JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 213–227

Applications of the direct exponential curve resolution algorithm (DECRA) to solid state nuclear magnetic resonance and mid-infrared spectra

Willem Windig[1]*†, Brian Antalek[1], Mark J. Robbins[1], Nicholas Zumbulyadis[1] and Charles E. Heckler[2]

[1]*Imaging Research and Advanced Development, Eastman Kodak Company, Rochester, NY 14650-2132, USA*
[2]*Computer and Experimental Analysis, Eastman Kodak Company, Rochester, NY 14650-2132, USA*

SUMMARY

DECRA (direct exponential curve resolution algorithm) is a fast multivariate method used to resolve spectral data with concentration profiles that are linear combinations of exponential functions. DECRA has been previously applied to a wide variety of spectroscopies. Results are presented in this paper for two new application areas: solid state nuclear magnetic resonance spectra of polymorphic crystal mixtures and mid-infrared spectroscopy of chemical reactions. Furthermore, the paper will show the effect of the way the data set is split, which is a part of the algorithm, on the results. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: direct exponential curve resolution algorithm (DECRA); generalized rank annihilation method (GRAM); MIR spectra; solid state NMR spectra; kinetics; polymorphs; exponentials

1. INTRODUCTION

Recently, a novel technique was developed to resolve linear combinations of exponential profiles. The technique is called DECRA (direct exponential curve resolution algorithm) and is based on GRAM (generalized rank annihilation method) [1,2]. The technique is multivariate, and it is possible to resolve spectral data sets with thousands of exponential variables in a fraction of a second with a personal computer. Because exponential behavior is a common phenomenon in many processes, DECRA can be applied in a wide variety of areas: pulsed gradient spin echo nuclear magnetic resonance (PGSE NMR) diffusion spectra of complex polymer mixtures [1–4], magnetic resonance images (MRIs) of the human brain [5,6] and kinetic spectral data in the near-infrared (NIR) [7] and ultraviolet-visible (UV-vis) region [4,8]. DECRA is able to resolve spectra with extreme overlap, and up to five components have been resolved [4]. It is interesting to note that basically the same approach was developed before in the field of signal processing [9].

This paper will introduce two new application areas: solid state NMR of polymorphic materials and mid-infrared (MIR) spectroscopy of several chemical reactions. In order to apply DECRA, a data set

* Correspondence to: W. Windig, Eastman Kodak Company, Rochester, NY 14650-2132, USA.
† E-mail: windig@kodak.com

needs to be split into two parts. The way in which the data set is split affects the accuracy and precision of DECRA. A simple model system is used to help characterize this phenomenon.

2. MATERIALS AND METHODS

2.1. Computerized analysis

2.1.1. DECRA. When two mixture data sets are available where the component concentrations in one are proportional to those in the other (unique to each component), it is possible to unambiguously resolve the data sets into a matrix containing the concentrations of the pure components and a matrix containing the pure component spectra [10]. This can be expressed in equations as follows:

$$\mathbf{A} = \mathbf{C}\mathbf{P}^{\mathrm{T}} \tag{1}$$

$$\mathbf{B} = \mathbf{C}\mathbf{X}\mathbf{P}^{\mathrm{T}} \tag{2}$$

In these equations, matrices **A** and **B** represent spectral data sets that are proportional. The spectra are organized in rows. Matrix **C** represents the pure component concentrations organized in columns, matrix **P** represents the pure spectra organized in rows, and superscript T stands for the matrix transpose. With the type of problems discussed in this paper, concentrations are not obtained, but profiles proportional to concentrations. In order to obtain concentrations, a calibration step would be needed. In order to indicate the nature of the profiles discussed in this paper, the term *contribution* will be used instead of concentration.

The diagonal matrix **X** contains the unique proportionality factors relating **A** with **B**. In other words, the mixtures have the same pure components and the same concentrations, but the contributions differ by a scale factor unique to each component. This problem can be expressed as the generalized eigenvector problem and is the basis of the generalized rank annihilation method (GRAM) [11,12].

When data are of an exponential character, it is possible to split one data set into two data sets that are proportional [1–8]. The principle is given in Table I. Under **D**, two exponential decays are listed. This is representative of the contribution profiles of a data set containing two components. As a next step, **D** will be split into two parts: **A** contains the first three rows of **D**, and **B** contains the last three rows of **D**. As defined above, data sets **A** and **B** are proportional. The proportionality factors (matrix **X** in Equation (2)) are 1/3 and 1/2. A data set in which the components have contribution profiles of exponential character can be resolved unambiguously, because it can be split into two proportional data sets. This adaptation of GRAM is the basis for DECRA.

The data set in Table I was split into two parts which differed by one spectrum. It is also possible to split the data set in other ways. The term *shift* will be used to identify how the data set was split up. For example, when a data set contains 30 spectra, a shift of one means that the first data set consists of spectra 1–29 and the second data set consists of spectra 2–30. A shift of five means that the two data sets consist of spectra 1–25 and 6–30, etc.

2.1.2. Kinetics. The use of the DECRA approach for kinetic data was first described by Bijlsma *et al* for a pseudo-first-order reaction with an intermediate product [7,8]:

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C}$$

The reaction constants are represented by k_1 and k_2, which have the dimension min^{-1}. The

Table I. Model for data set **D** with exponential decays, and how it can be split into the proportional data sets **A** and **B**

D		A		B	
Component 1	Component 2	Component 1	Component 2	Component 1	Component 2
27	8	27	8	9	4
9	4	9	4	3	2
3	2	3	2	1	1
1	1				

concentration profiles for the starting material A, the intermediate product B and the final product C are

$$C_{\mathrm{A},i} = C_{\mathrm{A},0}\mathrm{e}^{-k_1 t_i} \tag{3}$$

$$C_{\mathrm{B},i} = \frac{k_1 C_{\mathrm{A},0}}{k_2 - k_1}(\mathrm{e}^{-k_1 t_i} - \mathrm{e}^{-k_2 t_i}) \tag{4}$$

$$C_{\mathrm{C},i} = C_{\mathrm{A},0} - C_{\mathrm{A},i} - C_{\mathrm{B},i} \tag{5}$$

where i indexes the ith spectrum and t_i represents the time in minutes. $C_{\mathrm{A},0}$ represents the concentration of A at time zero. In this paper a simple first-order reaction without intermediate will also be shown:

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B}$$

The concentration profile for the starting material is described by Equation (3). The concentration of the product B is

$$C_{\mathrm{B},i} = 1 - C_{\mathrm{A},i} \tag{6}$$

An important fact is that the constant terms in these kinetic equations can be described as exponentials with the decay value zero, which enables the use of DECRA. As a consequence, Equations (3)–(5) can be described as linear combinations of exponentials, where the value one has been chosen for $C_{\mathrm{A},0}$:

$$C_{\mathrm{A},i} = \mathrm{e}^{-k_1 t_i} \tag{7}$$

$$C_{\mathrm{B},i} = \frac{k_1}{k_2 - k_1}(\mathrm{e}^{-k_1 t_i} - \mathrm{e}^{-k_2 t_i}) \tag{8}$$

$$C_{\mathrm{C},i} = \mathrm{e}^{0 t_i} - C_{\mathrm{A},i} - C_{\mathrm{B},i} \tag{9}$$

The concentration arrays are in columns. Similarly, Equation (6) can be expressed as a linear combination of exponential functions. For a first-order reaction with an intermediate product, DECRA is used to resolve three components. The contribution profiles extracted are exponential functions with decay values k_1, k_2 and zero. The decay values can be directly calculated from the eigenvalues resulting from the solution of the data sets by DECRA. The eigenvalues resulting from

the solution of the generalized eigenvector problem are the matrix **X** in Equation (2). The decay values can then be calculated simply by taking the log. The decay values are substituted in Equations (3)–(5) to give the desired kinetic profiles.

The scaling of the contribution profiles is arbitrary, since $C_{A,0}$ was chosen to be one. Furthermore, different components have different intensity responses in the spectra. In order to scale the contributions in such a way that they reflect the spectral signals, a scaling is applied so that the sum of the profiles reproduces the total signal **t** of the spectra. The total signal of the spectra is calculated as follows:

$$t_j = \sum_{i=1}^{nspec} D_{i,j} \tag{10}$$

D stands for the original (i.e. unsplit) data set with *nspec* spectra.

The contribution matrix **C** is defined as follows:

$$\mathbf{C} = [\mathbf{c}_A \quad \mathbf{c}_B \quad \mathbf{c}_C] \tag{11}$$

C has a row for each spectrum and three columns. The scaled contributions $\mathbf{C}^{scaled}$ are calculated as follows. The contribution profiles in **C** need to be scaled in such a way that they reproduce **t**, as expressed by the following relation:

$$\mathbf{Cx} = \mathbf{t} \tag{12}$$

The unknown **x** can be calculated as follows:

$$\mathbf{x} = \mathbf{C}^{+}\mathbf{t} \tag{13}$$

The properly scaled contribution profiles in $\mathbf{C}^{scaled}$ can now be calculated as follows:

$$\mathbf{C}^{scaled} = \mathbf{C}\mathrm{diag}(\mathbf{x}) \tag{14}$$

where diag(**x**) stands for a diagonal matrix with the three elements of **x**.

With these scaled contributions the spectra can be calculated by standard least squares procedures from the original data matrix (i.e. the matrix before splitting):

$$\mathbf{P}^{T} = \mathbf{C}^{scaled+}\mathbf{D} \tag{15}$$

where $\mathbf{C}^{scaled+}$ stands for the pseudoinverse of $\mathbf{C}^{scaled}$.

The contributions in $\mathbf{C}^{scaled}$ are based on ideal exponential profiles. In order to see how well these profiles match the original data, the projection of $\mathbf{C}^{scaled}$ in the original data is calculated as follows:

$$\mathbf{C}^{projected} = \mathbf{D}(\mathbf{P}^{T})^{+} \tag{16}$$

The combined plots of $\mathbf{C}^{scaled}$ and $\mathbf{C}^{projected}$ provide a good visual diagnostic tool. From the resolved data the original data set can be reconstructed:

$$\mathbf{D}^{reconstructed} = \mathbf{C}^{projected}\mathbf{P}^{+}. \tag{17}$$

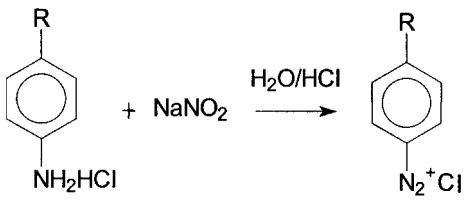
The relative root of sum squares of difference (RRSSQ) between the two data sets can be calculated as follows:

$$\mathrm{RRSSQ} = \sqrt{\frac{\sum_{i=1}^{nspec}\sum_{j=1}^{nvar}(d_{i,j} - d_{i,j}^{\mathrm{reconstructed}})^2}{\sum_{i=1}^{nspec}\sum_{j=1}^{nvar} d_{i,j}^2}} \tag{18}$$

2.1.3. Data analysis. For the data analysis, MATLAB software (v. 5·2) was used (The MathWorks, Natick, MA). The computer configuration is a 266 MHz Pentium processor with 128 MB RAM.

2.2. Kinetics of chemical reactions using MIR spectroscopy

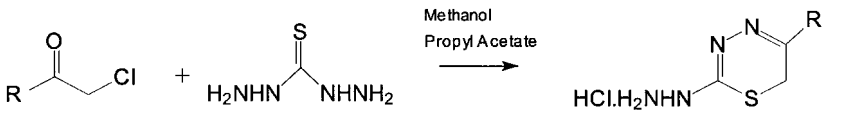
2.2.1. First-order reaction. The reaction studied involves the formation of an aromatic diazonium salt from a primary aromatic amine with sodium nitrite in aqueous mineral acid as follows:



The diazonium reaction was run in a 500 ml jacketed vessel equipped with a mechanical stirrer set to stir at 300 rpm. The starting amine was dissolved in aqueous mineral acid and cooled to 10 °C. A concentrated sodium nitrite/water solution was added over 20 min. The reaction was stirred for a total of 50 min. The diazonium product was then used immediately to form other products, which are not part of this discussion. The reaction is a second-order reaction; however, since the sodium nitrite was in excess, the reaction can be considered as a pseudo-first-order reaction.

The diazonium salt formation was monitored *in situ* using Fourier transform infrared (FTIR) spectroscopy. The FTIR instrument was a Bomem MB-100 equipped with an MCT detector and Axiom DP-210 immersion probe with an AMTIR ATR reflection element. Data collection was performed using Bomem GRAMS/386 software in the kinetics mode. Data were collected over the 50 min reaction time at the rate of one spectrum per minute. Thirty-two FTIR scans of 4 cm^{-1} resolution were averaged over each 60 s scan period.

2.2.2. First-order reaction with intermediate. The reaction involves the formation of a six-membered heterocycle via the reaction of thiocarbohydrazide with a β-chloro aliphatic ketone. The reaction proceeds through a non-isolated intermediate not shown here. The reaction is as follows:



The reaction was run in a 500 ml jacketed vessel equipped with a mechanical stirrer set to stir at 300 rpm. The starting carbohydrazide was mixed with the methanol/propyl acetate and formed an insoluble slurry. The starting ketone was added to this mixture and dissolved completely. The ketone material was in excess. As a result, the second-order reaction can be considered as a pseudo-first-order reaction. The reaction mixture was then warmed moderately and stirred for over 1 h before the product was isolated.

The reaction was monitored *in situ* using Fourier transform infrared (FTIR) spectroscopy. The FTIR instrument and collection parameters were the same as for the diazonium reaction. Data were collected over the entire reaction time at the rate of one spectrum per minute at a resolution of $4\ \mathrm{cm}^{-1}$.

2.3. Proton relaxation in solid state NMR

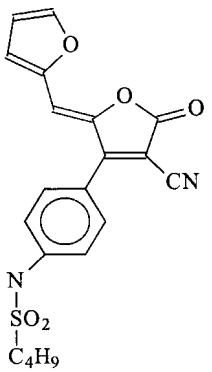
For a complete description of this study, see Reference [13]. Many organic solids can crystallize in more than one crystal structure, a behavior known as polymorphism. Polymorphs differ in their solubility, color and many other physical, chemical and mechanical properties that are key to their end use in the pharmaceutical or materials sciences. The ability to detect the presence of different polymorphs in a sample, assay their relative amount and ultimately obtain the desired form in very high yield are key issues in chemical technology.

Proton relaxation may be used to discriminate unique components (or phases) in a solid. Because they possess different crystal structures, polymorphs may exhibit different relaxation behavior caused by varied molecular ordering. Since there is fast spin diffusion in the solid state, all protons in a given phase will exhibit the same relaxation rate determined by the fastest-relaxing protons, usually methyl groups. Therefore the spectra of separate crystalline phases can in principle be resolved. The spin–lattice relaxation rate T_1 may be measured, or T_1-edited spectra acquired, using an inversion recovery cross-polarization/magic angle spinning (IR-CP/MAS) experiment [14]. In this experiment the protons on the molecule are excited using an inversion recovery scheme; then magnetization is transferred to carbon nuclei and the ^{13}C NMR spectrum is acquired. A delay period τ is varied in the experiment and a separate spectrum for each τ value is acquired. Each spectrum is then added to a spectrum acquired under the condition where $\tau = 0$. The relationship between τ and the acquired signal is exponential and is given below (k is a constant):

$$S = -k\mathrm{e}^{-\tau/T_1} \tag{19}$$

Hence for long τ values the signal approaches zero.

The NMR spectra were obtained on a Chemagnetics CMX 300 solid state NMR spectrometer using Spinsight 2.5 software. Two polymorphs (type I and type III) of the structure shown below were mixed in a 1:1 ratio:



Ten ^{13}C CP/MAS spectra were collected for values of τ ranging from 0·2 to 3·8 s in increments of 0·4 ns. Each spectrum contained 2048 real data points. The sign of the spectra was inverted and the data set was split into two: spectra 1–9 and spectra 2–10. DECRA requires that the number of components be defined prior to analysis. Because the analysis is so fast, typically a few seconds for a data set this size, it is no problem to try several permutations of spectral range and number of components. For this data set we tried both two and three components.

3. RESULTS AND DISCUSSION

3.1. Effect of the shift

In order to determine the best way to split the data set into two parts, ten replicate UV-vis data of a chemical reaction were analyzed by Bijlsma *et al.* [8] in order to determine the kinetics. The reaction constants, which can be derived directly from the eigenvalues, were calculated for all the data sets as a function of the shift. The standard deviation of the reaction constants was calculated, resulting from separate analyses of the data sets for different shifts. It appeared that the optimal shift, in terms of the minimal standard deviation for the reaction constants, was 30 [8]. However, when the diagnostic was based on similarity of the reference spectra with the extracted spectra, a shift of one appeared to give the best results [4].

In order to understand the apparent discrepancy in these results, the data sets were studied in more detail. It appeared that the DECRA results were slightly better when a baseline correction was applied [4]. Knowing that a positive noisy baseline problem is basically always present in the data, because of sources such as electronic noise, a simple model was generated by adding a positive noise baseline to an exponential. The exponential

$$y = e^{-bx} \tag{20}$$

was calculated for x values ranging from 0 to 40, with $b = 0{\cdot}1$. Random uniformly distributed noise r with values between 0 and 0·001 was added:

$$y^* = y + r \tag{21}$$

where y^* stands for an exponential function with the added noise. The exponential decay values for different y^* shifts Δ were estimated as follows:

$$b_\Delta = \frac{1}{\Delta}\log\left(\frac{y^{*T}y^*_\Delta}{y^{*T}_\Delta y^*_\Delta}\right) \tag{22}$$

This estimate is reasonable, because if there is no noise, $b_\Delta = b$ and corresponds to the DECRA estimate for univariate data. Furthermore, the decay values calculated by the exponential ratio model in Equation (22) were compared with the decay values calculated from the eigenvalues resulting from solving the generalized eigenvector problem for the same data sets. It appeared that the differences between the values were in the order of the machine precision, thus for all practical purposes the results were identical. The use of Equation (22) as a simplified model of the calculations allows the exploration of Taylor series to obtain a better understanding of the effect of the shift.

The shift Δ was varied from 1 to 20. The calculations were repeated 1000 times for newly generated noise distributions. The two exponentials y^* and y^*_Δ are of the same size, i.e. 20 elements, but shifted with respect to each other. This is different from the way actual data sets are split up for DECRA, where increasing shift values will result in smaller data sets. In order to simplify the comparison of exponentials for different shift values, the size of the range is kept constant. Simulation studies showed that the behaviors of the exponentials, as will be discussed below, are basically identical for exponentials with a constant number of elements and exponentials with different lengths occurring in practical situations.

One may wonder if this simple model properly represents the calculations to solve the generalized eigenvector problem when noise is present. A better choice would be simulated data sets with more than one component. This was not as simple as it might seem. For example, a simple noisy positive baseline added to a simulated three-component data set has the same effect as adding a noisy baseline with an average of zero, since there is already a constant contribution in the data set (the exponential with decay value zero; see Equation (9)). Determination of the proper baseline model is therefore a complicated and time-consuming matter that falls outside the scope of this application study. Furthermore, for real data sets the effect may be different for different components. One resolved component may be heavily affected by the shift while another is basically unaffected. Therefore an attempt was made to simplify the problem into one of an exponential model. In Figures 1(a) and 1(b) the mean values and the standard deviations for b_Δ are plotted. As can be seen, the best value for the decay constant, based on the mean values, is obtained when the shift $\Delta = 1$. On the other hand, the best value based on the minimal standard deviation value is for shift $\Delta = 8$. This simple simulation shows variance behavior similar to that reported by Bijlsma *et al.* [8]. In order to better understand the variance properties of this estimator of b, we attempted to derive an exact expression for the variance. It turned out to be intractable. A first-order approximation to the variance was derived, and it showed non-monotonic behavior of the variance as a function of the shift. This approximation did not agree well with the simulation, indicating that higher-order approximations are necessary. Nevertheless, the complexity of the first-order approximation suggests that the dependence of the variance on Δ is a complicated one.

More importantly, however, is the issue of what shift to use in everyday practice. As stated above, no theoretical guidance is available at this time. Practical experience and simulations provide some guidance. However, to correctly interpret them, it is important to realize that variability is only part of the issue. We must consider bias as well. Mean square error (MSE) is a commonly accepted measure of the overall accuracy of an estimator, and it takes both bias and variance into account:

$$\mathrm{MSE} = \mathrm{E}[(\hat{b} - b)^2] = \mathrm{bias}^2(\hat{b}) + \mathrm{var}(\hat{b}) \tag{23}$$

where E[ ] stands for the expected value and $\hat{b}$ is the estimator of b. There are many examples of estimators that are biased yet have smaller MSEs than their unbiased counterparts. PLS and ridge regression are well-known examples.

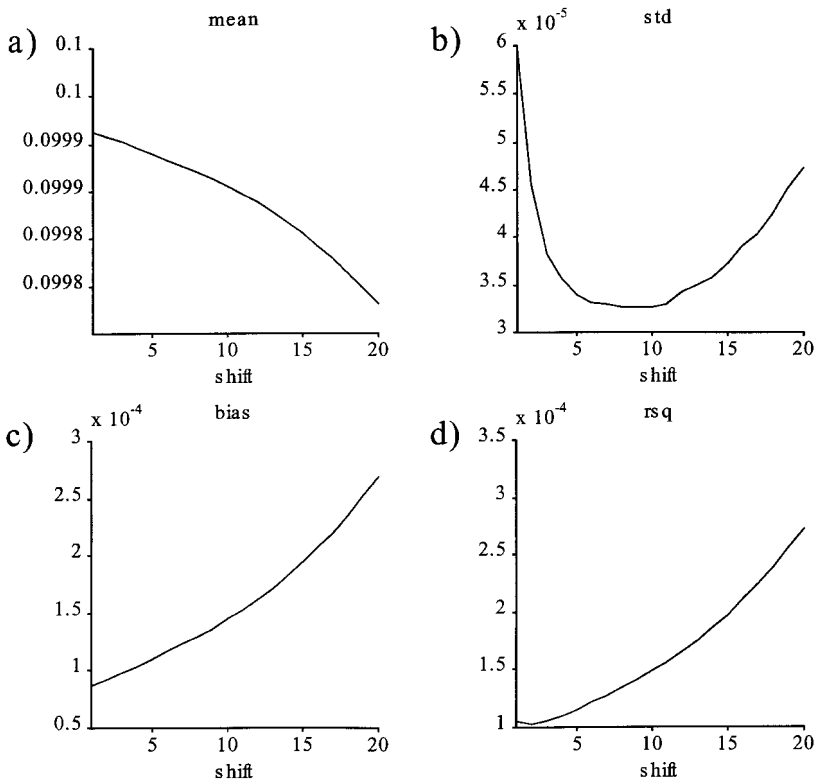


Figure 1. Results of a simulation study to determine the effect of shift.

Figure 1(c) shows the bias for the simulation example. It increases in magnitude as the shift increases. This is largely due to the fact that the added noise is positive. Offsetting the increased bias is decreased variance (standard deviation). Putting both together as (root) mean square error shows that shift $\Delta = 1$ is the best choice in terms of overall accuracy (Figure 1(d)). This is in concert with our practical experience. DECRA is used as a routine tool for PGSE NMR spectra. By now, more than 70 different data sets resulting from a wide variety of problems have been analyzed, and the results were always, with not a single exception, the best for a shift of one. The quality of the results was evaluated by an experienced spectroscopist (B. Antalek) by comparing the resolved results with known reference spectra (if available), known properties of spectra (even if reference spectra are not available, 'typical' peaks are often known, so it is possible to judge the quality of the resolved results as far as overlap is concerned, etc.) and the occurrence of negative intensities in spectra. Next to the PGSE NMR spectra, we have experience with MRIs, MIR, NIR and solid state NMR spectra. Again, with no exception, a shift of one gives the best results.

3.2. Kinetics of chemical reactions using MIR spectroscopy

3.2.1. First-order reaction. The first few spectra of the data set were deleted from the analysis because of problems with them, which were probably caused by incomplete mixing. Also, the results of the DECRA analysis were better when the spectra at the end of the reaction and after completion of the reaction were not included in the data analysis. In the discussion below the first and last

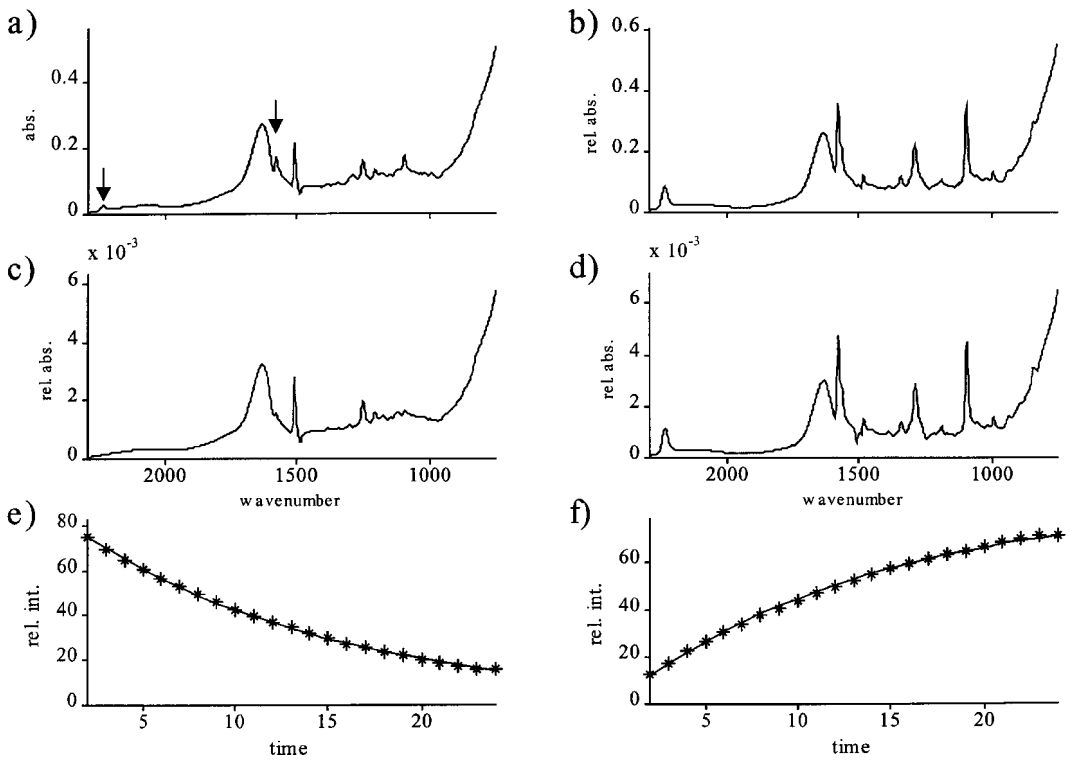


Figure 2. The first and last spectra of the data set are presented in (a) and (b). The resolved spectra are presented in (c), which represents the starting material, and (d), which represents the product. The presence of the product in the first spectrum is indicated by arrows. The exponential contribution profiles of the starting material and product are given in (e) and (f), where the curves represent the exact calculated exponential profiles (Equation (14)) and the asterisks represent the projected data points (Equation (16)).

spectra refer to the data set used for the data analysis. Because of the speed of the DECRA analysis, the proper data to analyze can be determined interactively within 1 min.

Figure 2 shows the results from the DECRA analysis for the diazonium reaction. The first spectrum (a) is representative of the starting material, but contains some contribution from the product as indicated by the arrows. Spectrum (b), while not a reference spectrum, is representative of the product. No reference spectrum of this component is available, since the material was not isolated. The DECRA-resolved spectra are presented in (c) and (d). It is evident that the DECRA-resolved spectrum (d) very closely matches the corresponding spectrum for the product (b). The resolved spectrum (c) very closely matches that of the starting material (a), but the contribution of the product (indicated with arrows in (a)) is negligible, which confirms that the spectrum in (c) represents the pure starting material. Of most significance is the presence of the absorption band at 2239 cm^{-1} in spectra (b) and (d), which is indicative of the diazonium $(—N_2^+)$ functionality and therefore product formation.

The exponential contribution profiles of the starting material and product are given in Figures 2(e) and 2(f). These profiles are consistent with the intensity changes seen for individual absorption bands for the reaction. Furthermore, these results are similar to those seen with SIMPLISMA [15–17] analysis of the data set.

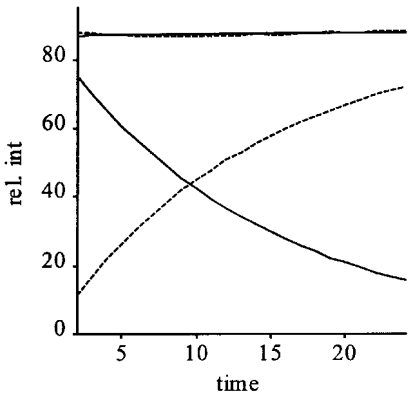


Figure 3. The starting (—) and product (---) profiles are presented again, together with their sum (—) and the total signal of the data set (---). The RRSSQ is 0·0069838.

In Figure 3 the resolved profiles are plotted again, together with their sum and the total signal intensity of the data set. A close match between the latter two profiles is confirmed by the very low RRSSQ (Equation (18)) value given in the figure caption.

3.2.2. First-order reaction with intermediate. The proper series of spectra to analyze was determined interactively as described above.

Figure 4 shows the data from the DECRA experiment for the heterocyclic reaction. Spectrum (a) is the first spectrum of the data set and is a good reference of the starting ketone. Spectrum (b) is the eleventh spectrum of the data set, at which point the intermediate is at its highest contribution. While representative of the intermediate, there are also minor contributions from the product and starting ketone. Spectrum (c) is the last spectrum of the data set and is representative of the product. Spectrum (g) is a reference spectrum of the product. The DECRA-resolved spectra are presented in (d)–(f). It is evident that the DECRA-resolved spectrum (d) very closely matches the corresponding spectrum for the ketone starting material (a), and the resolved spectrum (e) that of the intermediate (b), even with the trace presence of other components in (b). Also, resolved spectrum (f) compares favorably with the reaction product spectrum (g). However, it is apparent that spectrum (f) has some problems: slightly negative absorbencies (set to zero in the figure) are present across the entire spectrum. This is an indication that the spectral mixture data deviate somewhat from the ideal exponential mixture. Close examination of these bands shows that they have a contribution from the intermediate spectrum (e) and likely result from the fact that this component has an extremely high absorptivity. Even so, spectrum (f) is sufficiently resolved such that the major absorption bands can be identified. Furthermore, it is clear that these bands match up well with those in the reference spectrum of the product (g).

It should be noted here that while the strong methanol band at 1040 cm^{-1} has been deleted from the data set, other methanol bands are still present. These can be seen between 1350 and 1450 cm^{-1} in spectra (a), (c) and (d) and are less apparent in (b), (e) and (f). In the product reference spectrum (g), methanol has been spectrally subtracted out.

The exponential contribution profiles of the starting material, intermediate and product are given in (h)–(j). These represent the kinetic behavior of each component over the course of the reaction, as determined by DECRA. These profiles are consistent with those obtained from SIMPLISMA [15–17] analysis of the data set.

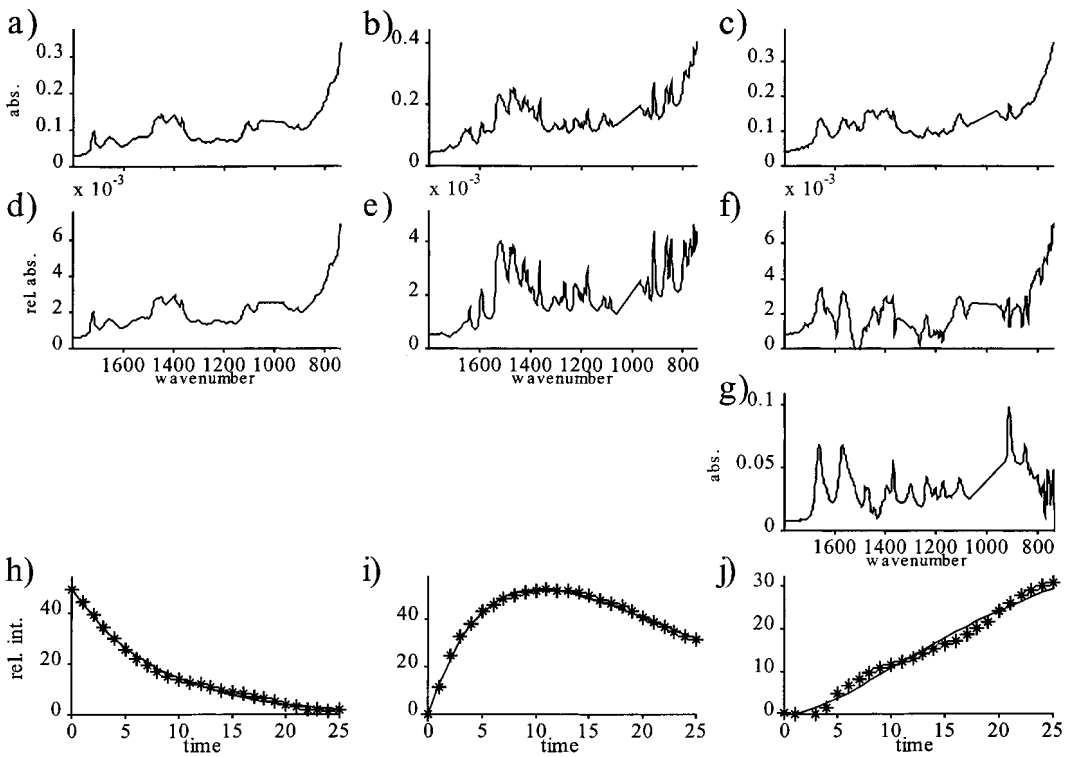


Figure 4. The first, eleventh and last spectra of the data set are presented in (a)–(c). The resolved spectra are presented in (d), which represents the starting material, (e), which represents the intermediate product, and (f), which represents the product. Spectrum (g) is a reference spectrum of the product. The exponential contribution profiles of the starting material, intermediate and product are given in (h)–(j), where the curves represent the exact exponential profiles (Equation (14)) and the asterisks represent the projected data points (Equation (16)).

Figure 5 shows how well the sum of the resolved contribution profiles matches the total signal of the original data set.

3.3. Solid state NMR data analysis

Figure 6 shows a portion of the data set as well as the results of DECRA compared with reference spectra. As can be seen in Figures 6(a) and 6(b), changes are observed in the spectra as a function of the τ value; note the change in the vertical scale. Some of the peaks decay (or have completely disappeared) with respect to the others. DECRA analysis using two components provides two very different spectra. These are shown in Figures 6(c) and 6(d). Reference spectra for comparison are shown in Figures 6(e) and 6(f) and are types I and III respectively. Analysis using three components provides two spectra that are identical to those of the two-component analysis, as well as a third spectrum consisting of only noise. This indicates that there are only two components present (based on T_1). The similarity between the resolved spectra and the reference spectra is excellent.

The resolved T_1 values for type I and type III forms are 1·16 and 7·18 s. Separate T_1 experiments were not performed, but as a check we determined T_1 values directly from two well-resolved

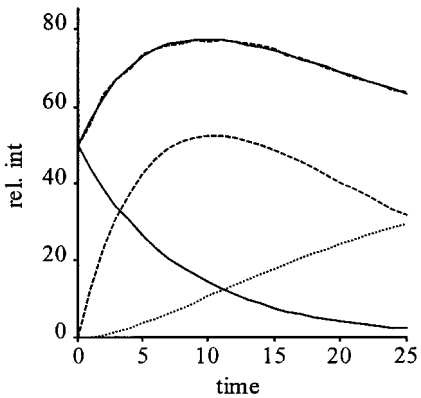


Figure 5. The starting (—), intermediate (---) and product (.....) profiles are presented again, together with their sum (—) and the total signal of the data set (---). The RRSSQ is 0·011896.

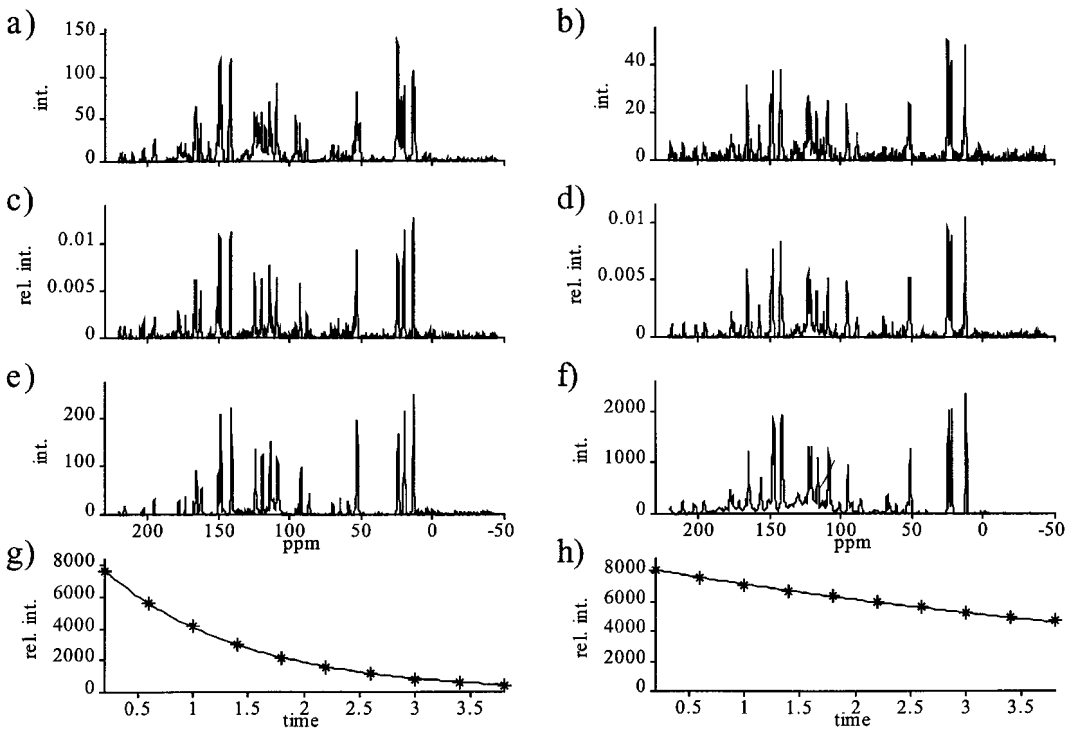


Figure 6. The first and last spectra of the data set are presented in (a) and (b). The resolved spectra are presented in (c), which represents the starting material, and (d), which represents the product. Reference spectra are presented in (e) and (f). The correlation coefficients of the resolved components with the model spectra are 0·93 and 0·92 respectively. The exponential contribution profiles of the starting material and product are given in (g) and (h), where the curves represent the exact exponential profiles (Equation (14)) and the asterisks represent the projected data points (Equation (16)).

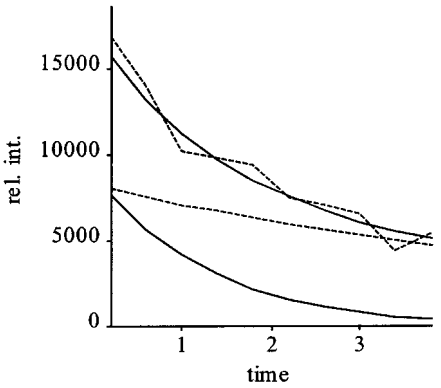


Figure 7. The starting (—) and product (---) profiles are presented again, together with their sum (—) and the total signal of the data set (---). The RRSSQ is 0·12491.

resonances in the data set. For the resonances at 13·7 and 12·3 ppm we obtain values of 1·36 and 6·13 s respectively for T_1.

Figure 7 shows how well the resolved results reproduce the original data. It can be seen that the match is not as close as with the other examples shown. The resolved spectra in Figure 6 show, however, that DECRA resolved the mixture data successfully.

4. CONCLUDING REMARKS

The results above showed that DECRA can be used to study chemical reactions monitored by MIR. It is important for DECRA that the data do not deviate significantly from the trilinear model. An interactive process is necessary to determine the proper series of spectra to analyze. Future research will focus on the automated determination of the proper part of the data set to analyze. It is also noteworthy that the data analysis improved when the solvent peak was deleted, for reasons that are not clear. Whereas the solvent peak is basically constant, it does not add to the rank of the data set, and one would expect to obtain resolved components intermixed with the solvent. A possible reason may be some effects on the solvent peak, such as peak broadening. Pre processing this type of data is an important subject for future research.

In order to resolve the data sets properly, it is important to know the starting time for the reaction, since different starting points will result in different contribution profiles for Equations (3)–(5). This is not always trivial, because of mixing problems, etc.

It also has to be noted that it is not clear from the DECRA results which one of the resulting decay constants is k_1 and which is k_2. This needs either to be known, or to be judged from the resolved spectra resulting from Equation (15). When the decay constants are exchanged, the resulting spectra often exhibit unrealistic features, such as negative intensities.

The solid state NMR application of DECRA was straightforward and without problems. As can be seen in Figure 7, the reconstructed total signal and the original total signal do not match as closely as for the MIR data, although the resolved spectra have an excellent match with the reference spectra.

The study of the effect of the shift was limited and a simplified model was used. A more thorough understanding of this issue is an open topic for further research.

REFERENCES

1. Antalek B, Windig W. Generalized rank annihilation method applied to a single multicomponent pulsed gradient spin echo NMR data set. *J. Am. Chem. Soc.* 1996; **118**: 10331.
2. Windig W, Antalek B. Direct exponential curve resolution (DECRA): a novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying profiles. *Chemometrics Intell. Lab. Syst.* 1997; **37**: 241.
3. Windig W, Antalek B. Resolving nuclear magnetic resonance data of mixtures by 3-way analysis. Examples of chemical solutions and the human brain. *Chemometrics Intell. Lab. Syst.* 1999; **46**: 207.
4. Windig W, Antalek B, Sorriero LJ, Bijlsma S, Louwerse DJ, Smilde AK. Applications and new developments of the direct exponential curve resolution algorithm (DECRA). Examples of spectra and magnetic resonance images. *J. Chemometrics* 1999; **13**: 95.
5. Windig W, Hornak JP, Antalek B. Multivariate image analysis of magnetic resonance images with the direct exponential curve resolution algorithm (DECRA), Part 1: Algorithm and model study. *J. Magn. Reson.* 1998; **132**: 298.
6. Antalek B, Hornak JP, Windig W. Multivariate image analysis of magnetic resonance images with the direct exponential curve resolution algorithm (DECRA), Part 2: Application to human brain images. *J. Magn. Reson.* 1998; **132**: 307.
7. Bijlsma S, Louwerse DJ, Windig W, Smilde AK. Rapid estimation of rate constants using on-line SW-NIR and trilinear models. *Anal. Chim. Acta* 1998; **376**: 339.
8. Bijlsma S, Louwerse DJ, Smilde AK. Estimating rate constants and pure UV-VIS spectra of a two-step reaction using trilinear models. *J. Chemometrics* 1999; **13**: 1.
9. Roy R, Kailath T. Estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust., Speech, Signal Processing* 1989; **37**: 984.
10. Kubista M. A new method for the analysis of correlated data using Procrustes rotation which is suitable for spectral analysis. *Chemometrics Intell. Lab. Syst.* 1990; **7**: 273.
11. Sanchez E, Kowalski BR. Generalized rank annihilation factor analysis. *Anal. Chem.* 1988; **58**: 496.
12. Wilson B, Sanchez E, Kowalski BR. An improved algorithm for the generalized rank annihilation method. *J. Chemometrics* 1989; **3**: 493.
13. Zumbulyadis N, Antalek B, Windig W, Scaringe RP, Lanzafame AM, Blanton T, Helber M. The elucidation of polymorph mixtures using solid state 13C CP/MAS NMR spectroscopy and direct exponential curve resolution. *J. Am. Chem. Soc.* 1999; **121**: 11554.
14. Freeman R, Hill HDW. Fourier transform study of NMR spin-lattice relaxation by "Progressive Saturation". *J. Chem. Phys.* 1971; **54**: 3367.
15. Windig W, Guilment J. Interactive self-modeling mixture analysis. *Anal. Chem.* 1991; **63**: 1425.
16. Windig W, Stephenson DA. Self-modeling mixture analysis of second derivative near infrared spectral data using the SIMPLISMA approach. *Anal. Chem.* 1992; **64**: 2735.
17. Windig W. Spectral data files for self-modeling curve resolution with examples using the SIMPLISMA approach. *Chemometrics Intell. Lab. Syst.* 1997; **36**: 3.

