

COMMUNICATION

Comment on a recently proposed resampling method

Nicolaas (Klaas) M. Faber*

Dunantsingel 28, NL-2806 JB Gouda, The Netherlands

SUMMARY

A recently introduced resampling method for determining the contribution of measurement errors to the standard error in bilinear and trilinear model estimates is critically examined. Using Monte Carlo simulations, it is shown that this method does not work as intended. A noise addition method is proposed as a generally applicable alternative. The principles involved in applied work are illustrated on real UV-vis data taken from the literature. Copyright © 2001 John Wiley & Sons, Ltd.

KEY WORDS: bilinear model; trilinear model; PARAFAC; measurement errors; resampling; jackknife; Monte Carlo simulations; noise addition; UV-vis

1. INTRODUCTION

Trilinear models of the PARAFAC type are of considerable interest in analytical chemistry because the parameters are identifiable under rather general conditions. The smallest PARAFAC model is obtained when two bilinear data matrices are available that have both modes in common but differ with respect to the contribution of the different factors. The classical example arises in the context of calibration when data matrices are collected for an unknown and a calibration sample. Sánchez and Kowalski [1] developed the generalized rank annihilation method (GRAM) for analyzing this type of data and recognized that a single calibration sample suffices for obtaining a valid prediction for the unknown sample. Thus, ideally, PARAFAC models overcome the need for extensive calibration, which often hampers applications of multivariate calibration methodology.

Antalek and Windig [2,3] made the clever observation that a bilinear data matrix, for which one mode consists of exponentially decaying functions, can yield the input for methods such as GRAM by applying a shift. In other words, with exponentially decaying functions one even overcomes the need for a 'calibration sample'. This is understood as follows. Assume that the data matrix is arranged such that the decaying functions span the column space. Let the data matrix be dimensioned $J_1 \times J_2$ and let S denote the shift. Then two smaller data matrices are formed from rows 1 through $J_1 - S$ and $1 + S$ through J_1 respectively. These matrices can be seen as being shifted with respect to each other over S units. Clearly, the column profiles (with $J_1 - S$ components, instead of J_1 for the original matrix) are

* Correspondence to: ATO, Department Production & Control Systems, P.O. Box 17, 6700AA Wageningen, The Netherlands.
E-mail: n.m.faber@ato.wag-ur.nl

identical for both matrices, since a shift does not affect the shape of exponentially decaying functions. However, shifting results in a constant proportionality between the matrices, which directly depends on the applied shift. In common applications of GRAM this proportionality is an implicit measure of analyte concentration in the unknown sample, whereas in this special case it is related to the exponent.

Bijlsma *et al.* [4,5] successfully exploited this elegant principle for estimating reaction rate constants from kinetic data. The fundamental relationship for their problem reads as

$$\hat{k} = \ln(\hat{\pi}) \frac{T}{S} \quad (1)$$

where \hat{k} denotes the estimated reaction rate constant, $\ln(\cdot)$ signifies the natural logarithm, $\hat{\pi}$ is the associated proportionality and T is the sampling rate. If \hat{k} is expressed in min^{-1} , T is the number of spectra measured per minute. It is noted that T is absent in the expressions given by Bijlsma *et al.* [4,5]. The confusion arises because they use S as the time and also as the number of spectra over which the data are shifted. Here S will always denote the number of spectra to keep consistency with the computations.

For quality assessment of the obtained constants, they proposed, among others, to use a resampling procedure referred to as the jackknife. The resampling method was developed for determining the contribution of measurement errors to the standard error in the results. This quantity was defined as the *lower error limit* because only measurement errors are involved. In contrast, the *upper error limit* has a contribution from other sources of error variance (model errors and experimental errors) and can be determined from several repeated individual batch processes. This paper focuses on the method for determining the lower error limit. The lower error limit is believed to be of interest for two reasons. First, the additional sources of error variance may be relatively small or even negligible for other problems. A good example is the work of McCue and Malinowski [6], who developed a liquid chromatograph with UV detection for rank annihilation and obtained excellent results. Second, Bijlsma *et al.* [4,5] mention that concentration errors and errors due to the start of the reaction contribute to the upper error limit. However, deleting spectra for which the reproducibility between batches is low effectively eliminates these errors (see Section 4.2 in Reference [4] and Figure 5 in Reference [5]). It follows from the derivation of (1) that the absolute values of concentrations as well as reaction time are irrelevant. These quantities, hence also their uncertainty, drop out of the relevant equations when dividing the profiles. Thus the lower error limit may often be closer to the upper error limit than previously assumed.

Elsewhere, Bijlsma and Smilde [7] report estimation of the reaction rate constants by performing curve resolution on the original data matrix, which amounts to estimating parameters of a bilinear model. The resampling method has been applied in this context too. In summary, the resampling method has been proposed to quantify the contribution of measurement errors to the standard error in bilinear as well as trilinear model estimates.

It is noted that the theoretical foundation of the proposed resampling method is not clear because the current data analytical problem is characterized by the fact that one has only a single observation. However, the jackknife is a resampling technique that is developed for use in situations where several observations are available [8]. By leaving out certain observations, estimates are obtained from the remaining data. This procedure is repeated and from the spread in the results one calculates a standard error. It follows that the proposed method cannot be a genuine jackknife. To avoid confusion with the generally accepted term jackknife, in what follows, the proposed method will be referred to as the 'resampling method'.

Monte Carlo simulations are conducted to determine its performance for data that follow a prescribed model. Characteristic for the current simulations, which closely resemble the ones

performed by Bijlsma *et al.* [4], is the condition that the measurement errors are independently and identically distributed (i.i.d.). This condition may greatly simplify the practical situation where measurement errors are likely to be heteroscedastic and correlated. However, this simplification should not affect any of the conclusions. It stands to reason that a method can only be expected to work in a practical situation if it works under 'ideal' circumstances. This also holds for the alternative method that is based on adding noise to experimental data.

A real UV-vis data set is analyzed to further illustrate the principles involved when applying the noise addition method in practice. The reaction rate constants are estimated using GRAM. Other curve resolution methods investigated by Bijlsma and co-workers [4,5,7] are not considered here.

2. METHODS

2.1. Repeated data generation

In principle, repeating the entire data generation process under *rigorously defined conditions* is the best method for determining the contribution of measurement errors to the standard error in the results. It is essential that the experimental conditions be rigorously controlled, otherwise it is not possible to assess the contribution of other sources of error to the standard error. Unfortunately, this method is seldom practical because it implies a substantial increase in workload.

For the purpose of testing the adequacy of the resampling method, Monte Carlo simulations are suitable. During these simulations one generates 'true' data according to a certain model and adds artificial noise. In other words, there are no other sources of error. Each replication yields an estimate and from the spread in these estimates a standard error can be calculated. Since computer time is cheap, a large number of replications can be generated (say 10000), which results in accurate standard errors.

2.2. Resampling method

Reduced data sets are formed by deleting a *fixed* number of spectra that are separated by a certain interval. Assume that the interval is five. Then the first reduced data set is formed by deleting spectra 1, 6, 11, etc., the second one by deleting spectra 2, 7, 12, etc., and so on. The deleted groups must all be of equal size, because the perturbations of the data must be comparable. In other words, if the last group is smaller than the others, it cannot be used to obtain an additional estimate. Thus at most five reduced data sets are generated in this case. This procedure yields a number of estimates from which a standard error can be calculated as [8]

$$\hat{\sigma}(\hat{k}) = \frac{(M-1)^{1/2}}{M} \times \left(\frac{1}{M-1} \sum_{m=1}^M (\hat{k}_m - \bar{k})^2 \right)^{1/2} \quad (2)$$

where M is the number of estimates, \hat{k}_m denotes the estimate for the m th reduced data set and \bar{k} is the average of the \hat{k}_m s. The second part of the right-hand side of (2) yields the standard error in the \hat{k}_m s. The factor $(M-1)^{1/2}/M$ is required because the standard error in \hat{k} follows from the spread in so-called pseudovalues, which are postulated to be distributed as $M^{1/2}\hat{k}$ (see Reference [8] for more details).

The results reported by Bijlsma and co-workers [4,5,7] are based on the second part of the right-hand side of (2) only. Thus a correction is required to enable a comparison with the current results. A plausible explanation for this error is that it is the appropriate formula for computing the average prediction error using cross-validation, which is similar to the jackknife. Further confusion arises because they report

incorrect numbers for M . It has been explained above that M is related to the interval. However, the numbers given by Bijlsma and co-workers are equal to the number of deleted spectra.

It is emphasized that the resampling method can be implemented in two ways. First, the spectra can be deleted *before* the shift is applied, i.e. the spectra are deleted from the original data matrix. This will be referred to as mode A resampling. It is the method used by Bijlsma and Smilde [7] when performing curve resolution on the original data matrix. Second, the spectra can be deleted *after* the shift is applied, i.e. the spectra are deleted from the matrices that enter the GRAM calculation. This is the method used by Bijlsma *et al.* [4,5]; it will be referred to as mode B resampling. Bijlsma *et al.* [4,5] do not motivate why the spectra should be deleted *after* application of the shift. However, this particular implementation implies that *different* spectra are deleted from the two data matrices. In addition, the number of estimates that yields the standard error now depends on the shift S . This is easily verified by considering the case where only a single spectrum is deleted: $M = J_1$ estimates are generated in mode A versus $M = J_1 - S$ in mode B. One would expect the number of estimates to depend on the amount of original data only. Finally, different implementations of the resampling method obscure a comparison of the results reported for the same UV-vis data [5,7]. All these properties seem to be undesirable.

The main problem with the resampling method is the nature of the deleted entities. The jackknife is based on deleting observations [8]. When applied to multivariate calibration, for example, one deletes rows from the predictor matrix and the associated components of the predictand vector. The format of the predictor matrix is observations times variables ($O \times V$). However, the format of the currently resampled data matrices is variables times variables ($V \times V$). It is conceivable that the jackknife methodology can be applied in situations where several observations of $V \times V$ matrices are available, but this possibility is not pursued here.

Another way of looking at this problem is the following. A key assumption of *all* resampling methods is that the data have been generated by *random* sampling from some distribution. This condition is required to ensure *independence* of the resampled entities. In equation form,

$$\mathbf{x}_j = \boldsymbol{\mu} + \mathbf{e}_j, \quad j = 1, \dots, J_1 \quad (3)$$

where \mathbf{x}_j ($J_2 \times 1$) is the j th spectrum, $\boldsymbol{\mu}$ ($J_2 \times 1$) denotes the expected value of *all* spectra and \mathbf{e}_j ($J_2 \times 1$) is a random vector. Application of the jackknife is appropriate if $E(\mathbf{e}_j^T \mathbf{e}_{j'}) = 0$ for $j \neq j'$.

However, for the problem considered by Bijlsma and co-workers, the resampled entities are spectra that are taken at constant time intervals while the absorbing species are involved in a chemical reaction. The model for the data matrix is

$$\mathbf{X} = \mathbf{F}\mathbf{D}^T + \mathbf{E} \quad (4)$$

where \mathbf{X} ($J_1 \times J_2$) contains the recorded spectra of the reaction mixture in time, \mathbf{F} ($J_1 \times N$) contains the concentration profiles of the reacting absorbing species, \mathbf{D} ($J_2 \times N$) contains the pure spectra of the species, \mathbf{E} ($J_1 \times J_2$) is the matrix of errors and N is the number of species. The matrices \mathbf{F} and \mathbf{D} are assumed to be of full rank.

Consequently, the j th spectrum is given by

$$\mathbf{x}_j = (\mathbf{X}_{j-\text{row}})^T = \sum_{n=1}^N c_n(t_j) \mathbf{d}_n + \mathbf{e}_j, \quad j = 1, \dots, J_1 \quad (5)$$

where $c_n(t_j)$ is the concentration of species n at time j , \mathbf{d}_n is the n th column of \mathbf{D} , and $\mathbf{e}_j = (\mathbf{E}_{j-\text{row}})^T$.

Clearly, the postulated kinetic model dictates that the *true* mixture spectra $\sum c_n(t_j)\mathbf{d}_n$ ($j = 1, \dots, J_1$) are exactly related. As a matter of fact, estimating the parameter that describes this relationship is the goal of the studies of Bijlsma and co-workers [4,5,7]. Alternatively, this exact relationship among the true spectra allows one to resolve a single bilinear data matrix using the shift procedure developed by Antalek and Windig [2,3]. Finally, it is noted that the true spectra are not only dependent within a batch but also between different 'independent' batches. Bijlsma and co-workers [4,5,7] observed a high reproducibility for the spectra taken after the mixing of the chemicals was complete. This is a direct consequence of the fundamental dependence of the spectra too.

Both reviewers commented that the resampling procedure proposed by Bijlsma and co-workers [4,5,7] is appropriate if the measurement errors are independent and identically distributed across spectra within the single array. The preceding discussion shows that the true spectra are dependent. Consequently, the nature of the spectral errors is not an issue: the *measured* spectra are dependent whether the measurement errors are independent across spectra or not. In particular, they are not randomly scattered around a common expectation as required by (3).

It is important to understand why the resampling procedure proposed by Bijlsma and co-workers [4,5,7] is inappropriate, because the reasoning holds in general. It can be applied, for example, to another important class of applications, namely the resolution of single spectrochromatograms. Completely analogous to the true spectra underlying the kinetic data analyzed by Bijlsma and co-workers [4,5,7], the true spectra are coupled through the chromatography. In other words, they do not form a random sample from a distribution. Clearly, the nature of the measurement errors is immaterial.

Finally, it is noted that application of the jackknife has been reported to yield incorrect estimates of the standard errors in the eigenvalues of principal component analysis (PCA) when focus is on the influence of measurement errors (see Tables 1 and 2 in Reference [9]). Similar results were obtained for the bootstrap, which is an alternative for the jackknife. Because of this similarity, the bootstrap is not considered here.

2.3. Noise addition

Repeatedly adding noise to the experimental data can be a reasonable alternative to the ideal method described above [10]. The basic idea is that the added noise has a similar effect on the final results as the original measurement noise. This is probably a useful hypothesis if the signal-to-noise ratio is good. The noise addition method has been introduced into chemometrics by Derks *et al.* [11], who reported satisfactory results for the sensitivity analysis of artificial neural networks. When applied to the current problem, the standard errors in the rate constants are estimated as

$$\hat{\sigma}(\hat{k}) = \zeta \cdot \sigma(\hat{k}_{\text{add}}) \quad (6)$$

where ζ denotes the ratio of the noise in the data and the noise that is artificially added in, and $\sigma(\hat{k}_{\text{add}})$ is the standard error of the rate constants obtained from the noise additions.

If the noise in the data is homoscedastic (constant variance), $\zeta = \hat{\sigma}_X / \sigma_{\text{add}}$, where $\hat{\sigma}_X$ denotes the estimated standard deviation of the measurement noise in \mathbf{X} , and σ_{add} is the standard deviation of the added noise. For the heteroscedastic case the ratio ζ applies to the standard deviation of each data point. A note seems to be in order with respect to the question of how the noise in the data should be estimated. Obviously, the best way to do this is from repeated measurements. In fact, many spectrophotometers generate spectra as averages so that the necessary information is readily available. (The UV-vis spectra analyzed by Bijlsma and co-workers [4,7] constitute the average over 10 spectra.) If repeating the measurement is not feasible, one could use the method recently introduced by Wang and Hopke [12]. The key requirement is that the size of the data matrix be

sufficiently large so that the noise is homoscedastic over a limited domain. It works by calculating residuals for a principal component (PC) model and averaging them in a root mean square fashion for a certain number of neighbors of the data point under consideration. Unfortunately, the method cannot be used to determine correlation between errors. For more details see Reference [12].

Equation (6) is a linear extrapolation formula. It assumes that the system responds linearly to noise, i.e. doubling σ_{add} will double $\sigma(\hat{k}_{\text{add}})$. However, methods such as GRAM are intrinsically non-linear. It is therefore recommended to add noise at different levels to investigate whether this curvature plays a significant role.

3. EXPERIMENTAL

3.1. Simulated data

The Monte Carlo simulations were patterned after the simulations carried out by Bijlsma *et al.* [4] Data matrices were generated that emulate bilinear instrument responses for $N=2$ species whose concentration profiles are decaying functions governed by the reaction rate constants $k_1=0.30$ and $k_2=0.05$ respectively. The time range was chosen from 0 to 20 min. While Bijlsma *et al.* [4] selected an increment of 1 min, it was preferred for the current study to simulate increments of 0.2 min (i.e. $T=5$), because generating a larger number of spectra allows for obtaining a larger number of deleted groups. (This modification should favor the resampling method.) Consequently, a single data set consists of $J_1=105$ spectra instead of the 21 generated by Bijlsma *et al.* [4] The time shift parameter S was equal to five, which is five times the value used by Bijlsma *et al.* [4] The decaying functions were convoluted with the UV spectra of adenine and cytidine [13], which contained $J_2=36$ absorbances (Figure 1). Normally distributed noise (i.i.d.) was added at two levels: σ_X was equal to 1% and 2% of the maximum absorbance of the simulated spectrum at time zero. The initialization of the pseudorandom number generator was identical for both noise levels.

3.2. Real data

Only a brief description follows, because details are given elsewhere [4,7]. The data consist of spectra recorded using a Hewlett Packard 8453 UV-vis spectrophotometer with diode array detection at 25 °C for the two-step consecutive reaction of 3-chlorophenylhydrazonopropane dinitrile (A), which is an uncoupler of oxidative phosphorylation in cells, with 2-mercaptoethanol (B). Schematically, the

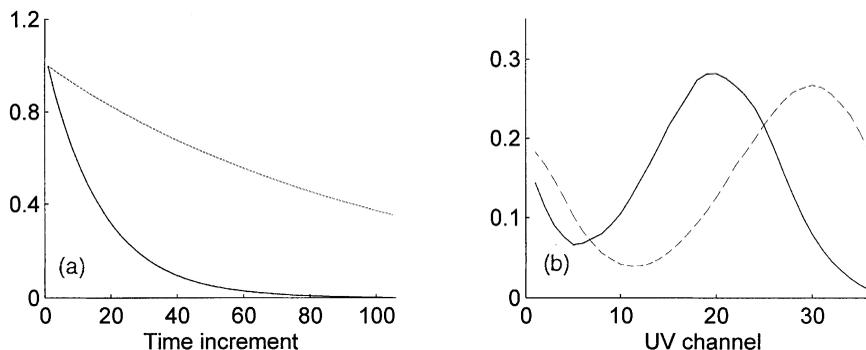


Figure 1. (a) Simulated concentration profiles and (b) real UV spectra of adenine (—) and cytidine (---).

reaction is given by



Pseudo-first-order kinetics can be assumed, because 2-mercaptoethanol is present in large excess. Six spectra are taken per minute, i.e. the sampling rate $T = 6$. The run time is 45 min, which results in a total of $J_1 = 271$ spectra. The spectra are taken between 200 and 600 nm with an interval of 1 nm. For the range below 300 nm, species B absorbs and by-product E shows an increasing absorbance. For the range above 500 nm the reacting species do not absorb significantly. Thus for the data analysis only the range between 300 and 500 nm is considered, i.e. $J_2 = 201$ (see Figure 2). Since only A, C and D absorb, $N = 3$. The experiment is repeated 10 times (at the same target temperature) to enable determining the upper error limit (see Section 1).

3.3. Calculations

All calculations were performed in Matlab, and copies of the computer programs are available on request.

4. RESULTS AND DISCUSSION

4.1. Simulated data

In most instances the resampling method and the ideal method of repeating the generation of the data yield substantially different results (Table I). The results for the resampling method are obtained by deleting 10 spectra. As a result, 10 groups are formed in mode A, after which five spectra are left. In mode B, 10 groups are formed without leaving spectra 'unused'. It is seen that, for this specific

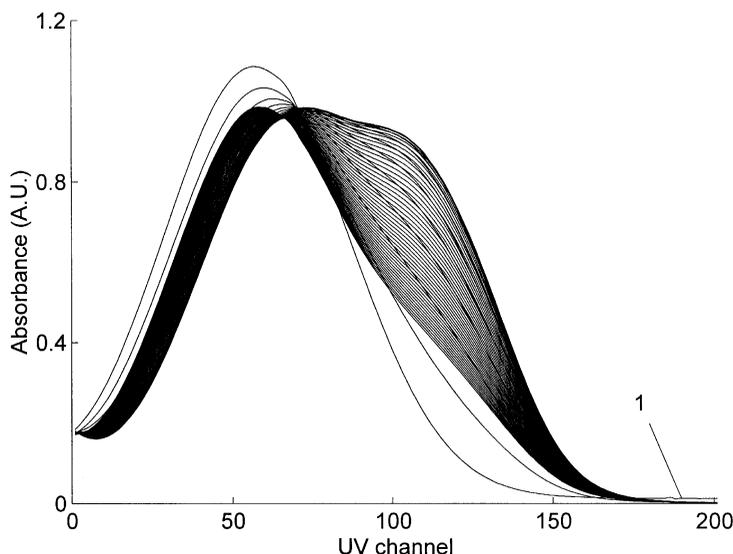


Figure 2. Blank-corrected UV-vis spectra for batch run 1. Only every fifth spectrum is shown for visual clarity. The poor reproducibility of the first spectrum is manifest: it does not reach the baseline in the high-wavelength region.

Table I. Standard errors in the reaction rate constants. The symbols are explained in the text

σ_x (%)	Reaction	\hat{k} (min^{-1})	Repetition	Standard error in \hat{k} (min^{-1})				Noise addition (σ_{add} in %)									
				Resampling spectra		Mode B		0.2		0.4		0.6		0.8		1.0	
				Mode A	Mode B	Mode A	Mode B	Mode A	Mode B	Mode A	Mode B	Mode A	Mode B	Mode A	Mode B	Mode A	Mode B
1	1	0.297	2.93×10^{-3}	2.33×10^{-2}	7.45×10^{-3}	2.93×10^{-3}	2.92×10^{-3}	2.91×10^{-3}	2.92×10^{-3}	2.92×10^{-3}	2.91×10^{-3}	2.92×10^{-3}	2.92×10^{-3}	2.94×10^{-3}			
	2	0.0495	6.14×10^{-4}	1.21×10^{-3}	7.36×10^{-4}	6.28×10^{-4}	6.28×10^{-4}	6.34×10^{-4}	6.28×10^{-4}	6.36×10^{-4}	6.34×10^{-4}	6.36×10^{-4}	6.36×10^{-4}	6.41×10^{-4}			
2	1	0.289	5.81×10^{-3}	2.00×10^{-2}	1.52×10^{-2}	5.91×10^{-3}	5.91×10^{-3}	5.90×10^{-3}	5.91×10^{-3}	5.91×10^{-3}	5.90×10^{-3}	5.91×10^{-3}	5.96×10^{-3}				
	2	0.0477	1.34×10^{-3}	1.99×10^{-3}	1.60×10^{-3}	1.42×10^{-3}	1.42×10^{-3}	1.44×10^{-3}	1.42×10^{-3}	1.43×10^{-3}	1.44×10^{-3}	1.43×10^{-3}	1.45×10^{-3}				

example, mode A and mode B yield the same number of estimates. In most cases the true standard error is severely overestimated. An exception is formed by the mode B results obtained for \hat{k}_2 , which are quite close. However, this agreement is believed to be a coincidence. After all, the results are off by 150% for \hat{k}_1 . The mode A results exhibit a peculiar behavior: the standard error for \hat{k}_1 decreases upon increasing the noise!

In contrast, the noise addition method seems to work well. The results listed in the last five columns of Table I are obtained using 10000 additions of normally distributed noise (i.i.d.) with σ_{add} equal to 0.2%, 0.4%, 0.6%, 0.8% and 1% of the maximum absorbance of the simulated spectrum at time zero. At the lowest level of the noise ($\sigma_X = 1\%$) the results are in excellent agreement with the ideal values for \hat{k}_1 , which is the largest rate constant. Moreover, these results seem to be independent of the standard deviation of the added noise. The results obtained for \hat{k}_2 show that such an excellent agreement cannot be expected in general: they are slightly biased upwards. This effect can be explained from a violation of the underlying assumption of this method, namely that adding noise to the 'experimental' data gives rise to the same fluctuations in the final results as adding noise to the errorless data. The analogy with simple univariate straight-line fitting may help to understand what is going on. Owing to estimation errors, the experimental line will not coincide with the true line. Depending on the slope being smaller or larger than the true slope, the experimental line will be more or less sensitive to noise than the true line. As a result, the noise addition method will either over- or underestimate the true prediction error. This analogy further illustrates that the currently obtained overestimation of the true error is certainly not a general rule. Moreover, this bias depends on the standard deviation of the added noise. The latter effect can be understood from the non-linear character of GRAM. Only in a small neighborhood around the true data (i.e. for low noise levels) is a linear approximation effective. However, the observed dependence on σ_{add} is weak enough for the results to be trustworthy. At the highest level of the noise ($\sigma_X = 2\%$) the noise addition method slightly overestimates the standard error for both reaction rate constants.

Recalling the extensive discussion in Section 2.2, the poor results obtained for the resampling method should not come as a surprise. The key assumption, namely independence among the resampled entities, is violated. The dependence among the spectra is conveniently visualized in a score plot (Figure 3). It is observed that the spectra are dependent within a data matrix as well as between different 'independently' generated data matrices.

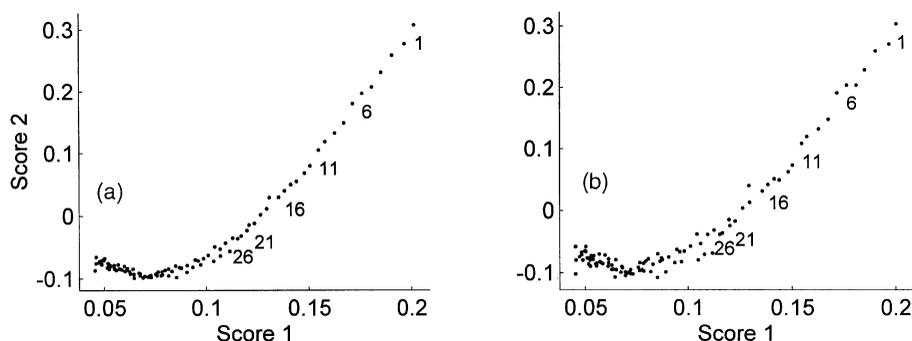


Figure 3. Score 1 versus score 2 at two noise levels: (a) $\sigma_X = 1\%$ and (b) $\sigma_X = 2\%$. The positions of six spectra (1, 6, 11, 16, 21 and 26) are labeled to illustrate that the spectra are coupled through the chemical reaction, hence dependent.

4.2. Real data

Bijlsma *et al.* [4] show that for the reaction depicted in (7) the resulting data matrix can be expressed as

$$\begin{aligned} \mathbf{X} &= \exp(-k_1 \mathbf{t})[\mathbf{d}_A + k\mathbf{d}_C - (k+1)\mathbf{d}_D]^T + \exp(-k_2 \mathbf{t})(-k\mathbf{d}_C + k\mathbf{d}_D)^T + \exp(-0\mathbf{t})(\mathbf{d}_D)^T \\ &= \mathbf{f}_1 \mathbf{d}_1^T + \mathbf{f}_2 \mathbf{d}_2^T + \mathbf{f}_3 \mathbf{d}_3^T + \mathbf{E} \\ &= \mathbf{F}\mathbf{D}^T + \mathbf{E} \end{aligned} \quad (8)$$

where $\exp(\cdot)$ signifies the exponential function, \mathbf{t} is a vector that contains the time points, \mathbf{d}_A , \mathbf{d}_C and \mathbf{d}_D denote the spectra of A, C and D respectively and $k = k_1/(k_2 - k_1)$. It is noted that Equation (12) in Reference [4] is slightly different from Equation (8) in that it includes an additional column of ones. The latter has been reported to be advantageous in the case of noisy data [14,15]. GRAM calculates the proportionality $\hat{\pi}$ in (1) from an eigenvalue problem. Given that $k_1 > k_2 > 0$ (see below), one finds the reaction rate constants in the same order as the eigenvalues. Hence results associated with the two largest eigenvalues are reported.

The remainder of this subsection is organized as follows. First we determine the optimum shift parameter S and the optimum number of spectra to be deleted. Next the reaction rate constants are estimated under these 'optimal' conditions. Then we determine the contribution of the measurement errors to the standard error. It is found that the standard error for \hat{k}_2 is too small to explain the difference from the value obtained by Bijlsma and Smilde [7] using their algorithm 2. It is known that random measurement errors induce a bias in the GRAM eigenvalues [16,17]. For this reason we also consider the contribution of the measurement errors to the bias in the reaction rate constants.

4.2.1. Determining the optimum shift parameter. Bijlsma *et al.* [5] determine the optimum shift parameter by calculating the means \bar{k}_1 and \bar{k}_2 and the standard errors $\hat{\sigma}(\hat{k}_1)$ and $\hat{\sigma}(\hat{k}_2)$ from the 10 repeated batch runs while varying S . The optimum value minimizes the standard errors. The first spectrum is deleted from all data sets because of the poor reproducibility (see Figure 2).

The current results deviate considerably from the numbers reported by Bijlsma *et al.* [5] (see Table 2).* The optimum value for S now follows as 35 instead of 30. It is noted in passing that Bijlsma *et al.* [5] select this value as a compromise, "because this gives the lowest standard deviation for k_1 and a compromise in standard deviation for k_2 ". However, there is no reason why one should use the same shift for both parameters. The current optimum value of 35 is not a compromise, but such a result cannot be expected in general. If different optimum values are found, it is recommended to perform different GRAM calculations. More importantly, however, are the substantially lower values for the standard error: for \hat{k}_1 it is reduced by 35% while for \hat{k}_2 it is even lower by a factor of four. The current results do not support the rather strong conclusion of Bijlsma and Smilde [7] that "From the results, it can be concluded that GRAM is not applicable if precise estimates for the reaction rate constants are

* It is somewhat disturbing that the results of Bijlsma *et al.* could not be reproduced. The reason for this may be that several variations of GRAM exist. In the original formulation of GRAM [1] a standard eigenvalue problem (SEP) is solved, whereas Wilson *et al.* [18] have proposed to solve a generalized eigenvalue problem (GEP) since this would be numerically more stable. In addition, several variations exist to implement these methods. It is not clear which variation is used by Bijlsma *et al.*, because their description involves solving a GEP while they refer to the work of Sánchez and Kowalski [1]. The current results are based on solving an SEP. Obviously, application of a different variation of GRAM does not affect any of the conclusions concerning the resampling method.

Table II. Mean and standard error for reaction rate constants obtained using various time shifts. The numbers in parentheses are copied from Table 2 in Reference [5]. The numbers printed bold indicate the optimal time shift. The symbols are explained in the text

S	\bar{k}_1 (min ⁻¹)	$\hat{\sigma}(\hat{k}_1)$ (min ⁻¹)	\bar{k}_2 (min ⁻¹)	$\hat{\sigma}(\hat{k}_2)$ (min ⁻¹)
5	0.3107 (0.3050)	0.0099 (0.0104)	0.0277 (0.0291)	0.00047 (0.0023)
10	0.3101 (0.3047)	0.0085 (0.0087)	0.0279 (0.0291)	0.00040 (0.0015)
15	0.3088	0.0076	0.0281	0.00035
20	0.3070 (0.3052)	0.0068 (0.0083)	0.0283 (0.0286)	0.00033 (0.0011)
25	0.3049	0.0061	0.0285	0.00030
30	0.3022 (0.3040)	0.0053 (0.0083)	0.0286 (0.0284)	0.00027 (0.0010)
35	0.2993	0.0046	0.0287	0.00025
40	0.2960 (0.3019)	0.0051 (0.0091)	0.0288 (0.0282)	0.00025 (0.0010)
45	0.2931	0.0052	0.0289	0.00028
50	0.2901 (0.3006)	0.0051 (0.0093)	0.0289 (0.0280)	0.00032 (0.0009)

desirable. On account of the non-iterative character of the algorithm, it leads to rough estimates of the reaction rate constants”.

Finally, it is noted that adding a column of ones had no palpable effect on the estimated reaction rate constants. In the remainder of the paper, only results are presented where this form of data pretreatment was omitted.

4.2.2. *Determining the optimum number of spectra to be deleted.* Bijlsma *et al.* [5] delete a single spectrum for all 10 individual batch runs. This decision was based on a low relative reproducibility. However, by examining the first non-significant singular vectors, one may arrive at different numbers (see Figure 4). Recall that $N=3$. Thus, ideally, one does not expect to find structure in the singular vectors associated with PCs 4 and 5. However, it is clear from Figures 4(a) and 4(c) that the first spectra have a background problem. In addition, Figures 4(a) and 4(b) suggest that the noise between spectra is heteroscedastic but uncorrelated. Finally, Figure 4(d) illustrates that the noise between wavelengths is heteroscedastic and correlated. It is unclear why Bijlsma and Smilde [7] report that “the fourth singular value did not show any structure”.

Deleting additional spectra may improve the fit of a PC model to the data, which is equivalent to improving the bilinear character of the data or, alternatively, the capability of GRAM to resolve them. The fit error is expressed as residual standard deviation (RSD) and calculated as

$$\text{RSD}_a = \left(\frac{\sum_{p=a+1}^A \hat{\lambda}_p}{(J_1 - a)(J_2 - a)} \right)^{1/2}, \quad a = 1, \dots, A \quad (9)$$

where a is the number of PCs extracted from the data, $\hat{\lambda}_p$ denotes the p th eigenvalue and A is the minimum of J_1 and J_2 . The numbers in Table III demonstrate that the fit can be markedly improved by deleting more than one spectrum: it seems that for batch run 1 it is best to delete three spectra. This reduces the average fit error obtained when deleting one spectrum by 28%, while deleting more spectra achieves little additional gain. Similarly, one finds that the optimum numbers of spectra to be deleted for batch runs two through 9 are one, one, one, two, two, one, one, two and two respectively.

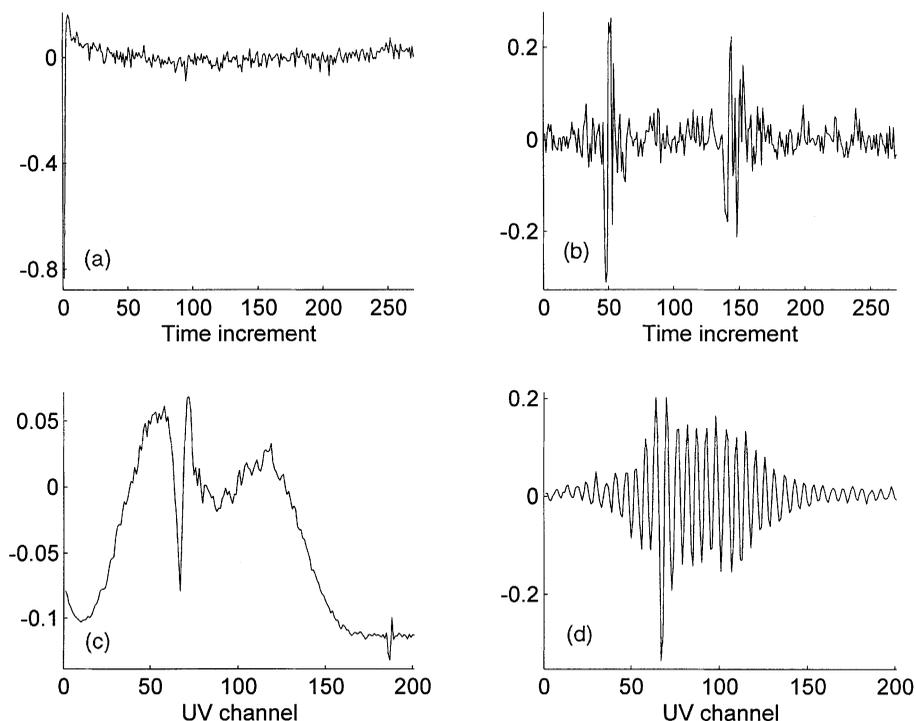


Figure 4. (a) Score vector 4, (b) score vector 5, (c) loading vector 4 and (d) loading vector 5 obtained for batch run 1 when only the first spectrum is deleted.

It is stressed that deleting additional spectra does not improve the standard errors given in Table II. On the contrary, the standard errors for \hat{k}_1 and \hat{k}_2 have increased from 0.0046 and 0.00025 to 0.0048 and 0.00027 respectively. This is believed to be a chance effect. The explanation for this and other seemingly contradictory results reported below is simple: the uncertainties associated with the data analysis give a negligible contribution to the standard errors listed in Table II (see Section 4.2.4). In other words, the data are of such excellent quality that data pretreatment has a marginal effect. Consequently, the resulting standard errors vary in a random fashion. This is certainly not a general

Table III. RSD as a function of the number of extracted PCs and spectra deleted for batch run 1. The numbers printed bold denote the fit error when equating the number of extracted PCs to the number of independent spectrally active constituents (N)

PCs extracted	Spectra deleted					
	0	1	2	3	4	5
1	6.773×10^{-2}	6.667×10^{-2}	6.568×10^{-2}	6.487×10^{-2}	6.420×10^{-2}	6.368×10^{-2}
2	5.012×10^{-3}	4.838×10^{-3}	4.644×10^{-3}	4.442×10^{-3}	4.237×10^{-3}	4.033×10^{-3}
3	3.865×10^{-4}	1.525×10^{-4}	1.150×10^{-4}	1.098×10^{-4}	1.098×10^{-4}	1.097×10^{-4}
4	1.038×10^{-4}	1.030×10^{-4}	1.026×10^{-4}	1.022×10^{-4}	1.022×10^{-4}	1.021×10^{-4}
5	9.739×10^{-5}	9.652×10^{-5}	9.614×10^{-5}	9.566×10^{-5}	9.562×10^{-5}	9.553×10^{-5}
6	9.325×10^{-5}	9.233×10^{-5}	9.194×10^{-5}	9.143×10^{-5}	9.138×10^{-5}	9.127×10^{-5}

Table IV. Results for individual batch runs. The numbers in parentheses are copied from Table 3 in Reference [5]. The symbols are explained in the text

Batch run	Deleted spectra	RSD ($\times 10^{-4}$)	\hat{k}_1 (min^{-1})	\hat{k}_2 (min^{-1})
1	3	1.098	0.2977 (0.3006)	0.0288 (0.0287)
2	1	1.259	0.2996 (0.3009)	0.0286 (0.0288)
3	1	1.201	0.3014 (0.3040)	0.0286 (0.0285)
4	1	1.155	0.3042 (0.3137)	0.0290 (0.0280)
5	2	1.102	0.2979 (0.3069)	0.0286 (0.0265)
6	2	1.200	0.2949 (0.2932)	0.0286 (0.0298)
7	1	1.060	0.2997 (0.2985)	0.0283 (0.0297)
8	1	1.611	0.3064 (0.3180)	0.0293 (0.0275)
9	2	1.177	0.2947 (0.3109)	0.0285 (0.0277)
10	2	1.006	0.2897 (0.2936)	0.0289 (0.0285)

result. Moreover, deleting additional spectra may be critical for obtaining dependable noise estimates (see Table III), and having dependable noise estimates is essential for correct application of the noise addition method.

4.2.3. *Estimating the reaction rate constants.* Applying the optimum shift and the optimum number of deleted spectra yields the results presented in Table IV. It is observed that the RSD varies more than one would expect on the basis of the degrees of freedom, $(J_1 - N)(J_2 - N) \approx 54000$. Making the common χ^2 assumption for the squared RSDs, one finds that they are significantly different in an F -test. The reason for this large variation is that the residuals are not entirely random: there are small substructures to be observed which are believed to be caused by a partially successful correction for the blank. This is particularly evident for batch run 8 (see Figure 5), which has an RSD value that greatly exceeds the value for the other batch runs. In contrast, the substructures are rather small for batch run 1 (see Figure 6).

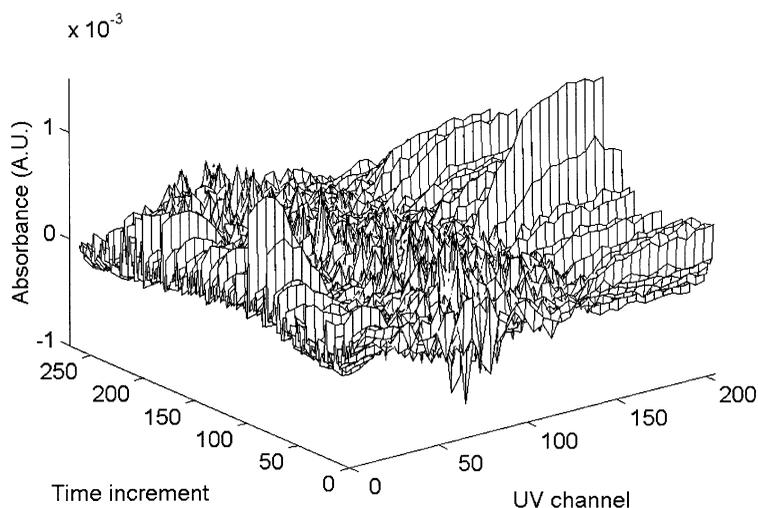


Figure 5. Residuals left after extracting three PCs for batch run 8. The first spectrum is deleted prior to performing PCA. Only every fifth UV channel is shown for visual clarity.

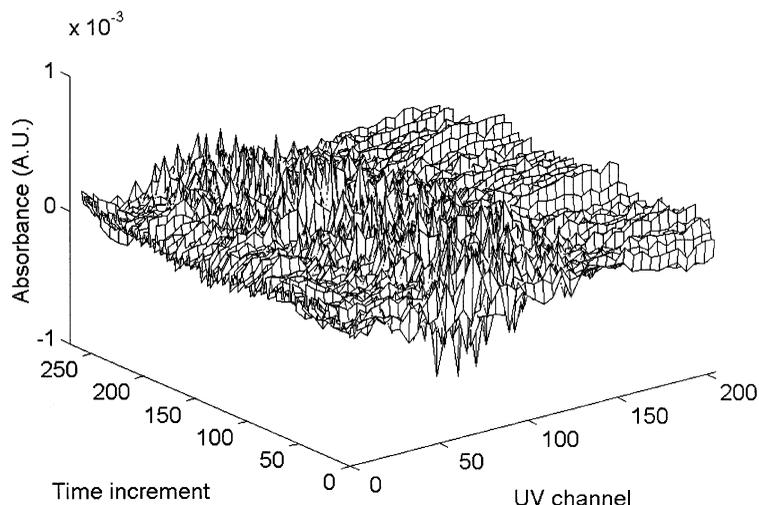


Figure 6. Residuals left after extracting three PCs for batch run 1. The first three spectra are deleted prior to performing PCA. Only every fifth UV channel is shown for visual clarity.

The difference from the numbers reported by Bijlsma *et al.* [5] is substantial. We believe that the currently obtained values are to be preferred because they are more precise.

4.2.4. *Estimating the contribution of measurement errors to the standard error in the reaction rate constant estimates.* Standard errors calculated using various procedures are summarized in Table V for batch run 1. The numbers obtained by repetition should estimate the *upper error limit*, while the *lower error limit* is targeted by resampling and adding noise. The numbers reported for mode A and mode B resampling are based on 26 and 23 estimates respectively. The numbers obtained by mode A resampling are twice as high as the numbers obtained by repeating the entire experiment. Thus they can hardly be seen as a lower error limit. The mode B results seem reasonable in that they are smaller than the upper error limit. However, this is a misleading result, which is evident from a comparison with the numbers obtained by adding noise. The ratio between the standard errors is 6.1 for \hat{k}_1 , whereas for \hat{k}_2 a ratio of 3.1 is found. This simple calculation reveals an internal inconsistency.

The noise is added in two ways. Inspection of the residuals has demonstrated that the noise is heteroscedastic (see Figures 4–6). Plotting the data points versus the absolute residuals gives further

Table V. Standard errors in the reaction rate constants obtained for batch run 1. The symbols are explained in the text

Reaction	\hat{k} (min^{-1})	Standard error in \hat{k} (min^{-1})				
		Repetition	Resampling spectra		Noise addition	
			Mode A	Mode B	Heteroscedastic	Homoscedastic
1	0.2977	4.83×10^{-3}	1.07×10^{-2}	1.20×10^{-3}	1.97×10^{-4}	1.52×10^{-4}
2	0.0288	2.71×10^{-4}	5.16×10^{-4}	8.99×10^{-5}	2.88×10^{-5}	2.61×10^{-5}

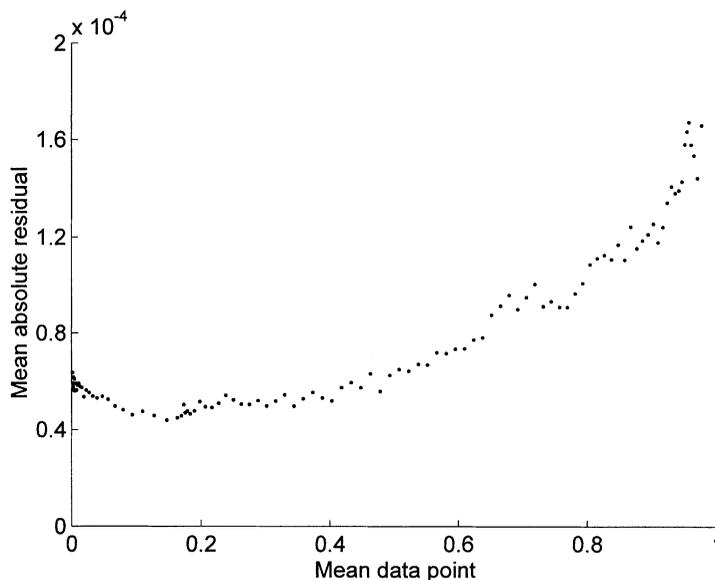


Figure 7. Mean data point versus mean absolute residual left after extracting three PCs for batch run 1. The first three spectra are deleted prior to performing PCA.

information about the nature of the heteroscedasticity (see Figure 7). To avoid an overcrowded plot, 500 data points are averaged together to yield a single point. It is seen that the noise increases with absorbance. Surprisingly, it is also relatively large for small absorbance values. An explanation for the latter observation could be the relatively poor correction for the blank at these wavelengths (compare Figures 2 and 5). These observations imply that adding homoscedastic noise would be an oversimplification here. The standard deviations of the individual data points are estimated using the method proposed by Wang and Hopke [12]; they are based on the residuals of 40 neighbors of the data point under consideration. Adding heteroscedastic noise yields the numbers presented in column 6 of Table V. These constitute an average for the results obtained for 100 noise additions at five levels, namely 0.2, 0.4, 0.6, 0.8 and 1.0 times the noise already present. The average is reported because no trends were observed. These numbers can be compared with the results obtained when adding homoscedastic noise with the standard deviation given in column 3 of Table IV. It is seen that the latter numbers are somewhat optimistic, but perhaps useful as a crude approximation. The noise addition results are based on sampling from a normal distribution. To investigate the effect of distribution, the computations were repeated with sampling from a uniform distribution. The results are very similar (not shown).

The main observation from Table V is that the standard errors obtained by noise addition are much smaller than the standard errors obtained by repeating the entire experiment. In other words, the uncertainty due to data analysis is much smaller than the uncertainty due to other sources of error. A plausible explanation for the relatively large *upper error limit* is a variation in the temperature. The temperature is a controlled variable: "a Pt-100 and a constant-temperature bath (Neslab) were used to control the temperature inside the cuvette" [5,7]. Unfortunately, Bijlsma and co-workers do not report

* The true temperature is a random variable, because the target temperature is fixed at 25°C.

information about the uncertainty in the true temperature.* In addition, they have not performed experiments at different target temperatures to investigate the sensitivity of the results to the expected fluctuation in the true temperature. It is unclear why they performed 10 experiments at the same target temperature. This procedure seems to be rather wasteful if a detailed error analysis is to be made.

It may come somewhat as a surprise that the uncertainty in a controlled variable (temperature) could have more effect on the results than the uncertainty in a random one (instrument responses). However, one should realize that the results are based on approximately 54 000 data points, and each of them has a low noise level. Under these circumstances, performing an experiment under *rigorously defined conditions* amounts to putting very tight limits on variables that are known to have an effect.

Adding noise at high levels supports the conjecture that the measurement errors give a small contribution to the upper error limit (see Table VI). The added noise is homoscedastic and normally distributed (i.i.d.). The reason for adding i.i.d. noise is that it allows for a simple interpretation in terms of analytical figures of merit (see below). It is emphasized that correlation is neglected during the noise additions that have led to the results in Table V, since the method of Wang and Hopke [12] only accounts for heteroscedasticity. However, by adding noise at high levels, one arrives at a region where the additional noise overwhelms the heteroscedastic, correlated noise that is already present. This is illustrated in Figure 8, where the standard deviation of the added noise is $\sigma_{\text{add}} = 5 \times 10^{-4}$, i.e. approximately five times the RSD: the mean absolute residual is at best weakly related to the value of the data point. The RSD found after adding the noise is 5.15×10^{-4} , which is close to the expected value of $\sqrt{(1.098)^2 + 5^2} \times 10^{-4} = 5.12 \times 10^{-4}$. The average of the mean absolute residuals is 4.19×10^{-4} , which is close to the expected value for the normal distribution, namely 0.80 times the standard deviation. One observes irregularities for small values of the data, but these can be attributed to the blank correction predominantly affecting small data values (see above). Even at this high level of the noise the upper error limit has changed very little. It has even decreased for \hat{k}_2 , but this is surely a chance effect.

The preceding results can be interpreted in terms of analytical figures of merit. These figures of merit are summary statistics that relate the quality of the data, i.e. the amount of overlap between the columns of **F** and **D** and the level of the noise, to the uncertainty in the results. Faber *et al.* [19] have given the required formulae for bilinear data. Important figures of merit are sensitivity, selectivity, signal-to-noise ratio (SNR) and limit of detection. In a rather crude approximation the SNR is inversely proportional to the relative error in the reaction rate constant [20,21]. With i.i.d. errors in the

Table VI. Mean and standard error for reaction rate constants obtained after adding noise to the data of batch run 1. The symbols are explained in the text

$\sigma_{\text{add}} (\times 10^{-4})$	$\bar{k}_1 (\text{min}^{-1})$	$\hat{\sigma}(\hat{k}_1) (\text{min}^{-1})$	$\bar{k}_2 (\text{min}^{-1})$	$\hat{\sigma}(\hat{k}_2) (\text{min}^{-1})$
0	0.2986	0.0048	0.0287	0.00027
1	0.2986	0.0049	0.0287	0.00027
2	0.2985	0.0049	0.0287	0.00027
3	0.2983	0.0048	0.0288	0.00029
4	0.2979	0.0048	0.0288	0.00030
5	0.2972	0.0048	0.0289	0.00025
6	0.2971	0.0054	0.0289	0.00024
7	0.2968	0.0055	0.0289	0.00027
8	0.2960	0.0047	0.0291	0.00036
9	0.2960	0.0049	0.0291	0.00033
10	0.2948	0.0054	0.0292	0.00038

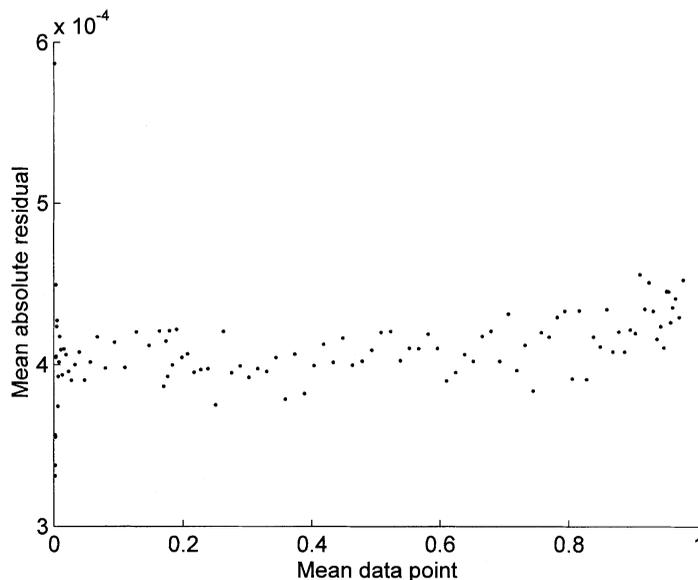


Figure 8. Mean data point versus mean absolute residual left after extracting three PCs for batch run 1 to which normally distributed noise with standard deviation $\sigma_{\text{add}} = 5 \times 10^{-4}$ is added. The first three spectra are deleted prior to performing PCA.

data it is defined as

$$\text{SNR} = \frac{\text{NAS}}{\sigma_X} \quad (10)$$

where NAS stands for net analyte signal. This simple formula is applicable in the region where the additional i.i.d. noise overwhelms the heteroscedastic, correlated noise that is already present. It is noted that the term net analyte signal is introduced in the context of analyte determination. Here the quantities of interest are reaction rate constants. Thus NAS should be seen as just an algebraic quantity that leads to an SNR for each reaction rate constant. The SNRs associated with \hat{k}_1 and \hat{k}_2 are 233 and 1350 respectively after adding noise with standard deviation $\sigma_{\text{add}} = 5 \times 10^{-4}$. Consequently, even at this severely exaggerated noise level the relative errors due to the noise in the data are expected to be as small as 0.4% and 0.07%. Comparing these numbers with the upper error limits of $100 \times 0.0048/0.2977 = 1.6\%$ and $100 \times 0.00027/0.0288 = 0.9\%$, one must conclude that the original measurement noise has a small effect indeed.*

Deleting wavelengths at constant intervals further supports the conjecture that the measurement errors give a small contribution to the upper error limit (see Table VII). It is seen that the loss in precision is marginal if as few as five wavelengths are retained.

* Assuming that the temperature causes the difference between the lower and upper limits (i.e. pH and other variables have a negligible effect) leads to an alternative interpretation of the lower error limit: it can be seen as the upper error limit at an unknown true temperature.

Table VII. Mean and standard error for reaction rate constants obtained after skipping wavelengths. The symbols are explained in the text

Wavelengths skipped	J_2	\bar{k}_1 (min ⁻¹)	$\hat{\sigma}(\hat{k}_1)$ (min ⁻¹)	\bar{k}_2 (min ⁻¹)	$\hat{\sigma}(\hat{k}_2)$ (min ⁻¹)
0	201	0.2986	0.0048	0.0287	0.00027
5	34	0.2981	0.0050	0.0288	0.00028
10	19	0.2981	0.0048	0.0288	0.00027
15	13	0.2976	0.0046	0.0288	0.00034
20	10	0.2981	0.0053	0.0288	0.00028
25	8	0.2975	0.0047	0.0289	0.00033
30	7	0.2971	0.0054	0.0290	0.00048
35	6	0.2933	0.0053	0.0296	0.00070
40	5	0.2989	0.0049	0.0286	0.00035

4.2.5. *Estimating the contribution of measurement errors to the bias in the reaction rate constant estimates.* Using their algorithm 2, Bijlsma and Smilde find the following average values and standard errors in the individual estimates for the 10 batch runs: $\bar{k}_1 = 0.3164$, $\bar{k}_2 = 0.0257$, $\hat{\sigma}(\hat{k}_1) = 0.0098$ and $\hat{\sigma}(\hat{k}_2) = 0.00017$ (see Table 3 in Reference [7]). They state that “the estimated reaction rate constants obtained with other algorithms for this batch process in the literature [21,23–25] are within the 95% confidence limit”. However, this is not true for the currently obtained value $\bar{k}_2 = 0.0287$, because the standard error $\hat{\sigma}(\hat{k}_2) = 0.00027$ is smaller than the old value by a factor of four. An obvious explanation for this discrepancy could be that the GRAM estimates are severely biased. (Bias has been completely neglected in the previous discussion.) This assertion is easily verified using noise addition too. In the current context the assumption is that adding noise will induce a similar bias in the estimates as the original measurement noise. Then adding noise at increasing levels will yield an increasing bias. This increasing bias causes a trend in the mean of the estimates that can be extrapolated back to the case where no noise is added. When applied to the quantification of bias, the noise addition method is known as simulation extrapolation (SIMEX) [22].

Table VIII gives the results obtained for batch run 1 by adding heteroscedastic normally distributed noise at 100% and 200% of the original noise respectively. The addition of homoscedastic noise gave similar results, and drawing the pseudorandom numbers from a uniform distribution had little effect as well (not shown). The so-called naive estimate is the estimate obtained without paying attention to bias; it is the estimate reported above. The average values listed in the next columns show a trend: they suggest that the naive estimate \hat{k}_1 is slightly biased low, while the naive estimate \hat{k}_2 is slightly biased upwards. (This explains the trends observed for the average values given in Table VI when adding high levels of noise.) Faber *et al.* [17] have derived that the bias in the GRAM eigenvalues

Table VIII. SIMEX results for batch run 1. The symbols are explained in the text

Reaction	\hat{k} (naive)	\bar{k} (100%)	\bar{k} (200%)	$\hat{B}(\hat{k})$ (SIMEX)	\hat{k} (SIMEX)
1	0.29769	0.29763	0.29744	-0.00006	0.29775
2	0.028845	0.028852	0.028875	+0.000008	0.028837

scales up with the variance of the measurement noise when the errors are i.i.d. Using a linear approximation, it can be shown that

$$B(\hat{k}) \approx B(\hat{\pi}) \frac{T}{\pi S} \quad (11)$$

where $B(\cdot)$ signifies the bias in the associated quantity.

Thus doubling the standard deviation of the noise should quadruple the bias in the reaction rate constants. Assuming for simplicity that this behavior also holds for heteroscedastic noise, one arrives at the estimates for bias given in column 5. (This assumption is not unreasonable when considering the general agreement between the results obtained when adding homoscedastic and heteroscedastic errors.) The SIMEX estimates in the last column follow by subtracting this bias from the naive estimates. It is seen that the (estimated) bias is very small. Thus it is believed that the rather crude assumption about the extrapolant function allows the conclusion that the current value for k_2 is not in agreement with the value reported by Bijlsma and Smilde [7]. Not being able to explain this discrepancy presents an obvious problem. It is conceivable that the value reported by Bijlsma and Smilde can be improved, as has been done here for their GRAM results, but pursuing this matter is considered to be outside the scope of this paper.

5. CONCLUSIONS

The resampling method has performed poorly, whereas the results obtained by the noise addition method seem to be promising. The analysis of simulated and real data has demonstrated that deleting spectra perturbs the data quite differently than the addition of measurement errors. Clearly, correct assessment of the uncertainty in the results requires that the perturbation should emulate the errors that are inherent to the data-generating process.

An interesting aspect of the noise addition method is that its implementation in more complex situations (multiblock, multiway, etc.) is fairly straightforward. However, the term 'straightforward' should by no means be interpreted as 'simple', because the method assumes that an adequate characterization of the measurement errors is possible. An adequate characterization implies that variances as well as covariances should be determined. This requirement may constitute a major challenge for practical chemometricians.

Finally, it is concluded that Monte Carlo simulations are highly convenient for testing the properties of a newly developed method. Unfortunately, chemometricians often tend to limit themselves to stressing the practical utility of their methods, and the added value of these 'computer experiments' is not always recognized.

REFERENCES

1. Sánchez E, Kowalski BR. Generalized rank annihilation factor analysis. *Anal. Chem.* 1986; **58**: 496–499.
2. 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**: 10331–10332.
3. Windig W, Antalek B. Direct exponential curve resolution algorithm (DECRA): A novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying contribution profiles. *Chemometrics Intell. Lab. Syst.* 1997; **37**: 241–254.
4. Bijlsma S, Louwse DJ, Windig W, Smilde AK. Rapid estimation of rate constant using on-line SW-NIR and trilinear models. *Anal. Chim. Acta* 1998; **376**: 339–355.
5. Bijlsma S, Louwse DJ, Smilde AK. Estimating rate constants and pure UV-VIS spectra of a two-step reaction using trilinear models. *J. Chemometrics* 1999; **13**: 311–329.
6. McCue M, Malinowski ER. Rank annihilation factor analysis of unresolved LC peaks. *J. Chromatogr. Sci.* 1983; **21**: 229–234.

7. Bijlsma S, Smilde AK. Application of curve resolution based methods to kinetic data. *Anal. Chim. Acta* 1999; **396**: 231–240.
8. Shao J, Tu D. *The Jackknife and Bootstrap*. Springer: New York, 1995.
9. Faber NM, Buydens LMC, Kateman G. Standard errors in the eigenvalues of a cross product matrix: theory and applications. *J. Chemometrics* 1993; **7**: 495–526.
10. Press WH, Flannery BP, Teukolski SA, Vetterling WT. *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press: Cambridge, 1988.
11. Derks EPPA, Sánchez Pastor MS, Buydens LMC. Robustness analysis of radial base function and multi-layered feed-forward neural network models. *Chemometrics Intell. Lab. Syst.* 1995; **28**: 49–60.
12. Wang J-H, Hopke PK. Estimation of the heteroscedastic noise in large data arrays. *Anal. Chim. Acta* 2000; **412**: 177–184.
13. Zscheile FP, Murray HC, Baker GA, Peddicord RG. Instability of linear systems derived from spectrophotometric analysis of multicomponent systems. *Anal. Chem.* 1962; **34**: 1776–1782.
14. 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. Res.* 1998; **132**: 298–306.
15. 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. Res.* 1998; **132**: 307–315.
16. Booksh K, Kowalski BR. Error analysis of the generalized rank annihilation method. *J. Chemometrics* 1994; **8**: 45–63.
17. Faber NM, Buydens LMC, Kateman G. Generalized rank annihilation method. II. Bias and variance in the estimated eigenvalues. *J. Chemometrics* 1994; **8**: 181–203.
18. Wilson BE, Sánchez E, Kowalski BR. An improved algorithm for the generalized rank annihilation method. *J. Chemometrics* 1989; **3**: 493–498.
19. Faber K, Lorber A, Kowalski BR. Analytical figures of merit for tensorial calibration. *J. Chemometrics* 1997; **11**: 419–461.
20. Lorber A. Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem.* 1986; **59**: 1167–1172.
21. Faber NM. Crude assessment of the contribution of measurement errors to prediction error when using a multiway calibration model *J. Chemometrics* submitted.
22. Carroll RJ, Ruppert D, Stefanski LA. *Measurement Error in Nonlinear Models*. Chapman and Hall: London, 1995.