

# Kinetic hard-modelling and spectral validation of rank-deficient spectroscopic data: A case study

Julien Billeter, Yorck-Michael Neuhold\*, Konrad Hungerbühler

Institute for Chemical and Bioengineering, Safety and Environmental Technology Group, ETH Zürich, Hönggerberg HCI G130, 8093 Zürich, Switzerland

## ARTICLE INFO

### Article history:

Received 5 February 2009

Received in revised form 30 June 2009

Accepted 3 July 2009

Available online 8 August 2009

### Keywords:

Kinetics

Spectroscopy

Hard-modelling

Concentration matrix

Rank deficiency and augmentation

A priori information and experimental design

## ABSTRACT

In this case study, we apply a recently developed method to systematically predict the linear dependencies in concentration profiles and identify minimum requirements to enable optimisation of rate constants and pure component spectra via direct multivariate kinetic hard-modelling of spectroscopic data. This systematic method was applied to the rank-deficient acid catalysed reaction of benzophenone with phenylhydrazine in THF. Various experimental conditions (different dosing and initial concentrations) and data treatments (defining uncoloured species, including known component spectra into the analysis) were considered. For all these conditions, the kinetic mechanism of this condensation reaction was successfully validated by the agreement between fitted and independently measured mid-IR and UV–vis pure component spectra and by the highly reproducible fitted rate constants. This case study particularly demonstrated the value of the direct spectral fitting as a tool for the validation of rank-deficient kinetic mechanisms, as inherent contributions within the fitted component spectra, due to the definition of uncoloured species, can be systematically addressed.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

Various chemometric methods capable to analyse time dependent multivariate data measured by spectroscopic techniques have been introduced in recent years [1–10]. Amongst these chemometric methods, kinetic hard-modelling, based on a hard model (the rate law), can be used to directly determine the kinetic parameters (e.g. rate constants) of chemical reactions [2,8,10–14].

In spectroscopy, multivariate data, i.e. time and wavelength/wave-number resolved absorbance measurements, can be decomposed according to Beer's law into the concentration profiles and the molar spectra of the pure components. In kinetic hard-modelling these concentration profiles are calculated by numerical integration of the rate laws describing the postulated kinetic hard-model. Maeder and Zuberbühler suggested to eliminate the pure component spectra from the non-linear optimisation and to linearly estimate them at each iteration by regressing the spectroscopic data on the concentration matrix via its pseudo-inverse [2]. This latter method, used in the present article and sometimes referred to as kinetic hard-modelling by implicit direct calibration [15,16], allows the validation of the rate law by comparing the estimated pure component spectra with

independently measured ones, or, if not fully accessible, at least with expected peaks.

For some kinetic models and experimental conditions, however, the concentration matrix is rank deficient and the pure component spectra cannot be computed, as the linear regression step cannot be performed. Different solutions have been proposed in order to circumvent this rank deficiency problem and to allow the fitting of pure component spectra: defining some absorbing species as uncoloured (Strategy 1) [10,17], including some independently known component spectra to the analysis (Strategy 2) [10,18], dosing one or more species (Strategy 3) [17] or varying some initial concentrations and analysing simultaneously the resulting experiments (Strategy 4) [14,19,20]. This last strategy is referred to as second order global analysis or 3-way analysis.

In a recent article [17], we have introduced a method for the systematic experimental and data analytical design of bi-linear spectroscopic kinetic measurements that allows identifying the species to be incorporated in strategies (1) to (4) to obtain partial or full spectral resolution. When only partial resolution is possible (Strategy 1), we have also presented a method for the calculation of the linear combinations in the fitted component spectra, i.e. the coefficients by which the true (resolved) pure spectra need to be weighted in order to obtain the reduced set of fitted (unresolved) component spectra corresponding to the species defined as coloured. Theoretical background of this approach was introduced and applied to simulated data based on various kinetic models.

\* Corresponding author. Tel.: +41 43 633 43 73; fax: +41 43 632 11 89.  
E-mail address: [bobby.neuhold@chem.ethz.ch](mailto:bobby.neuhold@chem.ethz.ch) (Y.-M. Neuhold).

In the present case study, we apply this systematic method to sets of kinetic data, measured by mid-IR and UV–vis spectroscopy simultaneously, recorded during the course of the reaction of benzophenone with phenylhydrazine in THF catalysed by acetic acid (see Fig. 1). The kinetic mechanism of this reaction has been discussed in previous articles [21,22] and is used here as a test reaction to experimentally validate the theoretical concepts presented previously [17]. Strategies (1) to (4) are applied to this reaction in order to break the rank deficiency in the concentration matrix without distorting the calculated rate constants. As a consequence of Strategy (1), linear dependencies in the concentration profiles translate into the fitted component spectra, and are compared to those theoretically predicted by our method. The underlying kinetic model is spectroscopically validated using different experimental conditions, e.g. dosing different species, via the reproducibility of the fitted rate constants and the accuracy of the fitted component spectra.

## 2. Theoretical background

In spectroscopy, Beer's law, i.e.  $\mathbf{Y} = \mathbf{C}\mathbf{A} + \mathbf{R}$ , is used to decompose a measured absorbance signal  $\mathbf{Y}$  ( $nt \times nw$ ) into the product of the concentrations  $\mathbf{C}$  ( $nt \times ns$ ) and the molar spectra  $\mathbf{A}$  ( $ns \times nw$ ) of the pure components. Deviations from the product of  $\mathbf{C}$  and  $\mathbf{A}$  are captured in a matrix  $\mathbf{R}$  ( $nt \times nw$ ) of residuals. Matrix dimensions are defined using  $nt$  as the number of reaction times,  $nw$  as the number of wavelengths and  $ns$  as the total number of species. For details regarding the notation, we refer to a previous publication introducing the mathematical concepts of our approach [17]. In the following, we provide a short summary.

Kinetic hard-modelling uses the kinetic rate law to define a system of ordinary differential equations (ODE) that depend on kinetic parameters, e.g. rate constants. This chemical model can be uniquely identified using the matrix of reactants coefficients,  $\mathbf{E}$  ( $nr \times ns$ ), and the matrix of stoichiometry,  $\mathbf{N}$  ( $nr \times ns$ ), for all species involved in the  $nr$  elementary reaction steps [23]. The matrix of concentration profiles  $\mathbf{C}$  can then be calculated for a given set of rate constants  $\mathbf{k}$  ( $nr \times 1$ ) by numerical integration of the ODE, using initial concentrations,  $\mathbf{c}_0$  ( $1 \times ns$ ), matrices  $\mathbf{E}$  and  $\mathbf{N}$ , as well as dosing conditions, i.e. dosing rates  $\mathbf{f}$  ( $nt \times 1$ ) and concentrations  $\mathbf{C}_{in}$  ( $nf \times ns$ ) of the  $nf$  dosing steps [14,24]. Note that in this somewhat simplified approach, changes in density or activity are not taken into account.

In direct kinetic hard-modelling of spectroscopic data, originally proposed by Maeder and Zuberbühler [2], the product of the integrated concentration profiles  $\mathbf{C}$  and of the pure component spectra  $\mathbf{A}$  is compared to the measured data matrix  $\mathbf{Y}$ , and results in the residuals  $\mathbf{R} = \mathbf{Y} - \mathbf{C}\mathbf{A}$ , capturing the differences between the measured and the modelled absorbances. In the least-squares analysis, the sum of all squared residuals  $\mathbf{R}$  ( $nt \times nw$ ) is used as the objective function to be minimised by iteratively optimising the rate constants  $\mathbf{k}$ . For this article, the Newton–Gauss–Levenberg/Marquardt algorithm (NGL/M) was used to solve this non-linear regression [2,10,24–27]. This gradient-based method allows estimating the uncertainties in the optimised rate constants from the variance/covariance matrix, including the propagation of errors, such as the uncertainties in initial concentrations and in dosing rate, as discussed previously [21].

As  $\mathbf{A}$  is comprised of linear parameters only, it can be eliminated from the non-linear optimisation and replaced by its linear least-squares estimate [10,14].

$$\mathbf{A} = \mathbf{C}^+ \mathbf{Y} = (\mathbf{C}^T \mathbf{C})^{-1} \mathbf{C}^T \mathbf{Y} \quad (1)$$

Eq. (1) is only applicable if  $\mathbf{C}$  is of full rank. If  $\mathbf{C}$  is rank deficient, one or more strategies (1 to 4) need to be used in order to allow the computation of  $\mathbf{A}$  by linear regression. The species to include in these four strategies can be identified from the linear dependencies in the  $ns$  concentration profiles of  $\mathbf{C}$ , obtained by the kernel of  $\mathbf{C}$ , denoted  $\ker \mathbf{C}$ . The kernel of  $\mathbf{C}$  is defined as the vector basis spanned by the vectors forming the null space  $\mathbf{0}$  when left-multiplied by  $\mathbf{C}$ , i.e.  $\mathbf{C}(\ker \mathbf{C}) = \mathbf{0}$  [28]. Note, if the kernel is not empty, it also defines the mass balance (closure) of the chemical system. The analogy between kernel and mass balance is demonstrated in Appendices A.6 and A.7. Due to its definition, the kernel inherently involves a rotational ambiguity and thus any linear combinations of the kernel also span a basis for this vector space. Importantly, a row only comprised by zeros in  $\ker \mathbf{C}$  indicates that the corresponding concentration profile is linearly independent from the others.

To predict the linear dependencies in the concentration profiles without numerical integration, we recently introduced a time-invariant augmented matrix  $\mathbf{\Omega}$  of general dimensions ( $ns + nf + ne + 1 \times ns$ ) having the same kernel as the time variant matrix of concentrations  $\mathbf{C}$  ( $nt \times ns$ ) [17]. In this notation,  $(ne + 1)$  defines the number of simultaneously analysed experiments. In Eq. (2), matrix  $\mathbf{\Omega}$  is presented for the case of one single batch experiment ( $nf = 0$  and  $ne = 0$ ), i.e.  $\mathbf{\Omega}$  has dimensions ( $ns + 1 \times ns$ ).

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1}) \cdot \mathbf{E}^T \text{DIAG}(\mathbf{k}) \mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} \quad (2)$$

With  $\mu$  being an arbitrary positive scalar different from 1, matrix  $\mathbf{1}$  ( $ns \times nr$ ) comprised of ones only and the superscript  $\cdot \mathbf{E}^T$  representing the element-wise raise to the power of  $\mathbf{E}^T$ .

The prediction for the linear dependencies in  $\mathbf{C}$  and in the fitted component spectra  $\mathbf{A}$ , based on the analysis of  $\mathbf{\Omega}$ , assumes that (i) the kinetic hard-model is correct, (ii) no reactions are identical, (iii) all species initiating the reactions are present initially or are dosed, (iv) the true pure component spectra are linearly independent, (v) each provided pure component spectrum is correct, (vi) all species absorb or the truly non-absorbing species are known and defined such that they do not contribute to Beer's law, and (vii) all reaction steps defined by the kinetic hard-model can be observed in  $\mathbf{Y}$ .

### 2.1. Strategy (1)—defining uncoloured species

When Strategy (1) is used to circumvent linear dependencies in the concentration profiles, Beer's law is reduced by defining ( $ns - nu$ ) coloured and  $nu$  uncoloured species, and by eliminating from the time invariant matrix  $\mathbf{\Omega}$  (and  $\mathbf{C}$ ) the columns corresponding to these  $nu$  uncoloured species. This leads to reduced matrices  $\mathbf{\Omega}$  and  $\mathbf{C}_c$  of dimensions ( $ns + 1 \times ns - nu$ ) and ( $nt \times ns - nu$ ) respectively. As a consequence of this reduction, the fitted component spectra of the

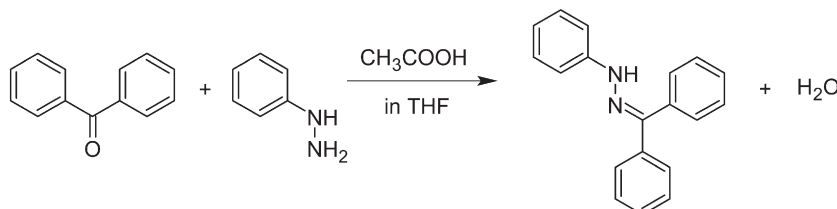


Fig. 1. Reaction of benzophenone (B) with phenylhydrazine (P) catalysed by acetic acid (Aa) forming benzophenone–phenylhydrazone (BP) and water.

coloured species  $\mathbf{A}_c$  ( $ns - nu \times nw$ ) are comprised of linear combinations of the true (resolved) pure spectra  $\mathbf{A}$  ( $ns \times nw$ ). To obtain the coefficients of these linear combinations, the linear dependencies in the rows of the non-reduced matrix  $\mathbf{\Omega}$  ( $ns + 1 \times ns$ ) have to be eliminated, leading to a matrix  $\mathbf{\Omega}_{\text{ind}}$  of dimensions ( $nr_i + 1 \times ns$ ), where  $nr_i$  denotes the number of independent reactions, and  $\ker \mathbf{\Omega}_{\text{ind}}^T = \mathbf{0}$ . Based on the coloured species and the linearly independent reactions, the coefficients for the linear combinations of the true pure component spectra,  $\Delta$  ( $ns - nu \times ns$ ), can be calculated as follows:

$$\mathbf{A}_c = \Delta \mathbf{A} \text{ with } \Delta = (\mathbf{\Omega}_{\text{ind}} |_{\text{comprised of coloured species}})^{-1} \mathbf{\Omega}_{\text{ind}} |_{\text{comprised of all species}} \quad (3)$$

## 2.2. Strategy (2)—including known spectra into the analysis

In this case,  $nks$  pure component spectra are included into the kinetic analysis in order to avoid linear dependencies in the concentration profiles. Here, the corresponding  $nks$  columns of  $\mathbf{\Omega}$  and  $\mathbf{C}$  are removed, leading to a reduced time-invariant matrix  $\mathbf{\Omega}$  of dimensions ( $ns + 1 \times ns - nks$ ) and the corresponding reduced matrix of concentrations  $\mathbf{C}_{\text{uk}}$  of dimensions ( $nt \times ns - nks$ ).

## 2.3. Strategy (3)—dosing one or more species

This strategy employs  $nf$  successive dosing steps, such that the rank of  $\mathbf{\Omega}$  (and  $\mathbf{C}$ ) becomes augmented. Thus matrix  $\mathbf{\Omega}$  is expanded to dimensions ( $ns + nf + 1 \times ns$ ) with the appropriate rows corresponding to the dosed concentrations of each dosing step. The dosed concentrations are collected in matrix  $\mathbf{C}_{\text{in}}$  ( $nf \times ns$ ).

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1}) \cdot \mathbf{E}^T \text{DIAG}(\mathbf{k}) \mathbf{N} \\ \mathbf{c}_0 \\ \mathbf{C}_{\text{in}} \end{bmatrix} \quad (4)$$

## 2.4. Strategy (4)—performing a second order global analysis

Strategy (4), also known as second order global analysis, is used to augment the rank of  $\mathbf{\Omega}$  (and  $\mathbf{C}$ ) by concatenating multiple kinetic experiments performed under various initial concentrations. When common pure component spectra are fitted (global mode), matrix  $\mathbf{\Omega}$  is expanded to dimensions ( $ns + ne + 1 \times ns$ ) by adding  $ne$  rows corresponding to the different initial concentrations of the  $ne$  additional experiments, which are collected in matrix  $\mathbf{C}_{\text{in}}^{\text{ne}}$  ( $ne \times ns$ ).

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1}) \cdot \mathbf{E}^T \text{DIAG}(\mathbf{k}) \mathbf{N} \\ \mathbf{c}_0 \\ \mathbf{C}_{\text{in}}^{\text{ne}} \end{bmatrix} \quad (5)$$

## 3. Experiments

The reaction of benzophenone (species  $B$ ) with phenylhydrazine ( $P$ ) in THF catalysed by acetic acid ( $Aa$ ) can be monitored in mid-IR and UV-vis spectral ranges. Products of the reaction are benzophenone-phenylhydrazone ( $BP$ ), water and the regenerated catalyst ( $Aa$ ). This model reaction has been extensively described for acetic acid being dosed into benzophenone and phenylhydrazine [21,22]. In a recent paper, we have also shown that this condensation reaction and the optimisation of its rate constant are highly reproducible [21], thus being an ideal case study to validate our systematic method to treat kinetic rank deficiencies.

## 3.1. Instruments

Experiments were carried out in the Combined Reaction Calorimeter (CRC.v4), a small-scale reaction calorimeter working under isothermal conditions and combining the principle of power compensation and heat balance [29]. As the reaction does not produce significant amounts of heat, the calorimetric heat signal was disregarded and the well-defined environment of the CRC.v4 was only used to maintain a constant temperature of  $25 (\pm 0.04)^\circ\text{C}$  for all experiments. For a detailed description of the reactor, we refer to [29].

Mid-IR signals were followed by ATR FT-IR spectroscopy (ReactIR 4000, K-4 conduit and ATR crystal from Mettler Toledo) in the range of  $1200$  to  $1650 \text{ cm}^{-1}$  at a resolution of  $4 \text{ cm}^{-1}$ . UV-vis signals were monitored by ATR UV-vis spectroscopy (Cary 50 from Varian coupled to an ATR dip-probe from Hellma) between  $240$  and  $400 \text{ nm}$  at  $1 \text{ nm}$  resolution. UV-vis and mid-IR signals were recorded simultaneously every minute during  $150 \text{ min}$  against THF as the reference background.

## 3.2. Sample preparation

A solution of phenylhydrazine in THF (Acros Organics for analysis, certified 99.99%) was prepared by weighing  $40.003 \text{ g}$  of phenylhydrazine (Aldrich-Fine Chemicals, certified 99.2%) into a volumetric flask and filling up to  $250 \text{ mL}$  with THF. A benzophenone solution was prepared by weighing  $36.261 \text{ g}$  of benzophenone (Fluka purum, certified 99.8%) and filling up to  $100 \text{ mL}$  with THF. This sample preparation led to stock concentrations of  $1.468 (\pm 0.001) \text{ mol L}^{-1}$  for the phenylhydrazine solution and  $1.986 (\pm 0.002) \text{ mol L}^{-1}$  for the benzophenone solution. Uncertainties were obtained from propagating errors due to sample preparation (weighing, filling and pipetting) [21]. Glacial acetic acid (Carlo Erba Reagents for analysis, certified 100.0%) was directly used from its original bottle ( $17.483 \text{ mol L}^{-1}$ ) and no sampling error needed to be propagated.

## 3.3. Independently measured pure component spectra

Mid-IR and UV-vis pure component spectra of benzophenone ( $1.986 \text{ mol L}^{-1}$  for mid-IR,  $0.993 \text{ mol L}^{-1}$  for UV-vis), phenylhydrazine ( $1.468 \text{ mol L}^{-1}$  for mid-IR and UV-vis), acetic acid ( $2.914 \text{ mol L}^{-1}$  for mid-IR,  $8.742 \text{ mol L}^{-1}$  for UV-vis), benzophenone-phenylhydrazine ( $1.000 \text{ mol L}^{-1}$  for mid-IR,  $0.500 \text{ mol L}^{-1}$  for UV-vis) and water ( $27.704 \text{ mol L}^{-1}$  for mid-IR,  $41.556 \text{ mol L}^{-1}$  for UV-vis) were independently measured at  $25^\circ\text{C}$  in THF. As not commercially available, the hydrazone product was synthesised by performing the reaction in acetonitrile (Acros Organics for analysis), in which it is not soluble and crystallises. Two successive re-crystallisations were performed to obtain the pure product. The product purity was assessed by GC to be more than 99% and the structure was confirmed by MS and NMR.

## 3.4. Experimental conditions (a)—dosing Aa

The reactor was initially charged with  $20 \text{ mL}$  of the phenylhydrazine solution and  $5 \text{ mL}$  of the benzophenone solution, then  $5 \text{ mL}$  of glacial acetic acid were quickly injected using a syringe. Note that for all experimental conditions (a to e), quick injections were done within  $24 \text{ s}$ . Experimental conditions (a) are similar to those used in previous papers [21,22]. The experiment performed under these conditions (Experiment #1) was analysed assuming batch and semi-batch conditions. Note, when semi-batch conditions are assumed, the first measured spectrum in data matrix  $\mathbf{Y}$  is the one recorded just before dosing, whereas, when batch conditions are considered, it is the one recorded just after dosing. For batch conditions, initial concentrations of  $0.979 (\pm 0.007) \text{ mol L}^{-1}$  in phenylhydrazine,  $0.331 (\pm 0.003) \text{ mol L}^{-1}$  in benzophenone and  $2.9 (\pm 0.1) \text{ mol L}^{-1}$  in acetic acid were used, accounting for the dilution due to the fast dosing of

Aa. Note that this batch mode of analysis, neglecting the conversion of about 0.7% during the fast dosing, was shown to be reasonable (see Table 2). For semi-batch conditions, initial concentrations of  $1.174 \text{ mol L}^{-1}$  ( $\pm 0.004$ ) and  $0.397$  ( $\pm 0.002$ )  $\text{mol L}^{-1}$  in phenylhydrazine and benzophenone respectively, a dosing concentration of  $17.483 \text{ mol L}^{-1}$  in acetic acid and a dosing rate of  $12.5$  ( $\pm 0.5$ )  $\text{mL min}^{-1}$  were considered. All experimental conditions are summarised in Table 1.

### 3.5. Experimental conditions (b)—dosing B ( $1.986 \text{ mol L}^{-1}$ )

In these experimental conditions, 5 mL of the benzophenone solution was quickly dosed into the reactor, which was initially filled with 20 mL of the phenylhydrazine solution and 5 mL of glacial acetic acid. Three experiments (Experiments #2 to #4) were carried out under these conditions (see Table 1) and were analysed assuming batch conditions, using initial concentrations of  $0.979$  ( $\pm 0.007$ )  $\text{mol L}^{-1}$  in phenylhydrazine,  $0.33$  ( $\pm 0.01$ )  $\text{mol L}^{-1}$  in benzophenone and  $2.91$  ( $\pm 0.02$ )  $\text{mol L}^{-1}$  in acetic acid. Experiment #2 was also analysed assuming semi-batch conditions, with initial concentrations of  $1.174$  ( $\pm 0.004$ )  $\text{mol L}^{-1}$  in phenylhydrazine,  $3.50$  ( $\pm 0.02$ )  $\text{mol L}^{-1}$  in acetic acid, a dosing concentration of  $1.986$  ( $\pm 0.002$ )  $\text{mol L}^{-1}$  in benzophenone and a dosing rate of  $12.5$  ( $\pm 0.5$ )  $\text{mL min}^{-1}$ .

### 3.6. Experimental conditions (c)—sequential dosing of B and Aa

The reactor was initially filled with 20 mL of the phenylhydrazine solution. A first quick dosing of 5 mL of benzophenone was performed. Subsequently, 5 mL of glacial acetic acid were quickly dosed. One experiment (Experiment #5) was carried out under these conditions and analysed assuming semi-batch conditions (see Table 1), with initial concentrations of  $1.468 \text{ mol L}^{-1}$  ( $\pm 0.001$ ) in phenylhydrazine, dosing concentrations of  $1.986$  ( $\pm 0.002$ )  $\text{mol L}^{-1}$  and  $17.483 \text{ mol L}^{-1}$  in benzophenone and acetic acid respectively, added at dosing rates of  $12.5$  ( $\pm 0.5$ )  $\text{mL min}^{-1}$ .

### 3.7. Experimental conditions (d)—dosing P

Here, 5 mL of the benzophenone solution was initially mixed in the reactor with 5 mL glacial acetic acid. Then, 20 mL of the phenylhydrazine solution was quickly dosed. The experiment performed under these conditions (Experiment #6) was analysed assuming batch and semi-

batch conditions (see Table 1). For batch conditions, initial concentrations of  $0.98$  ( $\pm 0.05$ )  $\text{mol L}^{-1}$  in phenylhydrazine,  $0.331$  ( $\pm 0.009$ )  $\text{mol L}^{-1}$  in benzophenone and  $2.91$  ( $\pm 0.08$ )  $\text{mol L}^{-1}$  in acetic acid were used. For semi-batch conditions, initial concentrations of  $0.993$  ( $\pm 0.006$ )  $\text{mol L}^{-1}$  in benzophenone,  $8.74$  ( $\pm 0.05$ )  $\text{mol L}^{-1}$  in acetic acid, a dosing concentration of  $1.468$  ( $\pm 0.001$ )  $\text{mol L}^{-1}$  in phenylhydrazine and a dosing rate of  $50$  ( $\pm 2$ )  $\text{mL min}^{-1}$  were considered.

### 3.8. Experimental conditions (e)—dosing B ( $1.192 \text{ mol L}^{-1}$ )

In this fifth experimental procedure, 20 mL of the phenylhydrazine solution and 5 mL of glacial acetic acid were initially charged in the reactor and a mixture of 3 mL of the benzophenone solution and additional 2 mL THF was quickly dosed. This was done in order to maintain an approximately constant THF concentration to avoid baseline changes due to the change of the reference spectrum (THF). The experiment performed under these conditions (Experiment #7) was analysed under batch conditions (see Table 1), with initial concentrations of  $0.98$  ( $\pm 0.01$ )  $\text{mol L}^{-1}$  in phenylhydrazine,  $0.20$  ( $\pm 0.01$ )  $\text{mol L}^{-1}$  in benzophenone and  $2.91$  ( $\pm 0.03$ )  $\text{mol L}^{-1}$  in acetic acid. These conditions resemble conditions (b) except that the dosing concentration of B represents only 3/5 compared to the one used in experimental conditions (b), i.e.  $1.192 \text{ mol L}^{-1}$ .

## 4. Results and discussion

As the reaction studied in this paper only involves one single kinetic step, several matrices and vectors collapse to vectors and scalars respectively. However, for brevity, the boldface capital and boldface lowercase notations for matrices and vectors respectively are maintained, e.g. vector  $\mathbf{k}$  is comprised by only one element  $k_1$ . The few Matlab lines required for the calculation of most equations of the following sections are given in the Appendix.

### 4.1. Analytical solution for $\Omega$

The reaction of benzophenone with phenylhydrazine catalysed by acetic acid, as described in Fig. 1, includes  $ns = 5$  species ( $B, P, Aa, BP$  and  $H_2O$ ) involved in  $nr = 1$  reaction. The reaction follows a third order rate law  $\Phi$  ( $nt \times 1$ ), with partial orders of one for each reactant ( $B, P, Aa$ ) [21,22]. As water is transparent in the followed range of mid-IR (see Fig. 2a) and UV-vis (see Fig. 2b), it can be omitted and considered as non-absorbing, formally including for this species a known spectrum

**Table 1**  
Experimental conditions (a) to (e), and related methods of analysis (batch, semi-batch analysis).

Experimental conditions <sup>a</sup>	Dosed volumes [mL]			Initial volumes [mL]			Data analysis <sup>b</sup>	Dosing concentrations [ $\text{mol L}^{-1}$ ] <sup>c</sup>			Initial concentrations [ $\text{mol L}^{-1}$ ] <sup>c</sup>			# Exp.
	B	P	Aa	B	P	Aa		B	P	Aa	B	P	Aa	
(a)	–	–	5 <sup>d</sup>	5	20	–	Semi-batch	–	–	17.483 <sup>e</sup>	0.397 ( $\pm 0.002$ )	1.174 ( $\pm 0.004$ )	–	1
	–	–	–	–	–	–	Batch	–	–	–	0.331 ( $\pm 0.003$ )	0.979 ( $\pm 0.007$ )	2.9 ( $\pm 0.1$ )	1
(b)	5 <sup>d</sup>	–	–	–	20	5	Semi-batch	1.986 ( $\pm 0.002$ )	–	–	–	1.174 ( $\pm 0.004$ )	3.50 ( $\pm 0.02$ )	2
	–	–	–	–	–	–	Batch	–	–	–	0.33 ( $\pm 0.01$ )	0.979 ( $\pm 0.007$ )	2.91 ( $\pm 0.02$ )	2–4
(c)	5 <sup>d, f</sup>	–	5 <sup>d, f</sup>	–	20	–	Semi-batch	1.986 ( $\pm 0.002$ )	–	17.483 <sup>e</sup>	–	1.468 ( $\pm 0.001$ )	–	5
(d)	–	20 <sup>g</sup>	–	5	–	5	Semi-batch	–	1.468 ( $\pm 0.001$ )	–	0.993 ( $\pm 0.006$ )	–	8.74 ( $\pm 0.05$ )	6
	–	–	–	–	–	–	Batch	–	–	–	0.331 ( $\pm 0.009$ )	0.98 ( $\pm 0.05$ )	2.91 ( $\pm 0.08$ )	6
(e)	3 <sup>d, h</sup>	–	–	–	20	5	Batch	–	–	–	0.20 ( $\pm 0.01$ )	0.98 ( $\pm 0.01$ )	2.91 ( $\pm 0.03$ )	7

<sup>a</sup> Experimental conditions are described in the following sections: (a) in Section 3.4, (b) in Section 3.5, (c) in Section 3.6, (d) in Section 3.7, (e) in Section 3.8.

<sup>b</sup> All experiments were performed under semi-batch conditions (fast dosing) but analysed assuming either batch or semi-batch conditions. When semi-batch conditions are assumed, the first measured spectrum in Y is the one recorded just before dosing, whereas, when batch conditions are considered, it is the one recorded just after dosing. For batch conditions, concentrations are given accounting for the dilution due to dosing.

<sup>c</sup> Concentrations are rounded to their last significant digit. Uncertainties are calculated from propagating errors due to sample preparation (weighing, filling and pipetting) [21].

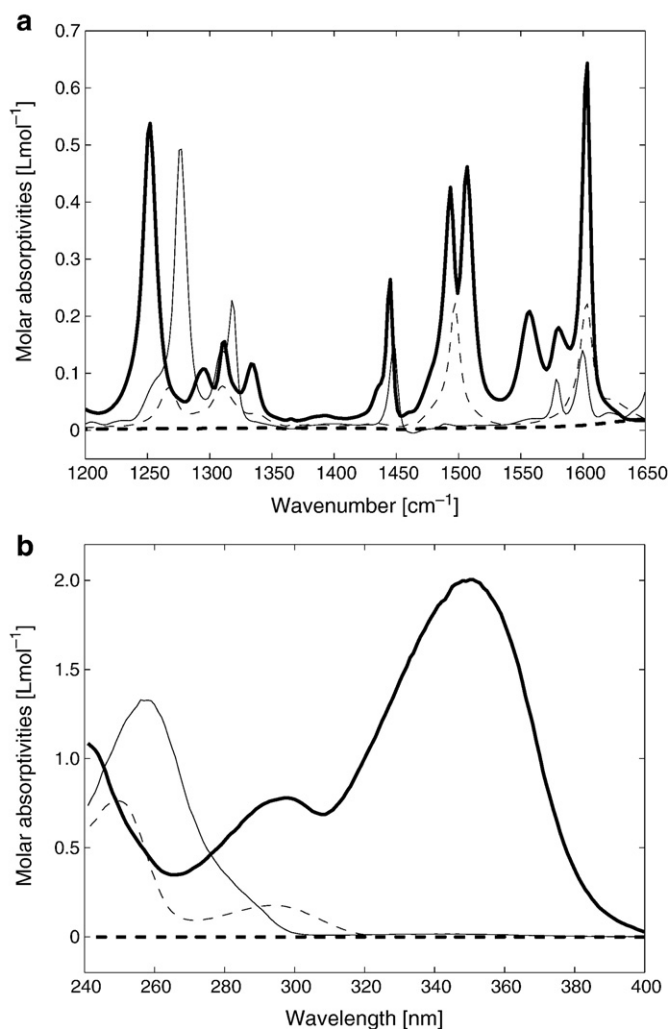
<sup>d</sup> Volume added at a dosing rate of  $12.5$  ( $\pm 0.5$ )  $\text{mL min}^{-1}$ .

<sup>e</sup> No sampling error was propagated for Aa as it was used from its original bottle.

<sup>f</sup> B and then Aa were dosed sequentially.

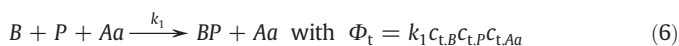
<sup>g</sup> Volume added at a dosing rate of  $50$  ( $\pm 2$ )  $\text{mL min}^{-1}$ .

<sup>h</sup> A mixture of 3 mL of the benzophenone stock solution and 2 mL THF was dosed.



**Fig. 2.** Independently measured pure component spectra of benzophenone *B* (—), phenylhydrazine *P* (---), benzophenone-phenylhydrazone *BP* (-·-) and water (···) in mid-IR (a) and in UV-vis (b) with THF as background.

comprised by zeros into the analysis (Strategy 2). Thus, the number of species is reduced to  $ns - nks = 5 - 1 = 4$ . For simplicity, this reduced number of species ( $ns - nks$ ) is re-defined as  $ns$ , i.e. we assume a kinetic model with  $ns = 4$  species (*B*, *P*, *Aa*, *BP*). For a description of the differential equations describing the concentration profiles, we refer to [21].



Matrices  $\mathbf{E}$  ( $1 \times 4$ ) and  $\mathbf{N}$  ( $1 \times 4$ ), as given in Eqs. (7) and (8), describe the kinetic hard-model of Eq. (6) in terms of reactant and stoichiometric coefficients respectively.

$$\mathbf{E} = \begin{bmatrix} B & P & Aa & BP \\ 1 & 1 & 1 & 0 \end{bmatrix} \quad (7)$$

$$\mathbf{N} = \begin{bmatrix} B & P & Aa & BP \\ -1 & -1 & 0 & 1 \end{bmatrix} \quad (8)$$

The initial concentrations  $\mathbf{c}_0$  can be defined symbolically in order to account for all different initial conditions used in Sections 3.4–3.8.

$$\mathbf{c}_0 = \begin{bmatrix} B & P & Aa & BP \\ c_{0,B} & c_{0,P} & c_{0,Aa} & 0 \end{bmatrix} \quad (9)$$

Using matrices  $\mathbf{E}$  and  $\mathbf{N}$ , matrix  $\mathbf{\Omega}$  ( $5 \times 4$ ), as introduced in Eq. (2), can be calculated as a function of  $k_1$ ,  $\mathbf{c}_0$  and  $\mu$ .

$$\mathbf{\Omega} = \begin{bmatrix} B & P & Aa & BP \\ (\mu \mathbf{1})^T \cdot \mathbf{E}^T k_1 \mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} = \begin{bmatrix} -\mu k_1 & -\mu k_1 & 0 & \mu k_1 \\ -\mu k_1 & -\mu k_1 & 0 & \mu k_1 \\ -\mu k_1 & -\mu k_1 & 0 & \mu k_1 \\ -k_1 & -k_1 & 0 & k_1 \\ c_{0,B} & c_{0,P} & c_{0,Aa} & 0 \end{bmatrix} \quad (10)$$

As can be easily seen in Eq. (10), row five of  $\mathbf{\Omega}$  is linearly independent from the others, and either of rows one to three are linearly dependent on row four. Consequently,  $\mathbf{\Omega}_{\text{ind}}$  can, for example, be comprised by rows four and five, for which  $\ker \mathbf{\Omega}_{\text{ind}}^T = \mathbf{0}$ .

$$\mathbf{\Omega}_{\text{ind}} = \begin{bmatrix} B & P & Aa & BP \\ -k_1 & -k_1 & 0 & k_1 \\ c_{0,B} & c_{0,P} & c_{0,Aa} & 0 \end{bmatrix} \quad (11)$$

The linear dependencies in  $\mathbf{C}$  can be predicted from  $\ker \mathbf{\Omega}$  ( $4 \times 2$ ), which does neither depend on  $\mathbf{k}$  nor on  $\mu$ . The matlab lines for Eqs. (7)–(12), making use of the symbolic toolbox [30], are given in Appendix A.1.

$$\ker \mathbf{\Omega} = \begin{bmatrix} -\alpha & -\beta \\ 1 & 0 \\ 0 & 1 \\ 1-\alpha & -\beta \end{bmatrix} \begin{matrix} B \\ P \\ Aa \\ BP \end{matrix} \quad \text{with } \alpha = \frac{c_{0,P}}{c_{0,B}} \text{ and } \beta = \frac{c_{0,Aa}}{c_{0,B}} \quad (12)$$

The dimension of the kernel indicates that the matrix of concentration profiles  $\mathbf{C}$  for this mechanism is two times rank-deficient. The absence of rows comprised by zeros only in  $\ker \mathbf{\Omega}$  (and in  $\ker \mathbf{C}$ ) leads to the conclusion that all reactive species (*B*, *P*, *Aa* and *BP*) are involved in the linear dependencies of  $\mathbf{C}$ . Two mass balances for this mechanism can be calculated according to the definition of the kernel, i.e.  $\mathbf{C}(\ker \mathbf{C}) = \mathbf{0}$ .

$$(-\alpha)c_{t,B} + (1)c_{t,P} + (0)c_{t,Aa} + (1-\alpha)c_{t,BP} = \mathbf{0} \quad (13)$$

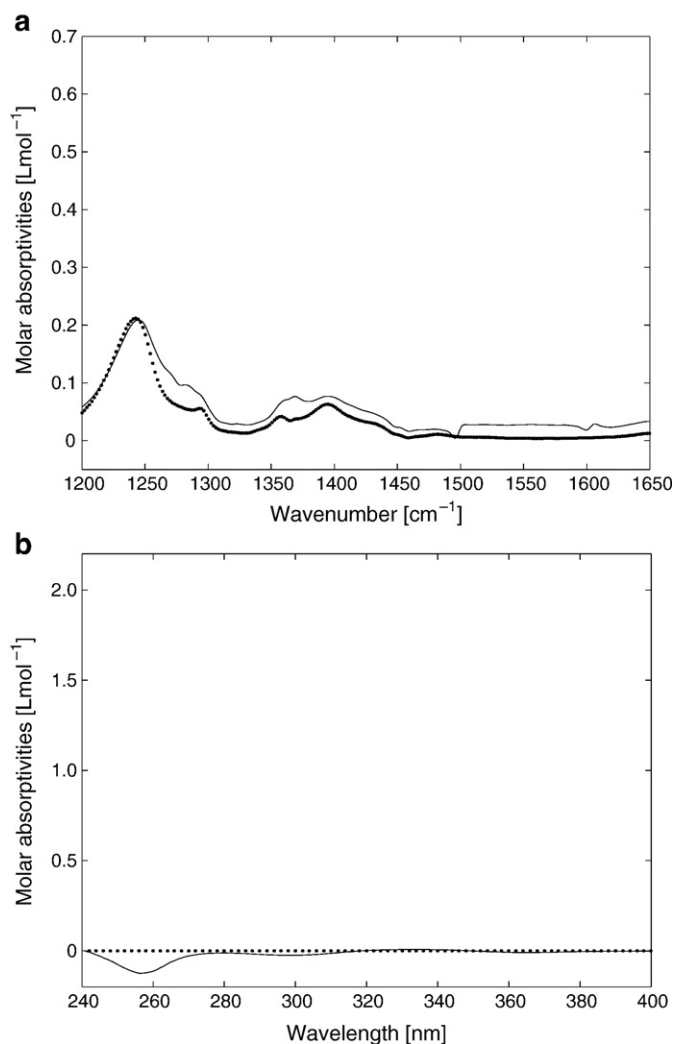
$$(-\beta)c_{t,B} + (0)c_{t,P} + (1)c_{t,Aa} + (-\beta)c_{t,BP} = \mathbf{0} \quad (14)$$

Coefficients in parenthesis, representing the values given by the kernel of  $\mathbf{\Omega}$  (see Eq. (12)), can alternatively be calculated by mass balance equations, as described in Appendix A.6.

To treat this rank deficiency problem in  $\mathbf{C}$  and to fit pure component spectra, one of the four strategies described in Sections 2.1–2.4 has to be applied. For reasons that will become obvious in Section 4.2, no pure Strategies (2), (3) and (4) were applied but they were combined with Strategy (1).

#### 4.2. Independently measured pure component spectra

Pure component spectra of species *B*, *P*, *BP* and  $H_2O$ , independently measured as outlined in Section 3.3, are shown in Fig. 2a (mid-IR) and Fig. 2b (UV-vis). As molar absorptivities strongly depend on the solvent and can be affected by baseline shifts, deviations have to be expected between these pure spectra (measured in plain THF as reference) and those measured in a mixture of solvents comprised by acetic acid and THF in a molar ratio of approximately 1:4, as described in the experimental procedures (a) to (e). For all experimental conditions described in Section 3, dosing *Aa*, *B* or *P* affects significantly the molar ratio of THF and acetic acid, i.e. the medium properties, and thus may also alter the pure component spectra.



**Fig. 3.** Fitted (—) and independently measured (•) mid-IR (a) and UV-vis (b) component spectrum of acetic acid (*Aa*) from the analysis of Experiment #1 assuming semi-batch conditions (*Aa* dosed) and using Strategy (1), setting species *B* and *P* uncoloured.

To counteract this solvent effect, a preliminary experiment (Experiment #1) was performed under experimental conditions (a). This experiment was analysed assuming semi-batch conditions using strategies (1), setting species *B* and *P* uncoloured, and Strategy (3), dosing *Aa*, and optimising the rate constant  $\mathbf{k}$ . As the dosing is fast and only occurs between the acquisition of two reaction spectra, solvent and baseline effects are expected to translate into the resolved fitted component spectrum of acetic acid, as its concentration stays constant with respect to THF after dosing (catalytic effect). This is shown in Fig. 3a (mid-IR) and Fig. 3b (UV-vis). Note that, despite this solvent effect, the fitted mid-IR spectrum of *Aa* still shows the distinct pattern of the independently measured one, and the fitted UV-vis spectrum of *Aa* just obtains some small contributions compared to the non-absorbing independently measured one.

The fitted pure spectrum of *Aa* (in mid-IR and UV-vis) obtained through the analysis of this preliminary experiment can subsequently be used, together with the independently measured pure spectra of *B*, *P* and *BP*, to reconstitute the fitted component spectra for the experiments of Sections 4.3, 4.4 and 4.6 obtained under different experimental conditions (b and e) and using different strategies (1, 2 and 4). Thus, in first approximation, the deviations due to the change of the solvent mixture of THF and acetic acid do not have to be modelled more deeply.

The optimised rate constants (in mid-IR and UV-vis) obtained for this preliminary experiment analysed under semi-batch conditions are close to literature values as shown in Table 2. Differences in rate constants compared to previous works [21,22] are likely due to differences in the purity of the chemicals. This was confirmed by a limited investigation showing that the fitted rate constants depend on the batch to batch variability in purity of the purchased THF and phenylhydrazine. For the sake of completeness, Experiment #1 was also analysed assuming batch conditions and using Strategy (1), treating species *P* and *Aa* as uncoloured. This analysis resulted in optimised rate constants that are very similar compared to the values obtained when the experiment was analysed assuming semi-batch conditions (see Table 2). This confirmed the applicability of a batch analysis for this fast semi-batch experiment.

#### 4.3. Application of Strategy (1)—defining uncoloured species

To analyse the reaction of benzophenone with phenylhydrazine by Strategy (1), experimental conditions (b) were used, i.e. species *B* was quickly dosed into a mixture of *P* and *Aa*. Three experiments (Experiments #2–#4) were performed and analysed individually assuming batch conditions. If the rate constants are not a priori known,  $\mathbf{\Omega}$  ( $5 \times 4$ ) and  $\ker \mathbf{\Omega}$  ( $4 \times 2$ ) can be calculated as a function of  $\mathbf{k}$ ,  $\mathbf{c}_0$  and  $\mu$ . This leads to the general solution given in Eqs. (10) and (12). Alternatively,  $\mathbf{\Omega}$  and  $\ker \mathbf{\Omega}$  can be calculated by defining  $\mathbf{k} = \mathbf{1}$ , as any strictly positive value does not affect rows of zero in the kernel of the augmented matrix [17]. Using, for example,  $\mu = 2$  and the initial conditions described in Section 3.5, i.e.  $\mathbf{c}_0 = [0.33 \ 0.979 \ 2.91 \ 0]$ ,  $\mathbf{\Omega}$  can be calculated numerically.

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1})^T \mathbf{E}^T \mathbf{1} \mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} = \begin{bmatrix} B & P & Aa & BP \\ -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -1 & -1 & 0 & 1 \\ 0.33 & 0.979 & 2.91 & 0 \end{bmatrix} \quad (15)$$

Now, the linear dependencies in  $\mathbf{C}$  can be predicted from the kernel of  $\mathbf{\Omega}$ .

$$\ker \mathbf{\Omega} = \begin{bmatrix} 0.8083 & 0.1098 \\ -0.3045 & -0.7356 \\ 0.0108 & 0.2350 \\ 0.5038 & -0.6258 \end{bmatrix} \begin{matrix} B \\ P \\ Aa \\ BP \end{matrix} \quad (16)$$

As the dimension of the kernel is two,  $ns\text{-rank}(\mathbf{\Omega}) = 4 - 2 = 2$  species have to be defined as uncoloured in Strategy (1). These two species can be freely chosen among the  $ns$  species, as no row in  $\ker \mathbf{\Omega}$  is comprised by zeros only, i.e. no concentration profile is linearly independent from the others. If, for example, species *Aa* is set uncoloured, its corresponding column is removed from  $\mathbf{\Omega}$  resulting in Eq. (17).

$$\mathbf{\Omega} = \begin{bmatrix} B & P & BP \\ -2 & -2 & 2 \\ -2 & -2 & 2 \\ -2 & -2 & 2 \\ -1 & -1 & 1 \\ 0.33 & 0.979 & 0 \end{bmatrix} \quad (17)$$

The kernel of  $\mathbf{\Omega}$  as written in Eq. (17) has dimensions ( $3 \times 1$ ).

$$\ker \mathbf{\Omega} = [0.8024 \ -0.2705 \ 0.5319]^T \quad (18)$$

**Table 2**  
Optimised rate constants for different experimental conditions (a to e) and strategies (1 to 4).

# Exp.	Experimental conditions <sup>a</sup>	Data analysis <sup>b</sup>	Dosed species <sup>b</sup>	Applied strategy <sup>c</sup>	Mid-IR		UV-vis	
					$k_1^{d,e}$	Published $k_1^d$	$k_1^{d,e}$	Published $k_1^d$
1	(a)	Semi-batch	<i>Aa</i>	(1 and 3)	1.60 (±0.04)	1.74 (±0.05) <sup>f</sup>	1.65 (±0.04)	1.77 (±0.03) <sup>f</sup>
1	(a)	Batch	<i>Aa</i>	(1)	1.59 (±0.07)	1.40 <sup>g</sup>	1.65 (±0.07)	1.51 <sup>g</sup>
2	(b)	Semi-batch	<i>B<sup>h</sup></i>	(1 and 3)	1.60 (±0.03)		1.64 (±0.03)	
2	(b)	Batch	<i>B<sup>h</sup></i>	(1)	1.58 (±0.02)		1.63 (±0.02)	
3	(b)	Batch	<i>B<sup>h</sup></i>	(1)	1.61 (±0.02)		1.67 (±0.02)	
4	(b)	Batch	<i>B<sup>h</sup></i>	(1)	1.58 (±0.02)		1.62 (±0.02)	
2	(b)	Batch	<i>B<sup>h</sup></i>	(1 and 2)	1.58 (±0.02) <sup>i</sup>		1.63 (±0.02) <sup>i</sup>	
5	(c)	Semi-batch	<i>B<sup>h</sup> + Aa</i>	(3)	1.59 (±0.06)		1.64 (±0.06)	
6	(d)	Semi-batch	<i>P</i>	(1 and 3)	1.62 (±0.06)		1.67 (±0.06)	
6	(d)	Batch	<i>P</i>	(1)	1.61 (±0.09)		1.66 (±0.09)	
7	(e)	Batch	<i>B<sup>j</sup></i>	(1)	1.55 (±0.02)		1.57 (±0.02)	
2 and 7	(b and e)	Batch	<i>B<sup>h,j</sup></i>	(1 and 4)	1.57 (±0.02) <sup>k</sup>		1.62 (±0.02) <sup>k</sup>	

<sup>a</sup> Experimental conditions are described in the following sections: (a) in Section 3.4, (b) in Section 3.5, (c) in Section 3.6, (d) in Section 3.7, (e) in Section 3.8.

<sup>b</sup> All experiments were performed under semi-batch conditions (fast dosing) but analysed assuming either batch or semi-batch conditions. When semi-batch conditions are assumed, the first measured spectrum in **Y** is the one recorded just before dosing, whereas, when batch conditions are considered, it is the one recorded just after dosing. For batch conditions, concentrations are given accounting for the dilution due to dosing.

<sup>c</sup> Strategies are described in the following sections: (1) in Section 2.1 (defining uncoloured species), (2) in Section 2.2 (including known pure spectra), (3) in Section 2.3 (dosing), (4) in Section 2.4 (second order global analysis).

<sup>d</sup> In  $L^2 mol^{-2} s^{-1} \times 10^{-4}$ .

<sup>e</sup> Uncertainties are calculated according to the propagation of errors developed in Reference [21].

<sup>f</sup> Reference [21].

<sup>g</sup> Reference [22]. Note that pseudo-second order rate constants from this reference have been recalculated to third order rate constants by division with the excess catalyst concentration.

<sup>h</sup> Dosing concentration in *B* of  $1.986 mol L^{-1}$ .

<sup>i</sup> Uncertainties in the provided pure spectrum are not propagated.

<sup>j</sup> Dosing concentration in *B* of  $1.192 mol L^{-1}$  obtained by dosing a mixture of 3 mL of the benzophenone stock solution and 2 mL THF.

<sup>k</sup> Uncertainty based on the residuals and on the errors in the initial concentrations of Experiment #7.

The kernel in Eq. (18) indicates that any of the remaining species (*B*, *P* or *BP*) can be defined as uncoloured. If, for example, species *P* is set uncoloured, its corresponding column is removed from  $\mathbf{\Omega}$ .

$$\mathbf{\Omega} = \begin{bmatrix} & B & BP \\ -2 & 2 & \\ -2 & 2 & \\ -2 & 2 & \\ -1 & 1 & \\ \hline 0.33 & 0 & \end{bmatrix} \quad (19)$$

The kernel of  $\mathbf{\Omega}$  as defined in Eq. (19) now only contains the trivial solution, i.e.  $\ker \mathbf{C} = \mathbf{0}$ . Thus, Experiments #2–#4 were analysed defining species *P* and *Aa* as uncoloured. Individual optimised rate constants are gathered in Table 2 and result in mean rate constants of  $1.59 \cdot 10^{-4} L^2 mol^{-2} s^{-1}$  (mid-IR) and  $1.64 \cdot 10^{-4} L^2 mol^{-2} s^{-1}$  (UV-vis). For comparison, very similar rate constants were obtained for Experiment #2 when analysed assuming semi-batch conditions (defining only species *Aa* as uncoloured since *B* is dosed, i.e. combining Strategies 1 and 3), as shown in Table 2.

Spectral validation of the kinetic model for Experiment #2 (analysed assuming batch conditions and using Strategy 1) is done by calculation of  $\mathbf{\Omega}$  employing the corresponding optimised rate constants obtained after fitting. As an example,  $\mathbf{\Omega}$  is calculated for mid-IR spectroscopy using  $k_1 = 1.58 \cdot 10^{-4} L^2 mol^{-2} s^{-1}$  (see Table 2).

$$\mathbf{\Omega} = \begin{bmatrix} (\mu\mathbf{1})^T \mathbf{E}^T k_1 \mathbf{N} \\ \mathbf{c}_0 \end{bmatrix} = \begin{bmatrix} B & P & Aa & BP \\ -3.16 \cdot 10^{-4} & -3.16 \cdot 10^{-4} & 0 & 3.16 \cdot 10^{-4} \\ -3.16 \cdot 10^{-4} & -3.16 \cdot 10^{-4} & 0 & 3.16 \cdot 10^{-4} \\ -3.16 \cdot 10^{-4} & -3.16 \cdot 10^{-4} & 0 & 3.16 \cdot 10^{-4} \\ -1.58 \cdot 10^{-4} & -1.58 \cdot 10^{-4} & 0 & 1.58 \cdot 10^{-4} \\ \hline 0.33 & 0.979 & 2.91 & 0 \end{bmatrix} \quad (20)$$

Keeping only rows four and five in  $\mathbf{\Omega}$  results in matrix  $\mathbf{\Omega}_{ind}$  ( $2 \times 4$ ), as described in Eq. (11) with  $k_1 = 1.58 \cdot 10^{-4} L^2 mol^{-2} s^{-1}$ ,  $c_{0,B} = 0.33 mol L^{-1}$ ,  $c_{0,P} = 0.979 mol L^{-1}$  and  $c_{0,Aa} = 2.91 mol L^{-1}$ . Knowing that

species *Aa* and *P* were set uncoloured, matrix  $\mathbf{\Omega}_{ind}$  can then be used in Eq. (3) to calculate matrix  $\mathbf{\Delta}$  ( $2 \times 4$ ), i.e. the coefficients for the linear combinations of the (resolved) pure component spectra leading to the (unresolved) fitted mid-IR component spectra of the coloured species ('*B*' and '*BP*').

$$\mathbf{\Delta} = \begin{bmatrix} B & P & Aa & BP \\ 1 & 2.967 & 8.818 & 0 \\ 0 & 1.967 & 8.818 & 1 \end{bmatrix} \begin{matrix} 'B' \\ 'BP' \end{matrix} \quad (21)$$

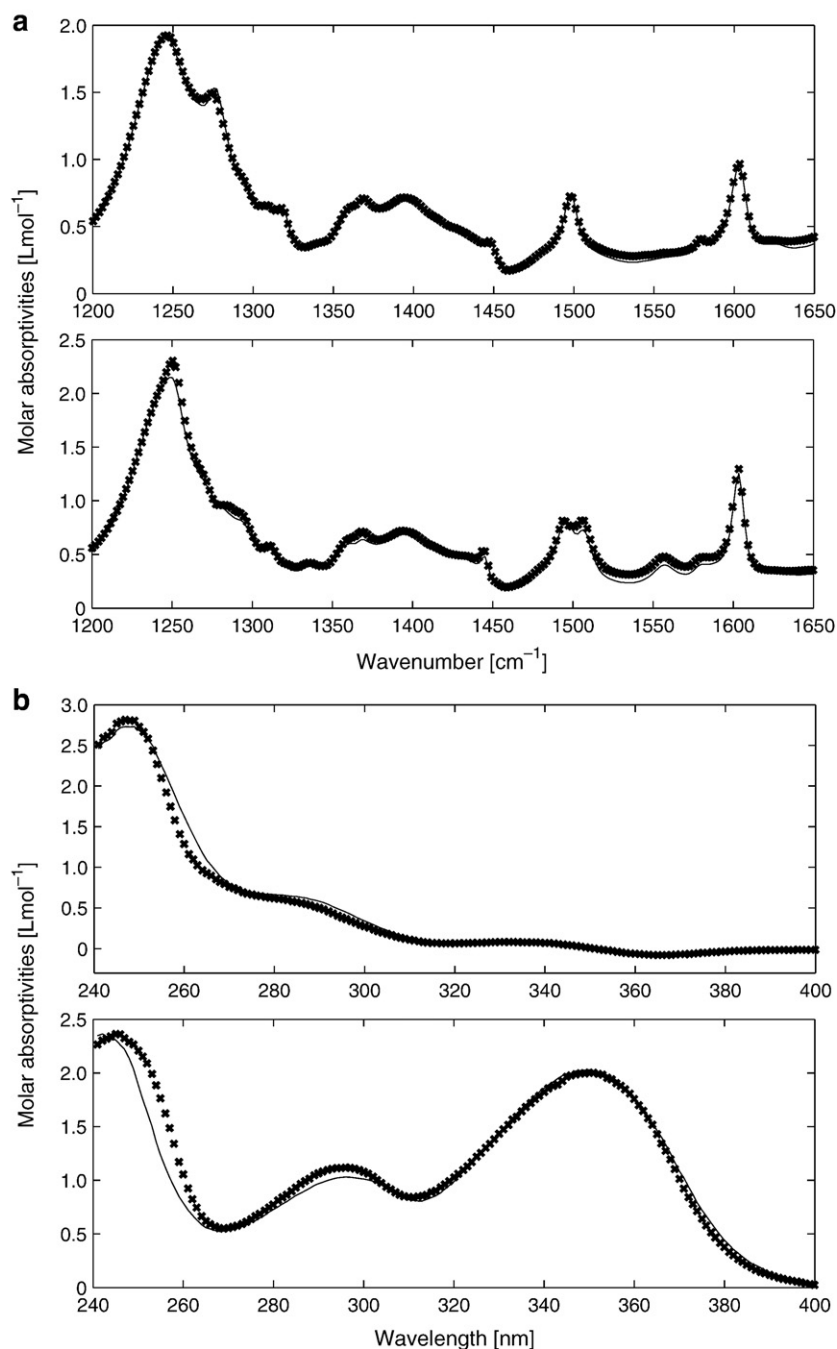
With the rows '*B*' and '*BP*' indicating the coloured species for which (unresolved) fitted component spectra are obtained, and the columns *B*, *P*, *Aa* and *BP* denoting the absorbing species for which true pure spectra exist. The matlab lines required to set up Eqs. (15)–(21) are given in Appendix A.2.

Eq. (21) indicates that the fitted component spectrum of the coloured species '*B*' is a linear combination of  $1 \times$  pure spectrum of *B*,  $2.967 \times$  pure spectrum of *P* and  $8.818 \times$  pure spectrum of *Aa*; the fitted component spectrum of the coloured species '*BP*' is a linear combination of  $1.967 \times$  pure spectrum of *P*,  $8.818 \times$  pure spectrum of *Aa* and  $1 \times$  pure spectrum of *BP*. As the rate constants determined for UV-vis and mid-IR as so similar, i.e.  $1.63 \cdot 10^{-4}$  compared to  $1.58 \cdot 10^{-4} L^2 mol^{-2} s^{-1}$  (Table 2), matrices  $\mathbf{\Delta}$  ( $2 \times 4$ ) for mid-IR and UV-vis are the same, within the precision of Eq. (21). Note that *Aa* was allowed to absorb in the UV-vis range to account for baseline and solvent effects, as described earlier.

Fig. 4a and b shows the good agreement between the fitted component spectra of the coloured species and the predicted linear combinations of the true (resolved) pure spectra for mid-IR and UV-vis.

#### 4.4. Application of Strategy (2)—including known spectra into the analysis

This strategy was also applied to Experiment #2 performed under experimental conditions (b) and analysed assuming batch conditions. If rate constants are not a priori known,  $\mathbf{\Omega}$  ( $5 \times 4$ ) and  $\ker \mathbf{\Omega}$  ( $4 \times 2$ ) can be calculated symbolically (Eqs. (10) and (12)) or calculated numerically by defining  $\mathbf{k} = \mathbf{1}$  (Eqs. (15) and (16)). As discussed in Section 4.3, two columns need to be removed from matrix  $\mathbf{\Omega}$  (or  $\mathbf{C}$ ) to



**Fig. 4.** Fitted (—) and predicted (x, Eq. (21)) mid-IR (a) and UV-vis (b) component spectra of coloured species 'B' (top) and 'BP' (bottom) from the analysis of Experiment #2 under batch conditions and using Strategy (1), defining species *P* and *Aa* as uncoloured.

enable the calculation of the pseudo-inverse by Eq. (1). In Strategy (2), these columns could be eliminated by including to the analysis two pure spectra of any of the *ns* species. If however just the pure spectrum of *B* is provided ( $nks = 1$ ), the kernel of  $\Omega$  is not empty, as numerically shown in Eq. (22) for  $\mathbf{k} = \mathbf{1}$ .

$$\ker \Omega = \begin{bmatrix} P & Aa & BP \\ 0.6879 & -0.2314 & 0.6879 \end{bmatrix}^T \quad (22)$$

As  $\ker \Omega \neq \mathbf{0}$ , instead of including a second known pure spectrum, a second species, e.g. *Aa*, can be defined as uncoloured using Strategy (1) in order to reach  $\ker \Omega = \mathbf{0}$ . So Strategies (1) and (2) are

combined. Although full resolution of the calculated spectra cannot be obtained, this puts less constraints and thus less error (due to the known spectra) on the analysis. Thus, Experiment #2 was analysed by providing the pure spectrum of *B* and defining species *Aa* as uncoloured. Fitted rate constants obtained using this Strategy (see Table 2) are close to values previously determined in Sections 4.2 and 4.3.

Spectral validation of the kinetic model for Experiment #2 (analysed under batch conditions and using Strategies 1 and 2) is performed by calculating  $\Omega$  ( $5 \times 3$ ) using the corresponding rate constant obtained after fitting. Note that the column dimension of  $\Omega$  is  $ns - nks = 4 - 1 = 3$ , as the pure component spectrum of species *B* was included, and thus this species must now be excluded from Beer's law



during the fitting. Keeping only rows four and five of  $\Omega$  results in matrix  $\Omega_{\text{ind}}$  ( $2 \times 3$ ), which is shown in Eq. (23) for UV-vis spectroscopy ( $k_1 = 1.63 \cdot 10^{-4} \text{ L}^2 \text{ mol}^{-2} \text{ s}^{-1}$  from Table 2).

$$\Omega_{\text{ind}} = \begin{bmatrix} P & Aa & BP \\ -1.63 \cdot 10^{-4} & 0 & 1.63 \cdot 10^{-4} \\ 0.979 & 2.91 & 0 \end{bmatrix} \quad (23)$$

Matrix  $\Omega_{\text{ind}}$  is then used to calculate matrix  $\Delta$  ( $2 \times 3$ ) and thus the implied linear dependencies of the fitted UV-vis component spectra corresponding to the coloured species 'P' and 'BP' (Aa set uncoloured).

The matlab lines needed to set up Eqs. (22)–(24) are given in Appendix A.3.

$$\Delta = \begin{bmatrix} P & Aa & BP \\ 1 & 2.972 & 0 \\ 0 & 2.972 & 1 \end{bmatrix} \begin{matrix} 'P' \\ 'BP' \end{matrix} \quad (24)$$

Due to the similarity between the fitted rate constants in UV-vis and mid-IR, matrices  $\Delta$  for mid-IR and UV-vis are again the same, within the precision of Eq. (24). Fig. 5a shows the good agreement between the fitted component spectra and the predicted linear combinations of the true (resolved) pure spectra for mid-IR. For UV-vis, there are some deviations between the fitted and the predicted

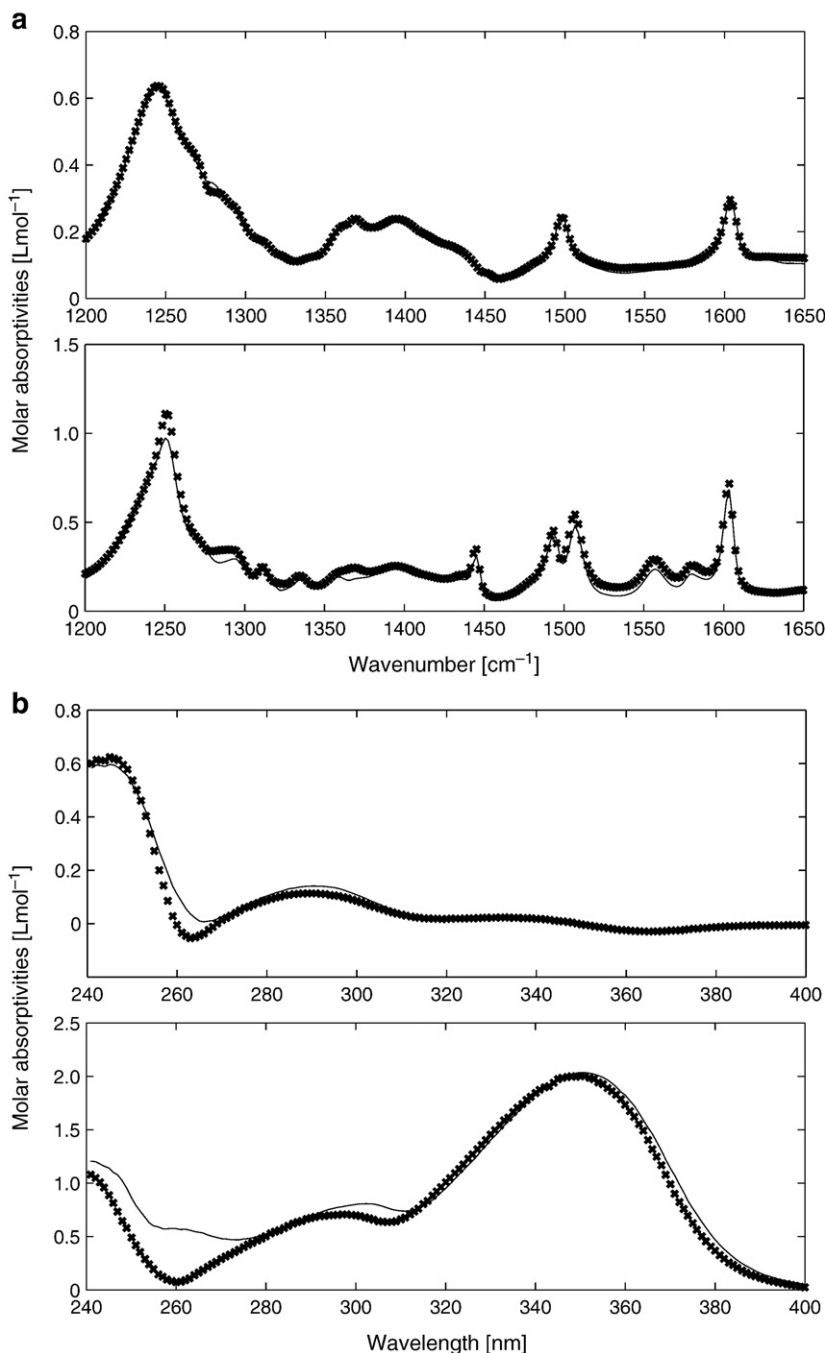


Fig. 5. Fitted (—) and predicted (x, Eq. (24)) mid-IR (a) and UV-vis (b) component spectra of coloured species 'P' (top) and 'BP' (bottom) from the analysis of Experiment #2 assuming batch conditions and combining Strategy (2), including the pure component spectrum of species B, with Strategy (1), setting species Aa uncoloured.

component spectra (see Fig. 5b), particularly in the region of maximum absorbance of species *B*. This indicates that subtracting the contribution of the known spectrum of species *B* from measurement data *Y* is slightly hampered by an imperfection of this known spectrum and does not fully remove the contribution of *B* in *Y*. This was to be expected as the fitted component spectrum of the coloured species '*B*' obtained by Strategy (1) in UV-vis already showed minor discrepancies when compared to the expected linear combinations of the pure spectra of *B*, *P* and *Aa* (see Section 4.3 and Fig. 4b). Naturally, these discrepancies can only increase when an additional constraint is used, such as the inclusion of a known spectrum.

#### 4.5. Application of Strategy (3)—dosing one or more species

Experimental conditions (c) were used to present the effects of applying Strategy (3). These experimental conditions involve a first quick dosing of species *B* in *P*, followed by a second dosing, in which species *Aa* is quickly added to the resulting mixture of *B* and *P*. Experiment #5, performed under these conditions, was analysed in semi-batch conditions. With such an experimental design, all species can be included in Beer's law and all fitted pure component spectra are resolved. The dosing conditions are given in Eq. (25).

$$\mathbf{C}_{\text{in}} = \begin{bmatrix} B & P & Aa & BP \\ 1.986 & 0 & 0 & 0 \\ 0 & 0 & 17.483 & 0 \end{bmatrix} \quad (25)$$

Using  $\mu=2$ , initial conditions  $\mathbf{c}_0 = [0 \ 1.468 \ 0 \ 0]$  and defining  $k_1$  symbolically,  $\mathbf{\Omega}$  ( $7 \times 4$ ) and  $\ker \mathbf{\Omega}$  can be calculated according to Eq. (26). Mass balance equations can alternatively be used to calculate the kernel, as outlined in Appendix A.7. The matlab lines required for the set up of Eqs. (25) and (26) are given in Appendix A.4.

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1})^T \mathbf{E}^T k_1 \mathbf{N} \\ \mathbf{c}_0 \\ \mathbf{C}_{\text{in}} \end{bmatrix} = \begin{bmatrix} B & P & Aa & BP \\ -2k_1 & -2k_1 & 0 & 2k_1 \\ -2k_1 & -2k_1 & 0 & 2k_1 \\ -2k_1 & -2k_1 & 0 & 2k_1 \\ -1k_1 & -1k_1 & 0 & 1k_1 \\ 0 & 1.468 & 0 & 0 \\ 1.986 & 0 & 0 & 0 \\ 0 & 0 & 17.483 & 0 \end{bmatrix}, \quad \ker \mathbf{\Omega} = \mathbf{0} \quad (26)$$

Matrix  $\mathbf{\Omega}_{\text{ind}}$  ( $4 \times 4$ ) is comprised by independent rows, such as four, five, six and seven of Eq. (26) and results in  $\ker \mathbf{\Omega}_{\text{ind}} = \mathbf{0}$ . Calculation of  $\Delta$  ( $4 \times 4$ ) leads to an identity matrix, i.e. all fitted pure component spectra are resolved.

As shown in Table 2, fitted rate constants obtained for Experiment #5 (using Strategy 3 and experimental conditions c) are very similar to values previously determined using other strategies. Fig. 6a (mid-IR) compares the fitted (resolved) pure component spectra with those independently measured in Section 4.2. For UV-vis, the fitted pure spectra of the dosed species (*P* and *Aa*) are in very good agreement with their independently measured pure spectra (see Fig. 6b). However, for species *B* and *BP* some deviations could be observed between the fitted and the independently measured pure spectra, suggesting that solvent effects and baseline shifts could not be fully corrected.

As also shown in Table 2, very similar rate constants are obtained for Experiment #6 performed under experimental conditions (d), i.e. when species *P* is dosed into a mixture of *B* and *Aa*, and when this experiment is analysed assuming batch (with *B* and *Aa* uncoloured) and semi-batch (with *Aa* uncoloured) conditions.

#### 4.6. Application of Strategy (4)—performing a second order global analysis

For this Strategy, Experiment #2 (Section 4.3), performed under experimental conditions (b), was combined with Experiment #7 performed under experimental conditions (e) and just differing from Experiment #2 by its initial concentration in species *B*. Both experiments were analysed individually and also simultaneously by second order global analysis (Strategy 4) assuming batch conditions. For Strategy (4),  $\mathbf{\Omega}$  ( $6 \times 4$ ) is calculated using  $\mu=2$ ,  $\mathbf{k}=\mathbf{1}$ , initial conditions  $\mathbf{c}_0 = [0.33 \ 0.979 \ 2.91 \ 0]$  and initial conditions  $\mathbf{C}_0^{\text{e}} = [0.20 \ 0.98 \ 2.91 \ 0]$  for the  $ne=1$  additional experiment (Experiment #7).

$$\mathbf{\Omega} = \begin{bmatrix} (\mu \mathbf{1})^T \mathbf{E}^T \mathbf{1N} \\ \mathbf{c}_0 \\ \mathbf{C}_0^{\text{e}} \end{bmatrix} = \begin{bmatrix} B & P & Aa & BP \\ -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -2 & -2 & 0 & 2 \\ -1 & -1 & 0 & 1 \\ 0.33 & 0.979 & 2.91 & 0 \\ 0.20 & 0.98 & 2.91 & 0 \end{bmatrix} \quad (27)$$

Linear dependencies in  $\mathbf{C}$  are predicted by the kernel of  $\mathbf{\Omega}$ , which has dimensions ( $1 \times 4$ ).

$$\ker \mathbf{\Omega} = [0 \ 0.6880 \ -0.2311 \ 0.6880]^T \quad (28)$$

The coefficients in the kernel indicate that the linear dependencies in matrix  $\mathbf{C}$ , i.e. the mass balance (see Eq. (29)), involve the concentrations of species *P*, *Aa* and *BP*. Species *B*, whose initial concentration was varied between the two experiments, is linearly independent from all other species concentrations.

$$\mathbf{C}(\ker \mathbf{C}) = (0)c_{t,B} + (0.6880)c_{t,P} + (-0.2311)c_{t,Aa} + (0.6880)c_{t,BP} = \mathbf{0} \quad (29)$$

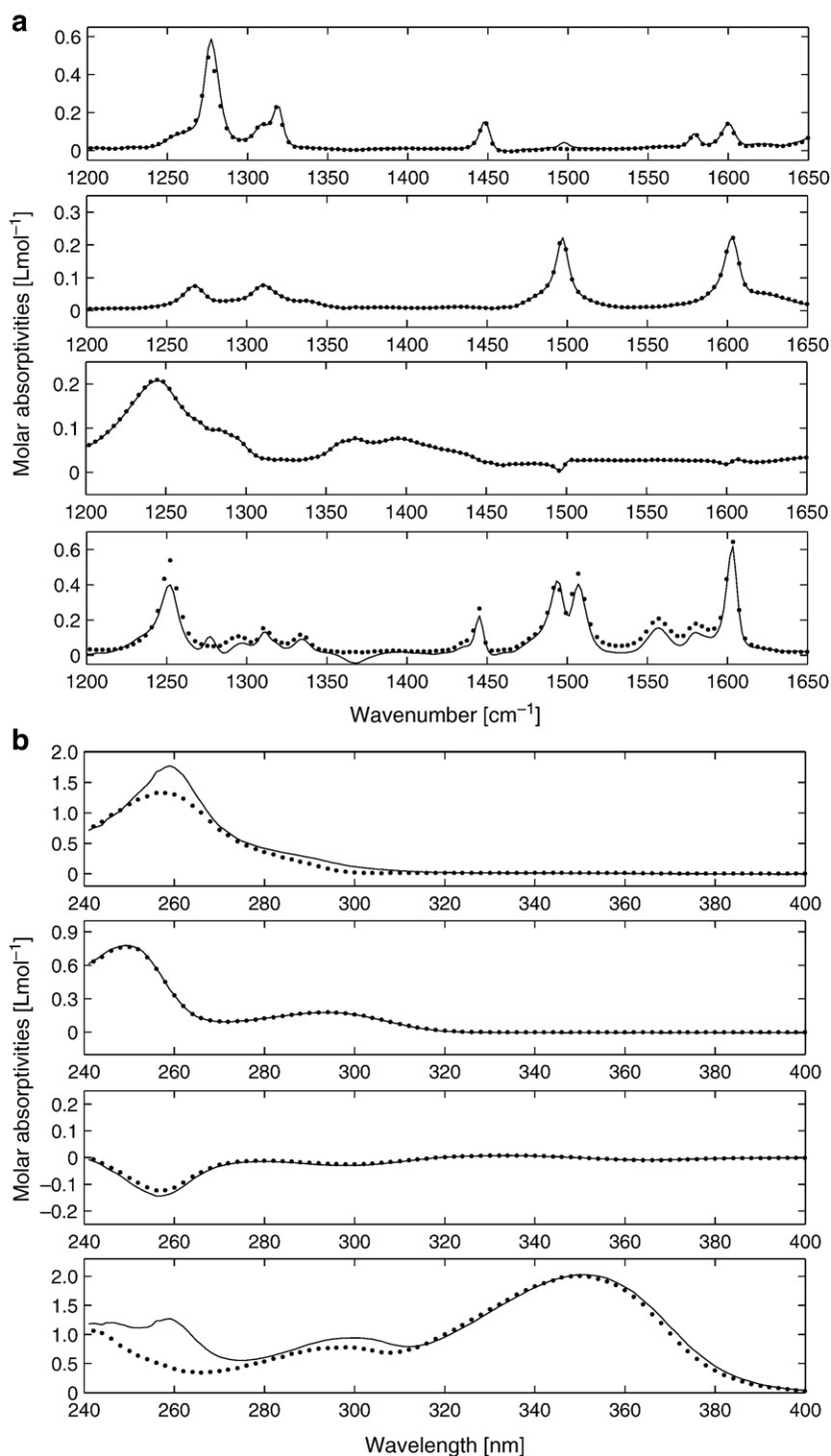
As the rank deficiency is not yet broken, another Strategy, e.g. Strategy (1), has to be applied. According to Eqs. (28) (kernel) and (29) (mass balance), any species among *P*, *Aa* or *BP* can be defined as uncoloured. If species *Aa* is treated as uncoloured using Strategy (1), its corresponding column is removed from  $\mathbf{\Omega}$  ( $6 \times 4$ ) of Eq. (27), leading to  $\mathbf{\Omega}$  ( $6 \times 3$ ) and  $\ker \mathbf{\Omega} = \mathbf{0}$ .

Optimised rate constants obtained for the fitting of Experiment #7 are slightly lower than those received for the fitting of Experiment #2 (see Table 2). This explains why the optimised rate constants obtained by Strategy (4) are also lower than the rate constants previously determined with other strategies. Using the rate constant fitted in mid-IR ( $k_1 = 1.57 \cdot 10^{-4} \text{ L}^2 \text{ mol}^{-2} \text{ s}^{-1}$ ),  $\mathbf{\Omega}$  is recalculated and reduced to  $\mathbf{\Omega}_{\text{ind}}$  ( $3 \times 4$ ), by keeping only rows four, five and six of  $\mathbf{\Omega}$ , which form a vector basis such that  $\ker \mathbf{\Omega}_{\text{ind}}^T = \mathbf{0}$ .

$$\mathbf{\Omega}_{\text{ind}} = \begin{bmatrix} B & P & Aa & BP \\ -1.57 \cdot 10^{-4} & -1.57 \cdot 10^{-4} & 0 & 1.57 \cdot 10^{-4} \\ 0.33 & 0.979 & 2.91 & 0 \\ 0.20 & 0.98 & 2.91 & 0 \end{bmatrix} \quad (30)$$

The coefficients for the linear combinations of the (resolved) pure component spectra leading to the (unresolved) mid-IR fitted component spectra,  $\Delta$  ( $3 \times 4$ ), are then computed with the constraint that species *Aa* was set uncoloured. The matlab lines needed for the set up of Eqs. (27), (28), (30), and (31) are given in Appendix A.5.

$$\Delta = \begin{bmatrix} B & P & Aa & BP \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 2.977 & 0 \\ 0 & 0 & 2.977 & 1 \end{bmatrix} \begin{matrix} 'B' \\ 'P' \\ 'BP' \end{matrix} \quad (31)$$



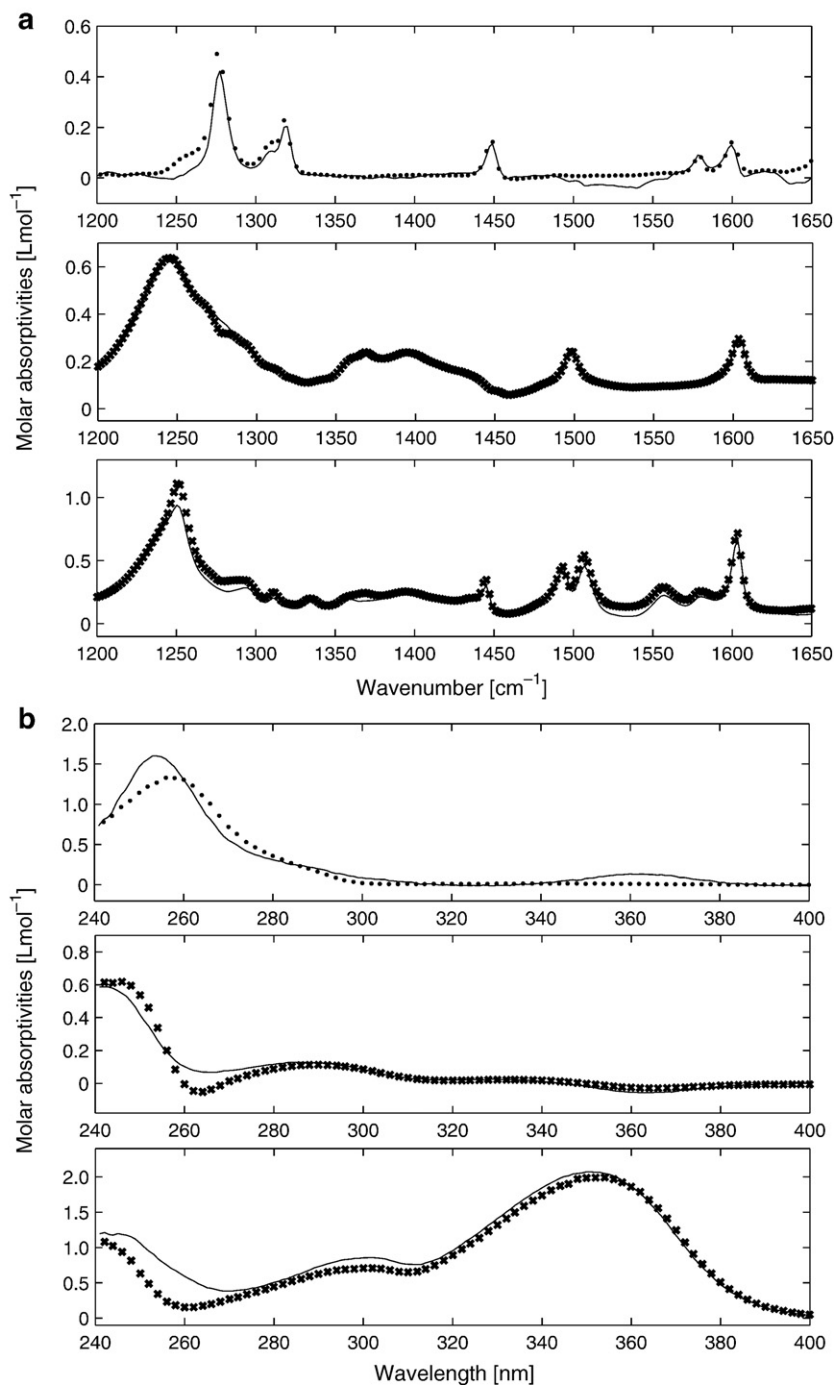
**Fig. 6.** Fitted (—) and independently measured (•) mid-IR (a) and UV-vis (b) pure spectra of *B* (1st row), *P* (2nd row), *Aa* (3rd row) and *BP* (4th row) from the analysis of Experiment #5 assuming semi-batch conditions (Strategy 3) and dosing successively species *B* and *Aa* in *P*. Note that the independently measured spectra of *Aa* have been corrected for solvent effects (see Section 4.2 and Fig. 3).

Eq. (31) indicates that, under these conditions and the combination of Strategies (1) and (4), the fitted component spectrum of *B* is resolved and the fitted component spectra of coloured species '*P*' and '*BP*' are mixed with the pure spectrum of acetic acid. Fig. 7a (mid-IR) and Fig. 7b (UV-vis) compare the fitted component spectra with those independently measured in Section 4.2. Component spectra fitted in mid-IR are in good agreement with the predicted linear combinations of the independently measured pure spectra. For UV-vis, some larger

deviations can be observed between fitted and predicted component spectra.

## 5. Conclusion

A very intrinsic and important outcome of the direct fitting of spectroscopic data employing a hard kinetic model is the computation of pure component spectra in addition to the kinetic parameters. It is



**Fig. 7.** Fitted (—), independently measured (\*, same as in Fig. 2) and predicted (x, Eq. (31)) mid-IR (a) and UV-vis (b) pure spectra of species *B* (top) and coloured species '*P*' (middle) and '*BP*' (bottom) from the second order analysis of the concatenated Experiments #2 and #7 (assuming batch conditions) measured with different initial concentrations in *B* (Strategy 4) and setting species *Aa* uncoloured (Strategy 1).

the reproducibility of the fitted kinetic parameters, as well as the correctness of the fitted component spectra, under different experimental conditions that allows the validation of a postulated kinetic mechanism.

However, for kinetic models leading to rank-deficient concentration profiles, the calculation of the component spectra and thus their validation can become very complex. One elegant way to deal with rank deficiency is to define uncoloured species (Strategy 1) as then no or only partial a priori information (known spectra, Strategy 2) or additional experimental design (semi-batch, Strategy 3; variation of initial concentrations, Strategy 4) is required that can potentially

perturb the fitting, for example, due to baseline effects and propagated errors. Strategy (1) necessarily leads to fitted component spectra that are complex linear combinations of the true ones.

In the present case study and based on previous theoretical considerations [17], we applied different strategies to allow the kinetic analysis of the rank-deficient acid catalysed reaction of benzophenone with phenylhydrazine in THF. For this condensation reaction, fitted rate constants were highly reproducible, when experimental conditions were varied, e.g. dosing any of the reactants. We have also shown that the coefficients for the linear combinations of the true pure component spectra (matrix  $\Delta$ ) can be correctly

predicted, even when Strategy (1) is combined with any other strategies. This indicates that the kinetics of this reaction is well-described by the given mechanism.

Knowing the coefficients for the linear combinations of the true pure component spectra allows a rationalisation of the fitted ones as they can now be assessed on the basis of independently measured pure component spectra (as long as all species can be isolated, as it was possible in this case study). In many other cases, however, isolation can be a tedious task or sometimes even impossible, particularly for intermediate species. Then, the assessment needs to be done by a peak to peak assignment. For this, it is even more important to know the individual possible contributions of each species to the fitted component spectra.

## Notation

As a convention for the notation, matrices are written in boldface capitals (e.g.  $\mathbf{R}$ ), vectors in boldface lowercase ( $\mathbf{r}$ ) and scalars in italics ( $r$ ). For indices, lowercase characters are used. Elements of a matrix  $\mathbf{R}$  are denoted as  $r_{ij}$  and elements of a vector  $\mathbf{r}$  as  $r_i$ . For a detailed list of all used symbols, we refer to our previous publication [17].

## Acknowledgements

We thank Andreas Dutly from ETH Zurich for the GC–MS measurements and the analysis of chromatograms and MS spectra, and Dr Heinz Rügger from ETH Zurich for the interpretation of NMR spectra.

## Appendix A

### A.1. Matlab code for Section 4.1

```
>> syms c0B c0P c0Aa mu k1;
>> E      = [ 1 1 1 0 ]; (Equation 7)
>> N      = [ -1 -1 0 1 ]; (Equation 8)
>> one    = ones(size(E'));
>> k      = k1;
>> c0     = [ c0A, c0P, c0Aa, 0 ]; (Equation 9)
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0 ]; (Equation 10)
>> omega_ind = omega([ 4 5], :); (Equation 11)
>> null(omega); (Equation 12)
```

### A.2. Matlab code for Section 4.2

```
>> E      = [ 1 1 1 0 ]; N=[ -1 -1 0 1 ]; one = ones(size(E'));
>> k      = 1;
>> mu     = 2;
>> c0     = [ 0.33, 0.979, 2.91, 0 ];
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0 ]; (Equation 15)
>> null(omega); (Equation 16)
>> omega(:, [ 1 2 4 ]); (Equation 17)
>> null(omega(:, [ 1 2 4 ])); (Equation 18)
>> omega(:, [ 1 4 ]); (Equation 19)
>> k = 1.58e-4;
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0 ]; (Equation 20)
>> omega_ind = omega([ 4 5], :);
>> delta  = inv(omega_ind(:, [ 1 4 ]))*omega_ind; (Equation 21)
```

### A.3. Matlab code for Section 4.4

```
>> E      = [ 1 1 1 0 ]; N=[ -1 -1 0 1 ]; one = ones(size(E'));
>> k      = 1;
>> mu     = 2;
>> c0     = [ 0.33, 0.979, 2.91, 0 ];
```

```
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0 ];
>> null(omega(:, [ 2 3 4 ])) (Equation 22)
>> k      = 1.63e-4;
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0 ];
>> omega_ind = omega([ 4 5], [ 2 3 4 ]); (Equation 23)
>> delta  = inv(omega_ind(:, [ 1 3 ]))*omega_ind; (Equation 24)
```

### A.4. Matlab code for Section 4.5

```
>> syms k1;
>> E      = [ 1 1 1 0 ]; N=[ -1 -1 0 1 ]; one = ones(size(E'));
>> k      = k1;
>> mu     = 2;
>> c0     = [ 0 1.468 0 0 ];
>> Cin    = [ 1.986 0 0 0; 0 0 17.483 0 ]; (Equation 25)
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0; Cin ]; (Equation 26)
```

### A.5. Matlab code for Section 4.6

```
>> E      = [ 1 1 1 0 ]; N=[ -1 -1 0 1 ]; one = ones(size(E'));
>> k      = 1;
>> mu     = 2;
>> c0     = [ 0.331, 0.979, 2.914, 0 ];
>> C0ne   = [ 0.199, 0.979, 2.914, 0 ];
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0; C0ne ]; (Equation 27)
>> null(omega); (Equation 28)
>> k      = 1.57e-4;
>> omega  = [ (mu*one) .^ (E')*diag(k)*N; c0; C0ne ];
>> omega_ind = omega([ 4 5 6], :); (Equation 30)
>> delta  = inv(omega_ind(:, [ 1 2 4 ]))*omega_ind; (Equation 31)
```

### A.6. Calculation of $\ker \Omega$ under batch conditions using mass balance equations

Under batch conditions, the mass balance can be written in concentration terms according to the following equation:

$$\mathbf{C}_0 = \mathbf{C} - \mathbf{XN} \quad (32)$$

The matrix of initial concentrations,  $\mathbf{C}_0$  ( $nt \times ns$ ), is calculated as  $\mathbf{1c}_0$ , with  $\mathbf{c}_0$  ( $1 \times ns$ ) being the vector of initial concentrations and  $\mathbf{1}$  a column vector of ones of dimensions ( $nt \times 1$ ). Matrix  $\mathbf{X}$  ( $nt \times nr$ ) denotes the extent of reaction and matrix  $\mathbf{N}$  is defined in Eq. (8). For details regarding Eq. (32), we refer to [17].

For the chemical system of Eq. (6), the application of Eq. (32) leads to the following system:

$$C_{0,B} = C_{t,B} + C_{t,BP} \quad (33)$$

$$C_{0,P} = C_{t,P} + C_{t,BP} \quad (34)$$

$$C_{0,Aa} = C_{t,Aa} \quad (35)$$

Importantly, concentration profiles  $C_{t,B}$ ,  $C_{t,P}$ ,  $C_{t,Aa}$  and  $C_{t,BP}$  are only linearly dependent if it is possible to write the relationship between these concentrations as homogeneous algebraic equations with time invariant coefficients. In this context, 'homogeneous' indicates that the equations equal the null vector  $\mathbf{0}$ . In such case, the number of these equations defines the dimension of  $\ker \Omega$  and the time invariant coefficients are the elements of the kernel, such that  $\mathbf{C}(\ker \mathbf{C}) = \mathbf{0}$ .

For the studied chemical system, a first homogeneous equation with time invariant coefficients (see Eq. (36)) can be calculated according to

the ratio between Eqs. (34) and (33), denoted (Eq. (34)/Eq. (33)). A second equation (see Eq. (37)) can be obtained by taking the ratio (Eq. (35)/Eq. (33)). Time invariant coefficients of the following equations define the elements of  $\ker \mathbf{\Omega}$ , as previously shown in Eq. (12).

$$(\text{Eq. 34} / \text{Eq. 33}) : (-\alpha)c_{t,B} + (1)c_{t,P} + (0)c_{t,Aa} + (1-\alpha)c_{t,BP} = \mathbf{0}$$

$$\text{with } \alpha = \frac{c_{0,P}}{c_{0,B}} \quad (36)$$

$$(\text{Eq. 35} / \text{Eq. 33}) : (-\beta)c_{t,B} + (0)c_{t,P} + (1)c_{t,Aa} + (-\beta)c_{t,BP} = \mathbf{0}$$

$$\text{with } \beta = \frac{c_{0,Aa}}{c_{0,B}} \quad (37)$$

These equations are identical to Eqs. (13) and (14). As the system of Eqs. (33)–(35) is comprised by three equations, two linearly independent ratios can be defined, i.e. the dimension of  $\ker \mathbf{\Omega}$  is two. For parallelism with Appendix A.7, the equation obtained from the ratio (Eq.(34)/Eq. (35)) is also calculated.

$$(\text{Eq. 34} / \text{Eq. 35}) : (0)c_{t,B} + (1)c_{t,P} + (-\gamma)c_{t,Aa} + (1)c_{t,BP} = \mathbf{0}$$

$$\text{with } \gamma = \frac{c_{0,P}}{c_{0,Aa}} \quad (38)$$

#### A.7. Calculation of $\ker \mathbf{\Omega}$ under semi-batch conditions using mass balance equations

Under semi-batch conditions, the mass balance described in Eq. (32) requires minor adaptations to take into account the dosed concentrations and the dilution effect. This is shown in the following equation:

$$\mathbf{D}\mathbf{C}_0 + (\mathbf{I}-\mathbf{D})\mathbf{C}_{\text{dos}} = \mathbf{C}-\mathbf{X}\mathbf{N} \quad (39)$$

Matrix  $\mathbf{D}$  ( $nt \times nt$ ) is the diagonal matrix of dilution calculated as  $v_0 \text{DIAG}(\mathbf{v})^{-1}$ , with  $\mathbf{v}$  ( $nt \times 1$ ) being the vector of volumes and  $v_0$  the initial volume such that  $v_0 = v_1$ .  $\mathbf{C}_{\text{dos}}$  ( $nt \times ns$ ) denotes the matrix of dosing concentrations obtained using  $\mathbf{C}_{\text{in}}$  ( $nf \times ns$ ), and  $\mathbf{I}$  ( $nt \times nt$ ) is the identity matrix. For details regarding Eq. (39), we refer to [17].

When species  $B$  is dosed alone, Eq. (39) leads to the following system of equations:

$$d_{t,t}c_{0,B} + (1-d_{t,t})c_{in,B} = c_{t,B} + c_{t,BP} \quad (40)$$

$$d_{t,t}c_{0,P} = c_{t,P} + c_{t,BP} \quad (41)$$

$$d_{t,t}c_{0,Aa} = c_{t,Aa} \quad (42)$$

Here, the ratio (Eq. (41)/Eq. (40)) also leads to a homogeneous equation however with time variant coefficients.

$$(\text{Eq. 41} / \text{Eq. 40}) : (-\alpha')c_{t,B} + (1)c_{t,P} + (0)c_{t,Aa} + (1-\alpha')c_{t,BP} = \mathbf{0}$$

$$\text{with } \alpha' = \frac{d_{t,t}c_{0,P}}{d_{t,t}c_{0,B} + (1-d_{t,t})c_{in,B}} \quad (43)$$

Coefficients obtained for the ratio (Eq. (41)/Eq. (42)) are still time invariant and the same as previously defined in Eq. (38) for batch conditions.

$$(\text{Eq. 41} / \text{Eq. 42}) : (0)c_{t,B} + (1)c_{t,P} + (-\gamma)c_{t,Aa} + (1)c_{t,BP} = \mathbf{0}$$

$$\text{with } \gamma = \frac{c_{0,P}}{c_{0,Aa}} \quad (44)$$

As there is only one remaining equation left that can be written as a function of time invariant coefficients,  $\ker \mathbf{\Omega}$  is reduced to dimension one, with its elements given in Eq. (44).

When species  $B$  and  $Aa$  are dosed sequentially (as described in Section 4.5), Eq. (42) has to be modified and replaced by Eq. (47) in order to take into account the dosed concentration of  $Aa$ .

$$d_{t,t}c_{0,B} + (1-d_{t,t})c_{in,B} = c_{t,B} + c_{t,BP} \quad (45)$$

$$d_{t,t}c_{0,P} = c_{t,P} + c_{t,BP} \quad (46)$$

$$d_{t,t}c_{0,Aa} + (1-d_{t,t})c_{in,Aa} = c_{t,Aa} \quad (47)$$

With this sequential dosing, it is now impossible to set up a homogeneous equation with time invariant coefficients, as any ratio of Eqs. (45)–(47) leads to time variant coefficients. This is, for example, the case for the ratio (Eq. (46)/Eq. (47)) when compared to Eq. (44).

$$(\text{Eq. 46} / \text{Eq. 47}) : (0)c_{t,B} + (1)c_{t,P} + (-\gamma')c_{t,Aa} + (1)c_{t,BP} = \mathbf{0}$$

$$\text{with } \gamma' = \frac{d_{t,t}c_{0,P}}{d_{t,t}c_{0,Aa} + (1-d_{t,t})c_{in,Aa}} \quad (48)$$

As a consequence, the dimension of  $\ker \mathbf{\Omega}$  is reduced to zero and the kernel is only comprised of the trivial solution ( $\mathbf{0}$ ), i.e. linear dependencies in  $\mathbf{C}$  are broken.

## References

- [1] P. Geladi, B.R. Kowalski, *Anal. Chim. Acta* 185 (1986) 1–17.
- [2] M. Maeder, A.D. Zuberbühler, *Anal. Chem.* 62 (1990) 2220–2224.
- [3] E. Sanchez, B.R. Kowalski, *J. Chemom.* 4 (1990) 29–45.
- [4] I.M. Galvan, J.M. Zaldivar, H. Hernandez, E. Molga, *Comput. Chem. Eng.* 20 (1996) 1451–1465.
- [5] R. Bro, *Chemom. Intell. Lab. Syst.* 38 (1997) 149–171.
- [6] E.R. Malinowski, *Factor Analysis in Chemistry*, Third Edition, John Wiley & Sons, Inc., New York, USA, 2002.
- [7] M. Maeder, Y.M. Neuhold, in: P. Gemperline (Ed.), Chapter 7 in *Practical Guide to Chemometrics*, Taylor and Francis, Boca Raton, USA, 2006, pp. 218–256.
- [8] R. Tauler, A. de Juan, in: P. Gemperline (Ed.), Chapter 11 in *Practical Guide to Chemometrics*, Taylor and Francis, Boca Raton, USA, 2006, pp. 417–474.
- [9] G.M. Escandar, N.K.M. Faber, H.C. Goicoechea, A.M. de la Pena, A.C. Olivieri, R.J. Poppi, *Trac-Trends Anal. Chem.* 26 (2007) 752–765.
- [10] M. Maeder, Y.M. Neuhold, *Practical Data Analysis in Chemistry*, Elsevier, Amsterdam, NL, 2007.
- [11] E. Furusjö, L.G. Danielsson, *Anal. Chim. Acta* 373 (1998) 83–94.
- [12] S. Bijlsma, D.J. Louwerse, A.K. Smilde, *AIChE J.* 44 (1998) 2713–2723.
- [13] E. Bezemer, S.C. Rutan, *Chemom. Intell. Lab. Syst.* 59 (2001) 19–31.
- [14] G. Puxty, M. Maeder, K. Hungerbühler, *Chemom. Intell. Lab. Syst.* 81 (2006) 149–164.
- [15] V.M. Taavitsainen, H. Haario, *J. Chemom.* 15 (2001) 215–239.
- [16] V.M. Taavitsainen, H. Haario, M. Laine, *J. Chemom.* 17 (2003) 140–150.
- [17] J. Billeter, Y.M. Neuhold, K. Hungerbühler, *Chemom. Intell. Lab. Syst.* 95 (2009) 170–187.
- [18] S. Bijlsma, A.K. Smilde, *Anal. Chim. Acta* 396 (1999) 231–240.
- [19] P. Bugnon, J.C. Chottard, J.L. Jestin, B. Jung, G. Laurenczy, M. Maeder, A.E. Merbach, A.D. Zuberbühler, *Anal. Chim. Acta* 298 (1994) 193–201.
- [20] R. Dyson, M. Maeder, Y.M. Neuhold, G. Puxty, *Anal. Chim. Acta* 490 (2003) 99–108.
- [21] J. Billeter, Y.M. Neuhold, L. Simon, G. Puxty, K. Hungerbühler, *Chemom. Intell. Lab. Syst.* 93 (2008) 120–131.
- [22] A.R. de Carvalho, M.D. Sanchez, J. Wattoom, R.G. Brereton, *Talanta* 68 (2006) 1190–1200.
- [23] R. Dyson, M. Maeder, G. Puxty, Y.M. Neuhold, *Inorg. React. Mech.* 5 (2003) 39–46.
- [24] W.H. Press, W.T. Vetterling, S.A. Teukolsky, B.P. Flannery, *Numerical Recipes in C++ – The Art of Scientific Computing*, Second Edition, Cambridge University Press, New York, USA, 2005.
- [25] K.Q. Levenberg, *Appl. Math.* 2 (1949) 164–168.
- [26] D.W. Marquardt, *J. Soc. Ind. Appl. Math.* 11 (1963) 431–441.
- [27] P.R. Bevington, D.K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, Third Edition, McGraw-Hill, New York, USA, 2003.
- [28] D.C. Lay, *Linear Algebra and Its Applications*, Third Edition, Addison Wesley Higher Education, Boston, USA, 2003.
- [29] F. Visentin, S.I. Gianoli, A. Zogg, O.M. Kut, K. Hungerbühler, *Org. Proc. Res. Dev.* 8 (2004) 725–737.
- [30] Matlab 7.5.0 (R2007b), The Mathworks, Natick, MA, USA, 2007 <http://www.mathworks.com>.