MAGNETIC RESONANCE IN CHEMISTRY, VOL. 35, S13–S21 (1997)

ANOVA and Factor Analysis Applied to Time Domain NMR Signals

D. N. Rutledge,* A. S. Barros and F. Gaudard
Laboratoire de Chimie Analytique, Institut National Agronomique, 16 rue Claude Bernard, 75005 Paris, France

Time domain (or low resolution pulse) NMR can generate a range of relaxation curves (CPMG, I–R, P–S, HSE, FID, etc.) which may vary depending on the characteristic of the product being controlled—water content, hydration state, solid fat content, iodine number or even authenticity of origin. Very often, the NMR signal is decomposed into a sum of exponential relaxation curves and the calculated NMR parameters (e.g. R_1, R_2, M_0) are correlated with the property under study. Chemometric techniques, such as analysis of variance (ANOVA) and factor analysis, were shown to be effective means of determining whether a given set of NMR signals contains any interesting information before proceeding to use fastidious and often uncertain signal decomposition procedures. These techniques were applied to NMR signals acquired for spreads and gelatines with different compositions, to mixtures of a cation (Cu^{2+}) and a ligand (tetraphenylporphin) and to glycine solutions at different pH values.
© 1997 John Wiley & Sons, Ltd.

Magn. Reson. Chem. **35**, S13–S21 (1997) No. of Figures: 12 No. of Tables: 0 No. of References: 15

Keywords: NMR; time domain NMR; chemometrics; factor analysis; ANOVA; evolving window ANOVA; evolving window factor analysis; outer product analysis

Received 3 April 1997; revised 2 July 1997; accepted 4 July 1997

INTRODUCTION

Time domain nuclear magnetic resonance (TD-NMR) is often used to quantify major proton-containing constituents in agro-food products or to monitor their evolution during processing. TD-NMR has two major advantages over other instrumental techniques such as Raman, FT-IR and near-IR (NIR) spectroscopy: it is possible to obtain a signal non-invasively from the whole of the sample, not just a superficial layer, and the nature of the signal observed depends on the particular r.f. pulse sequences applied to the sample to excite the protons. The second aspect is at the origin of the wide diversity of NMR spectroscopic, imaging and relaxation techniques that now exist.

So far as TD-NMR is concerned, where the objective is usually the development of a rapid instrumental method of quantification or characterization,[1] this apparently unlimited number of possible signals means that it is often difficult, or at least time consuming, to determine which pulse sequence, if any, produces a signal with information content.

Our objective here is to demonstrate that chemometric techniques, which are already widely used in NIR spectroscopy,[2] can facilitate (1) the detection of information in a TD-NMR signal and (2) the extraction and optimum utilization of that information. Chemometrics has been defined as the application of statistics and informatics to the collection, treatment and analysis of chemical data. The chemometric techniques that will be presented may be divided into univariate and multivariate techniques. Univariate techniques, as the name indicates, analyse only one variable at a time—if a 25 760 point signal is acquired for 33 samples, 25 760 separate univariate calculations will be performed using just 33 values. These univariate techniques therefore require little computer memory and may be very rapid. However, they have the disadvantage of neglecting any correlations or interactions that may exist between the variables.

Since in signals containing information the intensities of neighbouring points are usually strongly correlated, multivariate techniques are often preferred. However, these techniques have the disadvantage of being slower and requiring more computer memory. In the case given above, while an analysis of variance (ANOVA) on three groups of 11 samples each and 25 760 variables only takes 25 s on a Pentium 133 computer with 16 Mb of memory, a multivariate technique such as principal components analysis (PCA), which requires having all $25\,760 \times 33$ data values in memory simultaneously, takes 163 s using the NIPALS algorithm to calculate only the first two principal components. It should also be noted that most of the univariate techniques shown here may be carried out quickly and easily using the built-in functions of a spreadsheet such as Microsoft Excel, whereas factor analysis techniques require third-party add-ons or even a stand-alone statistics program.

Until now, little has been published on the application of chemometrics to TD-NMR signals and only a few teams have directly analysed the signals. Davenel *et al.*[3] applied multivariate statistics to the relaxation curves of doughs during cooking. Gerbanowski *et al.*[4] compared partial least squares (PLS) regression applied to relaxation curves with PLS and multiple linear regression (MLR) applied to the calculated relaxation

* *Correspondence to*: D. N. Rutledge.

© 1997 John Wiley & Sons, Ltd. CCC 0749–1581/97/SI0013–09 $17.50

parameters (T_1, T_2 and initial signal amplitudes). Vackier and Rutledge[5,6] applied univariate and multivariate statistical techniques to both relaxation curves and calculated parameters in order to study the influence of the physico-chemical characteristics of gelatines on their relaxation properties.

Here, two univariate techniques [ANOVA and evolving window (EW) ANOVA] and one multivariate technique [evolving window factor analysis (EW-FA)] are used to detect information in TD-NMR signals. These calculations are applied to TD-NMR signals acquired from various agro-food or model systems with the aim of replacing a classical analytical method by a rapid instrumental method: discrimination between butters and margarines, discrimination between 'light' and 'traditional' butters and margarines, detecting metal–ligand complexation, a study of the influence of pH on chemical exchange processes and the influence of gel strength, pH, sample temperature and water content on relaxation processes in gelatine gels.

METHODS

Analysis of variance

If the samples being examined can be classed into definite groups, it is possible to calculate the amount of the variation in the value of a measurement that is related to its belonging to a particular group.[7] Since one can also calculate the total variability of the measurement, by difference it is possible to calculate the amount of variability that is not due to the groups.

For each group one can then calculate three statistics:

Group variance:

$$V_G = \frac{\sum_{j=1}^{g} n_j(\bar{x}_j - \bar{\bar{x}})^2}{(g-1)} \quad (1)$$

Residual variance:

$$V_R = \frac{\sum_{j=1}^{g}\sum_{i=1}^{n_j}(x_{ji} - \bar{x}_j)^2}{\sum_{j=1}^{g}(n_j - 1)} \quad (2)$$

where

g = number of groups
n_j = number of samples in group j
x_{ij} = value for sample i in group j
$\bar{x}_j$ = mean value for group j
$\bar{\bar{x}}$ = grand mean value
N = total number of samples

The Fisher F value for the variable x may be calculated as

$$F_x = \frac{V_G}{V_R} \quad (3)$$

Although F values are usually used to estimate the level of significance of the influence of the variable x on the factor studied, very high values may be found even when group variance values are not very great, simply because of very small values for residual variance. For that reason, group variance (V_G) values are studied rather than F values.

If the variables studied are in fact points in a signal, such as a TD-NMR relaxation curve, it is interesting to plot these variance values as a function of their position in the signal. Regions in the signal that vary systematically from one group to another will give high group variance values. If there are no important differences between the samples, other than that due to the groups, the residual variance will be low and will not have a structured distribution as a function of position in the signal. The word 'structured' means that the intensities are not distributed in a random fashion but evolve as if determined by information in the signal. It is possible to have an objective measure of this structure or non-randomness. The Durbin–Watson D statistic is often used to determine whether the residuals after a regression are randomly distributed.[8] This statistic is given by

$$D = \frac{\sum_{i=2}^{n}(\delta x_i - \delta x_{i-1})^2}{\sum_{i=2}^{n}(\delta x_i \delta x_i)} \quad (4)$$

where δx_i and δx_{i-1} are the residuals for successive points in a series. For $n > 100$, the distribution is random with a 95% confidence interval for D between 1.7 and 2.3. If we do this calculation on the mean-centred V_G or V_R vectors, we can use D as an *indication* of the non-randomness of the values or, in other words, of the information content of the signals.

Evolving window analysis of variance

In the above studies, each group was compared with all the other groups combined. In some situations, where there is a progressive evolution from one group to the next, it may be interesting to see the incremental effect between adjacent groups. If one has, for example, ten sequential concentration groups, one can calculate nine group variance vectors that will highlight differences due to the progressive changes in concentration.

To do this, we developed a variant of the ANOVA process, EW-ANOVA, where the calculations are performed within a two-group high window that is progressively moved down the data matrix (Fig. 1). The main advantage here is that with this analysis we are able to trace the changes, in certain important variables, as a function of the level of the criterion used to distinguish the groups.

Outer product ANOVA

One sometimes wishes to determine the relationships that exist between two types of signal: near-IR and mid-IR, mid-IR and Raman, UV–visible, NMR, etc. To

© 1997 John Wiley & Sons, Ltd.

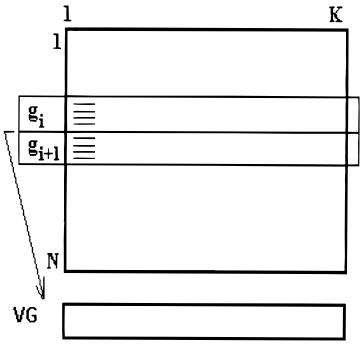

Figure 1. EW-ANOVA procedure.

do this it may be useful to acquire two sets of signals for the same samples and analyze how they vary simultaneously as a function of some property, such as concentration. One possibility is to apply statistical techniques, such as ANOVA, to the n outer product matrices calculated, for each of the n samples, by outer product multiplication of its two signal vectors, of lengths r and c. The n (r rows by c columns) matrices are then unfolded to give n ($r \times c$) long row-vectors.

An ANOVA or EW-ANOVA can be applied to this set of n row-vectors, based on a classification criterion. In the case of EW-ANOVA, each of the $(n-1)$ V_G vectors is then folded back to give $(n-1)$ (r rows by c columns) group variance matrices that may be visualized as V_G contour plots.

Since in this case the ANOVA is being applied to the product of two intensities:

$$z = xy \tag{5}$$

then

$$V_G = \frac{\sum_{j=1}^{g} n_j(\bar{z}_j - \bar{\bar{z}})^2}{(g-1)} \tag{6}$$

The V_G values therefore reflect the simultaneous variations of the x and y variables between groups.

One may assume that the value x_{ij} for individual i in group j is the result of a contribution X_j due to i being a member of group j, and a contribution ε_{ij} due to random noise and the fact that the individual i has other properties which may influence the value of x_{ij}. Similarly, y_{ij} has contributions Y_j and δ_{ij}. Then,

$$z_{ij} = x_{ij}y_{ij} \tag{7}$$

or

$$z_{ij} = (X_j + \varepsilon_{ij})(Y_j + \delta_{ij}) \tag{8}$$

If it is assumed that the contributions ε_{ij} and δ_{ij} are normally distributed, random variables, then $\bar{z}_j$, the average of z for the group j, is given by

$$\bar{z}_j = \frac{\sum_{i=1}^{n_j} X_j Y_j}{n_j} = X_j Y_j \tag{9}$$

and $\bar{\bar{z}}$, the grand average of z for all groups, is given by

$$\bar{\bar{z}} = \frac{\sum_{j=1}^{g}\sum_{i=1}^{n_j} z_{ij}}{\sum_{j=1}^{g} n_j} = \frac{\sum_{j=1}^{g}\sum_{i=1}^{n_j}(X_j Y_j + \varepsilon_{ij} Y_j + \delta_{ij} X_j + \varepsilon_{ij}\delta_{ij})}{\sum_{j=1}^{g} n_j} \tag{10}$$

$$\bar{\bar{z}} = \frac{\sum_{j=1}^{g}\sum_{i=1}^{n_j} X_j Y_j}{\sum_{j=1}^{g} n_j} = \frac{\sum_{j=1}^{g} X_j Y_j}{g} = \overline{XY} \tag{11}$$

$$V_G = \frac{\sum_{j=1}^{g} n_j(\bar{z}_j - \bar{\bar{z}})^2}{(g-1)} = \frac{\sum_{j=1}^{g} n_j(X_j Y_j - \overline{XY})^2}{(g-1)} \tag{12}$$

A high V_G value will be observed when both X_j and Y_j vary significantly from group to group, an intermediate value when only one of them varies and a low value when neither does.

By plotting the group variance matrix as a contour plot or surface, it is possible to highlight those variables in the two domains, e.g. visible spectroscopy and TD-NMR, which vary simultaneously between groups.

Evolving factor analysis

Evolving factor analysis is a multivariate technique which may be used to study the structure of the information contained in the data matrix **X** when it is made up of an ordered sequence of objects.[9]

After transposing the matrix, the rows designate signal intervals and the columns represent the ordered series of objects. Such a matrix may in theory be decomposed by factor analysis into an abstract signal matrix **S*** and an abstract concentration matrix C^*:

$$\mathbf{X} = \mathbf{S}^*\mathbf{C}^* \tag{13}$$

The columns s_i^* of **S*** are mutually orthogonal, linear combinations of the true signal vectors s_j:

$$s_j^* = \sum_{j=1}^{n} \alpha_{ij} s_j \tag{14}$$

while the rows of **C*** are linear combinations of the true concentration vectors. The eigenvalues calculated by factor analysis of the variance/covariance matrix of **X** are of two types: significant eigenvalues associated with the observed variations in the experimental data and other eigenvalues associated with the experimental noise or errors. The eigenvalues associated with significant factors are higher than those associated with noise. When the number of true signal vectors increases, owing to the presence of more chemical components, the significant eigenvalues increase.

© 1997 John Wiley & Sons, Ltd.

Cartwright[10] proposed applying factor analysis to ordered subsets **X*** of the data matrix **X**. The first eigenvalue is calculated when **X*** is just the first three object columns, then the fourth object column is added to give a second eigenvalue. This forward evolving factor analysis process is continued until all the column objects are included. The same process is repeated in the opposite direction (backward evolving factor analysis), starting with the last three column objects. The eigenvalues increase with the dimensionality of $\mathbf{X}^{*T} \cdot \mathbf{X}^{*}$. However, this increase may be non-linear if certain objects have signals that are significantly different from the others. This can be the case when the composition of the sequence of objects is not uniform, such as when a new compound appears. By plotting the evolution of the eigenvalues as new column objects are added, it is possible to monitor the appearance and disappearance of significant factors, and therefore of chemical components. This technique has already been largely used in the study of complexations using UV–visible and ESR spectral data.[11]

Evolving window factor analysis

This variant of EFA consists in performing successive factor analyses on submatrices of fixed size corresponding to the columns within a window that is progressively moved across the data matrix.[12] As in ordinary EFA, the calculated eigenvalues are sensitive indicators of dimensionality, and therefore of the presence of new chemical species or of changes in the behaviour of the samples.

RESULTS AND DISCUSSION

Analysis of variance

The ANOVA procedure was applied to a series of free induction decay (FID), Carr–Purcell–Meiboom–Gill (CPMG) and inversion–recovery (I–R) curves for 'light' and 'traditional' butters and margarines. Signals were acquired on a Minispec pc120 instrument (Bruker) for three samples of each type of spread and were concatenated to give a row vector of 160 intensity values for each sample. A representative vector is shown in Fig. 2. The abscissa is not a time scale and so the 22 equally spaced FID points acquired between 12 and 107 μs are not in scale with the 120 CPMG points between 4 and 480 ms and the 20 inversion–recovery points acquired in the liquid part of the signal, 70 μs after the second pulse, with a variable 180–90° pulse spacing of from 1.54 to 896 ms. The exponential form of this curve is distorted in Fig. 2, giving it a sigmoid shape.

To eliminate the variability in signal intensities due simply to NMR tube filling, all concatenated signals (row vectors) were adjusted by dividing each value by the maximum intensity of that row. The samples were

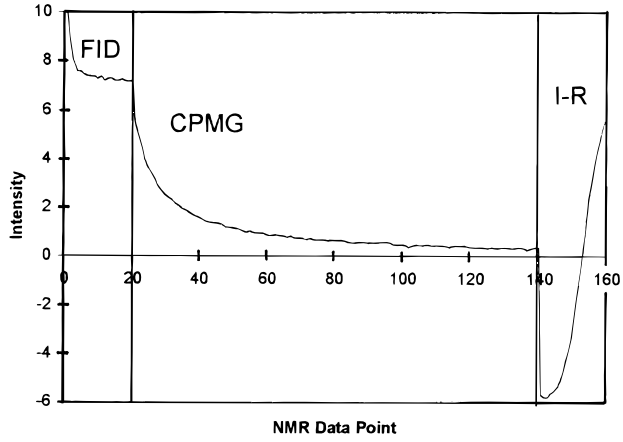

Figure 2. Typical concatenated FID, CPMG and inversion–recovery curves for spreads.

classified into two groups based on whether they were light (high moisture content) or traditional (low moisture content) butters and margarines.

The calculated group and residual variances are plotted in Fig. 3. It is clear that the FID signal is not greatly influenced by this water content-based grouping of the samples whereas the initial part of the CPMG curve and especially the central part of the I–R curve are very strongly influenced. As a first conclusion, one can already say that both the CPMG and I–R signals contain information on the water content of the butters and margarines. By examining the shape of the V_G plot, one can conclude that the difference in water content influences the whole of the CPMG curve. After an initial section close to zero, because the signals have been adjusted to the same maximum, the V_G plot decreases regularly as the CPMG signal intensity decreases. This evolution would be the result of a change in the proportions of the lipidic and aqueous phases. The V_G plot of the I–R curve is maximum in the intermediate region, indicating a shift in the curve due to a change in the relaxation rates or proportions of the longitudinal relaxation components. A measure of the structure in the V_G plots is given by its mean-centred D value of 0.0732. The V_R plot in Fig. 3 has a mean-centred D value of 0.1636, meaning that there is still

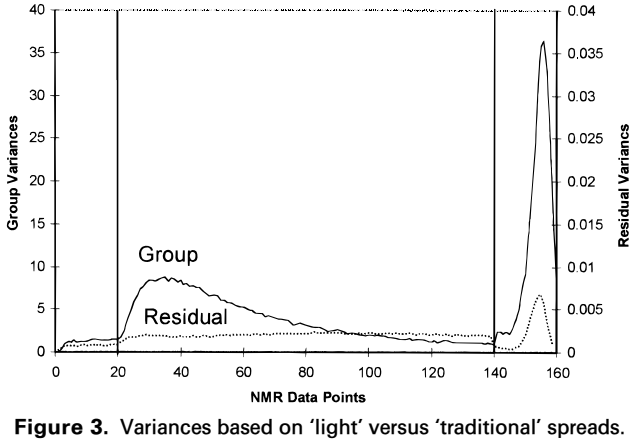

Figure 3. Variances based on 'light' versus 'traditional' spreads.

© 1997 John Wiley & Sons, Ltd.

much information in the signal that is not explained by the water content.

The same set of concatenated signal vectors was classified into two groups based on whether they were margarine (mainly partially hydrogenated plant lipids) or butter (dairy lipids) samples.

The calculated group and residual variances are plotted in Fig. 4, which shows that the FID signal is not influenced by this grouping of the spreads. The group variance values of the CPMG and I–R curves are much lower than before, and the signal-to-noise ratio is lower. The same intermediate part of the I–R curve is important, while it is now the final part of the CPMG which is most influenced by this grouping of the samples. One can therefore conclude that the differences in composition of the samples influence the slower transverse relaxation components (end of the CPMG curve) and the same longitudinal relaxation components. Multiexponential decomposition of the CPMG curves for these samples and the separated lipidic and aqueous phases, using the Marquardt non-linear regression method[13,14] and CONTIN,[15] showed significantly higher T_2 values for the aqueous phase of the margarines (results not shown).

Although the D value of 0.1159 for the mean-centred V_G vector is far from 2, the D value of the mean-centred V_R vector is even lower (0.0813), which means that there is more structure left in the V_R vector than in the V_G vector.

There is a significant resemblance between the V_G vector of one ANOVA and the V_R vector of the other, indicating that these two ways of grouping the samples contain the most part of all the sources of variability among the samples.

Similar calculations were performed on a range of gelatine gels, prepared according to an experimental design where the parameters water content, sample temperature, pH and gel strength were set at the following levels:

Temperature (°C)	15	19.23	25	30.77
pH	4	4.85	6	8
Water content [%(wet basis)]	70	74.23	80	85.77
Gel strength (Bloom)	60	300		

The longitudinal relaxation curve was acquired for the liquid components using a progressive saturation sequence:

$$[90°\text{–}t*(1.23)^{N-1}\text{–}90°\text{–delay–measure}]_N$$

with $N = 50$, $t = 0.5$ ms and delay = 70 μs.

The transverse relaxation curves were acquired using two CPMG sequences with a 90–180° pulse spacing (τ) of 0.625 and 7.5 ms:

$$90°\text{–}\tau\text{–}\{[180°\text{–}2\tau\text{–}]_M\text{–}180°\text{–}\tau\text{–measure–}\tau\}_N$$

The values of M and N for the two CPMG sequences were set to 9 and 168, respectively, for $\tau = 0.625$ ms and to 1 and 56, respectively, for $\tau = 7.5$ ms, in order to have a constant total duration of 1680 ms, and 56 points in common. The long-τ CPMG allows the observation of possible diffusion and chemical exchange effects, while the short-τ CPMG eliminates the effect of these phenomena on the transverse relaxation. The experimental details have been published elsewhere.[5]

ANOVAs based on the four classification criteria (pH, temperature, Bloom and water content) were applied to the concatenated NMR signals (Fig. 5). The group variances calculated for these ANOVAs are presented in Fig. 6.

A number of comments can be made here: (1) Bloom has little effect on any of these relaxation curves; (2) pH has little effect on the P–S relaxation curve and a slightly greater influence on the long-τ than on the short-τ

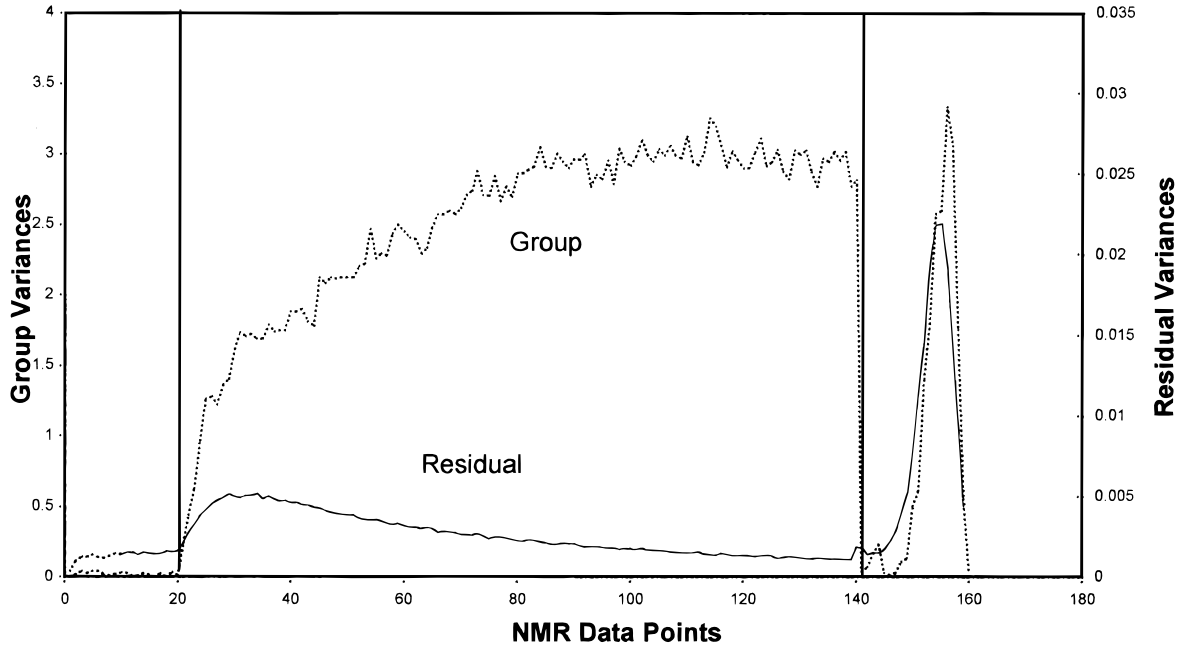

Figure 4. Variances based on butter versus margarine.

© 1997 John Wiley & Sons, Ltd.

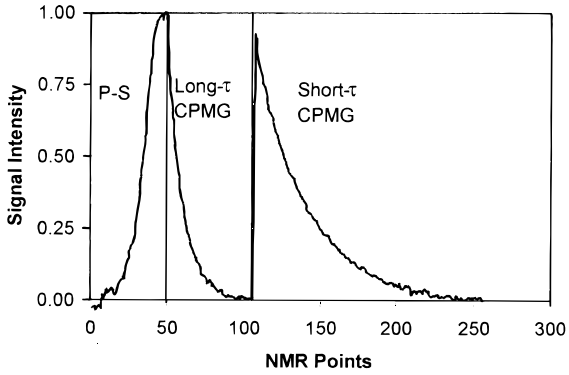

Figure 5. Typical concatenated P–S, long-τ CPMG and short-τ CPMG signals.

CPMG curve; (3) temperature influences the short-τ CPMG much more than the long-τ CPMG and the P–S curves; and (4) water content has a large effect on all signals.

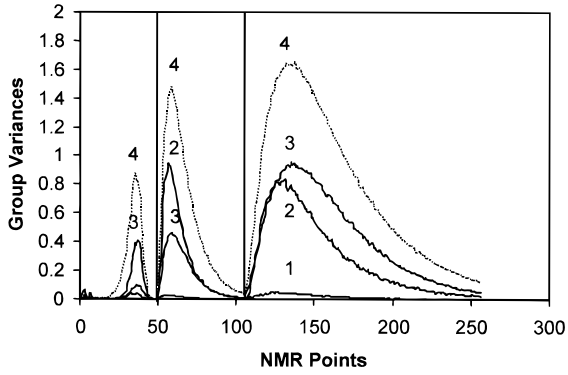

Figure 6. Group variances of ANOVAs based on (1) Bloom ($D = 0.0447$); (2) pH ($D = 0.0198$), (3) temperature ($D = 0.0130$) and (4) water content ($D = 0.0182$).

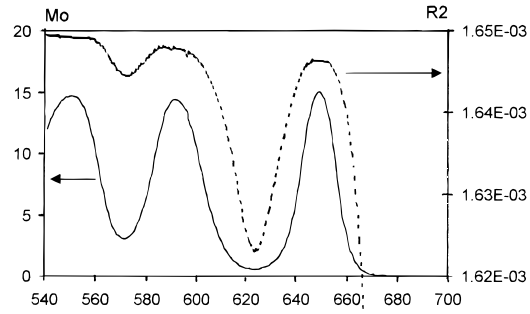

Figure 8. 'Relaxation rates' (R_2, dashed line) and initial intensities (M_0, solid line) estimated by log-linear regression on the V_G curves between the copper and TPP solutions.

These observations may be explained fairly easily; (1) gel strength may modify macroscopic rheological properties but does not influence molecular processes as significantly as do the other factors; (2) chemical exchange processes are catalysed by high and low pHs, which may also increase the intrinsic transverse relaxation rate of the gelatine macromolecules; (3) increasing temperature modifies the transverse and longitudinal relaxation rates of both the water and gelatine molecules; (4) increasing the water content reduces the rigidity of the gel again modifying the relaxation rates of both the water and gelatine molecules.

This rapid exploratory ANOVA analysis of the NMR signals has allowed us to determine which physicochemical or compositional parameters of the gelatine gels may be studied by TD-NMR and has given us some preliminary information on the nature of the phenomena involved.

Outer product evolving window analysis of variance

The example given here is a complexation between a metallic cation, copper and tetraphenylporphin (TPP), a

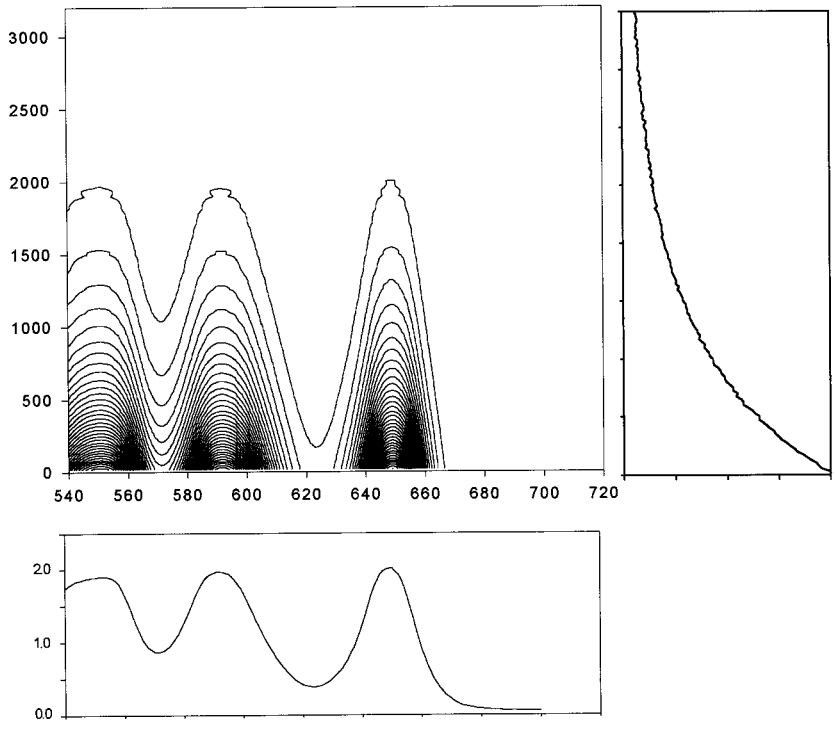

Figure 7. Evolving window ANOVA group variance plot for visible × TD-NMR signals between copper and TPP solutions.

© 1997 John Wiley & Sons, Ltd.

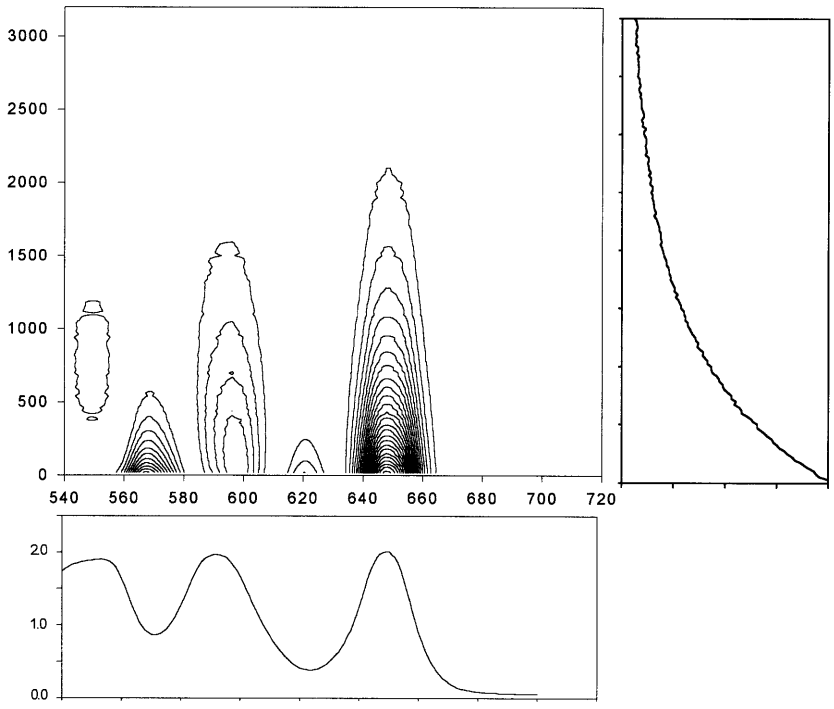

Figure 9. Evolving window ANOVA group variance plot for visible × TD-NMR signals between TPP and copper_TPP solutions.

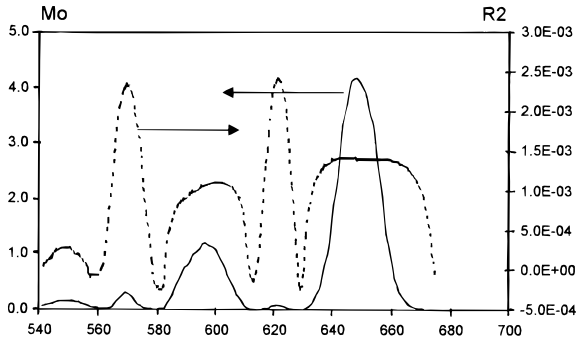

Figure 10. 'Relaxation rates' (R_2, dashed line) in ms^{-1} and initial intensities (M_0, solid line) estimated by log-linear regression on the V_G curves between the TPP and copper_TPP solutions.

ligand with a structure very similar to that of chlorophyll or haemoglobin. Two of the objectives of this study were to determine which regions in the TD-NMR relaxation signals and visible spectra were most influenced by the changes in concentration of the TPP, the copper and the complex, and the relationships between these two types of signals. Three sets of 11 pyridine solutions were prepared at equally spaced concentrations: copper from 0.226 to 0.678 mM, TPP from 0.226 to 0.678 mM, both copper and TPP varying from 0.226 to 0.678 mM but with the total concentration maintained constant at 0.94 mM.

Two data sets were acquired: (1) 160 point CPMG curves from 20 to 3200 ms with a 90–180° interpulse

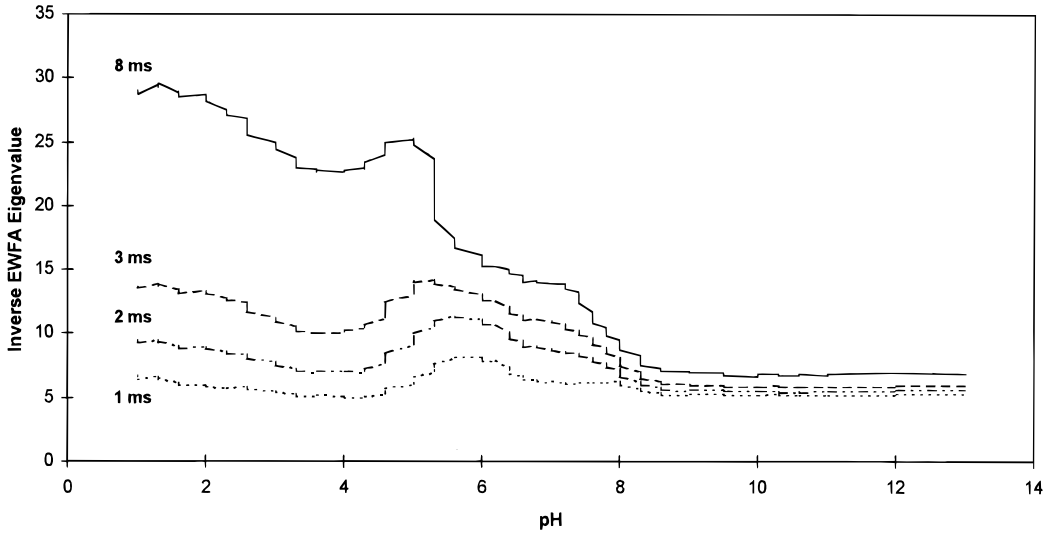

Figure 11. Inverse eigenvalue of first factor from EWFA on CPMG curves of glycine solutions at variables pH, for four different interpulse delays.

© 1997 John Wiley & Sons, Ltd.

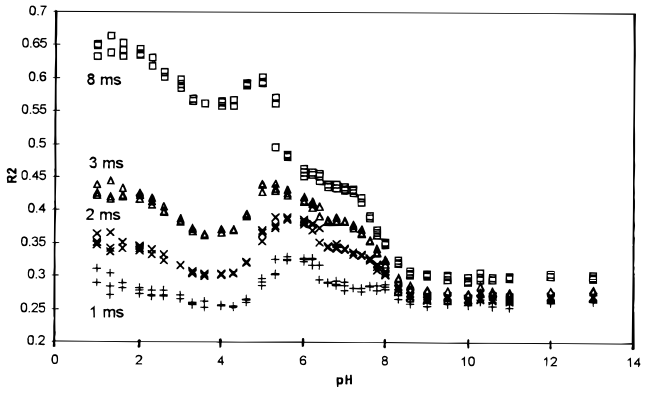

Figure 12. Variation in transverse relaxation rate (R_2) as a function of 180–180° interpulse delay for glycine solutions at different pH values.

delay of 1 ms and a point acquired every 10 echoes; (2) 161 point visible spectra between 700 and 540 nm. The TD-NMR relaxation curves were adjusted by dividing each value in a row by the maximum value in that row. In order to highlight the changes due to the solutes, the matrices of visible spectra and of TD-NMR curves were column centred.

For each sample, the outer product of the visible spectrum vector and TD-NMR vector was calculated to give a 160 rows by 161 columns matrix, which was then unfolded to give a $160 \times 161 = 25\,760$ long row-vector.

An EW-ANOVA was applied to this set of 33 row-vectors, based on the three groups copper, TPP and copper_TPP giving two group variance vectors: copper to TPP and TPP to copper_TPP. Each V_G vector was then folded back to give 2 (160 rows by 161 columns) group variance matrices that were then visualized as V_G contour plots.

The presence of TPP in one set of solutions leads to a variation in the signal intensity at characteristic frequencies. The replacement of copper by TPP also modifies the relaxation properties of the solvent and this is reflected in the high group variance at the beginning of the CPMGs. It is clear from the 2-D group variance plot in Fig. 7 that the TPP peaks at 550, 591 and 648 nm in the visible spectra have relaxation properties which vary in the same way between the group of 11 copper solutions and the group of 11 TPP solutions.

Each 'relaxation curve' in the V_G surface was analysed by log-linear regression, using Microsoft Excel, to estimate the initial amplitude (M_0) and an apparent transverse 'relaxation rate' (R_2) at each frequency in the visible spectrum. As can be seen from Fig. 8, the TPP peaks all have much the same relaxation rates, about 1.645×10^{-3} ms^{-1}.

The 2-D group variance plot in Fig. 9 shows that the differences between the group of 11 TPP solutions and the 11 copper_TPP solutions is not very great at 550 nm, slightly greater at 591 nm and fairly significant at 648 nm. This last EDTA peak is therefore most significantly modified by the complexation with copper. At the same time, two new peaks appear at 619 and 566 nm. As before, each 'relaxation curve' in the V_G surfaces was analysed by log-linear regression to estimate M_0 and R_2 at each frequency in the visible spectrum. As can be seen from Fig. 10, the two new peaks have higher apparent 'relaxation rates' than the TPP peaks (2.4×10^{-3} ms^{-1}), confirming that they are due to the formation of a copper_TPP complex with characteristic absorption bands in the visible region, and which relaxes the solvent more quickly than does the TPP.

It should be stressed, however, that although they facilitate the detection of UV–visible peaks due to the complex, these apparent R_2 values, calculated from the V_G surfaces, are not the true transverse relaxation rates of the various compounds in the solutions. The true transverse relaxation rates were estimated, using Excel, by linear regression on the log-transformed CPMG relaxation curves. Only the faster relaxing components, in the first 600 ms, were considered. The values for all the TPP solutions are stable at about 0.45×10^{-3} ms^{-1}, for copper they increase regularly from 0.9×10^{-3} to 1.7×10^{-3} ms^{-1}, while for the copper_TPP solutions the values remain stable at about 0.5×10^{-3} ms^{-1} up to the equimolar solution and then, once the copper is in excess, they increase regularly to attain 1.20×10^{-3} ms^{-1}.

Evolving window factor analysis

CPMG curves were acquired in triplicate, using four different 90–180° interpulse delays (τ) for glycine solutions between pH 1 and 13. For each value of τ (1, 2, 3 and 8 ms), the ordered matrix of CPMG curves was analysed by evolving window factor analysis with a three-column window and the inverse eigenvalue of the first factor was plotted against the pH of the central column (Fig. 11). Not only does this plot show the evolution of the variability in the relaxation curves as a function of pH, but, as can be seen from Fig. 12, it is also very similar to the standard plot of the transverse relaxation rate, R_2 calculated by non-linear regression,[14] against pH. The jagged shape of the inverse eigenvalue plot is due to the fact that when the three-column window is applied to triplicate repetitions of each pH, the first eigenvalue is influenced by the preceding pH, the last by the following pH and only one includes all three repetitions at the same time. This problem can be eliminated by only retaining the central eigenvalue.

It is possible (results not shown) to calculate the transverse relaxation rates for a limited number of selected CPMG curves and then establish a linear correlation model between the R_2 values and the inverse eigenvalues. This model can then be used to predict all the other R_2 values. This procedure is not only much quicker and less prone to error than performing non-linear decomposition on all of the curves, it is also a way of smoothing the data.

CONCLUSION

The application of chemometrics to TD-NMR signals can be an efficient and rapid method of determining the

© 1997 John Wiley & Sons, Ltd.

information content of the signals. As NMR is a very versatile technique that can produce an almost unlimited range of signals depending on the pulse sequences used, unsupervised procedures based on these statistical methods should simplify the development of analytical procedures.

Many new developments are to be expected in this field, as is exemplified by the outer product matrix where the analysis of combined signals can give information on the simultaneous variations in the two signal domains.

Acknowledgements

We thank the INRA for partial financing of the TD-NMR instrument and Dr Max Feinberg for fruitful discussions and help with the relaxation parameter calculations.

REFERENCES

1. D. N. Rutledge, *J. Chim. Phys.* **89**, 273 (1992).
2. B. G. Osborn and T. Fearn, *Near Infrared Spectroscopy in Food Analysis*. Longman, New York (1986).
3. A. Davenel, P. Marchal and J. P. Guillement, in *Magnetic Resonance in Food Science*, edited by P. S. Belton, I. Delgadillo, A. M. Gil, G. A Webb, pp. 146–155. Royal Society of Chemistry, Cambridge (1995).
4. A. Gerbanowski, D. N. Rutledge, M. Feinberg and C. Ducauze, *Sci. Aliments* **17**, 309 (1997).
5. M. C. Vackier and D. N. Rutledge; *J. Magn. Reson. Anal.* **2**, 321 (1996).
6. M. C. Vackier and D. N. Rutledge, *J. Magne. Reson. Anal.* **2**, 311 (1996).
7. J. Czerminski, A. Iwasiewicz, Z. Paszk and A. Sikorski, *Statistical Methods in Applied Chemistry*, pp. 186–207 Elsevier, Amsterdam (1990).
8. J. Durbin and G. S. Watson, *Biometrika* **37**, 409 (1950).
9. H. Gampp, M. Maeder, C. J. Meyer and A. D. Zuberbühler, *Talanta* **33**, 943 (1986).
10. H. Cartwright, *J. Chemom.* **1**, 111 (1987).
11. R. Tauler and E. Casassas, *J. Chemom.* **3**, 151 (1988).
12. E. R. Malinowski, *J. Chemom.* **6**, 29–40 (1992).
13. D. W. Marquardt, *J. Soc. Ind. Appl. Math.* **11**, 431 (1963).
14. D. N. Rutledge, in *Signal Treatment and Signal Analysis in NMR*, edited by D. N. Rutledge, pp. 191–217. Elsevier, Amsterdam (1996).
15. S. W. Provencher, *Comput. Phys. Commun.* **27**, 213 (1982).

© 1997 John Wiley & Sons, Ltd.