estimated concentrations for *o*-ABA/*m*-ABA mixtures by different methods (except the Powell algorithm) are shown in Fig. 6, together with the nominal values of the concentrations. It can be deduced that PLS gave better results than the Kalman filters, and that the extended Kalman filter performed better than the linear one for these experimental data. The results of *o*-ABA/*m*-ABA mixtures suggested that more wavelengths should be used in practice for the linear Kalman filter compared to the extended filter. The more wavelengths used, the less the influence of the errors in the rate constants. Moreover, the influence of the number of wavelengths on the extended Kalman filter is not so significant as in the linear Kalman filter (also referring to Fig. 3B), which was attributed to the correction of the rate constants.

5. Conclusions

Firstly, accuracy improves with the number of wavelengths used, especially when the difference of rate constants is small. Secondly, the Powell algorithm gives the worst results. The different performance of the Powell algorithm and the Kalman filter might be attributed to the different minimization criteria adopted (*sum of the squares of the residuals of the signals* for the former and *mean square error of the concentrations* for the latter) and to the different manner in the treatment of the data (*batch* for the former and *recursive* for the latter). Modification in the one dimensional searching and use of some constraints to limit the parameters to be optimized might be helpful to obtain good results and to reduce the computation time. Comparing the linear and extended Kalman filters, the latter performs somewhat better than the former one. This could be attributed to the correction of the rate constants in the extended Kalman filter. The linear Kalman filter also showed a certain potential to compensate the variation of the rate constants when multiple wavelengths were used. This is due to the decrease of the influence of the error in the rate constants as more spectral information is incorporated. Though the hard modelling methods like the Kalman filter can offer reliable results under certain conditions, the requirement of relatively accurate values of the initial estimates of the parameters and of other initial values to invoke the algorithm may bring some difficulty in practice. On the other hand, the soft modelling methods like PLS can provide quite favourable results almost without any spectral and kinetic knowledge about the system. The only requirement is that the relationships between the signal and the evaluated parameters should be linear. With PLS, when adequate calibration is performed, the results in terms of accuracy and computation time are as good as with the Kalman filters. Thus, they would be more convenient for the practical application.

Acknowledgements

Work supported by the DGICYT of Spain, Project PB93/355. Y.L. Xie is grateful for a postdoctoral grant from the Ministry of Education and Science of Spain.

References

[1] B.M. Quencer and S.R. Crouch, CRC Crit. Rev. Anal. Chem., 24 (1993) 243.
[2] M. Silva, Analyst, 118 (1993) 681.
[3] H.B. Mark, Jr. and G.A. Rechnitz, Kinetic in Analytical Chemistry, Interscience, New York, 1968.
[4] H.A. Mottola, Kinetic Aspects of Analytical Chemistry, Wiley, New York, 1988.
[5] J. Havel, J.L. González and M.N. Moreno, React. Kinet. Catal. Lett., 39 (1989) 41.
[6] J.J. Baeza-Baeza, G. Ramis-Ramos, F. Pérez Plá and R. Valero Molina, Analyst, 115 (1990) 721.
[7] F. Pérez Plá, J.J. Baeza-Baeza, G. Ramis-Ramos and J. Palou, Comput. Chem., 12 (1991) 283.
[8] S.D. Brown, Anal. Chim. Acta, 181 (1986) 1.
[9] S.C. Rutan and S.D. Brown, Anal. Chim. Acta, 160 (1985) 99.
[10] Y.L. Xie, J.H. Wang and R.Q. Yu, Anal. Chim. Acta, 269 (1992) 307.
[11] P.D. Wentzell, M.I. Kazagannis and S.R. Crouch, Anal. Chim. Acta, 224 (1989) 263.

[12] W.H. Lewis, Jr. and S.C. Rutan, Anal. Chem., 63 (1991) 627.
[13] E. Förster, M. Silva, M. Otto and D. Pérez-Bendito, Anal. Chim. Acta, 274 (1993) 109.
[14] R. Xiong, A. Velasco, M. Silva and D. Pérez-Bendito, Anal. Chim. Acta, 251 (1991) 313.
[15] A. Velasco, R. Xiong, M. Silva and D. Pérez-Bendito, Talanta, 40 (1993) 1505.
[16] S.C. Rutan and S.D. Brown, Anal. Chim. Acta, 167 (1985) 23.
[17] C.A. Corcoran and S.C. Rutan, Anal. Chem., 60 (1988) 1146.
[18] C.A. Corcoran and S.C. Rutan, Anal. Chem., 60 (1988) 2450.
[19] S.C. Rutan, C.P. Fitzpatrick, J.W. Skoug, W.E. Weiser and H. Pardue, Anal. Chim. Acta, 224 (1989) 243.
[20] R. Jiménez-Prieto, A. Velasco, M. Silva and D. Pérez-Bendito, Talanta, 40 (1993) 1731.
[21] M. Gui and S.C. Rutan, Anal. Chem., 66 (1994) 1513.
[22] B.M. Quencer and S.R. Crouch, Analyst, 118 (1993) 695.
[23] B.M. Quencer and S.R. Crouch, Anal. Chem., 66 (1994) 458.
[24] V. González, B. Moreno, D. Sicilia, S. Rubio and D. Pérez-Bendito, Anal. Chem., 65 (1993) 1897.
[25] A.N. Díaz and J.A.G. García, Anal. Chem., 66 (1994) 998.
[26] M. Blanco, J. Coello, H. Iturriaga, S. Maspoch, J. Riba and E. Rovira, Talanta, 40 (1993) 261.
[27] J. Havel, F. Jiménez, R.D. Bautista and J.J. Arias-León, Analyst, 118 (1993) 1355.
[28] Y.L. Xie, J.J. Baeza-Baeza and G. Ramis-Ramos, Chemom. Intell. Lab. Syst., 27 (1995) 211.
[29] S. Wold, P. Geladi, K. Esbensen and J. Ohman, J. Chemom., 1 (1987) 41.
[30] A.K. Smilde, Chemom. Intell. Lab. Syst., 15 (1992) 143.
[31] A.K. Smilde and D.A. Doornbos, J. Chemom., 5 (1991) 345.
[32] A.K. Smilde, P.H. Van D. Graaf, D.A. Doornbos, T. Steerneman and A. Sleuriak, Anal. Chim. Acta, 235 (1990) 41.
[33] A.K. Smilde and D.A. Doornbos, J. Chemom., 6 (1992) 11.
[34] S.S. Rao, Optimization, Theory and Applications, 2nd edn., Wiley Eastern Limited, 1985.
[35] S. Wold, K. Esbensen and P. Geladi, Chemom. Intell. Lab. Syst., 2 (1987) 37.
[36] G. Ramis-Ramos, J.S. Esteve-Romero and M.C. García Alvarez-Coque, Anal. Chim. Acta, 223 (1989) 327.
[37] J.S. Esteve-Romero, M.C. García Alvarez-Coque and G. Ramis-Ramos, Talanta, 38 (1991) 1285.