

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/6251211>

# Application of Chemometric Analysis to Complexity in Isothermal Calorimetric Data

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · AUGUST 2007

Impact Factor: 3.3 · DOI: 10.1021/jp0700985 · Source: PubMed

---

CITATIONS

9

---

READS

30

5 AUTHORS, INCLUDING:



Michael A. A. O'Neill

DLA Piper

31 PUBLICATIONS 384 CITATIONS

SEE PROFILE



Simon Gaisford

The School of Pharmacy

92 PUBLICATIONS 807 CITATIONS

SEE PROFILE

## Application of Chemometric Analysis to Complexity in Isothermal Calorimetric Data

Michael A.A. O'Neill,<sup>\*,†</sup> Anthony E Beezer,<sup>‡</sup> John Tetteh,<sup>§</sup> Simon Gaisford,<sup>‡</sup> and Meena Dhuna<sup>‡</sup>

Department of Pharmacy and Pharmacology, University of Bath, Claverton Down, Bath BA2 7AY, U.K., School of Pharmacy, University of London, 29-39 Brunswick Square, London WC1N 1AX, U.K., and University of Greenwich, School of Sciences, Central Avenue, Chatham Maritime, ME4 4TB, U.K.

Received: January 5, 2007; In Final Form: April 16, 2007

The interpretation of complexity in isothermal calorimetric data is demanding. The observed power signal is a composite of the powers arising from each of the individual events occurring (which can involve physical, as well as chemical, change). The challenge, therefore, lies in deconvoluting the observed data into their component parts. Here, we discuss the potential use of chemometric analysis, because it offers the significant advantage of being model-free, using principal component analysis to deconvolute data. Using model data, we discovered that the software required a minimum trivariate data matrix to be constructed. Two variables, power and time, were available from the raw data. Selection of a third variable was more problematic, but it was found that by running multiple experiments the small variation in the number of moles of compound in each experiment was sufficient to allow a successful analysis. In general we noted that it required a minimum  $2n + 2$  repeat experiments to allow analysis (where  $n$  is the number of reaction processes). The data outputted from the chemometric software were of the form intensity (arbitrary units) versus time, reflecting the fact that the software was written for analysis of spectroscopic data. We provide a mathematical treatment of the data that allows recovery of both reaction enthalpy and rate constants. The study demonstrates that chemometric analysis is a promising approach for the interpretation of complex calorimetric data.

## Introduction

Quantitatively analyzing isothermal calorimetric data is not straightforward when more than one event contributes to the measured signal, because heat is ubiquitous and the calorimeter simply measures the net heat change. In some instances experimental design can be used to isolate (and hence investigate) each process, but increasingly it is desirable to study “whole” systems. Thus, methods of dealing with complexity in calorimetric data must be developed.

We have proposed a number of methods for quantitative analysis of complexity, both in solution<sup>1–3</sup> and solid states.<sup>4</sup> We have also discussed a method for the direct calculation of reaction parameters in complex systems.<sup>5</sup> However, these approaches shared the common drawback of requiring knowledge of the reaction mechanism. An alternative approach to the problem of complexity in data is the application of chemometric procedures via techniques such as principal component analysis (PCA).<sup>6–7</sup> Chemometric analyses are well established for spectroscopic data, and there are many commercially available software packages for this purpose (from, for example, DiKnow Ltd, U.K. and Infometrix Inc, U.S.A.). Chemometric analysis has been applied to the interpretation of non-isothermal calorimetric data (in particular to solid-state systems) where the complexity arises from the underlying (applied) change in temperature over the course of the experiment as well as any multistep chemical/physical processes. Such analyses are designed to recover the “kinetic trilogy” (rate constant,  $k$ , reaction

order,  $n$ , and activation energy,  $E_a$ ). Debate exists as to the best chemometric approach for the analysis of scanning calorimetric data, with a study dedicated to the comparison of the different methods being available: the ICTAC Kinetics Project.<sup>8–13</sup> However, currently there exists no chemometric package written specifically for the analysis of isothermal calorimetric data; addressing this significant knowledge gap is the focus of this paper.

We discuss here the potential application of chemometric analysis for interpretation of complex isothermal calorimetric data. The primary issue to be addressed is to determine unequivocally the number of reaction processes; since there is no (direct) molecular information in isothermal calorimetric data there is a need to find a matrix that can yield the number of processes and the reaction order. Subsequent analysis can be performed to recover kinetic and thermodynamic parameters. As a check on the appropriateness of analysis, there is a requirement that all reaction orders are integral. This paper is concerned with showing that this analysis is possible. For the purposes of discussion, this paper focuses on consecutive reaction schemes, but we will return to the subject of more complex pathways in a future publication.

## Experimental Methods

**Chemometric Analysis.** The measured power from the calorimeter is the sum of the power outputs for each constituent reaction at defined times. The initial purpose of a meaningful chemometric analysis must then be to determine the *number* of reaction processes that comprise the overall data. A test of such deconvolution is that the reaction order of each process is *integral*. Once the number and orders of individual processes

\* Corresponding author. E-mail: mon20@bath.ac.uk. Telephone: +44 (0) 1225 386797. Fax: +44 (0) 1225 386114.

<sup>†</sup> University of Bath.

<sup>‡</sup> University of London.

<sup>§</sup> University of Greenwich.

are known, then reaction parameters can be recovered by classical analyses such as those we have already proposed.<sup>1–4</sup>

From the perspective of pattern analysis, each component in the system will exhibit a unique maximum during the course of the reaction at a defined time which suggests that the data may be analyzed by a multivariate method if a suitable matrix can be generated. Multivariate data are easily generated in spectroscopic systems<sup>14</sup> (for instance, intensity as a function of wavenumber, time, pH, temperature, etc.). Calorimetric data recorded at a fixed temperature are univariate, and there is, therefore, the need to engineer a third (intensity, or  $S$ ) factor of the matrix.  $S$  could, for instance, be obtained from replicate runs of the same experiment. These runs may be designed such that each one varies from the other in the total power output (for example, variation in sample loading or weight for repeated experiments provides  $S$  sets of data). A system in which  $X$  species evolve after time  $t$  will require at least  $2X + 2$  replicate runs (i.e.,  $S = 2X + 2$ ) to create a suitable matrix,  $M$ , which can then be analyzed to determine the rank of the data through multivariate data decomposition.

In the work reported here, factor analysis was used to deconvolute the data matrix on the assumption that (i) the power signal at time  $t$  is a linear sum of all species present at that moment and (ii) that the power signal is proportional to the sample load in the calorimeter. The data deconvolution and further treatments required to furnish the underlining evolution profile of a suitably simulated experimental data set are presented in the data analysis section of this paper. The algorithms and methods are embodied in software called InSight<sup>15</sup> coded in Matlab.<sup>16</sup> All chemometric data analysis was performed through this software.

**Principles of Analysis.** Calorimetric power–time data, at a fixed temperature, for the same sample with variable weights generate a data matrix  $M$  whose rows, in the work reported in this paper, represent the time ( $t$ ) points. The columns represent variation in sample weight ( $w$ ), and it is assumed that the reaction's kinetic parameters are not changed by sample size and that the instrument exhibits random noise for each sample run ( $S$ , 1 to  $m$ ;  $m$  is the number of runs). The deconvolution of  $M$  without prior knowledge will at least provide a basic understanding of the underlying factors or species in the reaction over a time period,  $t$ . In this work, we used a modified target factor analysis (TFA) technique to analyze the data. TFA is used to determine whether or not a hypothetical vector, gleaned from chemical principles or heuristic intuition, lies inside the factor space and thus contributes to the phenomenon.<sup>17–18</sup> The analytical strength of target testing lies in the fact that each hypothetical vector can be tested individually for significance in the presence of a host of other unknown factors. When a data matrix has been deconvoluted into abstract factors (also known as principal components) in the row and column space of the data  $M$  and the numbers of significant factors have been determined, these significant factors can be subjected to various forms of mathematical scrutiny to determine if they have real chemical or physical meaning. In this work the mathematical technique of singular value decomposition (SVD)<sup>19</sup> was used for the deconvolution process involved in the TFA process.

Equations 1–6 summarize the mathematical principles used to deconvolute the “spectral” data,  $M$ . The principal factor matrices describing the row and column information are  $R$  and  $C$  respectively.  $R$  and  $C$  are the significant components of the matrix after excluding  $E$ , the error or noise added to the data from the instrument to data,  $M$ . Recombination of the significant row and columns  $R$  and  $C$  produces a new matrix  $M_x$ , eq 2.  $X_i$

are hypothetical  $j$  sets of vectors that are considered fully to describe the data  $M_x$ . Various schemes have been proposed to determine  $X_i$ . In this paper the so-called needle search method (fully described elsewhere<sup>14–15</sup>) was employed. Essentially, initial hypothetical test vectors are generated by setting all time points to zero except those points that show significant power intensity for each of the  $j$  factors identified. From these test vectors, prototype profiles are generated after target testing. An iterative or interactive method is then used to select the unique set of  $X_i$  that best describes the data. Operating the pseudoinverse of  $R$ , i.e.  $R^+$  with  $X_i$  yields  $T$ , the transformation matrix indicated in eq 3. This parameter,  $T$ , is used to obtain  $X_j$  and  $Y$ . The parameter  $X_n$  represents the predicted test matrix based on the  $X_i$  (see eq 4) and describes the row domain real information, which are the predicted evolution profiles for each of the significant  $j$  factors identified in the data matrix,  $M$ .  $Y$  is the column domain real information matrix, equivalent to the predicted power output profiles obtained from eq 5. Recombination of  $X_n$  and  $Y$  produces  $M_n$  the significant part of the data matrix,  $M$ . The differences between  $M_n$  and  $M_x$  are minimized. An extended treatment of theory and applications of TFA can be found elsewhere.<sup>14–15</sup>

$$M = [RC] + E \quad (1)$$

$$M_x = RC \quad (2)$$

$$T = R^+ X_i \quad (3)$$

$$X_n = RT \quad (4)$$

$$Y = T^+ C \quad (5)$$

$$M_n = X_n Y \quad (6)$$

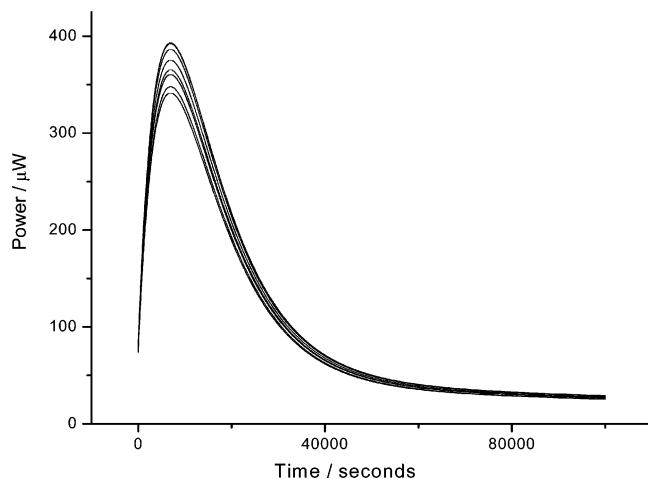
**Data Simulation.** Simulated data were created to allow exploration of the range of applicability of the proposed method. Simulated data offer the advantage that the reaction mechanism and reaction parameters are known unequivocally. It was regarded as essential for success that the methods described herein should require no information other than the sample load placed into the calorimeter and the associated power–time outputs observed over a reasonable time period (24 h in this instance). Thus, no model was imposed on the data during chemometric analysis for interpretation of mechanism. The first objective was to determine whether the analysis could correctly identify the number of processes involved in the reaction and their concentration profiles. Values for the target parameters (first-order rate constant,  $k$ , and reaction enthalpy  $\Delta H$ ) were selected for a consecutive reaction system ( $A$  to  $B$  to  $C$  to  $D$ ). The power outputs as a function of time for each component reaction were then calculated using eqs 7–9<sup>20</sup>

$$P_{A \rightarrow B} = H_1 k_1 A_0 e^{-k_1 t} \quad (7)$$

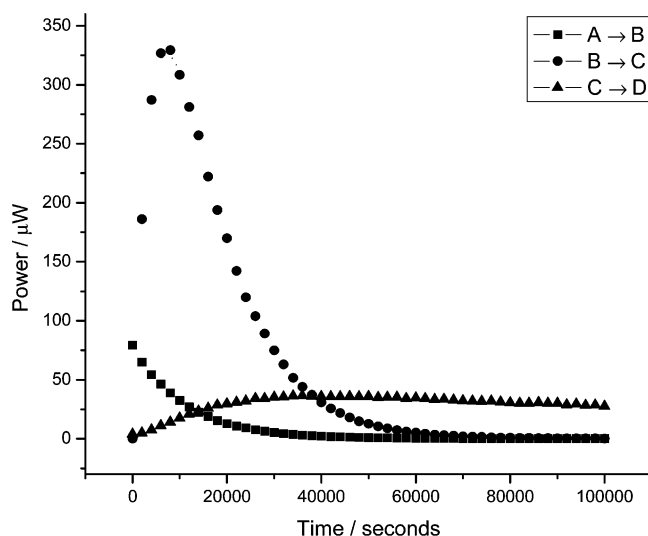
$$P_{B \rightarrow C} = H_2 \left[ A_0 k_1 k_2 \frac{(-e^{k_1 t} + e^{-k_2 t})}{(k_1 - k_2)} \right] \quad (8)$$

$$P_{C \rightarrow D} = \Delta H_3 \left[ A_0 k_1 k_2 k_3 \left( \frac{e^{-k_1 t}}{(k_2 - k_1)(k_3 - k_1)} + \frac{e^{-k_2 t}}{(k_1 - k_2)(k_3 - k_2)} + \frac{e^{-k_3 t}}{(k_1 - k_3)(k_2 - k_3)} \right) \right] \quad (9)$$

where  $P$  ( $J s^{-1}$ ) is the calorimetric output (power) and  $A_0$  is the initial quantity of reactant (mol). The subscripted numbers denote the reaction step.



**Figure 1.** Set of simulated data entered into the chemometric software for PCA analysis.



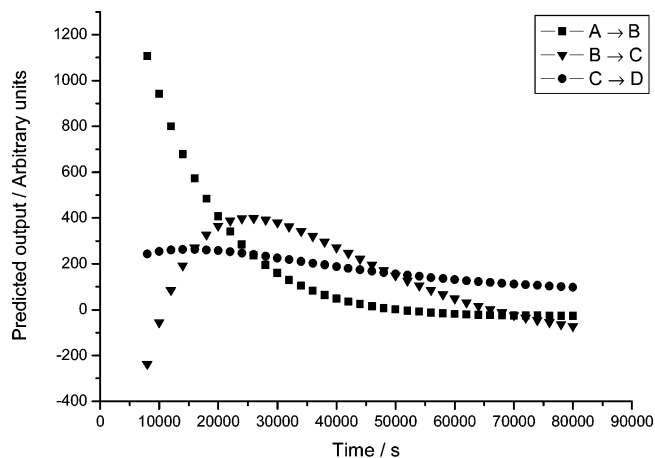
**Figure 2.** Component steps of the simulated data shown in Figure 1.

The simulated observed calorimetric output is then simply a summation of these reaction steps.

It was noted earlier that successful chemometric analysis usually requires a multi- (at least tri-) variate data matrix. The raw calorimetric data (power and time) provide two variables, and hence, a third parameter must be generated. Since in reality there is a contribution to the calorimetric signal from inherent noise within the system (electrical, thermal etc) as well as slight variations in experimental parameters (variations in sample weighing for instance), we opted to add a random error function to the simulated data as the third variable. This was achieved by (i) adding a random noise value to the enthalpy value used when creating data and (ii) varying the value of  $A_0$  between data sets.

The final data matrix thus consisted simply of files of power-time data for 8 different (arbitrary) initial concentration values of  $A_0$  with overlaid noise; an example data set can be seen in Figure 1. The set of individual power-time data which comprise the “overall” simulated data can be found in Figure 2. Only the overall data files were submitted to the chemometric program; thus, as noted above, no information on mechanism was provided prior to analysis.

The  $k$  values used to generate the simulated data for the three reaction steps were  $9 \times 10^{-5} \text{ s}^{-1}$ ,  $2 \times 10^{-4} \text{ s}^{-1}$ , and  $5 \times 10^{-6} \text{ s}^{-1}$  respectively. The  $\Delta H$  values used for the simulation were 10, 80 and 90  $\text{kJ mol}^{-1}$  (the percentage variation imposed was



**Figure 3.** Three principal components outputted from the chemometric analysis (further factors were indistinguishable from noise, data not shown).

a random value between +2% and -2% of the enthalpy value) respectively. The values of  $A_0$  ranged from  $7.5 \times 10^{-5}$  to  $9 \times 10^{-5}$  moles.

### Results/Discussion

The simulated data were entered into the InSight program and subjected to PCA analysis. The first test of the software was the determination of the number of reaction steps (or principal components). In order to effect this statistical examination three criteria were used; the statistical F-test, the percentage cumulative variance in the data and a visual examination of the factors in both row and column domains. The InSight software package performs the statistical tests as part of the deconvolution process and comprehensively described three principal factors, co-incidental with the model used to generate the data. Additional factors were indistinguishable from noise. This was further confirmed by visual inspection of the output, which also revealed three principal factors (shown in Figure 3). Thus, the first objective stated earlier, the ability of the software to determine the number of principal components, was demonstrated (note that although we have used a three-step model as an example, similar analyses of other consecutive data sets with varying numbers of steps always produces the correct number of principal factors; data not shown).

The next step was to confirm that each process identified by the analysis was integral order because, as noted above, this acts to demonstrate appropriateness of the output. There are several ways to accomplish this but we opted to use the data pairing technique derived by Willson.<sup>21</sup> In brief this algorithm takes pairs of data points and uses the derived ratios to predict an order of reaction. While this works perfectly with simulated data (i.e., integral numbers are returned for the predicted order of reaction) for our data set we returned values for reaction orders which were not integral but did return values for each step of  $1 \pm 0.2$ , i.e., the returned order values could be rounded up or down with reasonable confidence. It is perhaps unsurprising that this is so as real data have an element of random noise interspersed within the signal. This has the effect of introducing some uncertainty to the returned order values. Moreover the returned chemometric data sets do not reflect a “real” calorimetric output but are a composite response of the number of repeat data sets submitted to the software. Hence it is likely that further error will be incorporated into the analysis. Thus far we have demonstrated that complex isothermal calorimetric data are amenable to chemometric analysis to yield,

as a minimum, the number of reaction processes and their respective reaction orders. Once these values are known it is possible to select an appropriate reaction mechanism and analyze the original calorimetric data using the techniques we have already established.<sup>1–4,22</sup>

One interesting, and unexpected, issue that arose from this study is that the deconvoluted data had no defined y-axis units (the software plots the principal components as an arbitrary intensity; a result of the fact that the software package used was designed for spectroscopic analyses). This is a consequence of a deconvolution process in which no calibration curve is used (it is impossible to generate such a curve for calorimetric data). Although it is possible, once the mechanism is known, to return to the original data and analyze them using established techniques, it is interesting to consider whether quantitative reaction parameters can be determined from the chemometric output, both from an academic standpoint and because the values obtained would give good estimates for further analyses. While the rate constant values should be easy to ascertain (because they are time dependent, and the time axis is unaffected by chemometric deconvolution) the enthalpy values are not (because they are power dependent). One approach is to consider analysis based on fractional extents of reaction, since this normalizes the y-axis data. We show below a method for determining both parameters, considering up to three reaction steps; extension to longer schemes is easily achieved by following the same principles.

**Calculation of Rate Constants.** The determination of  $k_1$  is straightforward; a plot of  $\ln(\text{intensity})$  versus time for the first principal component should be linear, and the slope gives the rate constant. An alternative, approximate, method would be to plot  $\ln(\text{power})$  versus time (i.e., the observed calorimetric data) and take the slope over the initial few hours of data (where the reaction  $A$  to  $B$  predominates).

In the case of a two-step reaction, the value of  $k_2$  can be calculated as follows because there is only one intermediate compound. The concentration of product  $C$  as a function of time is given by

$$[C] = A_0 \left[ 1 + \frac{1}{k_1 - k_2} (k_2 e^{-k_1 t} - k_1 e^{-k_2 t}) \right] \quad (10)$$

Assuming that all of  $A$  will degrade to  $C$ , then when 50% of the reaction has occurred

$$[C] = \frac{A_0}{2} \quad (11)$$

and hence

$$F = 0.5 = 1 + \frac{1}{k_1 - k_2} [(k_2 e^{-k_1 t_{50}}) - (k_1 e^{-k_2 t_{50}})] \quad (12)$$

where  $F$  is the fraction of reaction that has occurred and  $t_{50}$  is the time at which 50% of reaction has occurred. ( $t_{50}$  can be calculated a number of ways; if the reaction has progressed to completion during the experimental time frame then fractional areas suffice. Alternatively, the amount of degradation that has occurred during the experiment can be quantified post measurement by a complementary analysis. Note also that there is no requirement for 50% to be the required fraction of reaction; the method is equally valid for any fraction of reaction.) The value of  $k_2$  is then easily calculated by iteration.

If there are more than two consecutive steps, then an alternative method must be used to determine the rate constants,

although the rate constant for the final step can be obtained as above (if one remembers to substitute the appropriate rate constants and concentration terms). Here, it is required that the time for maximum rate ( $t_{\max}$ ) to be attained is known; this corresponds to the peak power output in the power-time curve. It can be shown from classical reaction kinetics that

$$\frac{k_1 e^{-k_1 t_{\max}}}{k_2 - k_1} + \frac{k_2 e^{-k_2 t_{\max}}}{k_1 - k_2} = 0 \quad (13)$$

As before, the value of  $k_2$  is easily determined by iteration.

**Calculation of Reaction Enthalpies.** Once the rate constants are determined, reaction enthalpies are easily calculated. We present two methods: one based upon two power-time data and one upon two heat-time data. While the two methods are equivalent when applied to simulated data, it is likely that, in practice, the heat-time method will be more robust, because it averages data over the time range selected.

**Calculation of Reaction Enthalpies from Two Power-Time Points.** The power output for a reaction that proceeds through two consecutive, first-order steps is given by summation of eqs 7 and 8

$$\frac{dq}{dt} = k_1 H_1 A_0 e^{-k_1 t} + k_1 k_2 H_2 A_0 \frac{e^{-k_1 t} - e^{-k_2 t}}{k_2 - k_1} \quad (14)$$

It is therefore possible to select from the experimental data the instantaneous power outputs ( $P_1$  and  $P_2$ ) at two respective time points ( $t_1$  and  $t_2$ ) and write

$$P_1 = k_1 H_1 A_0 e^{-k_1 t_1} + k_1 k_2 H_2 A_0 \frac{e^{-k_1 t_1} - e^{-k_2 t_1}}{k_2 - k_1} \quad (15)$$

$$P_2 = k_1 H_1 A_0 e^{-k_1 t_2} + k_1 k_2 H_2 A_0 \frac{e^{-k_1 t_2} - e^{-k_2 t_2}}{k_2 - k_1} \quad (16)$$

Rewriting eq 15 to make  $H_1$  the subject

$$H_1 = \frac{P_1(k_2 - k_1) - k_1 k_2 A_0 H_2 (e^{-k_1 t_1} - e^{-k_2 t_1})}{k_1 A_0 (k_2 - k_1) e^{-k_1 t_1}} \quad (17)$$

Substituting eq 17 into eq 16 and rearranging for  $H_2$  gives

$$H_2 = \frac{(k_2 - k_1)(P_1 e^{-k_1 t_2} - P_2 e^{-k_1 t_1})}{k_1 k_2 A_0 (e^{-k_1 t_1 - k_2 t_2} - e^{-k_1 t_2 - k_2 t_1})} \quad (18)$$

Thus, eq 18 can be used to calculate the value of  $H_2$ . Knowledge of  $H_2$  then allows  $H_1$  to be calculated. If  $t_1$  is selected to be zero, then  $P_1$  becomes equal to the initial power output of the reaction, and eq 18 reduces to

$$H_2 = \frac{(k_2 - k_1)(P_1 e^{-k_1 t_2} - P_2)}{k_1 k_2 A_0 (e^{-k_2 t_2} - e^{-k_1 t_2})} \quad (19)$$

A further discussion of the practicality of the use of  $t_0$  as a reference point can be found below.

**Calculation of Reaction Enthalpies from Two Heat-Time Points.** In this case, the heat outputs ( $q_1$  and  $q_2$ ) at two time points ( $t_1$  and  $t_2$ ) are required. The heats correspond to the area under the curve up to the two time points and can easily be determined using a mathematical analysis package (such as Origin, Microcal Software Inc.). However, in order to show the derivation of this approach, it is necessary to consider the



mathematical equations that give these values; these are the integrals of eq 14 between  $t_i$  and  $t_1$  and  $t_i$  and  $t_2$  (where  $t_i$  is some initial time which does not have to be zero) and are represented by

$$q_1 = A_0 H_1 (k_1 e^{-t_1 k_1} - k_2 e^{-t_1 k_1} + k_2 e^{-t_i k_1} - k_1 e^{-t_i k_1}) + A_0 H_2 (k_1 e^{-t_1 k_2} - k_1 e^{-t_i k_2} + k_2 e^{-t_1 k_1} - k_2 e^{-t_i k_1}) / (k_2 - k_1) \quad (20)$$

$$q_2 = A_0 H_1 (k_1 e^{-t_2 k_1} - k_2 e^{-t_2 k_1} + k_2 e^{-t_i k_1} - k_1 e^{-t_i k_1}) + A_0 H_2 (k_1 e^{-t_2 k_2} - k_1 e^{-t_i k_2} + k_2 e^{-t_2 k_1} - k_2 e^{-t_i k_1}) / (k_2 - k_1) \quad (21)$$

Rewriting eq 20 to make  $H_1$  the subject gives

$$H_1 = \left[ \frac{q_1 (k_2 - k_1)}{A_0} - H_2 k_2 e^{-t_1 k_1} + H_2 k_1 e^{-t_i k_1} + H_2 k_2 e^{-t_1 k_1} - H_2 k_1 e^{-t_i k_1} \right] / (-k_2 e^{-t_1 k_1} + k_1 e^{-t_i k_1} + k_2 e^{-t_1 k_1} - k_1 e^{-t_i k_1}) \quad (22)$$

Substituting eq 22 into eq 21 and rearranging for  $H_2$  gives

$$H_2 = -q_2 [(-e^{-t_i k_1} + e^{-t_1 k_1})(k_2 - k_1)] - q_1 [(k_2 - k_1) e^{-2t_i k_1} + ((k_1 - k_2) e^{-(t_1 + t_2) k_1})] / A_0 (k_2 \alpha + k_1 \beta) \quad (23)$$

where

$$\alpha = e^{-t_2 k_1} - e^{-k_1(2t_1 + t_2)} - e^{-t_i k_1} + e^{-3t_i k_1} \quad (24)$$

and

$$\beta = e^{-t_i k_1 - t_1 k_2 - t_1 k_1} + e^{-t_1 k_2 - t_1 k_1 - t_2 k_1} + e^{-2t_i k_1 - t_2 k_2} - e^{-t_2 k_2 - t_1 k_1 - t_1 k_1} - e^{-t_1 k_2 - 2t_i k_1} - e^{-t_i k_1 - t_1 k_2 - t_2 k_1} \quad (25)$$

Thus, eq 23 can be used to calculate the value of  $H_2$ . Knowledge of  $H_2$ , as before, allows  $H_1$  to be calculated from eq 14. In the special case where  $t_i = t_0$  (i.e., the time at which the reaction was initiated), eq 23 reduces to

$$H_2 = \frac{-q_2 [(e^{-t_1 k_1} - 1)(k_2 - k_1)] - q_1 [(k_2 - k_1) + (k_1 - k_2) e^{-t_2 k_1}]}{A_0 k_1 (e^{-t_1 k_1} + e^{-t_1 k_2 - t_1 k_1} + e^{-t_2 k_2} - e^{-t_2 k_2 - t_1 k_1} - e^{-t_1 k_2} - e^{-t_2 k_1})} \quad (26)$$

In most calorimetric experiments there is a time delay between the initiation of reaction and the commencement of data measurement, commensurate with the fact that ampoules are usually prepared externally from the calorimeter. However, it is also usually the case that the percentage of data lost is negligible (for long-term reactions) or that the data can be extrapolated to the initiation time (for short-term reactions). We have shown previously<sup>5,22</sup> using simulated data for single-step processes that extrapolation to  $t_0$  by fitting the measured data to a suitable exponential or polynomial equation returns a value

that is entirely consistent with that expected. In either case, eq 19 or 26 can be used to determine  $H_2$ .

## Summary

The purpose of this study was to ascertain whether PCA analysis was capable of analyzing complex data where it was difficult to construct a trivariate data matrix. By incorporating small differences in the quantity of material studied we have shown that PCA does indeed have merit. In particular, it is clear that the deconvolution process reveals the correct number of processes, although it is also apparent that the deconvoluted data do not mimic exactly those that were generated initially (they are however, very close). The reasons for this are unclear although several possibilities present themselves; the output from the chemometric deconvolution is an average of the data sets (8 in this case) inputted and hence do not exactly describe any one of them individually; another reason may be scarcity of the data. Nevertheless, the fact remains that the software allows distinction of the contributing factors to the calorimetric signal. At the very least this reduces the burden on any further kinetic/thermodynamic analysis of the data by allowing an appropriate kinetic model to be employed in an iterative analysis. From the successful deconvolution of simulated data there is an increased confidence in any results returned from data that describe a real system. Although this discussion has centered on consecutive reaction schemes, and we have proposed methods for quantitative analysis of such data, we do not envisage problems with extending the methodology to other, more complex schemes.

## References and Notes

- (1) Willson, R. J.; Beezer, A. E.; Mitchell, J. C.; Loh, W. *J. Phys. Chem. B* **1995**, 99, 7108.
- (2) O'Neill, M. A. A. *Curr. Pharm. Biotechnol.* **2005**, 6, 251.
- (3) Skaria, C. V.; Gaisford, S.; O'Neill, M. A. A.; Buckton, G.; Beezer, A. E. *Int. J. Pharm.* **2005**, 292, 127.
- (4) O'Neill, M. A. A.; Beezer, A. E.; Morris, A. C.; Urakami, K.; Willson, R. J.; Connor, J. A. *J. Therm. Anal. Calorim.* **2003**, 73, 709.
- (5) Beezer, A. E.; Morris, A. C.; O'Neill, M. A. A.; Willson, R. J.; Hills, A.; Mitchell, J. C.; Connor, J. A. *J. Phys. Chem. B* **2001**, 105, 1212.
- (6) Galwey, A. K. *Thermochim. Acta* **2003**, 407, 93.
- (7) Dias, M.; Hadgraft, J.; Raghavan, S. L.; Tetteh, J. *J. Pharm. Sci.* **2004**, 93, 186.
- (8) Brown, M. E.; Galwey, A. K.; *Thermochim. Acta* **2000**, 387, 173.
- (9) Brown, M. E.; Maciejewski, M.; Vyazovkin, S.; Nomen, R.; Sempere, J.; Burnham, A.; Opfermann, J.; Strey, R.; Anderson, H. L.; Kemmler, A.; Keuleers, R.; Janssens, J.; Desseyn, H. O.; Chao-Rui Li; Tang, T. B.; Roduit, B.; Malek, J.; Mitsunashi, T. *Thermochim. Acta* **2000**, 355, 125.
- (10) Maciejewski, M. *Thermochim. Acta* **2000**, 355, 145.
- (11) Vyazovkin, S. *Thermochim. Acta* **2000**, 355, 155.
- (12) Burnham, A. K. *Thermochim. Acta* **2000**, 355, 165.
- (13) Roduit, B. *Thermochim. Acta* **2000**, 355, 171.
- (14) Geladi, P. *Spectrochim. Acta, Part B* **2003**, 58, 767.
- (15) InSight User Manual; DiKnow Ltd.: London, 2000.
- (16) Matlab User Manual; Mathworks Inc. U.S.A., 1999.
- (17) Malinowski, E. *Factor Analysis in Chemistry*, 2nd ed.; Wiley: New York, 1991.
- (18) Tetteh, J. Enhanced Target Factor Analysis and Radial Basis Function Neural Network for Analytical Spectroscopy. Ph.D. Thesis, University of Greenwich, London, U.K., 1997.
- (19) Johnson, R. M. *Psychometrika* **1963**, 28, 259.
- (20) Gaisford, S.; Hills, A. K.; Beezer, A. E.; Mitchell, J. C. *Thermochim. Acta* **1999**, 328, 39.
- (21) Willson, R. J.; Beezer, A. E. *Thermochim. Acta* **2003**, 402, 75.
- (22) O'Neill, M. A. A. Ph.D. Thesis, University of Greenwich, London, U.K., 2002.