JOURNAL OF CHEMOMETRICS
*J. Chemometrics* 2006; **20**: 198–208
Published online 21 November 2006 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/cem.991

# An exploratory chemometric study of $^1$H NMR spectra of table wines

**Flemming H. Larsen, Frans van den Berg and Søren B. Engelsen***
Quality & Technology, Department of Food Science, The Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

Received 30 September 2005; Revised 30 May 2006; Accepted 7 June 2006

**In this paper an exploratory study of 40 table wines by proton nuclear magnetic resonance spectroscopy and chemometric region-selection methods is presented. Several components of wine have been identified and quantified. It is demonstrated how signal alignment procedures were utilized to compensate for pH effects in the NMR spectra prior to the chemometric data modeling. The analysis included region selection by interval partial least squares for regression to reference data obtained from infrared spectroscopy. Accurate calibration models to the contents of ethanol, glycerol, lactic acid, methanol and malic acid were established. For the more general combined glucose/fructose infrared reference backwards interval partial least squares was introduced for optimal interval selection in calibration. The aim of the paper is to show how pre-processing and region-selection methods can assist in interpretation and quantification of chemical shift multiplets in $^1$H NMR spectra of complex biological systems. The extension to NMR metabolomics is straightforward.** 

**KEYWORDS:** $^1$H NMR spectroscopy; wine; chemometrics; warping; interval PLS

## 1. INTRODUCTION

Wine is a relatively complex biochemical system primarily consisting of water, ethanol, organic acids, and carbohydrates. In addition to the main constituents a range of aromatic compounds, polyphenols and colorants are present in small quantities. They are important ingredients with respect to quality characteristics such as taste and color [1]. In recent years the composition of wine has been the subject of increased focus because of demands from consumers regarding the quality and authenticity. With respect to authenticity, the isotope-based SNIF-Nuclear Magnetic Resonance (NMR) method [2], by which the deuterium ($^2$H) to hydrogen ($^1$H) ratio is determined with high accuracy, is used in order to obtain an isotopic fingerprint of ethanol. This ratio depends on the origin of the ethanol as well as on the type and geographical origin of the grape. No information about other components in the wine is obtained via this method. $^1$H NMR spectroscopy is a strong tool for assessing additional wine constituents, such as aromatics, carbohydrates and acids can be detected and quantified by this method. However $^1$H NMR spectra of even rather simple single-phase foods often result in complex spectra. For this reason it is advantageous to analyze the spectra by multivariate methods like those developed in the field of chemometrics. A quality parameter of particular interest for wine is the ratio of malic to lactic acid, essential for the sourness or flatness. To adjust the ratio between these acids the wine producers employ malolactic fermentation. This process is usually run to completion, as it is difficult to stop, and malic acid is thus completely converted into lactic acid. By mixing malolactic fermented wine with non-fermented wine the desired ratio of these acids is obtained.

The ability of chemometric methods to align NMR spectra enables easier and more reliable identification of components and thereby improved quantification using for example, Partial Least Squares (PLS) factor models. Furthermore, advanced versions of PLS like interval PLS (iPLS) and backward interval PLS (biPLS) serve as efficient tools for region selection. For the exploration of NMR data, region selection is mandatory in order to include a full multiplet (due to J-couplings) and is thus preferable to single variable/sample-points selection algorithms. In NMR spectroscopy one single chemical component can manifest itself at different locations on the chemical shift axis. NMR spectroscopists can interpret the dominant and well-defined spectral features but not necessarily less intense resonances. Advanced chemometric methods such as iPLS and biPLS can assist in interpretation of such diverse and information

*Correspondence to: S. B. Engelsen, Quality & Technology, Department of Food Science, The Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark.
E-mail: se@kvl.dk

rich signals, especially in case of a limited number of samples or when the relation between the independent block (NMR spectra) and the dependent block (IR measured reference values) is not strong. In this context iPLS is to be considered as an extension of the exploratory nature of chemometric data analysis.

Previously, the combination of chemometrics and $^1$H NMR spectroscopy has successfully been employed for analysis of beer [3,4], table wine [5] and fruit juice [6]. Wine and beer are more complex systems than fruit juice, but slightly less complex compared to blood and urine in regards to information contents and diversity [7,8].

The present work is an explorative study of $^1$H high-resolution NMR spectra of 40 different wines. Exploratory analysis of the data will be used to confirm our (*a priori*) knowledge about the chemical shifts of the various components. In order to remove the major part of chemical shift scatter induced by pH differences peak alignment procedures have to be employed [9–14]. Specifically we will employ Co-shifting and Correlation Optimized Warping (COW) [9,10] for signal alignment in this paper, removing pH-induced chemical shift dependencies in the NMR spectrum. Principal component analysis (PCA) and PLS factor models will be used for analysis, and (backwards) interval PLS [15,16] will be used for range/variable identification and selection. Several components in wine will be calibrated against reference measurements from infrared (IR) spectroscopy and a few representative examples will be treated in detail to demonstrate the capabilities of the methods employed.

**Table I.** List of wines

| No. | Name | Country | Type |
|---|---|---|---|
| 1 | Gewurztraminer, Martin Zahn, Vin Dálsace. | Germany | White |
| 2 | Maison Guillot, Merlot, Vin de Pays Dóc. | France | Red |
| 3 | Col di Livido, Copertino Riserva, (Negroamaro/Malvasia/Montepulciano). | Italy | Red |
| 4 | Sylvaner, Dietrich, Vin Dálsace. | Germany | White |
| 5 | Sabika, Tempranillo/Cabernet sauvignon, Utiel-Requena. | Spain | Red |
| 6 | Muscadet de Sevre et Maine, Loire. | France | White |
| 7 | Santa Rosana, Mendoza, Cabernet sauvignon/Merlot. | Argentina | Red |
| 8 | Vendange, Saint-Chinian rosé. | France | Rosé |
| 9 | Hardy VR, Shiraz. | Australia | Red |
| 10 | Pinot Noir, Martin Zahn, Vin D'Alsace. | Germany | Red |
| 11 | Il Tasso Chianti, Toscana, Sangiovese/Canaiolo. | Italy | Red |
| 12 | Torres, Viña Esmeraldi. | Spain | White |
| 13 | False Bay, Cabernet sauvignon/Cinsault, Stellenbosch. | S. Africa | Red |
| 14 | Savanha, Sulanga, Chardonnay. | S. Africa | White |
| 15 | Santa Ana, Cabernet sauvignon/Merlot, Mendoza. | Argentina | Red |
| 16 | Quintana, Vinho de mesa tinto. | Portugal | Red |
| 17 | Sylvaner, Laugel, Vin Dálsace. | Germany | White |
| 18 | Hardy VR, Cabernet sauvignon rosé. | Australia | Rosé |
| 19 | Riesling, Laugel, Vin Dálsace. | Germany | White |
| 20 | Château Fourcas Dupre, Appellation Listrac Médoc-Contrôlée | France | Red |
| 21 | Pierre André, Appellation Bourgogne Contrôlée 100% Pinot-Noir grapes | France | Red |
| 22 | A.Amarone della Valpoicella Clássico. Denominazione di origine controllata. Mix of Molinara Corvina Rondinella | Italy | Red |
| 23 | Domaine Guillaumette. Appellation Costières de Nîmes Contrôlée. Mic of 70% Syrah and 30% Grenache. | France | Red |
| 24 | Finca La Celia. Mix of Cabernet Sauvignom y Tempranillo | Argentina | Red |
| 25 | Los Molinos. Felix-Solis. CRDO Valdepeñas. | Spain | Red |
| 26 | Chateau Toutigeac. Appellation Bordeaux Contrôlée. Mix of Cabernet Franc, Cabernet Sauvignon and Merlot. | France | Red |
| 27 | Tegole. Indicazione Geografica Tipica Toscana. | Italy | Red |
| 28 | Farina. Amarone della Valpoicella Classico. Denominazione di origine controllata | Italy | Red |
| 29 | Same as wine 21 | France | Red |
| 30 | Otonal. Bodegas Olarra. Denominacion Origen Calificada RIOJA | Spain | Red |
| 31 | Glen Ellen. California. Cabernet Sauvignon | USA | Red |
| 32 | Puteus. Denominazione di origine controllata Salice Salentino. | Italy | Red |
| 33 | Castillo de Molina. Cabernet Sauvignon. | Chile | Red |
| 34 | Baron Charles- Louis. Appellation Mercurey Contrôlée (Bourgogne) | France | Red |
| 35 | Warburn. Shiraz | Australia | Red |
| 36 | Agramont. Denominacion de Origen Navarra. Mix of Cabernet Sauvignon-Tempranillo. | Spain | Red |
| 37 | Plovdiv Region. Cabernet Sauvignon | Bulgary | Red |
| 38 | Vidigal. Vinho Regional Estremadura. Reserva | Portugal | Red |
| 39 | Same as wine 22 | France | Red |
| 40 | Zalze. Wine of Origin Western Cape. Mix of Cinsault, Ruby Cabernet and Cabernet Sauvignon. | S. Africa | Red |



DOI: 10.1002/cem

## 2. EXPERIMENTAL

$^1$H NMR spectra were acquired on a Bruker Avance 400 spectrometer, operating at 400.13 MHz for protons, equipped with a 5 mm BBI probe with Z-gradients. All experiments were performed at 298 K using the *water* pulse sequence [17] that allows for efficient suppression of the water resonance by pre-saturation followed by a composite pulse. All spectra were acquired using a recycle delay of 5 s, 1536 scans and a dwell time of 60.4 μs for acquisition of 32 k data points resulting in a total acquisition time of 1.979 s. Before Fourier transformation the data set were zero filled to 64 k points and apodized by Lorentzian line-broadening of 0.3 Hz. Spectra were processed in Xwin-NMR and thereafter analyzed using Matlab with in-house routines and unscrambler. Samples were prepared from 495 μl wine and 55 μl of TSP-d4 (5.8 mM) in $D_2O$. All spectra are referenced to TSP-d4 (0.0 ppm) prior to warping. The IR-data are obtained using the WineScan instrument from Foss Analytical A/S. A complete list of the 40 wines is given in Table I.

To correct for spectral misalignments due to pH differences in the wines two signal pre-processing methods were employed: Correlation optimized shifting (called Co-shifting) and Correlation Optimized Warping (COW) [9,10]. In Co-shifting the sample spectra are aligned towards a reference spectrum by simple 'left-right' shifting of the signal vector in discrete steps on the equidistant, discrete signal index within a predetermined window. For each position the correlation between sample and reference is computed and the optimal alignment pre-processing is expressed by the maximum correlation coefficient. The same approach is taken in COW, except that sample and reference signal are divided into segments (usually of equal length) and a slack window around the segment boundaries is defined. The optimal alignment pre-processing is found via a linear programming optimization [10]. Co-shifting could be categorized as a linear correction, whereas COW can be seen as a more flexible non-linear (or segment wise) pre-processing method.

iPLS is employed for region selection in the information-rich $^1$H NMR spectrum [11]. This is a variation of the normal PLS regression method aimed at variable interval or region selection with a strong emphasis on visual exploration of spectroscopic data. When using iPLS, the NMR spectrum is divided into rather arbitrary segments, usually of equal size. The next step builds regression models—typically represented by leave-one-object-out cross-validation prediction errors—and compares the predictive performance of the full-spectrum model with all the local one-segment models. This gives the analyst an immediate indication of which regions are correlated with the dependent/response variable in the regression equation. However, since the segments are chosen on an arbitrary basis, user interaction is required to safeguard against chance correlations or to replace the segment boundaries slightly into more sensible positions, for example, to capture a full multiplet in NMR spectroscopy. Note that the iPLS approach differs from the so-called binning method for data reduction. The segmentation step in iPLS is not a reduction of the number of variables; it is merely a technique to obtain an overview of a large number of

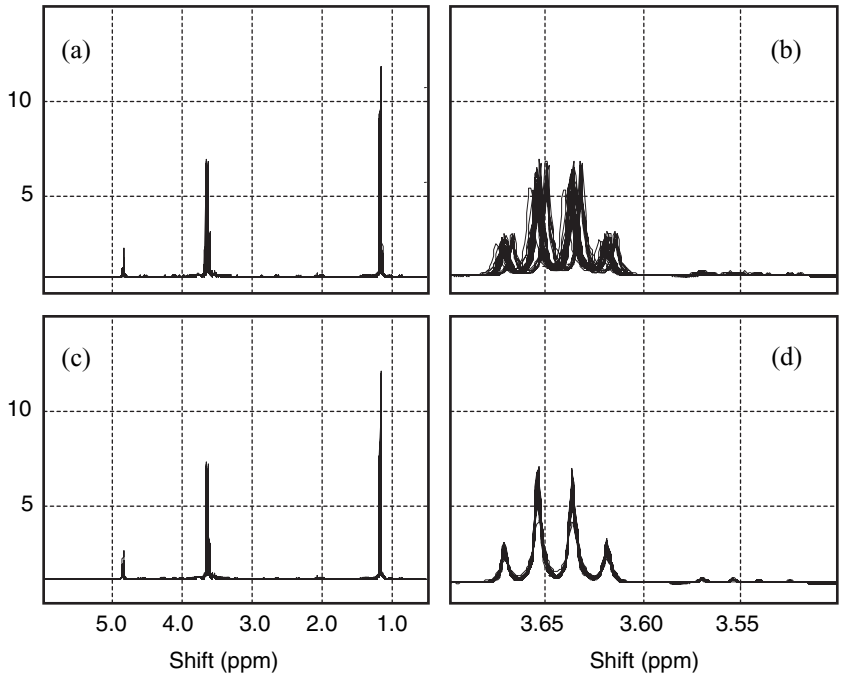


**Figure 1.** (a) Raw (uncorrected) $^1$H NMR spectra of 40 table wines. (b) Zoom in raw spectra to show the pH-related shift problem. (c) Correlation Optimized Warping (COW)-corrected (segment = 100 points = 0.0631 ppm; slack = 25 points = 0.0158 ppm) $^1$H NMR spectra. (d) Zoom in COW-corrected spectra to show improvement.



DOI: 10.1002/cem

(possibly diverse) variables in complex data sets such as high-resolution NMR spectra and to avoid unnecessary interferences. Once the interesting region(s) is (are) detected, all spectroscopic information is available for example, alignment by Co-shifting or COW to improve the interpretation could be the next step, a route that is pursued in this manuscript.

Backward interval PLS (biPLS) [16] is a further development of the concepts behind iPLS, designed to assist the user in the selection or optimization of several spectral segments for prediction. It functions similarly to iPLS, but discards one segment at a time and employs genetic algorithm to select an optimal combination of a predefined number of segments. We illustrate the potential of biPLS in this $^1$H NMR study of wines to improve the quantitative model for combined glucose/fructose content.

## 3. RESULTS AND DISCUSSION

The table wines in this study are analyzed 'as is', except for the addition of $D_2O$ and chemical shift reference (TSP-d4). The experiments have therefore been performed at the wines' natural pH value. A wide range of highly specific information is obtained by $^1$H NMR spectroscopy. One of the problems in analyzing complex foods and beverages, or biological samples in general, is to assign and quantify the resonances of individual protons. By far the most dominant signal in $^1$H NMR of wines stems from ethanol and water. Figure 1a shows the full spectral range used in this paper (0.5–6.0 ppm) and Figure 1b zooms in on the ethanol quartet at 3.65 ppm. A large shift in the peak positions, which is related to the pH differences, is observed. To correct this alignment problem, COW was applied to the whole spectral

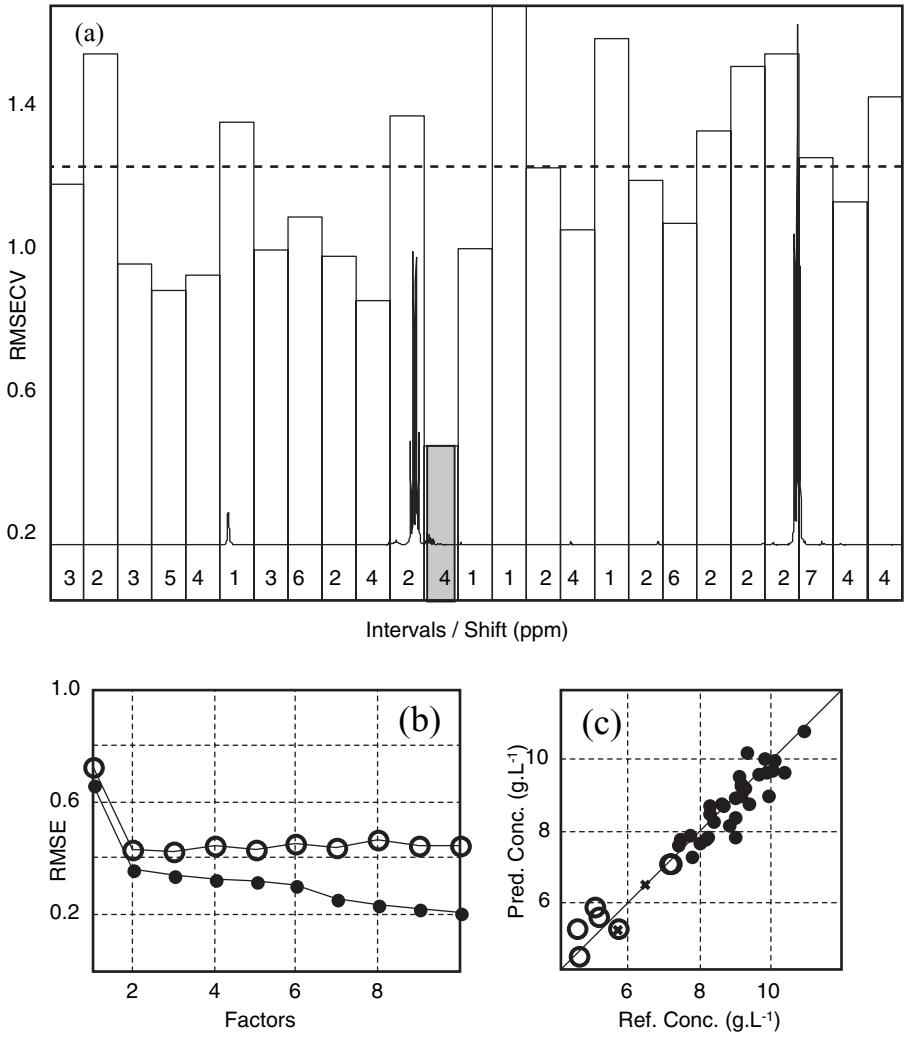


**Figure 2.** (a) Interval Partial Least Squares (iPLS) plot Root Mean Square Error of Cross-Validation prediction (RMSECV) versus interval number (COW-corrected spectra), superimposed on the average spectrum for glycerol prediction; digits indicate number of factors used in local PLS model. (b) Leave-one-sample-out cross-validation (○) and fit (●) RMSE as a function of PLS model complexity predicting glycerol contents in wine from spectral range 3.5–3.6 ppm. (c) Cross-validation predicted versus reference glycerol contents COW-corrected spectra for red (○), white (●) and (×) rosé wines.



DOI: 10.1002/cem

**Table II.** PLS calibration model results for various components in the wines, based on leave-one-out cross-validation. F = number of Factors used, $R^2$ = squared correlation coefficient × 100%, RMSECV = Root Mean Squared Error of Cross-Validation prediction, Mean/std. = mean and standard deviation of the reference values

| Component Mean/std ($g \cdot L^{-1}$) | Model type | Spectral range (ppm) | F | RMSECV ($g \cdot L^{-1}$) | $R^2$ (%) | Assignment |
|---|---|---|---|---|---|---|
| Glycerol 8.12/1.66 | Aligned data | 0.5–6.0 | 2 | 1.22 | 41 | $CH_2 + CH$ |
| | iPLS* | 3.5–3.6 | 2 | 0.43 | 92 | $CH_2$ |
| Ethanol 12.72/1.00 | Raw data | 0.5–6.0 | 7 | 0.18 | 94 | $CH_3 + CH_2$ |
| | Aligned data | 0.5–6.0 | 3 | 0.44 | 69 | $CH_3 + CH_2$ |
| | iPLS* | 3.6–3.7 | 6 | 0.36 | 79 | $CH_2$ |
| | iPLS* | 1.10–1.25 | 6 | 0.31 | 84 | $CH_3$ |
| Lactic acid 1.03/0.51 | Aligned data | 0.5–6.0 | 4 | 0.44 | 27 | $CH_3 + CH$ |
| | iPLS* | 1.3–1.6 | 3 | 0.10 | 96 | $CH_3$ |
| Malic acid 0.67/0.76 | Aligned data | 0.5–6.0 | 1 | 0.70 | 12 | $CH_2 + CH$ |
| | iPLS* | 2.7–2.9 | 1 | 0.17 | 94 | $CH_2$ |
| Methanol 0.13/0.07 | Aligned data | 0.5–6.0 | 2 | 0.05 | 48 | $CH_3$ |
| | iPLS* | 3.3–3.4 | 4 | 0.03 | 75 | $CH_3$ |
| Glucose/fructose 2.90/4.20 | Aligned data | 3.9–4.7 | 2 | 1.22 | 73 | $CH_2 + CH$ |
| | biPLS** | 1: 4.09–4.12<br>2: 4.22–4.25<br>3: 4.12–4.15<br>4: 3.96–3.99 | 4 | 0.96 | 83 | $CH_2 + CH$ |

* Model based on ppm-range selected by iPLS for full range, adjusted by visual inspection around the interval selected.
** Model based on ppm-ranges automatically selected by biPLS; ordered according to importance.

range. A clear improvement in appearance is evident in Figure 1c–d.

To illustrate the potential of iPLS in exploratory data analysis a model for glycerol prediction in the 40 wines is made using 25 equidistant intervals. Figure 2a shows the result; using the entire spectral range (0.5–6.0 ppm) a PLS model to glycerol with two factors gives a leave-one-sample-out cross-validation error of 1.22 $g \cdot L^{-1}$ (Figure 2a, horizontal broken line) and a squared correlation coefficient between prediction and IR reference of 0.41 (see Table II). However, the results indicate a better prediction result for segment number 12, and visual inspection of the iPLS model results and data suggests the use of the interval 3.5–3.6 ppm for glycerol quantification. This interval is in accordance with the anticipated chemical shift for half of the $CH_2$-protons in glycerol around 3.56 ppm [8]. Using this interval resulted in a

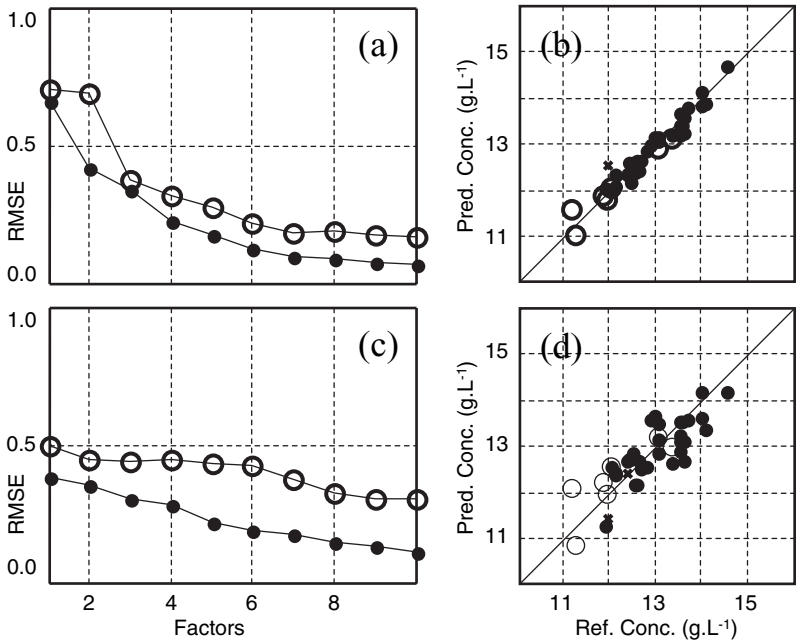


**Figure 3.** (a) Leave-one-sample-out cross-validation and fit RMSE as a function of PLS model complexity for raw (uncorrected) spectra predicting ethanol contents in wine. (b) Cross-validation predicted versus reference ethanol contents for raw spectra. (c) Cross-validation and fit RMSE as a function of PLS model complexity for COW-corrected spectra predicting ethanol contents. (d) Cross-validation predicted versus reference ethanol contents for COW-corrected spectra.



DOI: 10.1002/cem

significantly improved 2-factor model with an error of $0.43\,g \cdot L^{-1}$ (Figure 2b) and a squared correlation coefficient of 0.92 (Table II). It is noted that red wine contains more glycerol than white and rosé wines—in some cases almost twice as much (Figure 2c).

To illustrate the difference between exploratory data analysis and pure quantification Figure 3 shows the results of PLS calibration models for ethanol on raw (un-aligned) and COW-aligned data. The 7-factor model for the raw data has a leave-one-sample-out cross-validation error of $0.18\,g \cdot L^{-1}$ and a squared correlation coefficient between predicted and IR reference of 0.94 (Table II). In contrast, the alignment-corrected data yielded a 3-factor prediction model that performs worse: $0.44\,g \cdot L^{-1}$, 0.69 (Table II). For this major/

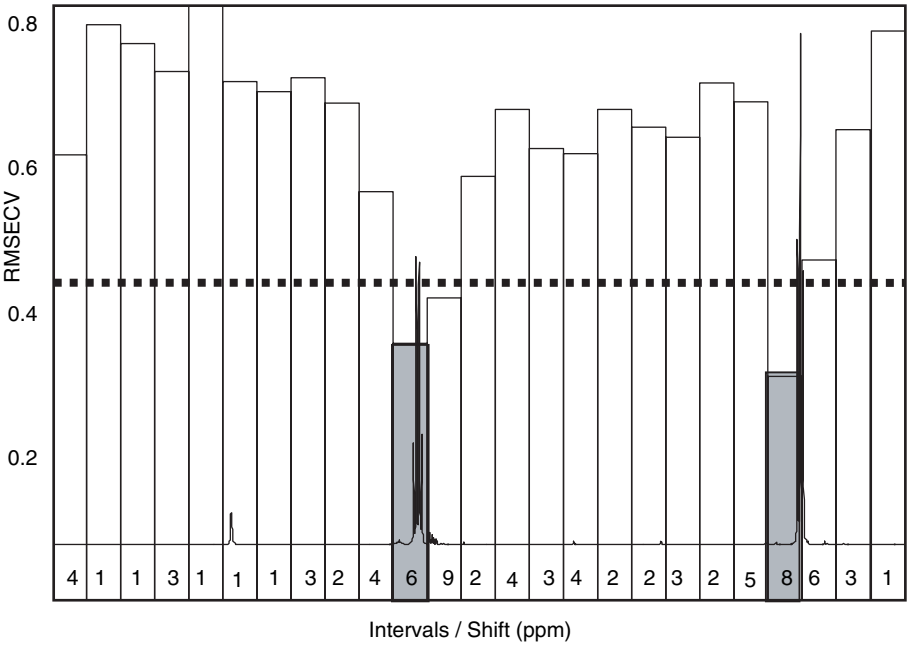


**Figure 4.** iPLS plot RMSECV versus interval number (COW corrected spectra), superimposed on the average spectrum for ethanol prediction. In each interval the number of factors for the automatically selected PLS model is displayed. Horizontal line is global error for 3-factor model.

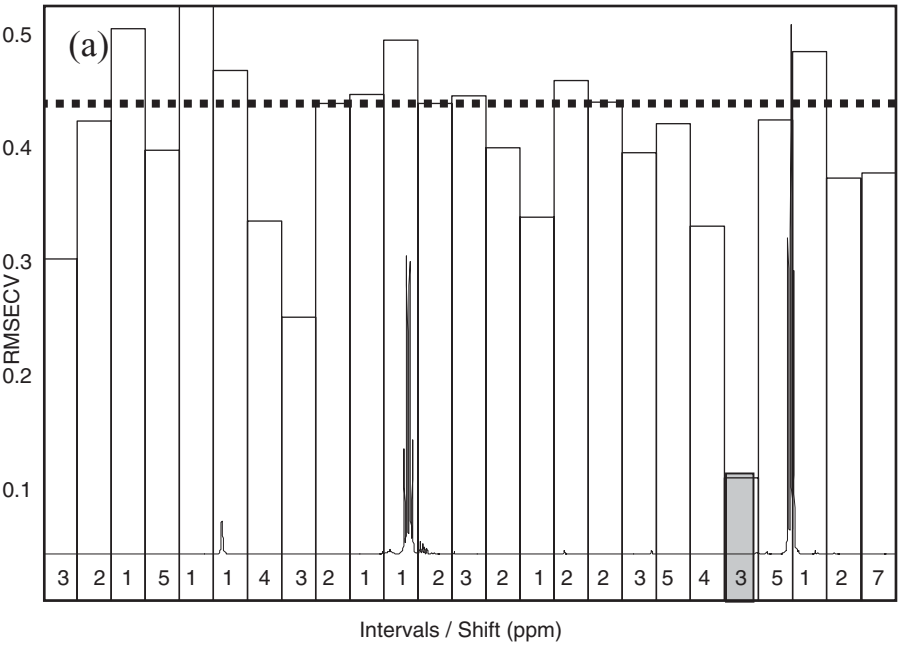


**Figure 5.** (a) iPLS plot RMSECV versus interval number (COW-corrected spectra), superimposed on the average spectrum for lactic acid prediction. (b) Leave-one-sample-out cross-validation and fit RMSE as a function of PLS model complexity predicting lactic acid contents in wine from spectral range 1.3–1.6 ppm. (c) Cross-validation predicted versus reference lactic acid contents.



DOI: 10.1002/cem

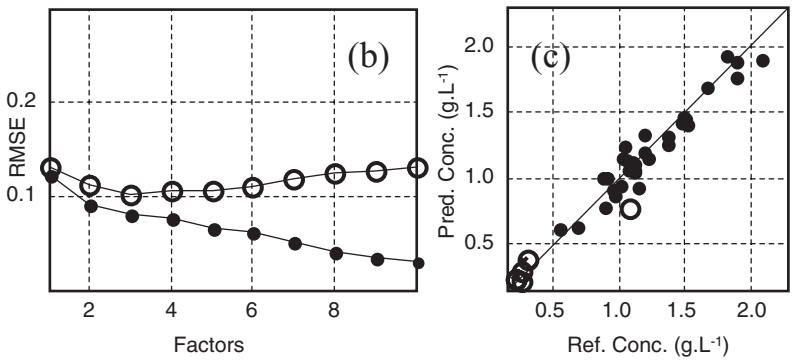


**Figure 5.** (Continued)

dominant component the alignment was thus not beneficial. The many-factor model for the raw spectra was required to correct for the non-linearity in the signal introduced primarily by peak misalignment. This non-linearity is supported by an iPLS analysis for ethanol on the alignment-corrected data set (see Figure 4). The important spectral areas are readily identified, but the local models for the quartet (3.65 ppm) and triplet (1.17 ppm) still have a high model complexity of six PLS factors (Table II). This indicates that the leave-one-out cross-validation method might not be sufficiently rigorous as a validation method in order to prevent overfitting for the ethanol signals in these wine samples. However, this subject will not be further pursued in this paper. We will only focus on minor components like glycerol for which the signal contributions of the aligned COW spectra were more parsimonious and better interpretable.

Besides glycerol and ethanol, other components in wine include small organic acids that are very important for the taste profile of the wine. One of these is lactic acid, and the results of PLS quantification are shown in Figure 5 and Table II, a significant improvement obtained by alignment of the spectra. A detailed picture of the methyl resonance in lactic acid is shown in Figure 6. Even though the full data set has been warped, the doublet of $CH_3$ is still not well aligned due to its strong dependence on pH. Next to this resonance is the triplet originating from the $^{13}C$ coupled $CH_3$-group in ethanol which is very well aligned (Figure 6c,d). By additional Co-shifting of the lactic acid part of the spectrum an excellent alignment is obtained. Using *a priori* chemical

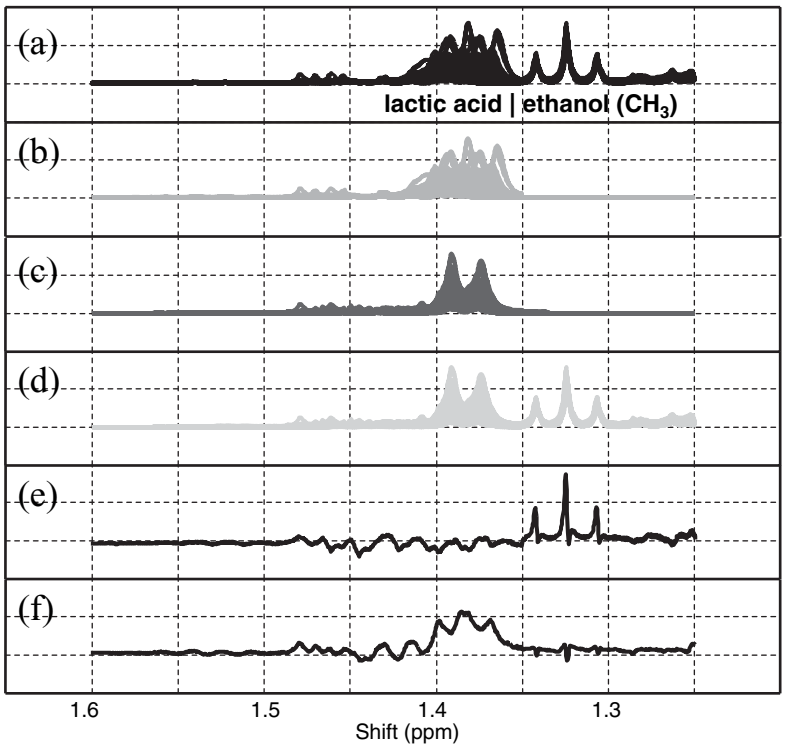


**Figure 6.** (a) Lactic acid and ethanol region before a global alignment correction. (b) Region below 1.35 ppm is set to zero before local co-shift correction by 'cutting' ethanol triplet. (c) Local co-shift alignment of the lactic acid region. (d) Merging of the two corrected regions. (e) PLS regression vector for ethanol. (f) PLS regression vector for lactic acid.



DOI: 10.1002/cem

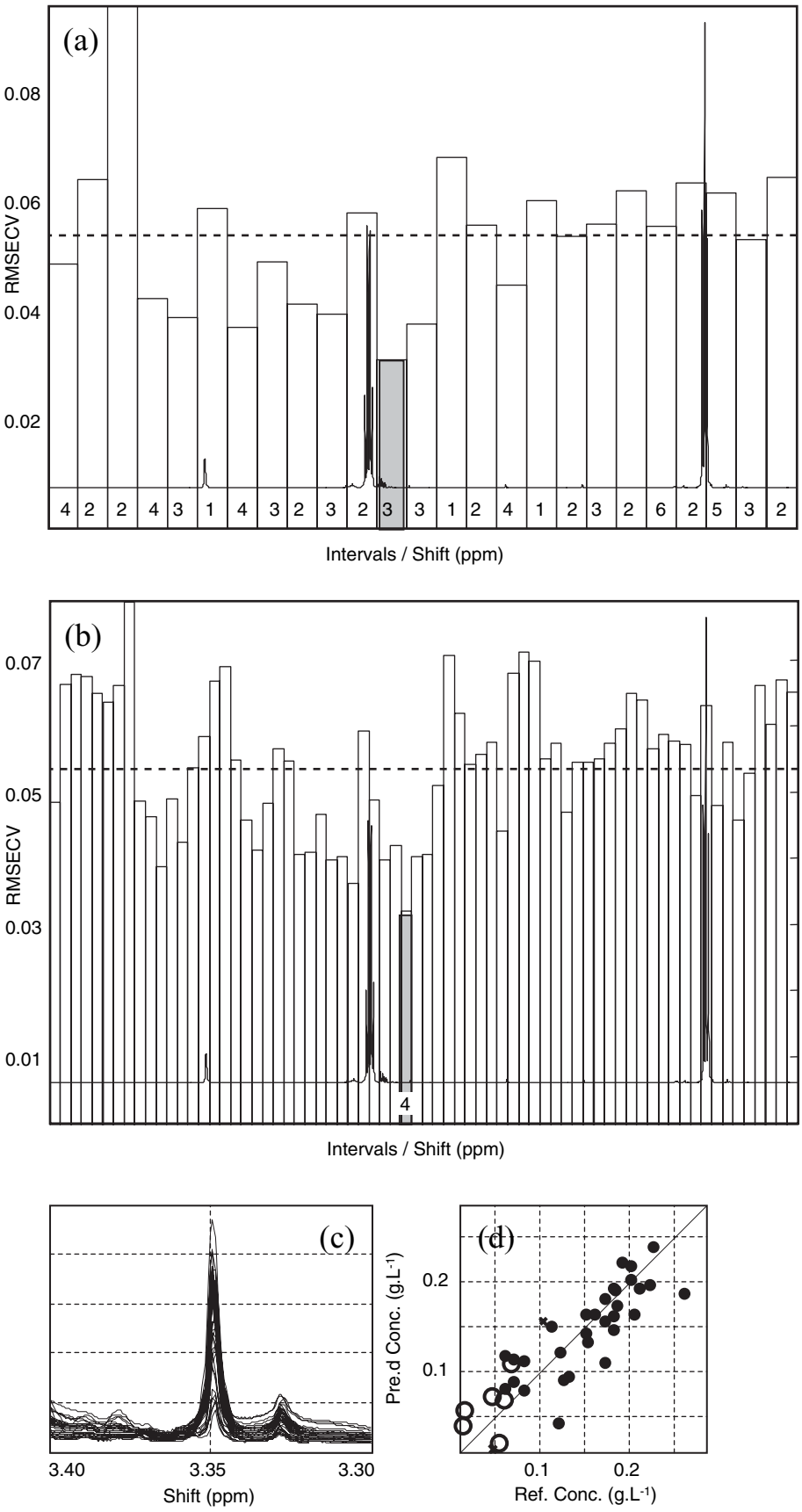


**Figure 7.** (a) iPLS plot RMSECV versus interval number (COW-corrected spectra), superimposed on the average spectrum for methanol prediction for 25 intervals. (b) iPLS-plot for 70 intervals. (c) Zoom in on methanol spectral range 3.3–3.4 ppm. (d) Cross-validation predicted versus reference methanol contents.



DOI: 10.1002/cem

knowledge the resonance doublet and triplet can be isolated and near optimally aligned, which is supported by the fact that the regression vectors for the two separate models show very little interference (Figure 6e,f). A similar procedure was applied for malic acid (Table II), the other component of interest in the malolactic fermentation of wines. It is noted that the content of malic acid is generally below $0.5\,g \cdot L^{-1}$ in red wines and around $2\,g \cdot L^{-1}$ in white and rosé wines. For lactic acid the content is usually a factor of 3–4 higher in red wines than the $0.4\,g \cdot L^{-1}$ level found in white and rosé wines.

To further illustrate the exploratory nature of iPLS a model for methanol is developed. Thus far, 25 intervals have proven sufficient to assist the spectroscopic knowledge in selection of the appropriate multiplet. As shown in Figure 7a, this is not the case for methanol in which case the model identifies the same interval as glycerol (see Figure 2) with a relatively large prediction error (taking into consideration a factor of 10 difference in concentration for glycerol and methanol for table wines). A subsequent 'zoom' into the problem was investigated by adding more intervals (here subjectively set to 70) over the same spectral range, requiring only a (modest) increase in computation time. Based on Figure 7b the methanol singlet (Figure 7c) was easily found making quantification of this component in the wines possible (Figure 7d), and no additional manual specification was necessary in this case. This iterative/interactive data exploration nicely illustrates the potential of iPLS when reference values are available. As an example, the same search refinement led to the identification of a second glycerol resonance around 3.56 ppm by isolating it from the strong ethanol signal in the neighboring region.

In the IR reference data available for our 40 table wines an indicator variable 'combined content of glucose and fructose' was available. Information about these compounds is

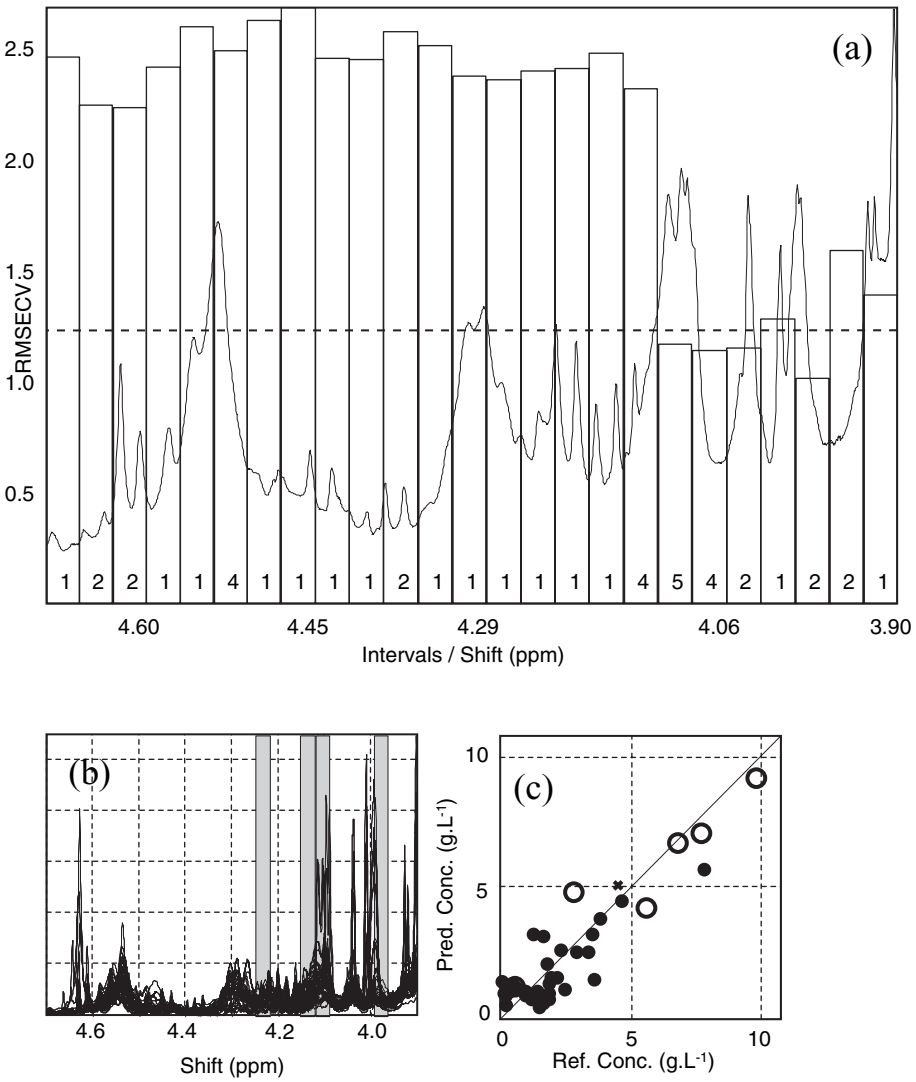


**Figure 8.** (a) iPLS plot RMSECV versus interval number (COW corrected spectra), superimposed on the average spectrum for combined glucose/fructose prediction for the carbohydrate spectral range. (b) Backward interval PLS-plot for combined glucose/fructose prediction for the carbohydrate spectral range with four intervals selected. (c) Cross-validation predicted versus reference combined glucose/fructose contents.



DOI: 10.1002/cem

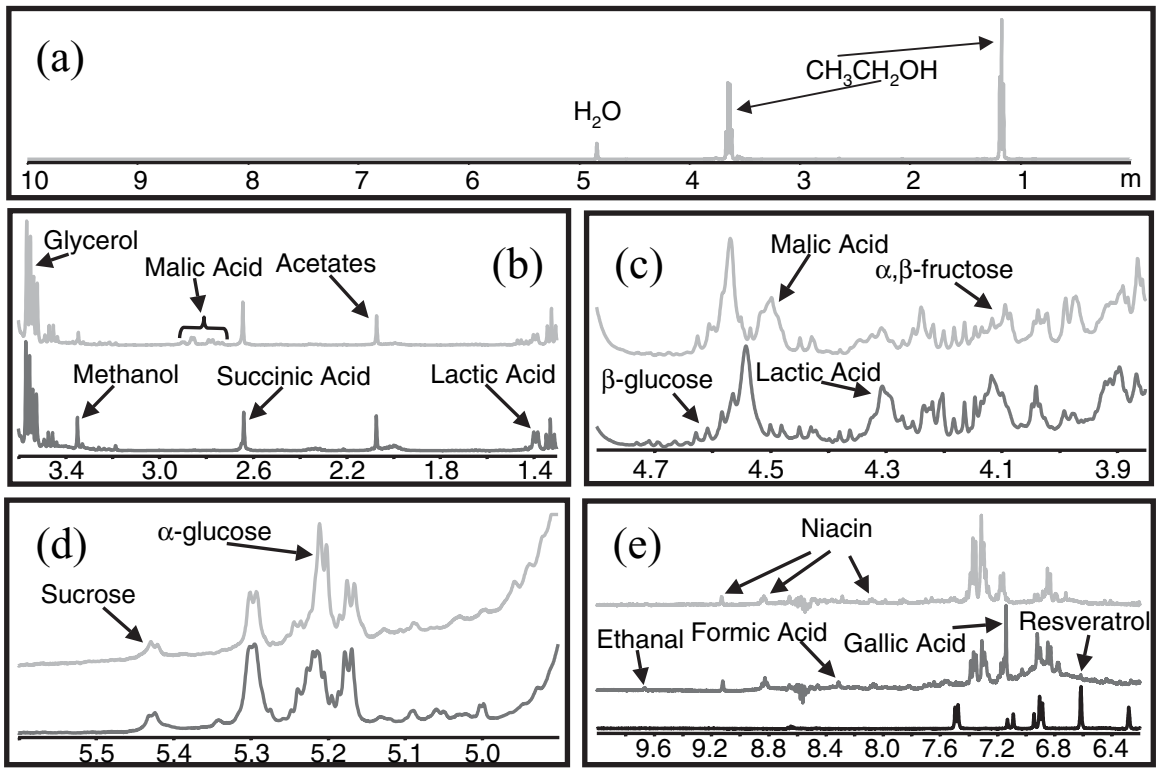


**Figure 9.** Components identified in $^1$H NMR spectra of a red (bottom line) and a white (top line) wine. (a) Whole spectrum, (b) Alifatic region, (c) Carbohydrates, (d) Anomers, (e) Aromatic compounds and proteins plus the spectrum of pure trans-resveratrol (lowest line).

expected to reside in slightly different regions of the carbohydrate region 3.9–4.7 ppm. As mentioned in the introduction, wines can be considered as relatively complex biological systems, and the number of possible and expected compounds in this region is considerable. Figure 8a and Table II show the results for iPLS; no obvious best interval is found. biPLS is an automated chemometric tool used to improve predictive performance, which may lead to better interpretability of the simpler/reduced models. The biPLS model results for combined glucose and fructose are shown in Figure 8b,c; four intervals are selected, giving a prediction error of 0.96 $g \cdot L^{-1}$ and a squared correlation coefficient of 0.83.

Besides the components analyzed in the previous parts of this work, additional chemical components can be identified in the NMR spectrum as displayed in Figure 9. The full NMR spectra is dominated by the peaks from ethanol and water (partly suppressed), but a detailed view of specific regions reveals resonances from other organic acids such as succinic and gallic acid as well as various carbohydrates such as $\alpha$- and $\beta$-anomers of fructose and glucose. All these resonance peaks can be used for direct calibration as they are quantitatively accurate. In IR-spectroscopy limited spectral resolution in a large part of the spectrum (i.e., the fingerprint region from approximately 400–1500 $cm^{-1}$) prevents accurate assignment without prior calibration. Moreover, the performance of multivariate calibration models applied to vibrational spectroscopy will sometimes rely on indirect correlations and thus depend on the composition and homogeneity of the data set. With complex and varying samples like our wines a direct, chemically based technology like NMR provides complementary information. Even a minor peak that is tentatively assigned [18] to *trans*-resveratrol can be observed. This stilbene compound has been related to health-beneficial properties, including the so-called French paradox [19].

## 4. CONCLUSIONS

In this study a collection of table wines has been subjected to analysis by $^1$H NMR spectroscopy and various PLS regression variations. Since the pH value of the wines were not adjusted prior to NMR analysis, the spectra required pre-treatment to correct for shift differences induced by variations in pH. Several important components such as malic acid and glycerol were identified and their contents calibrated against results from IR spectroscopy using PLS methods. In general, the combination of NMR spectroscopy and chemometrics is a promising partnership for detailed wine analysis. Such analysis may be performed in any stage of the wine manufacturing allowing for thorough evaluation of every step in the process.

The next obvious step, automatic NMR region/multiplet selection by algorithmic repositioning of the intervals, is not pursued in this paper. However, the complexity of most biological systems will make this a challenging task, placing high demands on factors such as the quality of the data and the number of samples available. The automatic selection will also be a function of proper multiplet alignment, requiring the merger of, for instance, iPLS and COW. This algorithmically very challenging concept will be the subject of future research.



DOI: 10.1002/cem

## Acknowledgements

The authors acknowledge the Danish Dairy Foundation, the Danish Directorate for Fisheries and Agro Business (FSK 03–5) and the Danish Research Councils for generous support to our Bromatonomics project (the use of quantitative NMR in Danish dairy and food science). Gilda Kischinovsky is acknowledged for proofreading the manuscript. Foss Analytical A/S, Denmark, is kindly acknowledged for making available IR reference measurements.

## REFERENCES

1. Soleas GJ, Diamandis EP, Goldberg DM. Wine as a biological fluid: history, production, and role in disease prevention. *J. Clin. Lab. Anal*. 1997; **11**: 287–313.
2. Martin GJ, Martin ML, Mabon F, Michon MJ. Identification of the origin of natural alcohols by natural abundance hydrogen-2 nuclear magnetic resonance. *Anal. Chem*. 1982; **54**: 2380–2382.
3. Nord LI, Vaag P, Duus JØ. Quantification of organic and amino acids in beer by $^{1}$H NMR spectroscopy. *Anal. Chem*. 2004; **76**: 4790–4798.
4. Duarte I, Barros A, Belton PS, Righelato R, Spraul M, Humpfer E, Gil AM. High-resolution nuclear magnetic resonance spectroscopy and multivariate analysis for the characterization of beer. *J. Agric. Food Chem*. 2002; **50**: 2475–2481.
5. Bescia MA, Košir IJ, Caldarola V, Kidrič J, Sacco A. Chemometric classification of apulian and slovenian wines using $^{1}$H NMR and ICP-OES together with HPICE data. *J. Agic. Food Chem*. 2003; **51**: 21–26.
6. Belton PS, Colquhoun IJ, Kemsley EK, Delgadillo I, Roma P, Dennis MJ, Sharman M, Holmes E, Nicholson JK, Spraul M. Application of chemometrics to the $^{1}$H NMR spectra of apple juices: discrimination between apple varieties. *Food Chemistry*. 1998; **61**: 207–213.
7. Holmes E, Foxall PJD, Nicholson JK, Neild GH, Brown SM, Beddell CR, Sweatman BC, Rahr E, Lindon JC, Spraul M, Neidig P. Automatic data reduction and pattern recognition methods for analysis of $^{1}$H Nuclear Magnetic Resonance spectra of human urine from normal and pathological states. *Anal. Biochem*. 1994; **220**: 284–296.
8. Nicholson JK, Foxall PJD, Spraul M, Farrant RD, Lindon JC. 750 MHz $^{1}$H and $^{1}$H-$^{13}$C NMR spectroscopy of human blood plasma. *Anal. Chem*. 1995; **67**: 793–811.
9. Tomasi G, van den Berg F, Andersson C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemometrics* 2004; **18**: 231–241.
10. Berg van den F, Tomasi G, Viereck N. Warping: investigation of NMR pre-processing and correction. In *Magnetic Resonance in Food Science: The Multivariate Challenge*, Engelsen SB, Belton PS, Jakobsen HJ (eds). The Royal Society of Chemistry: Cambridge, 2005; 131–138.
11. Torgrip RJO, Åberg M, Karlberg B, Jacobsson SP. Peak alignment using reduced set mapping. *J. Chemometrics* 2003; **17**: 573–582.
12. Forshed J, Torgrip RJO, Åberg KM, Karlberg B, Lindberg J, Jacobsson SP. A comparison of methods for alignment of NMR peaks in the context of cluster analysis. *J. Pharm. Biomed. Anal*. 2005; **38**: 824–832.
13. Forshed J, Schuppe-Koistinen I, Jacobsson SP. Peak alignment of NMR signals by means of a genetic algorithm. *Anal. Chim. Acta* 2003; **487**: 189–199.
14. Wong JW, Durante C, Cartwright HM. Application of fast Fourier transform cross-correlation for the alignment of large chromatographic and spectral datasets. *Anal. Chem*. 2005; **77**: 5655–5661.
15. Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. Interval Partial Least Squares Regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy. *Applied Spectroscopy* 2000; **54**: 413–419.
16. Leardi R, Nørgaard L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *J. Chemometrics* 2004; **18**: 486–497.
17. Bax A. A spatially selective composite radiofrequency pulse. *J. Magn. Reson*. 1985; **65**: 142–145.
18. Mattivi F, Reniero F, Korhammer S. Isolation, characterization, and evolution in red wine vinification of resveratrol monomers. *J. Agric. Food Chem*. 1995; **43**: 1820–1823.
19. Kopp P. Resveratrol, a phytoestrogen found in red wine. A possible explanation for the conundrum of the 'French paradox'? *Eur. J. Endocrinology* 1998; **138**: 619–620.



DOI: 10.1002/cem