Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Magnetic Resonance 172 (2005) 346–358
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Diagnostic analysis of experimental artefacts in DOSY NMR data by covariance matrix of the residuals

R. Huo[a], R.A. van de Molengraaf[b], J.A. Pikkemaat[b], R. Wehrens[a], L.M.C. Buydens[a,*]

[a] *Department of Analytical Chemistry, Institute for Molecules and Materials, Radboud University Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands*
[b] *Philips Research, Materials Analysis, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands*

Received 28 September 2004; revised 3 November 2004
Available online 8 December 2004

Abstract

Multivariate curve resolution (MCR) has been applied to separate pure spectra and pure decay profiles of DOSY NMR data. Given good initial guesses of the pure decay profiles, and combined with the nonlinear least square regression (NLR), MCR can result in good separation of the pure components. Nevertheless, due to the presence of artefacts in experimental data, validation of a MCR model is still necessary. In this paper, the covariance matrix of the residuals (CMR), obtained by postmultiplying the residual matrix with its transpose, is proposed to evaluate the quality of the results of an experimental data set. Plots of the rows of this matrix give a general impression of the covariance in the frequency domain of the residual matrix. Different patterns in the plot indicate possible causes of experimental imperfections. This new criterion can be used as diagnosis in order to improve experimental settings as well as suggest appropriate preprocessing of DOSY NMR data.

Keywords: DOSY NMR; Covariance matrix of residuals; Diagnosis; Experimental artefacts; MCR–NLR

1. Introduction

Diffusion ordered spectroscopy (DOSY) NMR has been regarded as a potential non-destructive alternative to LC–NMR to identify pure components in a mixture [1–3]. Basically, the individual spectra of the components in a mixture can be pseudo-separated according to their respective diffusion coefficients. Previous research has indicated that multivariate curve resolution integrated with nonlinear least square regression (MCR–NLR) and good initial decay profile estimation is a general method to provide reasonably well-resolved spectra and decay profiles [4,5]. However, the performance of MCR is still quite sensitive to the quality of the data. Antalek [6] describes that there are three main error sources during DOSY data acquisition which can cause artefacts and hence can affect the quality of DOSY NMR data. These are eddy currents, a non-uniform magnetic-field gradient, and convection. In addition, common artefacts like line broadening during the experiment, e.g., introduced by time-varying inhomogeneity of magnetic field (large) peak shifts caused by time-dependent temperature variation, and baseline artefacts can also decrease the quality of a data set. This paper will focus on analysing the artefacts induced by eddy currents, time-dependent temperature variation, baseline distortion, and increasing inhomogeneity of magnetic field during the experiment. Eddy currents are mainly caused by fast switching (on and off) of a magnetic-field gradient. If eddy currents persist during acquisition of the FID, they can produce distortions in the spectra such as phase distortion and line broadening. The extent of

* Corresponding author. Fax: +31 24 3652653.
E-mail address: L.Budyens@science.ru.nl (L.M.C. Buydens).

1090-7807/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2004.11.011

the distortion is dependent on the strength of the applied gradient pulse. This effect leads to incorrect diffusion constants [6]. Also time-varying sample temperatures may lead to serious problems, because in some cases even very small changes of the temperature during the experiment can cause relatively large frequency shifts in the spectra [7]. In addition, time-dependent inhomogeneity in the B_0 magnetic field, i.e., decreasing shim during the experiment, can lead to line broadening and lineshape distortion [7]. This can result in deviation of pure exponential decay of the measured signals in a DOSY experiment. Phase and baseline distortion are commonly present and may have several causes, e.g., a disparity between both echo times in the DOSY pulse sequence and hardware imperfection. After phase correction, which should always be done, baseline errors appear. This is a well-known shortcoming of the discrete Fourier transform (DFT) algorithm [8]. These experimental artefacts can have a negative effect on the results of data processing and hence they should be diagnosed and minimised as much as possible. Although convection can also cause major artefacts in DOSY data, it will not be discussed in this paper because it is very difficult to do experiments with a controlled amount of convection without introducing other artefacts [9,10].

Validation of a model is an important part to evaluate the reliability of the resolved pure spectra and decay profiles. When a model goes wrong, it may have two reasons. It can be the result of either the misuse of the MCR algorithm or from the presence of significant experimental artefacts in the data set. In general, residuals can give an indication of the goodness of a model. Usually, the relative root of sum of square differences (RRSSQ) is used to describe the quality of a MCR model [11]. However, RRSSQ only provides a general error value of the model and does not present diagnostic information. Moreover, in some cases the value of RRSSQ does not change but the resulting models are different. The residual matrix, the difference between the original data and the reconstructed data, on the other hand, contains a lot of valuable information to evaluate the model. Therefore, it is useful to plot the residual matrix for visual inspection. There are a few ways to do it [12,13]. The residuals can be plotted versus the chemical shift (plot each row, called residual spectrum), or versus the gradient square values that are used to create a series of decaying NMR spectra (plot each column, called the residual decay profile). Ideally, the plots should show random structure (corresponding to the experimental noise) if the model is correct. Nevertheless, due to the existence of experimental artefacts, there are deviations from the pure exponential decay in the data when the intensities of the NMR signals are plotted against the gradient strengths. As a result, the residual plot can still show non-random structure, even for simple cases. This paper proposes a new graphical diagnosis tool to identify possible sources of experimental artefacts in the DOSY NMR data, if there are any. This plot is constructed by covariance matrix of residuals (CMR), in which each row (or column, which are the same) is plotted versus the values of squared gradients or the sequence of the spectra. It can provide a summary of the covariance of residuals in the frequency dimension at a glance. In this paper, five computer-simulated data sets and four experimental data sets were generated with deliberately adding different kinds of artefacts. The choice of these artefacts is based on the common problems encountered in the acquisition of DOSY data as mentioned above. Since the intention of this paper is to use a diagnostic tool to optimise the experimental settings, the data are relatively simple, i.e., the simulated data have at least one pure variable for each of the component and the experimental data contain no overlapping peaks. Hence, good initial guesses of the pure decay profiles can be easily obtained by a pure variable method such as the orthogonal projection approach (OPA) [14], which means that pure components in the original data can be well resolved by MCR–NLR easily too. The residuals of each calculated MCR model are calculated and the corresponding CMR plots will be shown. It is shown here that the CMR plot shows a characteristic pattern for each particular artefact. Thus, by studying the CMR plots, the possible experimental artefacts in the data set can be identified. This can provide a diagnostic way for experimental optimisation. In addition, the CMR plot can also suggest what preprocessing techniques should be applied to improve the quality of the data.

2. Theory

2.1. The MCR model of the DOSY NMR data

The procedures of processing DOSY NMR data have been described previously [4,5]. Here only a brief review is given. The initial guesses of decay profiles are obtained by the pure variable method OPA. Pure decay profiles and pure spectra are resolved by the alternating least square (ALS) in MCR [15]. Due to the presence of artefacts and rotational ambiguities, the solutions of MCR are not always unique, which may make the iteration stop in a wrong solution without changing the fitting errors. Therefore, constraints need to be applied in each iteration of MCR. Non-negativity constraints are used to get the meaningful chemical and physical information and nonlinear least square regression is used to force pure exponential decay profiles and presents a better solution for MCR [16,17]. The signals of nth component are obtained as described in Eq. (1):

$$I(n, g^2) = I_0(n)\exp[-D(n)(\Delta - \delta/3)K^2],$$
$$K = \gamma g \delta, \quad (1)$$

where $D(n)$ is the diffusion coefficient of the nth component ($m^2\ s^{-1}$). δ is the duration of gradient pulses (s) and Δ is the diffusion time (s), both of which are experiment constants. K is multiplication of δ, γ, the gyromagnetic ratio of the ^1H nucleus (rad $s^{-1}\ T^{-1}$), and g, the gradient strength ($T\ m^{-1}$). $I_0(n)$ is the intensity when the gradient strength g is zero. With several components in the sample, the recorded signals (I) are the summation of Eq. (1).

The model of MCR is represented by Eq. (2):

$$I = C \cdot S^T + E, \quad (2)$$

where C represents resolved pure decay profiles of the exponential term in Eq. (1) and S contains the resolved pure spectra which is $I_0(n)$ in Eq. (1). E is the difference between original data and the calculated data, i.e., the residual matrix.

2.2. Diagnosis of DOSY NMR data by CMR

The conventional criterion to evaluate an MCR model quality is the relative root of sum of square differences (RRSSQ) [11]. It can be calculated by Eq. (3):

$$\%\mathrm{RRSSQ} = 100 \times \sqrt{\frac{\sum(I_{\mathrm{original}} - I_{\mathrm{predicted}})^2}{\sum(I_{\mathrm{original}})^2}}. \quad (3)$$

The ideal prediction should lead to the value of RRSSQ close to zero. This criterion gives an overall assessment of the goodness of fit of the model but does not provide diagnostic information.

The residual matrix E contains diagnostic information of a data set. It is the difference between the original data matrix and the predicted data matrix

$$E_{\mathrm{resi}} = I_{\mathrm{original}} - I_{\mathrm{predicted}}. \quad (4)$$

A plot of residuals can evaluate the MCR model of DOSY NMR data visually. One type of plot is obtained by plotting the residuals versus the chemical shift, i.e., the rows of the residual matrix. This will be called residual spectrum in the following. In the residual spectra of the experimental data, there are always some sharp peaks present in the signal part of the residual spectra because the variance of a region containing a peak is much higher than that of an area without signal. Therefore, the residual spectra hardly ever show a random noise pattern, even if the artefacts in the experimental data are very small.

The residuals can also be plotted as a function of the gradient strength, i.e. the columns of the residual matrix (variables of the matrix). However, with a large number of variables, plotting each of them is very cumbersome and the diagnostic information behind the graphs is difficult to be seen. In this paper we propose to plot the covariance matrix of residuals (CMR), by which one can visualise the structure of residuals more easily. Hence, it can provide diagnostic information better. Since it is the covariance between the residual spectra that is of interest, CMR is obtained by postmultiplying the residual matrix by its transpose, as described in Eq. (5)

$$Z_{\mathrm{resi}} = E_{\mathrm{resi}}E'_{\mathrm{resi}}. \quad (5)$$

The residual matrix contains r rows and c columns, which has the same size as the original data set (r = number of gradient values; c = number of spectral data points). Therefore, CMR, which is Z_{resi}, contains the size of $r \times r$ and the plots of its rows (or columns) summarise the information of the covariance of the residual spectra. If the data are of good quality and the model is correct, then there is no correlation or covariance in any row of the residual matrix and the plot of CMR shows random structure with unique variance. When there is high covariance between two of the residual spectra or there is a gross error in an individual spectrum (outlier), the corresponding elements of the CMR will have a very high value. Also, if there is a trend with the variables in the gradient dimension of the residual matrix, then the CMR plot will show it clearly.

3. Experimental section

3.1. Simulated data

A simulated data set containing three components is used to examine the response of CMR to the simulated experiment artefacts. The stacked plot of the data set and the pure spectra of the components (frequency domain) are shown in Fig. 1. The spectral width is 10 ppm in 1000 data points. The first component contains one Lorentzian peak at 5.01 ppm. The maximum intensity of the peak is 100 arbitrary units (U). The second component contains two Lorentzian peaks at 2.51 and 5.21 ppm, with the maximum intensities of 100 and 50 U, respectively. The third component consists of three peaks at 2.51, 4.81, and 7.51 ppm. The maximum intensities of the peaks are 100, 50, and 50 U. The linewidths (half-widths at half maximum) of the peaks are the same, which is 3.3 data points. The diffusion coefficients of the three components are 5.00×10^{-10}, 5.00×10^{-9}, and $1.00 \times 10^{-9}\ m^2\ s^{-1}$, respectively. The simulated data set contains 32 spectra with equally spaced K^2 values ranging from 1.0143×10^8 to $3.0622 \times 10^{10}\ m^{-2}$, and $\Delta = 0.1$ s, $\delta = 5$ ms. The data set contains white noise with the standard deviation of 0.04% of the lowest peak intensity. Five different kinds of simulated artefacts, as mentioned above, are added to this data set. The simulation is analogical to the

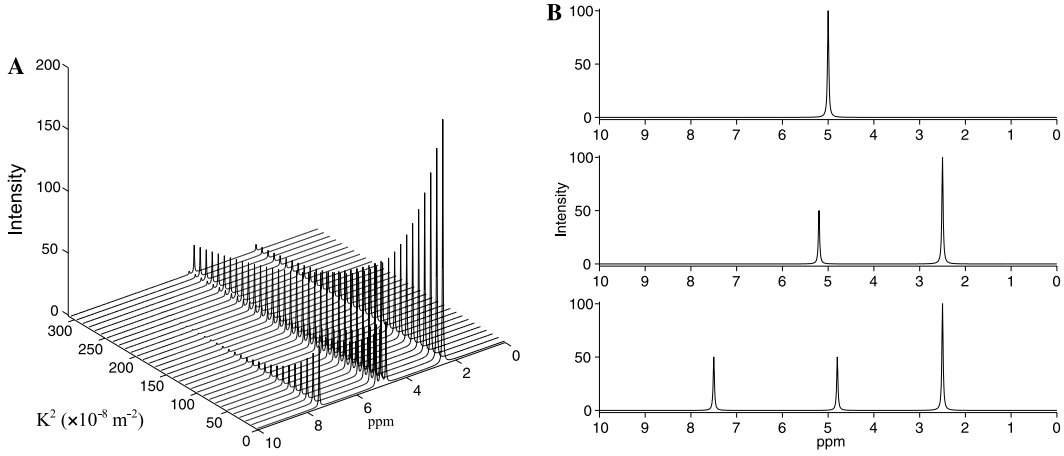

Fig. 1. (A) A stacked plot of the simulated data set; (B) the pure spectra of the components.

experimental situations under which experimental artefacts are deliberately created (see below).

The first data set, named *Sim-no-artefacts*, contains small random phase and frequency shifts between spectra that are typically present in measured DOSY data of high quality. Peak shifts ranging from −0.05 to 0.05 data points and phase shifts −0.5–0.5° are introduced randomly in each spectrum. This spectral shift can be corrected by a preprocessing method (combination of FIDDLE and Witjes' method [5]). To make the data more realistic, small global random shifts (which means that the peaks shift relative to one another in the same spectrum and cannot be corrected by the preprocessing method mentioned above) are added to the data set as well. The amount of the global random peak shifts is between −0.002 and 0.002 data points and that of random phase shifts varies between −0.005° and 0.005°. These global random shifts and the same noise level will be introduced to the following data sets as well.

The second data set, named *Sim-eddy*, simulates the effects of eddy currents. Because eddy currents can create first order phase shifts and line broadening, two data sets will be simulated (*Sim-eddy-phase* and *Sim-eddy-width*) with the two effects separately so as to interpret the CMR plot more easily. For this data analysis, the magnitude of the artefacts is taken to increase linearly as a function of gradient strength. First, in addition to the same level of white noise and the global random shifts as present in the previous data set, each spectrum in *Sim-eddy-phase* contains a phase distortion varying from 0° to 17.8° linearly at the peak of 2.5 ppm (here only the phase of the peak at 2.5 ppm is varied for easy illustration, but the CMR plot displays more or less the same pattern if all the peaks in each spectrum are varied with the same phase distortion). Here we assume that the phase distortion increases linearly as a function of the gradient strength. Likewise, the first order line broadening is produced in *Sim-eddy-width* with the values linearly changing from 0 to 0.015 data points through all the increments.

The third data set, named *Sim-temp*, simulates the artefacts resulting from temperature variation during a DOSY experiment. In addition to the global random shifts and noise as present in Sim-no-artefacts, *Sim-temp* contains linearly changing (time-dependent) frequency shifts. It is supposed that there are no temperature-dependent frequency shifts in the first 10 spectra because of unobvious temperature variation. Then from the 11th to the 32nd spectrum, the peak of 2.51 ppm starts to shift with equal steps of 0.03 data points up to a maximum shift of 0.23 data points in the last spectrum.

The fourth data set, *Sim-baseline*, simulates the effects of baseline distortion. To do this, a phase distortion of 3.4° is introduced to all the peaks of this data set. Then the data are corrected for the phase distortion so that baseline distortions are created.

The fifth data set, *Sim-field*, simulates an increasing inhomogeneity of the magnetic field during the acquisition of a DOSY data set, which results in an increase of the linewidth. A change in the line shape, an additional effect of field inhomogeneities, is not simulated. This effect may arise many different forms. Here, it is assumed that the main increase in the linewidth occurs during the first 21 spectra. The linewidths of all peaks in the first 21 spectra increase with steps changing from 0.66 data points to 0.132 data points linearly. Then for the rest of the spectra (22–32) the linewidths become more stable and the steps change from 0.11 data points to 0 data points in the last spectrum.

3.2. Experimental data

The experimental data are measured from a mixture containing 5 μl acetone (C_3H_6O), 5 μl water (H_2O), and 5 mg DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid) in 0.5 ml D_2O. The DSS-signal was calibrated at

0 ppm. The ^{1}H NMR spectra were obtained on a Bruker Avance 300-MHz spectrometer, making use of a 5 mm inverse broad band probe, equipped with a Z-gradient (maximum gradient strength of 54 G/cm). A stimulated spin-echo [18] sequence with bipolar gradients [19] and a spoiler gradient between the 2nd and 3rd 90° pulse (to dephase unwanted magnetisation) was used. All experiments were done at 298 K (except data *Exp-temp*). The amount of diffusion weighting was varied by increasing the diffusion gradient from 2 to 95% of the maximum gradient strength in 32 linear increments. The DOSY experiments were performed with a δ (time duration of the applied gradient pulse) of 2 ms and a Δ (diffusion time) of 80 ms. The residual signal intensities at the last increment are 1% for H_2O, 7% for C_3H_6O, and 25% for DSS, relative to the signal intensities of the first experiment. All spectra were recorded with a spectral width of 15 ppm with 32K data points.

The first experimental data set, *Exp-no-artefact*, is recorded with fully optimised experimental settings (no deliberate artefacts were introduced).

At the second experimental data set, *Exp-temp*, the temperature is decreased from 298.0 to 297.8 K in a linear way during the first 20 min of the experiment. With this experiment, time-dependent temperature variation is introduced.

The third experimental data set, *Exp-baseline*, is recorded with a deliberate disparity between both echo times in the pulse sequence. In this way, after discrete Fourier transformation and phase correction, the spectra show baseline artefacts [7] (non-zero offset and curvature).

The fourth experiment, *Exp-field*, the Z1-shim was linearly reduced to approximately 95% of the optimal intensity of the ^{2}H-lock signal. This experiment introduces time-dependent variation of the magnetic field homogeneity.

3.3. Data analysis

The data analysis is performed by MCR–NLR and then the CMR of the models are calculated. All calculations were done in MATLAB 6.5.1 on a SUN Unix workstation. The software package used for data preprocessing (combination of FIDDLE and Witjes' method) and data analysis (MCR–NLR) is available on our website: http://www.cac.science.ru.nl/.

4. Results and discussions

4.1. CMR plots of simulated data

4.1.1. Spectral shifts artefacts and no significant artefacts

The data Sim-no-artefacts only contain random spectral shifts which can be corrected by the preprocessing method (FIDDLE and Witjes' method). The CMR plots obtained from Sim-no-artefacts before and after correction for the spectral shifts are presented in Figs. 2A and B, respectively. Since CMR is a symmetric matrix (i.e., a matrix equal to its transpose) with a size of $r \times r$ (r is the number of the rows, or spectra in the original data), one can plot either its rows or columns. In the CMR plots shown in this paper, the elements of the rows are plotted against the number of the spectra, i.e., the number of the gradient increments. In the examples shown here this number goes from 1 to 32. As can be seen, the CMR plot in Fig. 2A displays diamond shapes. The diamond shapes are the result of the spectral shifts predominant in the original data, which can be caused by the imperfect experimental conditions. Fig. 2B shows a CMR plot of the corrected data leaving out the diagonal elements. A diagonal element of CMR represents the covariance of a residual spectrum with itself, which is always high compared with the values of other elements when there

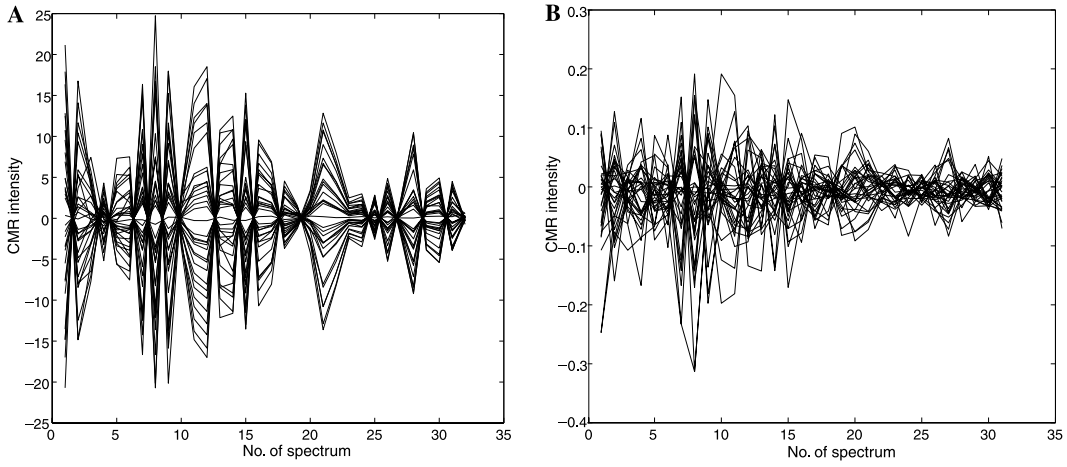

Fig. 2. The CMR plots of the data Sim-no-artefact. (A) Before data correction; (B) after data correction (leaving out the diagonal elements).

are no significant artefacts in the original data. If there are not many artefacts in the original data, the relatively high values of the diagonal elements in the CMR can hinder visualisation of the structure of the CMR, while if the original data contain a considerable amount of artefacts, a CMR plot will look more or less the same with or without diagonal elements. Therefore, it is recommended to leave out the diagonal elements. This can eliminate the interference caused by them without losing useful information. Thus, the diagonal elements are left out and one can see that the CMR plot shows random structure. This indicates that the model is well resolved and there are no obvious experimental artefacts in the original data.

Fig. 3 explains the diamond shapes of the CMR plot in Fig. 2A. Fig. 3A displays a small spectral region of the first five different rows of the residual matrix around the signal at 2.5 ppm. After reconstruction with MCR–NLR, signals in a DOSY data set appear at a constant-average-frequency for every row. If, for a certain row, the original signal was at a lower frequency as compared to the same signal in the reconstructed data set (as is the case for the signal at 2.5 ppm in rows 2 and 5, see Fig. 3A), the residual plot around the peak at 2.5 ppm shows (from left to right) first negative values and then positive values. Oppositely, if the original signal was at a higher frequency, then the residual spectrum will look as the mirror image (as is the case for the first residual spectrum in Fig. 3A). Fig. 3B shows the corresponding zoomed-in CMR plot containing the first five rows, where only the first seven matrix elements of each row are displayed. In Fig. 3B, one can see that there are some high negative and some high positive covariance values displayed in the CMR plot. The values of covariance describe the relationship of each pair of the residual spectra. For example, in the line representing the first row of CMR (row 1), each point illustrates the covariance between the first residual spectrum and each of other spectra. Therefore, a negative value at the second point (indicated with a circle in Fig. 3B) means that there is a negative relationship between the first and the second

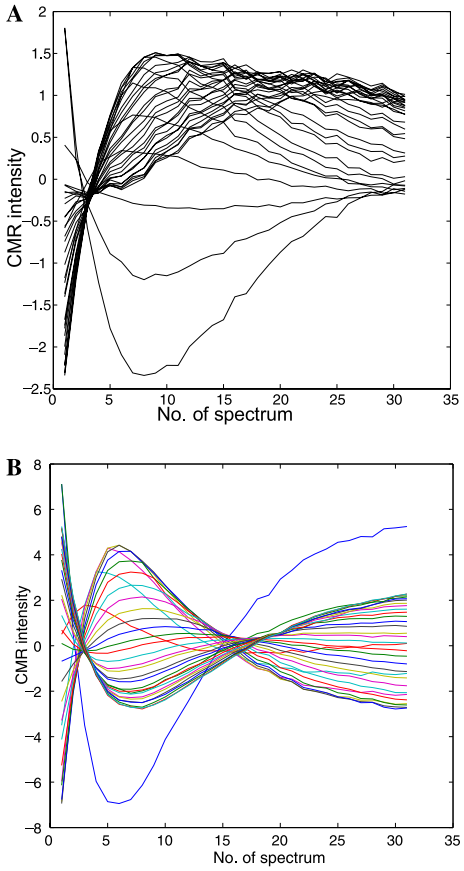

Fig. 4. The CMR plot of the data Sim-eddy-phase (A) with the effect of first order phase shift; and Sim-eddy-width (B) with the effect of first order line broadening.

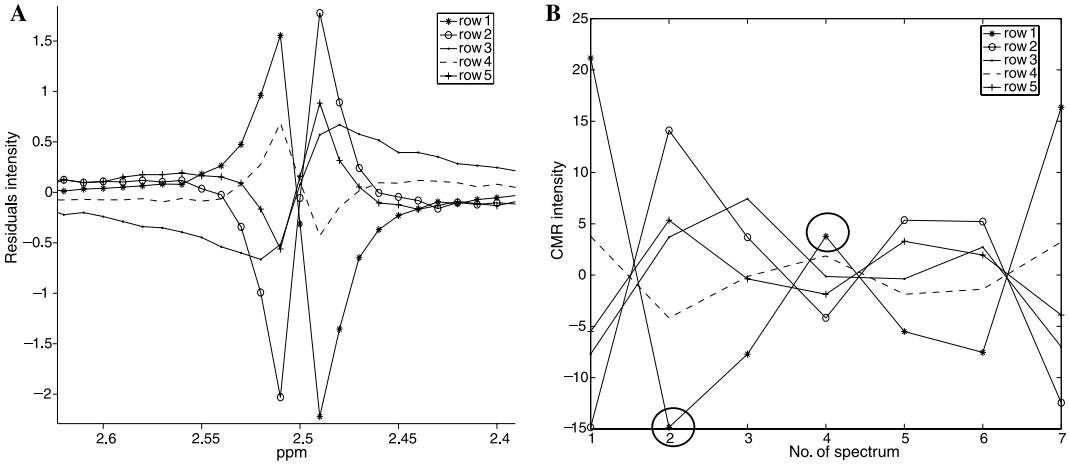

Fig. 3. Graphical explanation of the diamond shape of CMR from spectral peak and phase shifts. (A) Zoomed-in image of the first five residual spectra (rows) at around 2.5 ppm; (B) Zoomed-in image of the first five rows of the CMR plot containing seven elements.

residual spectrum. This can be compared to the corresponding residual spectra in Fig. 3A. Likewise, the positive value at the fourth point (in row 1) indicates a positive relationship between the first and the fourth residual spectrum. It is the positive and negative relationships of the residual spectra that lead to the diamond shapes appearing in the CMR plot.

4.1.2. Eddy-current artefacts

The CMR plots of Sim-eddy-phase and Sim-eddy-width are given by Figs. 4A and B, respectively. Fig. 4A clearly contains two kinds of curves: one kind has broad extrema (local minima or maxima) and the other kind shows a mirror image of an exponential decay shape. The structure of the curves in Fig. 4B contains broad extrema, which is similar to the first kind of structure in Fig. 4A. Fig. 5 explains these two kinds of structure in CMR plot using the first and the 18th rows of the CMR plot in Fig. 4A as examples. Fig. 5A is a plot of the first residual spectrum, i.e., first row of the residual matrix. It can be seen that the residual intensity is highest at 2.51 ppm. The sharp peak around 2.5 ppm in the residual spectrum is due to the simulated eddy-current artefacts occurring in the same position of the spectra in the original DOSY data (Sim-eddy). The plot of the column of the residual matrix at 2.51 ppm is shown in Fig. 5C. In this curve the intensity reaches to maximum and then decreases gradually. As defined [20], the covariance not only depends on the association between pairs of residual spectra, but also the intensities of the residuals. Therefore, if there is a trend in the variables with high intensity in the residual matrix, then the CMR plot will show the same structure. Because of the highest intensity at 2.51 ppm in the first residual spectrum, the first row of the CMR plot will display the same structure as that at 2.51 ppm. Fig. 5E represents the first row of the CMR plot and it displays a mirror image (same structure) of that in Fig. 5C, which is also a curve with an extremum. Similarly, Figs. 5B, D, and F illustrate the decay curves in the CMR plot. The 18th row of the CMR plot shows the same structure as the residual plot at 2.38 ppm because in the 18th

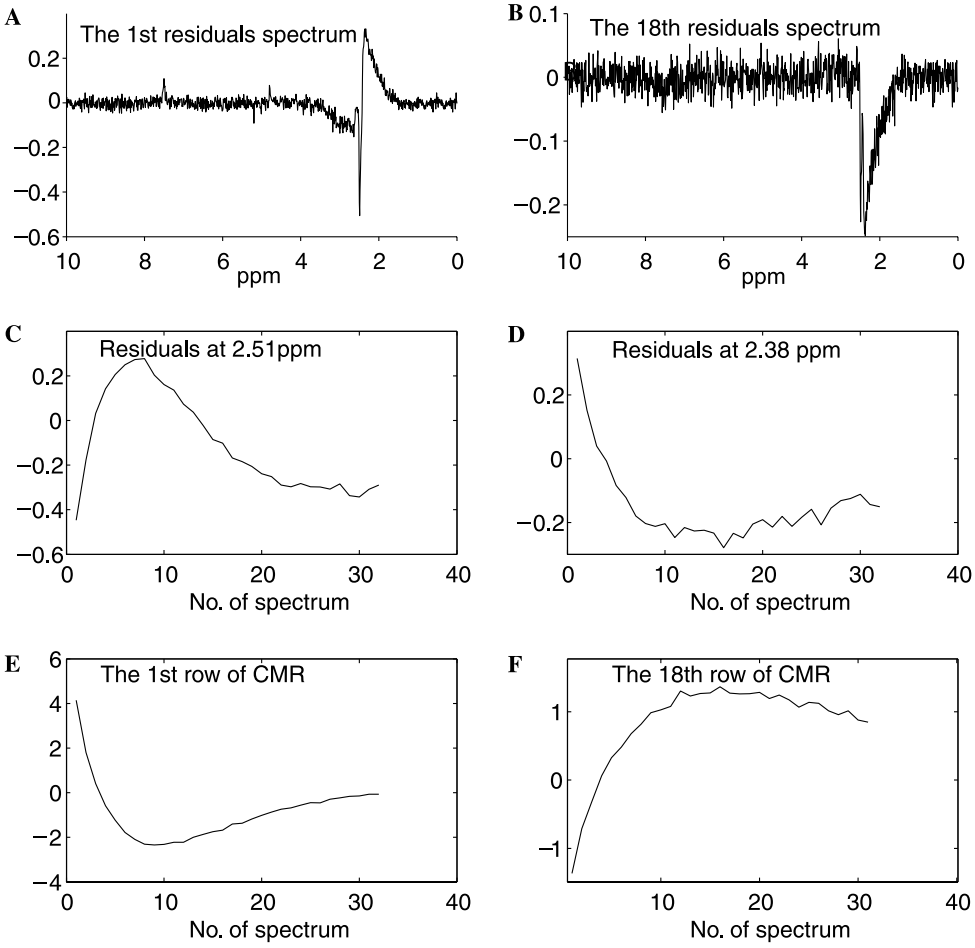

Fig. 5. Graphical explanation of the CMR structure of the data Sim-eddy-phase. (A) Plot of the first residual spectrum; (B) plot of the 18th residual spectrum; (C) plot of the residuals at 2.51 ppm; (D) plot of the residual matrix at 2.38 ppm; (E) plot of the first row of CMR; (F) plot of the 18th row of the CMR.

residual spectrum (Fig. 5B) there are more intensities contributing to this structure than in the first residual spectrum. From Figs. 4 and 5, the decay curves in the CMR plot become more distinct as the phase distortion increases.

From Figs. 4A and B, it can be seen that the effect of eddy currents can produce at least two different structures in the CMR plot, depending on the relative contributions of phase distortion and line broadening.

4.1.3. Time-dependent temperature variation artefacts

For certain components, a typical characteristic of a data set under the time-dependent temperature variation is time-dependent peak shifts. The CMR plot in Fig. 6A, obtained from data set Sim-temp, can be used to diagnose the time-dependent peak shifts. Figs. 6B and C explains the shape displayed in Fig. 6A under the effect of the time-dependent peak shifts. Fig. 6B shows the original values and the calculated values at 2.51 ppm in the logarithm coordinate and Fig. 6C shows plots of the residuals at 2.51 ppm. Due to the time-dependent peak shifts starting from the 11th spectrum, the highest peaks in the CMR plot (Fig. 6A) can be regarded as a big diamond shape resulting from global random shifts. Then the intensities at 2.51 ppm from 11th spectrum become smaller gradually and hence deviate from linearity. Therefore, curvature is created from the 11th spectrum in Sim-temp, which leads to the structure of the corresponding residual plot with a high peak in the beginning and a small parabolic shape in the following (see Fig. 6C). As a result, the CMR plot also shows the same pattern.

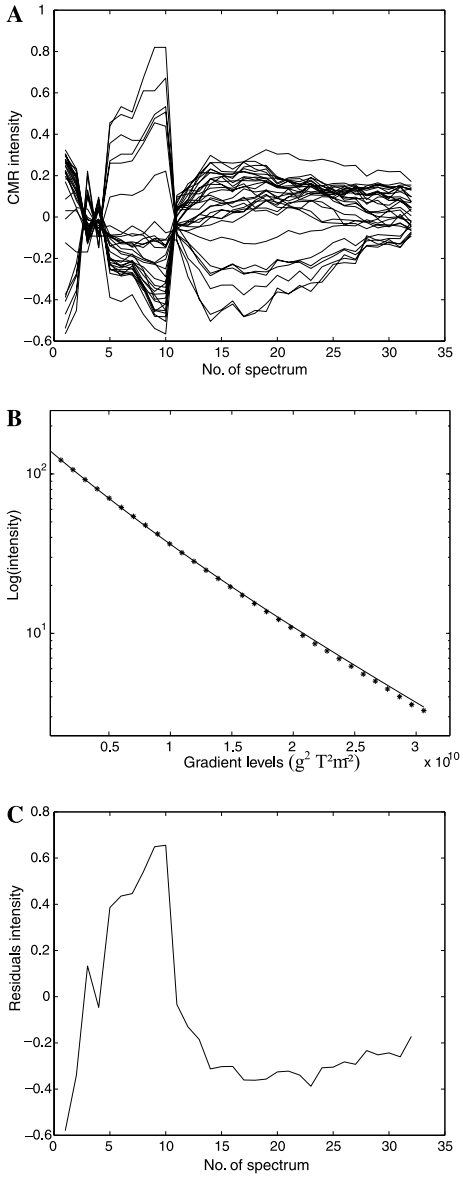

Fig. 6. (A) The CMR plot of the data Sim-temp; (B) plots of the logarithm of the original values (stars "*") and the calculated values (lines) at 2.51 ppm; (C) plot of the residuals at 2.51 ppm.

4.1.4. Baseline artefacts

In Fig. 7A, the CMR plot obtained from Sim-baseline shows a series of curves with decay tendency. This indicates there is a bias in the original data. In Fig. 7B, the original data and the reconstructed data at 4.81 ppm are plotted. It is difficult to see the difference between the two curves. However, because baseline distortion occurs in the original data while it is minimised or eliminated in the reconstructed data, there is a bias between the original data and the reconstructed data. As a result, the residual plot shows a trend of an inverse decay curve, as shown in Fig. 7C. Therefore, this trend is also displayed in the CMR plot.

4.1.5. Inhomogeneity of the magnetic field

The CMR plot of Sim-field is given in Fig. 8A. As can be seen, a parabolic shape is present at the beginning of the CMR plot. When the change of the linewidths in Sim-field is getting stable, the curves become flat. This structure of CMR plot can be considered as an indication of inhomogeneity of the magnetic field varying during the data acquisition. This is explained by Figs. 8B and C. In Fig. 8B, the original values and the corresponding fitting curve at 5.03 ppm of Sim-field are plotted. Due to the systematic line broadening effects caused by inhomogeneity of the magnetic field, the attenuation of the original values deviates from pure exponential decay. In this simulation the original values have a systematic change crossing the fitting curve for the first part and then the deviation of the pure exponential decay becomes more stable. Thus, this results in a CMR plot with a shape shown in Fig. 8C, in which the parabolic shape indicates the first part due to the systematic change of the exponential decay while the curve becomes flat as the signals follow pure exponential decay better.

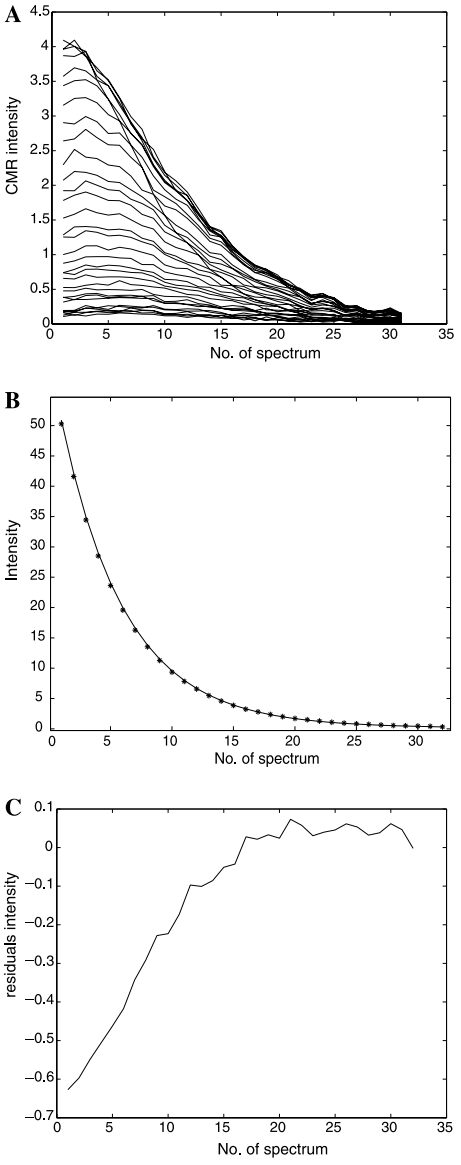

Fig. 7. (A) The CMR plot of the data Sim-baseline; (B) plots of the original values (stars "*") and the calculated values (line) at 4.81 ppm; and (C) plot of the residuals at 4.81 ppm.

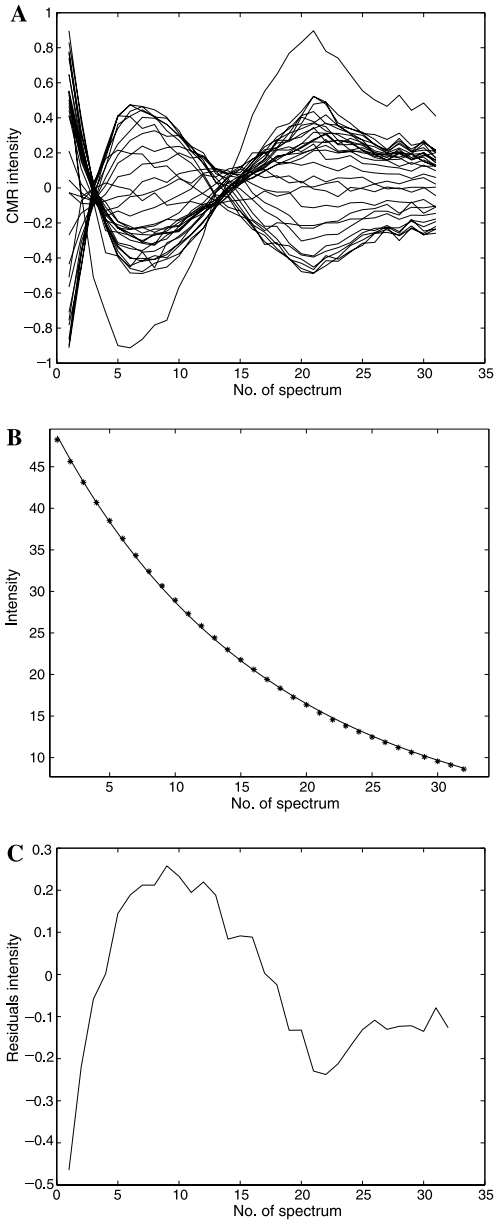

Fig. 8. (A) The CMR plot of the data Sim-field; (B) plots of the original values (stars "*") and the calculated values (line) at 5.03 ppm; (C) plot of the residuals at 5.03 ppm.

4.2. Experimental data

To illustrate the performance of the CMR plots in real data, experimental data recorded with and without introducing deliberate artefacts are obtained, similar to the computer-simulated data mentioned above. Fig. 9A shows the DOSY NMR data of a mixture containing water, acetone, and DSS. The pure spectra of the compounds are given in Fig. 9B. Four data sets are acquired, including one without any deliberate artefacts (Exp-no-artefact), one with time-dependent temperature variation effects (Exp-temp), one with baseline artefacts (Exp-baseline), and finally one recorded with a time-dependent decrease of the optimal shim setting (Exp-field). The effect of eddy currents is not analysed because it appeared difficult to introduce obvious eddy currents in a controlled way during the DOSY experiment with the equipment and the available gradient strength. In addition, the pulse sequence used for the experiments is designed to reduce eddy currents as much as possible [21].

The CMR plots obtained from the data set Exp-no-artefact before and after spectral shifts correction are displayed in Figs. 10A and B, respectively. As can be seen, the CMR plot in Fig. 10A shows diamond shapes, which relates to the presence of spectral shifts. After the original data are corrected, the corresponding CMR plot shows random structure and the values of residuals are lowered with a considerable amount, as showed in

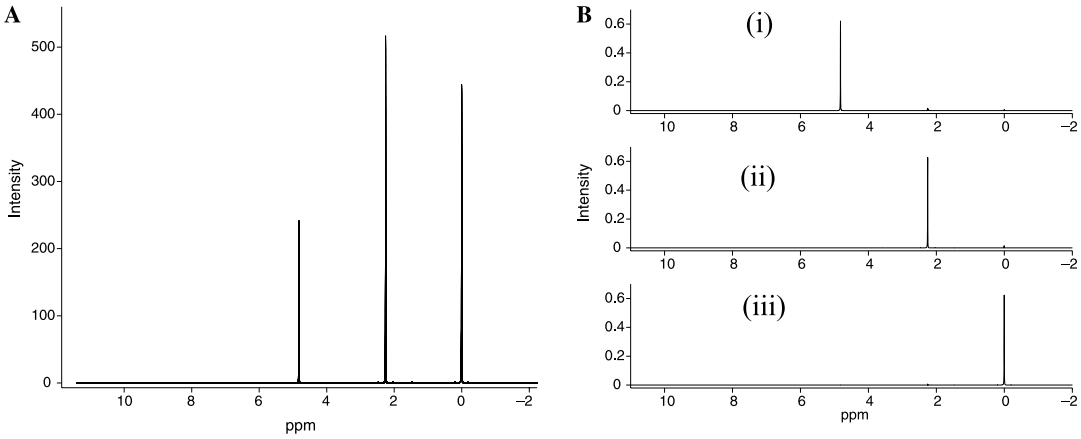

Fig. 9. Plots of the data Exp-no-artefacts. (A) The original data set; (B) the corresponding pure spectra: (i) H_2O; (ii) acetone; (iii) DSS.

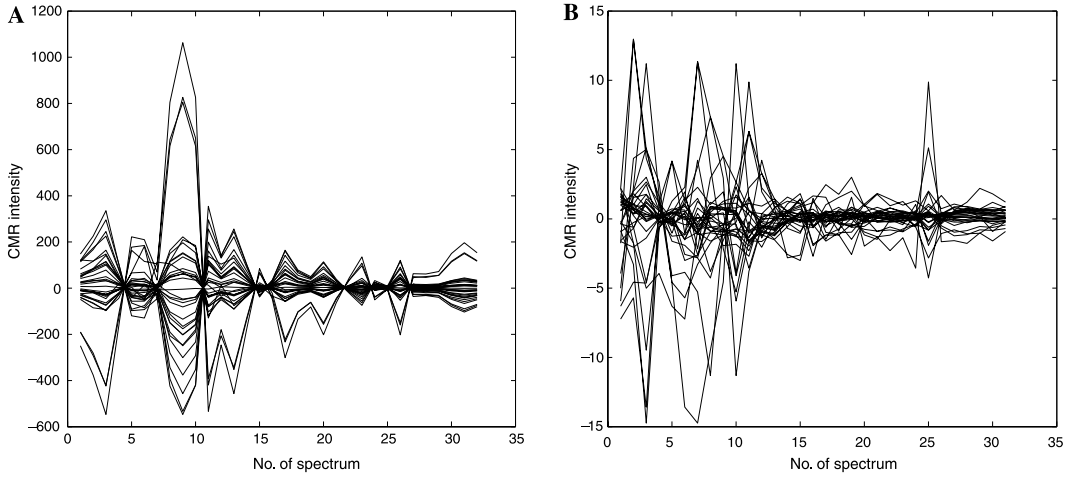

Fig. 10. The CMR plots of the data Exp-no-artefacts. (A) Before correction; (B) after spectral shifts correction.

Fig. 10B. One can see that there are still some sharp peaks in the CMR plot obtained from the data after correction. This can be caused from the global random shifts of the peaks that are not amendable by using the preprocessing correction method. This example illustrates that spectral shifts as well as the global shifts are very common in a DOSY data set even without any intended artefacts. Therefore, it is recommended to correct DOSY NMR data for the spectral shifts before it is analysed by MCR–NLR. The CMR plots of the following data are all obtained from the data corrected for spectral shifts.

In the second experimental data, Exp-temp, time-dependent peak shifts can be expected. In Fig. 11A the expanded peaks of water are shown. Only minor time-dependent frequency shifts are observed in the first five spectra and then the shifts increase in the following spectra as the temperature become more unstable with time. The CMR plot, shown in Fig. 11B, is comparable to the one shown in Fig. 6A. In addition, the CMR plot indicates the presence of global random shifts.

To illustrate the effects of baseline artefacts in data set Exp-baseline, the expanded first spectrum of the original data and that of the reconstructed data are presented in Fig. 12A. One can see the baseline distortion in the original data set, whereas it is eliminated in the reconstructed data. The corresponding CMR plot, resembling a decay curve, is given in Fig. 12B. The decay curve of the CMR plot can be compared to the CMR plot in Fig. 7A, obtained from the data Sim-baseline with baseline distortion. The CMR plot with a decay trend indicates there is a bias existing in the original data, which is caused by baseline distortion.

To investigate the effect of a varying inhomogeneity of the magnetic field during the experiment, the CMR plots obtained from the data Exp-field are depicted in Fig. 13A. Because the random shifts also exist in the data, it is difficult to see the specific structure of CMR resulting from the imperfection of Z1-shim only. For a comprehensive visualisation of the CMR structure, the first five rows of the CMR are also plotted in Fig. 13B. If compared with Fig. 8A, it can be seen that the

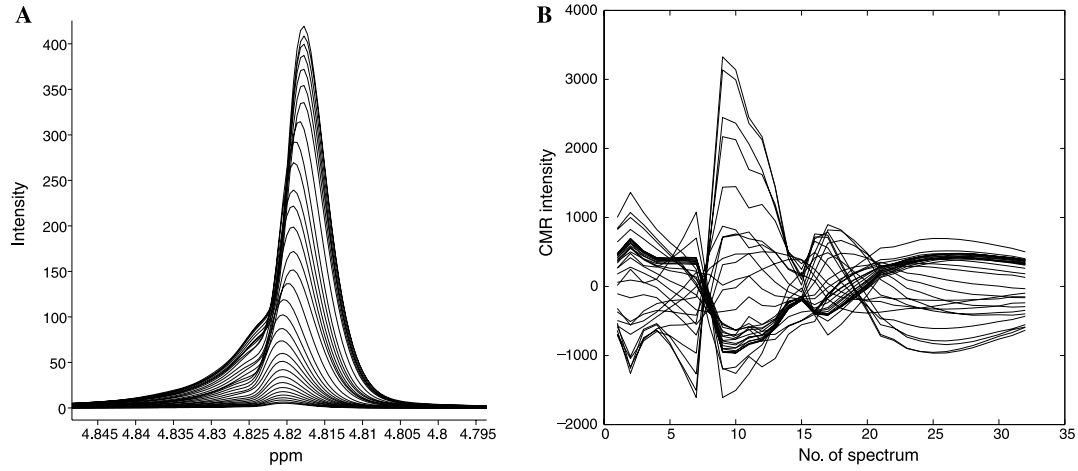

Fig. 11. Plots of the data Exp-temp. (A) The zoomed-in image of the peak of water; (B) the resulting CMR plot.

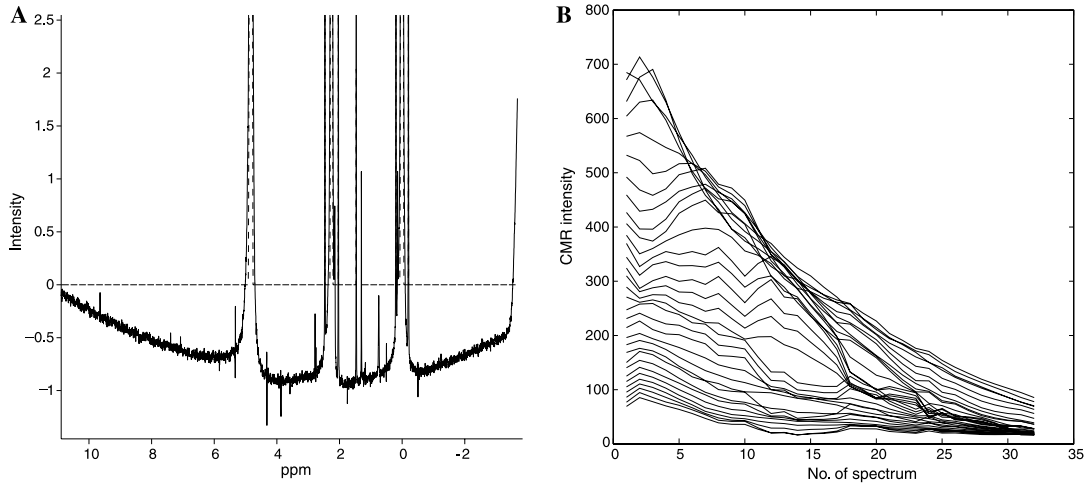

Fig. 12. Plots of the data Exp-baseline. (A) The zoomed-in image of original spectrum (solid line) and the reconstructed spectrum (dash line); (B) the resulting CMR plot.

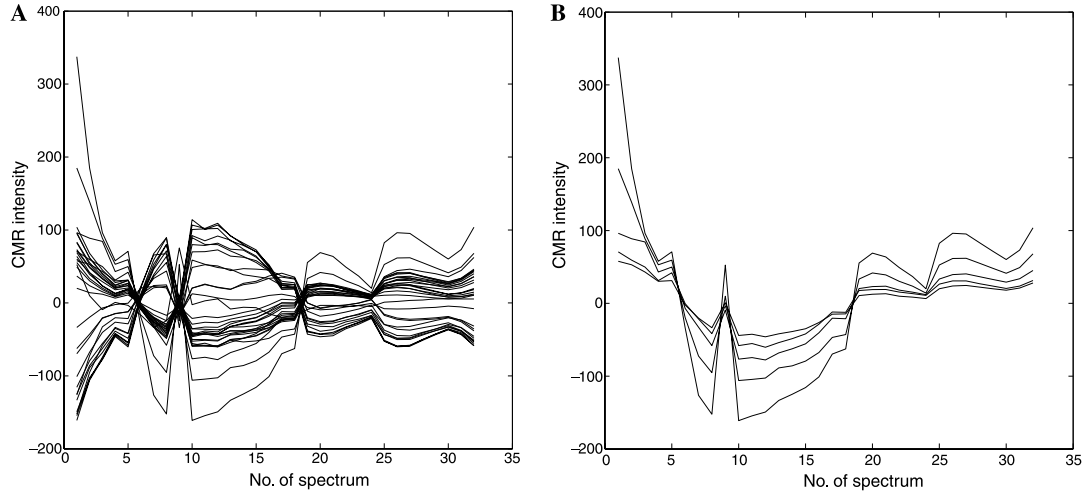

Fig. 13. Plots of the data Exp-field. (A) The CMR plot; (B) the first five rows of the CMR plot.

plot in Fig. 13A shows a similar structure that is related to the line broadening artefacts. This structure is more clearly visible in Fig. 13B. According to Fig. 13B, the data Exp-field may have various lineshape broadening trends throughout the whole data. The lineshape or linewidth of the first 18 spectra (except the 9th spectrum) seems to be broadened steadily according to the parabolic trend appearing in the first 18 elements in the CMR. After spectrum 18 the effect of the broadening becomes less intense. This is similar to the way that is used to simulate a data set with line broadening effects described above (Sim-field).

5. Conclusion

Artefacts are very commonly present in experimental DOSY data sets, even with "optimal" experimental settings. Diagnosis of the experimental artefacts is essential for successfully resolving and interpreting DOSY NMR data. The CMR plot provides a visual tool to diagnose artefacts in DOSY data because it summarises the covariance of the residual spectra in a residual matrix. Different shapes of the CMR plot reveal common suboptimal experimental conditions. A CMR plot with a random structure indicates the original data are completely resolved and contains no significant experimental artefacts. If a CMR plot shows a structure of diamond shapes, it suggests that the original data set contains random peak and phase shifts (spectral shifts or global shifts). The diamond shapes are present in most of the experimental data, implying that the global random shifts, e.g., caused by unwanted echoes are always present even with the best experimental settings. Eddy currents can cause line broadening, which can lead to curves with a broad extremum in the CMR plot. In addition, phase distortion may be caused. This leads to decay-like curves as well as curves with a broad extremum displaying in the CMR plot simultaneously. If a big diamond shape appears in the beginning of a CMR plot and then a parabolic-like shape appears in the following, it reveals that time-dependent peak shifts are present in the data and suggests that measures should be employed to control the sample temperature. If a CMR plot shows a trend of decay curves, then it can be the result of unbalanced echo times, or other sources of initial-phase errors. In this case, one needs to look into the possible causes of eddy currents as well as the quality of the baselines in the data. On the other hand, if a CMR plot displays a structure of first parabolic shape and the curves becoming more horizontal, line broadening and lineshape effects can be present in the data because the shimming has changed during experiment. In practice, various experimental artefacts often exist in a data set simultaneously. This can make it difficult to see one kind of shapes in the CMR plots. Plotting a part of the CMR (e.g., the first five rows) can allow visualising the CMR structure more easily. In case there is more than one kind of artefacts in a data set, the CMR plot will be mixed with several structures and will generally show the structure corresponding to the main effect. We would recommend implementing CMR in (commercial) DOSY software packages in order to provide the spectroscopist feedback on experimental artefacts and on appropriate data preprocessing.

Acknowledgments

The authors are grateful to the Dutch Technology Foundation STW (790.35.393) for the financial support of this research. The authors also thank J. van Duynhoven (Unilever research and development Vlaardingen, The Netherlands) for his valuable remarks on the manuscript.

References

[1] C.S. Johnson, Diffusion ordered nuclear magnetic resonance spectroscopy: principles and applications, Prog. Nucl. Magn. Reson. Spectrosc. 34 (1999) 203–256.
[2] K.F. Morris, P. Stilbs, C.S. Johnson Jr., Analysis of mixtures based on molecular size and hydrophobicity by means of diffusion-ordered 2D NMR, Anal. Chem. 66 (1994) 211–215.
[3] D.A. Jayawickrama, C.K. Larive, E.F. McCord, D. Christopher Roe, Polymer additives mixture analysis using pulsed-field gradient NMR spectroscopy, Magn. Reson. Chem. 36 (1998) 755–760.
[4] R. Huo, R. Wehrens, J. Van Duyhoven, L.M.C. Buydens, Assessment of techniques for DOSY NMR data processing, Anal. Chim. Acta 490 (2003) 231–251.
[5] R. Huo, R. Wehrens, L.M.C. Buydens, Improved DOSY NMR data processing by data enhancement and combination of multivariate curve resolution with non-linear least square fitting, J. Magn. Reson. 169 (2004) 257–269.
[6] B. Antalek, Using pulsed gradient spin echo NMR for chemical mixture analysis: how to obtain optimum results, Concepts Magn. Reson. 14 (2002) 225–258.
[7] J.K.M. Sanders, B.K. Hunter, Modern NMR Spectroscopy, second ed., Oxford University Press, Oxford, 1993.
[8] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, Protein NMR Spectroscopy: Principles and Practice, Academic Press, SanDiego, U.S.A., 1995.
[9] W.J. Goux, L.A. Verkruyse, S.J. Salter, The impact of Rayleigh–Bernand convection on NMR pulsed-field-gradient diffusion measurements, J. Magn. Reson. 88 (1990) 609–614.
[10] X.A. Mau, O. Kohlmann, Diffusion-broadened velocity spectra of convection in variable BP-LED experiments, J. Magn. Reson. 150 (2001) 35–38.
[11] W. Windig, B. Antalek, J.L. Lippert, Y. Batonneau, C. Brémard, Combined use of conventional and second-derivative data in the SIMPLISMA self-modeling mixture analysis approach, Anal. Chem. 74 (2002) 1371–1379.
[12] A.C. Atkinson, Plot, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis, Clarendon, Oxford, UK, 1985.

[13] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997.
[14] F.C. Sánchez, B. van de Bogaert, S.C. Rutan, D.L. Massart, Multivariate peak purity approaches, Chemometrics Intell. Lab. Syst. 34 (1996) 139–171.
[15] A. Garrido Frenich, D. Picón Zamora, J.L. Martínez Vidal, M. Martínez Galera, Resolution (and quantitation) of mixtures with overlapped spectra by orthogonal projection approach and alternating least squares, Anal. Chim. Acta 449 (2001) 143–155.
[16] S. Bijlsma, H.F.M. Boelens, H.C.J. Hoefsloot, A.K. Smilde, Constrained least squares methods for estimating reaction rate constants from spectroscopic data, J. Chemometrics 16 (2002) 28–40.
[17] A.K. Smilde, H.C.J. Hoefsloot, H.A.L. Kiers, S. Bijlsma, H.F.M. Boelens, Sufficient conditions for unique solutions within a certain class of curve resolution models, J. Chemometrics 15 (2001) 405–411.
[18] J.E. Tanner, The use of the stimulated echo in NMR diffusion studies, J. Chem. Phys. 52 (1970) 2523–2526.
[19] G. Wider, V. Dötsch, K. Wüthrich, Self-compensating pulsed magnetic-field gradients for short recovery times, J. Magn. Reson. 108 (Ser. A) (1994) 255–258.
[20] E.R. Malinowski, Factor Analysis in Chemistry, third ed., Wiley, New York, 2002, pp. 33–34.
[21] D. Wu, J. Chen, C.S. Johnson, An improved diffusion-ordered spectroscopy experiment incorporating bipolar-gradient pulses, J. Magn. Reson. A 115 (1995) 260–264.