Direct decomposition of NMR relaxation profiles and prediction of sensory attributes of potato samples

V.T. Povlsen\textsuperscript{a}, Å. Rinnan\textsuperscript{a,*}, F. van den Berg\textsuperscript{a}, H.J. Andersen\textsuperscript{b}, A.K. Thybo\textsuperscript{c}

\textsuperscript{a}Department of Dairy and Food Science, The Royal Veterinary and Agricultural University (KVL), Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

\textsuperscript{b}Department of Animal Quality, Danish Institute of Agricultural Sciences (DJF), DK-8830 Tjele, Denmark

\textsuperscript{c}Department of Horticulture, Danish Institute of Agricultural Sciences (DJF), DK-5792 Aarslev, Denmark

Received 13 March 2002; accepted 12 January 2003

Abstract

In this paper the decomposition of low-field Carr–Purcell–Meiboom–Gill (CPMG) NMR relaxation measurements on 23 raw potato categories was investigated. The potato categories were formed from five different cultivars, each binned in 2 or 3 dry matter intervals, sampled at two storage times. A novel data analytical tool—called SLICING—revealed that different amounts of four distinct proton relaxation profiles could describe the main variation in the data set. Magnitudes (scores) of the third and fourth profile separated the potato cultivars, storage times, and dry matter content indicating that properties related to fast relaxation times explain the differences between cultivars and storage times for the potatoes. The concept of direct decomposition using SLICING on low-resolution NMR data is a new approach in potato analysis and a promising tool for obtaining more information about the structure and water distribution in food products.

Furthermore, the texture-related sensory attributes, hardness, cohesiveness, adhesiveness, mealiness, graininess, and moistness of cooked potatoes were predicted by partial least-squares regression (PLSR). Four different types of predictor variables derived from the NMR relaxation curves were compared in the regression models: (i) the raw CPMG curves, (ii) the parameters from the traditional bi-exponential fitting, (iii) the results from a distribution analysis, and (iv) the scores from the SLICING model. The predictions based on the distribution analysis performed worse than the first three procedures, which all showed similar prediction ability. The advantage of the SLICING approach is in the possibility to interpret physical properties, e.g. water distribution of the potato samples.

Keywords: Potato; Low-field NMR; NMR relaxation; PARAFAC; PLSR

1. Introduction

The texture of cooked potatoes is an important quality attribute when assessing potato quality. In the potato industry great interest lies in both improving and developing rapid methods to determine this quality. Special interest lies in assessing the raw potato samples and relating them to the sensory quality of cooked potatoes. The potential perspective could be an early sorting of the raw material according to quality prior to packaging or processing. The texture of cooked potatoes is related to the size and amount of starch, rigidity and chemistry of the cell walls, enzyme activities, minerals, heating, water content as well as the subsequent heating process (Gould, 1999). Evaluation of potato texture and quality can be performed by mechanical, analytical and/or sensory methods (VanMarle, DeVries, Wilkinson, & Yuksel, 1997; Thybo & Martens, 1999; Ulrich, Hoberg, Neugebauer, Tiemann, & Darsow, 2000). Using sensory evaluation, information about the human perception of potato quality is obtained, as the senses of sight, smell, taste, touch and hearing are studied. In sensory analysis, the texture is evaluated in terms of moistness, adhesiveness, mealleness, etc. In addition, mechanical measurements, for example uniaxial compression and nuclear magnetic resonance (NMR) relaxation, have been applied in the texture analysis of vegetables (Tang, Belton, Ng, Waldron, & Ryden, 1999; Thybo & Martens, 1999; Tang, Godward,
NMR has been shown to provide useful information about molecular structure within a sample and has become a powerful nondestructive analytical tool in chemistry (Hemminga, 1992; Ruan & Chen, 1998). In food science, NMR techniques have been used to study the texture and the state of water in food samples (Hills & Le Floc’h, 1994; Seow & Teo, 1996; Hills, Goncalves, Harrison, & Godward, 1997; Ruan et al., 1997; Tang et al., 1999; Tang et al., 2000) and for the analysis of fats and oils (Pedersen, Munck, & Engelsen, 2000). Previous work by Thybo and Martens (1999) showed a higher correlation between sensory quality of cooked potatoes and 1H NMR on raw potatoes compared to using 1H NMR on cooked potatoes. This work forms the basis for the present study, where the objective was to compare the SLICING method (Pedersen, Bro, & Engelsen, 2001) to existing methods for analysing low-field proton NMR signals (1H-NMR) from Carr–Purcell–Meiboom–Gill (CPMG) pulse relaxation curves of raw potatoes. The comparison was based on the interpretability in data analysis and the predictive performance of sensory quality on cooked potatoes using multivariate regression. The SLICING procedure has previously shown good results for data analysis purposes when estimating the underlying relaxation curves of fish (Andersen & Rinnan, 2002). When handling low-field NMR data, these underlying relaxation curves ideally correspond to the different chemical states of water in the measured samples. Thus, SLICING makes it possible to interpret the data directly on a physical basis because the model separates the measured signal, a mixture of exponential curves, into physically meaningful uni-exponential contributions. It is noted that the SLICING method assumes that a fairly low number of such curves are sufficient for describing the actual measurement signal, in contrary to, e.g. distribution analysis, where it is assumed that the data consist of a sufficiently large number of distinguishable exponentials, such that a distribution of these can be computed. In the bi-exponential fitting method two contributing exponentials are assumed sufficient to describe the measured signal. However, the two last methods assume no relationship between samples—treating each sample individually—and thus differ from the factor-based SLICING method. The discussion as to which of these alternative decomposition methods is most appropriate will not be the main issue of this paper. Rather, it will be shown that the SLICING approach as such provides a solution, which is scientifically sound and useful for interpretation and further modelling.

The relation between the NMR relaxation curves on raw potatoes and sensory attributes evaluated on cooked potatoes was studied by regression modelling. The prediction performance based on partial least-squares regression (PLSR, Martens & Næs, 1989) using the SLICING scores as predictors were compared to modelling on the raw low-field 1H-NMR curves (CPMG PLSR). PLSR has previously been used on raw low-field 1H-NMR curves for prediction of fish and potatoes sensory attributes, showing good performance (Thybo, Bechmann, Martens, & Engelsen, 2000; Thygesen, Thybo, & Engelsen, 2001). However, the CPMG PLSR results are less interpretable because the loadings do not have a direct physical meaning. The regression performance based on the model parameters retrieved from bi-exponential fitting and distribution analysis was also compared to the regression methods based on the SLICING scores. Bi-exponential fitting was applied because it constitutes one of the main alternatives to the SLICING approach, while distribution analysis was applied because it has been used with success in previous potato studies (Hills & Le Floc’h, 1994; Hills, Goncalves, Harrison, & Godward, 1997), as well as other areas of research (Tang et al., 2000).

2. Materials and methods

2.1. Potatoes

The material used in the experiments included five potato cultivars grown at an experimental field at the Danish Institute of Agricultural Sciences. Within the five cultivars the potatoes were graded in salt solutions according to 1% dry matter bins (Burton, 1989) in the range of 18.0–22.9%, as described by Thybo and Martens (1999). Potato samples harvested in September 1999 were analysed in November 1999, and in May 2000 after being stored at 4°C at 95% relative humidity. This selection procedure gave a total of 23 different potato samples (see Table 1).

2.2. Sensory analysis

The potatoes were peeled and boiled in water for 20–25 min until they were cooked through. The sensory analysis was performed on the cooked potatoes by a trained panel of ten assessors and evaluated on a scale from 0 to 15. The measurements were performed as described by Thybo and Martens (1999) using the average of the ten assessors times four sensory replicates. The sensory variables hardness, cohesiveness, adhesiveness, mealiness, graininess, and moistness were evaluated.

2.3. NMR measurements

The relaxation measurements of the water protons were performed on a Maran Bench top Pulsed 1H-NMR Analyser (Resonance Instruments Ltd., Witney, UK) with a magnetic field strength of 0.47T, corresponding
to a resonance frequency of 23.2 MHz. The instrument was equipped with an 18-mm temperature variable probe. The samples were sized in cylinders of \( C_{2}d = 40 \times 14 \) mm². They were stamped longitudinally from the stem end of the potato, and placed in a cylindrical glass tube (14 mm in diameter and 50 mm in height). This tube fitted into the NMR temperature variable probe 18 mm in diameter. Before the measurement was performed, the sample was temperature controlled to 25°C in a water-bath for 15–20 min.

Transverse relaxation (\( T_{2} \)) was measured using the CPMG sequence (Carr & Purcell, 1954; Meiboom & Gil, 1958). The transversal relaxation measurements were performed with a \( \tau \) value (time between 90° and 180° pulse) of 1000 μs. The data were acquired as four scan repetitions. The repetition delay between two succeeding scans was 4 s. The signal amplitude was measured every echo and the relaxation measurements were performed at 25°C.

### 2.4. Data treatment

Each potato sample (bin) was measured by NMR in a number of replicates (tubers) ranging from 12 to 15. If outliers were detected in any of the replicate series, they were removed before the computations. Outliers were defined as replicates that were significantly different from the other replicates in any of the following attributes: low initial value, slower relaxing curve or faster relaxing curve. The initial data consisted of a total of 324 measurements, which was reduced to 295 after removing the outliers. Each sample was now represented by 11–14 NMR measurement replicates. The sensory analysis was performed on only four replicates with no direct link to the tubers used in the NMR measurements. To compensate for differences between tubers from one category, the average of the sensory analysis was used together with the average of the NMR curves for each bin. In this study the difference between cultivars, and not between tubers, was of interest, hence using the average reduces the natural variety within the bins.

### 3. Data analysis and modelling

#### 3.1. Description of the NMR curves

NMR relaxation signals can be expressed mathematically as a sum of exponential decays (see Eq. (1)):

\[
I(t) = \sum_{n=1}^{N} M_{0,n} \exp\left(-\frac{t}{T_{2,n}}\right).
\]  

In this equation the profile \( I(t) \) is parameterized such that \( N \) is the (expected) number of uni-exponentials, \( M_{0} \) holds the \( N \) magnitude values, \( t \) is time, and \( T_{2} \) is the time constants associated with each uni-exponential decay. For a set of curves, it is assumed that the quantitative information, amount of a specific proton signal, is carried by the \( M_{0} \) values and the qualitative information, the type of proton signal, by the \( T_{2} \) values. There are several methods to find these parameters. Three methods are evaluated in this article: bi-exponential fitting, distribution analysis and SLICING. In bi-exponential fitting, the assumption is that \( N \) in Eq. (1) is two for any sample and that the \( T_{2} \) value can vary from sample to sample. In the SLICING \( N \) is not known beforehand but determined as part of the modelling step. It is assumed that all samples can be described by the same set of \( T_{2} \) values. In distribution analysis, it is assumed that a distribution of \( T_{2} \) values generates each profile. Hence, \( N \) is assumed to be very large indicating that each proton has its own distinct value. This assumption appears reasonable at first glance, but in practice distribution analysis can be hampered by numerical instabilities caused by the high amount of parameters to be determined from a limited data set with finite signal-to-noise ratio. The discrete methods, bi-exponential fitting and SLICING, on the other hand, assume an approximation, which may be valid in practice due to this limited signal-to-noise ratio and the similarity of the individual proton relaxations over samples. Hence, it is not possible on theoretical grounds to reject any of the proposed methods. One purpose of this investigation is to show empirically to what extent, these methods can provide reliable information on the current data. In the following the different modelling approaches for NMR data and regression are described.

### Table 1

Tuber samples used in the experiments

<table>
<thead>
<tr>
<th>Cultivar</th>
<th>Dry matter bins (%)</th>
<th>Storage time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>November 1999</td>
<td>May 2000</td>
</tr>
<tr>
<td>Ditta</td>
<td>20.0–20.9</td>
<td>21.0–21.9</td>
</tr>
<tr>
<td></td>
<td>21.0–21.9</td>
<td>22.0–22.9</td>
</tr>
<tr>
<td>Sava</td>
<td>18.0–18.9</td>
<td>18.0–18.9</td>
</tr>
<tr>
<td></td>
<td>19.0–19.9</td>
<td>19.0–20.9</td>
</tr>
<tr>
<td></td>
<td>20.0–20.9</td>
<td>21.0–21.9</td>
</tr>
<tr>
<td>Bintje, low dry matter</td>
<td>19.0–19.9</td>
<td>20.0–20.9</td>
</tr>
<tr>
<td></td>
<td>20.0–20.9</td>
<td>21.0–21.9</td>
</tr>
<tr>
<td></td>
<td>21.0–21.9</td>
<td>—</td>
</tr>
<tr>
<td>Bintje, high dry matter</td>
<td>21.0–21.9</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>22.0–22.9</td>
<td>—</td>
</tr>
<tr>
<td>Berber</td>
<td>18.0–18.9</td>
<td>18.0–18.9</td>
</tr>
<tr>
<td></td>
<td>19.0–20.9</td>
<td>19.0–20.9</td>
</tr>
<tr>
<td></td>
<td>21.0–21.9</td>
<td>21.0–21.9</td>
</tr>
</tbody>
</table>
3.2. Regression by PLS on the raw CPMG curves

One of the advantages of multivariate methods such as PLS regression (Martens & Næs, 1989) is that they handle correlated variables well. This feature makes them suitable for handling data such as NMR relaxation curves, where neighbouring time points are highly correlated. Using PLSR on raw data, focus is on the prediction ability of the model, but the interpretation of the models might not be as straightforward as the other methods described in this paper.

3.3. Bi-exponential fitting

A common approach to model NMR curves is bi-exponential fitting, yielding for each sample individual values for parameters $M_0$, $M_0$, $T_{2,1}$, and $T_{2,2}$ in Eq. (1). This approach is based on the assumption that any sample can be described as a weighted sum of two exponentials and the $T_2$ values are specific for this sample. The $M_0$ and $T_2$ values may be used for the prediction of the sensory attributes by the use of PLSR.

3.4. Distribution analysis

Another method for describing the NMR curves is by the use of distribution analysis. Distributed exponential fitting analysis was performed on $T_2$ relaxation data using the Win-DXP program for Matlab (Butler, Reeds, & Dawson, 1981). A continuous distribution of exponentials for a CPMG experiment can be defined by Eq. (1), setting $N$ to a large number. To use this distribution information for regression analysis the results need to be transformed into a suitable set of variables. In this paper, the position and the amplitude of the peaks in the distribution were used for regression analysis.

4. SLICING

SLICING is a novel method for exploring NMR relaxation curves (Pedersen et al., 2001). The method decomposes the relaxation curves from NMR measurements into a few individual archetype proton contributions. It is based on increasing the dimensionality of the data from a two-way to a three-way array by a proper rearrangement. The rearranged data cube (three dimensional) will ideally follow the so-called tri-linear model. Performing a tri-linear decomposition of the rearranged data will directly yield a set of normalized exponential decays (i.e. $T_2$ values) as well as the corresponding amounts/magnitudes of these decays for each sample ($M_0$ values).

In SLICING the assumption is that all samples can be represented by a weighted sum of a number of exponentials, conforming Eq. (1). Thus, there is no predefined number of exponentials as in the bi-exponential fitting. On the other hand, it is assumed that all samples are sums of the same exponentials, which is not the case for bi-exponential fitting.

The SLICING algorithm uses the principles of direct exponential curve resolution algorithm (DECRA, Windig & Antalek, 1997). The idea is to split the CPMG relaxation curves (see Fig. 1a) into two (or more) overlapping parts (slabs), where the size of the overlap is determined by the lag term, generating a three-dimensional array. Most of the original relaxation curve is present in both slabs. This operation is illustrated in Fig. 1a. Next, PARAllel FACtor analysis (PARAFAC) is performed on the three-dimensional array (Bro, 1997). The PARAFAC model is described by the following equation:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk}$$

$$(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K).$$

Fig. 1. Going from NMR signal to the data cube for PARAFAC modelling: (a) illustrating the principles of creating a three-way array from NMR relaxation curves; (b) the data cube $X [23 \times 1991 \times 3]$ is decomposed into four triads with sample scores $a$ (23 exponential loadings), $b$ (1991) and slab loadings $c$ (3), plus residual cube $E$ (‘noise’).
The element $x_{ijk}$ is the original value in the position $(i,j,k)$ of the data cube $X$. The parameter $a_f$ is the object score (magnitude) for factor $f$ (first mode), $b_f$ is the exponential decay curve for the pure component $f$ (second mode), and loading $c_{ij}$ gives the ratio between the different slabs (third mode). The term $e_{ijk}$ contains residual variation not captured by the model. The data cube $X$ is decomposed into $F$ different components (triads) and a residual cube $E$ (Fig. 1b). In the PARAFAC algorithm used here, the factors (triads) are found simultaneously via an alternating least-squares algorithm (Bro, 1997). If the model is correctly specified, the residual of the exponential loadings indicates how much structural information remains unmodelled. If the residuals show random behaviour and no systematic trend, only noise is left unexplained and hence the $N$ estimated profiles explain the variation in the data up to the noise. Furthermore, if the model is adequate each loading is described by a single exponential. If too many components are extracted, the estimated curves will reflect this (one or more being nonexponential). The residuals were used together with the appearance of the relaxation loadings to estimate the correct number of components. The object scores from the SLICING were then used for prediction of the sensory attributes.

In this study the data matrix $X$ held the CPMG relaxation curves of the 23 samples. The SLICING was performed by splitting the relaxation curves into three slabs; with a lag of 0, 1, and 4 data points, respectively. This choice of lags was based on a subjective selection from initial investigations. The dimension of the rearranged data cube was 23 objects $\times$ 1991 relaxation variables $\times$ 3 slabs.

4.1. Validation

The validation of the regression models for the CPMG PLSR, the SLICING, the bi-exponential, and the distribution analysis predictions were all performed by the leave one subset out cross validation (Eastman & Kranowski, 1982; Martens & Næs, 1989). In this method the data are split into equally sized, randomly selected subsets. One subset is left out and a model is built from the remaining data. The properties of the left-out objects are then predicted using this model, and the residuals are calculated for models of increasing model complexity (number of factors). In the next step a new subset is removed and the procedure is repeated until every subset has been left out once. The root mean square error of cross validation (RMSECV, see Eq. (3)) indicates the difference between the predicted and the measured values. In the following equation, $y$ is the measured values, $\hat{y}$ is the predicted value, while $n$ represents the number of samples:

$$\text{RMSECV} = \sqrt{\frac{\sum (y - \hat{y})^2}{n}}$$

In this study the data sets were divided into four subsets. RMSECV and the correlation coefficients ($r$, upon plotting measured versus predicted) were used as indicators of the model’s predictive ability.

All data analysis and modelling were performed using Matlab 5.3 software (Mathworks) for Windows with algorithms taken from the PLS-Toolbox (www.eigenvector.com) and the N-way Toolbox (Andersson & Bro, 2000). A dedicated SLICING toolbox is available at www.models.kvl.dk, but was not yet available at the start of this investigation.

5. Results and discussion

To get an impression on the way the sensory attributes discriminate potato cultivars a principal component analysis (PCA) is performed (Martens & Næs, 1989). Fig. 2 shows the bi-plot of sample scores and attribute loadings. In this figure clear grouping of cultivars and storage times are observed, as well as for the dry matter bins. This proves that the data set contains information which can distinguish these design variables. It was of interest to investigate the possibility to extract the same information from the NMR measurements via multivariate data analysis, without the requirement of sensory panel input.

In the present work the region from 12 to 4000 ms of the NMR measurement signal was used in the analysis. The first five data points were considered unreliable due to noise and the last 2000 points had a signal close to zero, not contributing any significant information. The average CPMG relaxation curves of the raw potatoes were investigated prior to any analysis. Upon studying the raw data, a variation in the decays for the five potato cultivars and dry matter bins was observed (not shown). The potato samples of the cultivar Berber were distinct from the rest of the cultivars showing a slower exponential decay. The difference was observed throughout the entire signal, and indicates a deviation in the composition and distribution of water compared to the other four cultivars. Within the five cultivars, the two storage times, November 1999 and May 2000, appeared different, where the storage time May 2000 showed a faster decay. This implies changes in the water distribution due to storage time.

5.1. Data analysis using SLICING

A SLICING model of the CPMG curves was computed. The NMR profile loadings for the optimal SLICING model, consisting of four factors, are shown...
in Fig. 3. The four loadings are all exponentials as expected. This was further verified using a Monte Carlo approach, where 97% of 1000 randomly selected split-half tests resulted in the same four exponential loadings (Harshman & De Sarbo, 1994). In each split-half run, the data set was split into two parts, each part containing 12 and 11 samples, respectively. Both of these data sets were then modelled individually. Obtaining similar results from two such completely independent sets of data implies that the results are reproducible in a scientific sound way. I.e. the components are not merely an arbitrary result from a specific set of samples, but rather a fundamental property of all similar samples. This indicates that a valid estimate of the CPMG relaxation curves was derived from SLICING and hence the loadings could be associated with the water distribution in the potatoes. Previous studies have made use of bi-exponential fitting of the raw CPMG

Fig. 2. Bi-plot from PCA on the sensory data. Ditta (⊙), Sava (∆), Bintje—low dry matter (□), Bintje—high dry matter (★), and Berber (○) for storage 1999 (open) and 2000 (filled). The numbers from 18 to 22 represent the % dry matter bins (see Table 1) where the range of the % dry matter bin is 18: 18.0–18.9, 19: 19.0–19.9, 20: 20.0–20.9, 21: 21.0–21.9 and 22: 22.0–22.9.

Fig. 3. Exponential loadings for components one to four from the PARAFAC model.
relaxation curves to explain the different states of water in potatoes (Thygesen et al., 2001). The $T_{2,1}$ and $T_{2,2}$ relaxation times from bi-exponential of the raw CPMG relaxation curves using Eq. (1) ($N = 2$) are listed in Table 2, together with the uni-exponential fitting of the four loadings from the SLICING model.

The transversal relaxation times $T_{2,1}$ and $T_{2,2}$ from bi-exponential fitting for the 23 objects range from 130 to 180 and 430 to 540 ms, respectively. The relaxation times for the four exponential loadings in Fig. 3 show that the fourth loading has the fastest decay with a $T_{2,B4}$ of 52 ms followed by the third loading $T_{2,B3}$ of 192 ms ("B" indicating that these $T_2$ values are calculated from the B-loadings of PARAFAC). The $T_{2,B}$ for the first two exponential loadings are 378 and 646 ms, respectively. The results from the distribution analysis gave two peaks, where the first peak ranged from 56 to 82 ms and the second peak from 404 to 540 ms (not shown). A comparison of the four relaxation time constants showed that the relaxation times $T_{2,B}$ for the SLICING model span wider than the $T_{2,1}$ and $T_{2,2}$ from the bi-exponential fitting. Both the bi-exponential fitting and the distribution analysis have a peak around 480 ms, while the SLICING estimated decays at 378 and 646 ms. The first peak from the distribution analysis corresponds approximately with the fastest relaxing component from the SLICING. In the bi-exponential fitting, the fastest component lies in between the two fastest components from the SLICING.

By looking at the average residual over time for each of the three decomposition methods, it became clear that the residual from bi-exponential fitting was roughly four times larger than the residual from distribution analysis, which was twice as large as the average residual from the SLICING model. The reason may be caused by a smoothing constrain in the distribution algorithm. These observations imply that loadings derived from the SLICING model provide more information about the data than a simple bi-exponential fitting or a distribution analysis. In a four-factor SLICING model, the best description of the potato cultivars was given by the sample scores for factors 3 and 4 (the two fastest decays), where a clear distinction of the five cultivars was seen. This is shown in the score plot in Fig. 4, where the samples are marked due to cultivar, storage, and dry

<table>
<thead>
<tr>
<th>Curves</th>
<th>Fitted against $T_2$ (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential loadings ($T_{2n}$) from SLICING</td>
<td>$T_{2B1}$ 378</td>
</tr>
<tr>
<td>Fitted by uni-exponentials</td>
<td>$T_{2B2}$ 646</td>
</tr>
<tr>
<td>Raw curves fitted by bi-exponentials</td>
<td>$T_{2B3}$ 192</td>
</tr>
<tr>
<td>The range of the 23 potato samples</td>
<td>$T_{2B4}$ 52</td>
</tr>
</tbody>
</table>

$T_{2,B}$ represent the relaxation times of the uni-exponential fitting of the four relaxation slicing loadings and the $T_{2,1}$ and $T_{2,2}$ represent the relaxation times of the bi-exponential fitting of the raw CPMG relaxation curves. The raw curves show the range of the 23 potato samples.

Table 2
Overview of the transversal relaxation time ($T_2$)

![Score plot of sample score 3 and 4 for the PARAFAC](image)

Fig. 4. Score plot of sample score 3 and 4 for the PARAFAC. Ditta (⊙), Sava (△), Bintje—low dry matter (□), Bintje—high dry matter (★), and Berber (○) for storage 1999 (open) and 2000 (filled). The numbers from 18 to 22 represent the % dry matter bins (see Table 1) where the range of the % dry matter bin is 18: 18.0–18.9, 19: 19.0–19.9, 20: 20.0–20.9, 21: 21.0–21.9 and 22: 22.0–22.9.
matter bins. Within the cultivars the storage times for each bin are explained in the direction from the upper right corner to the lower left corner. The opposite direction from the left lower corner to the upper right corner describes the increase in dry matter bins within the same cultivar. The separation can be related to the diversity in the structure of the five cultivars and the varying water content and distribution of water (Table 1). There is no such clear distinction between the cultivars using the results from either the bi-exponential fitting or the distribution analysis (not shown).

Relaxation times can be related to the distribution of water within the samples. A high mobility of water makes it more available, and it will take a long time before it reaches the equilibrium state, giving rise to a high $T_2$. Thus, the highest $T_2$ value ($T_{2B2}$) may reflect the water used for gelatinization and can be expected to be of major importance for texture differences. However, the variation in the potatoes is not captured by the two slower decaying loadings, indicating that this type of water is not important for the description of the differences in the five cultivars. The description of the cultivars in the SLICING scores 3 versus 4 indicates that the clear difference in the cultivars is caused by the distribution of the water with low mobility in the potato tubers. These low-mobile water components are assumed to describe the less mobile diffusion-hindered water hypothesized to be located in, e.g. the cell walls, entrapped in pectin, in sites with high ionic strengths, and in the vascular tissue (water transport tissue).

Several states and locations of water are possible within potatoes. Water compartments may be found in the cytoplasm in the cells and in the pectin network in the cell walls. Furthermore, very different tissue segments within a potato tuber exist. This makes the investigation of the distribution of water in potatoes very complex. Hills and Le Floc'h (1994) made a thorough study of the water in potatoes as they froze them down. Their study give an explanation to three of the four components found using SLICING. The first one is similar to a peak they find at about 50 ms, coming from water in cell walls, while the next two resembles peaks they find at around 200 and 400 ms, which they state is from water in the cytoplasm. However, they do not find any component higher than ca. 400 ms. Tang et al. (2000), on the other hand, found a peak at around 50 ms upon studying water saturated starch granules, so the exact cause of the fastest decaying component cannot be given. By the application of the SLICING a more direct method is introduced. This makes it possible to get a quick estimate of the parameters related to the quality, instead of high-cost laboratory analyses.

5.2. Regression models

PLSR has previously been used for the prediction of sensory attributes and potato quality from CPMG relaxation curves (Thybo et al., 2000; Thygesen et al., 2001). In this work six texture-related sensory attributes—hardness, cohesiveness, adhesiveness, mealiness, graininess, and moistness—of cooked potatoes were predicted using four different types of predictor variables. First, the CPMG PLSR was performed on the raw data set. The right number of components—four—was selected using the RMSECV values, the exponential loadings, and the exponential residuals as diagnostics. The second approach was the bi-exponential fitting predictions, which was based on the $M_0$ and $T_2$ values as independent variables in a PLSR model, and the third was the predictions using the results from distribution analysis. Two peaks were found from the distribution analysis, and the predictions were based upon the position and the amplitude of these two peaks. The last type of predictors was the four scores from the SLICING model referred to as SLICING prediction. The model complexity for prediction based on bi-exponential fitting and distribution analysis ranges from one to four components, depending on the sensory attribute being regressed. In Table 3, the RMSECV and correlation coefficients (predicted versus reference

<table>
<thead>
<tr>
<th>Sensory variables</th>
<th>PLSR$^a$</th>
<th>Bi-exp. fit$^b$</th>
<th>Distrib. anal.$^b$</th>
<th>SLICING$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributes</td>
<td>Range$^c$</td>
<td>RMSECV</td>
<td>$r$</td>
<td>RMSECV</td>
</tr>
<tr>
<td>Hardness</td>
<td>4.9</td>
<td>1.19</td>
<td>0.69</td>
<td>1.22</td>
</tr>
<tr>
<td>Cohesiveness</td>
<td>5.7</td>
<td>1.10</td>
<td>0.78</td>
<td>1.01</td>
</tr>
<tr>
<td>Adhesiveness</td>
<td>4.7</td>
<td>1.27</td>
<td>0.58</td>
<td>1.01</td>
</tr>
<tr>
<td>Mealiness</td>
<td>7.4</td>
<td>1.49</td>
<td>0.74</td>
<td>1.26</td>
</tr>
<tr>
<td>Graininess</td>
<td>5.2</td>
<td>1.25</td>
<td>0.54</td>
<td>1.07</td>
</tr>
<tr>
<td>Moistness</td>
<td>7.0</td>
<td>1.11</td>
<td>0.76</td>
<td>0.86</td>
</tr>
</tbody>
</table>

$^a$ Four-factor models.
$^b$ Both $M_0$ and $T_2$ values used. Optimal regression results shown.
$^c$ Effective range on a scale from 0 to 15.
values) for the four regression models predicting the six sensory attributes are shown. The correlation coefficients are in the range of 0.29–0.91 and the RMSECV is between 0.71 and 2.00. The CPMG PLSR and the SLICING prediction show almost equal predicting performance, whereas the bi-exponential prediction in some cases gave a slightly better result. Distribution analysis gave the most varying results, ranging from the worst to the best predictions. In general, the six sensory attributes are not well predicted by any of the four methods except for the moistness attribute where the bi-exponential model gives a correlation coefficient of $r = 0.84$ and an acceptable RMSECV is observed. This is sensible as the relaxation curves express the water content and distribution within the potato starch cells, whereby the predictions indicate that this attribute was expressed in the CPMG relaxation curves. The correlation coefficient of the attributes cohesiveness and mealliness are also acceptable for all four methods, but taking into consideration the RMSECV and the range of the scale used by the assessors, the overall prediction is not impressive.

6. Conclusion

For the investigation of the differences in potatoes and potato texture by low-field NMR, this study compared new and established modelling methods to analyse NMR data: CPMG PLSR, bi-exponential fitting, distribution analysis, and SLICING. The work consists of two parts: a qualitative data analysis of the potato samples where the interpretation of the loadings was of special interest. Secondly regression analysis was performed using six sensory attributes as predictor variables.

In the data analysis part, the results show that the SLICING method is superior to CPMG PLSR, bi-exponential fitting, and distribution analysis. The SLICING method decomposed the CPMG relaxation curve into four uni-exponential components describing all the variation in the data set up to the noise. It is possible to interpret the exponential decaying loadings, and directly relate them to the design variables: cultivar, dry matter and storage time. The distinction between the five potato cultivars is caused by properties related to the fast decaying loadings, as the properties of water related to a long transversal relaxation time do not seem to have the same influence on the separation of the groups. To understand the role of the water component more research is required.

In the regression analysis the predictions from CPMG PLSR and the SLICING scores were very similar. There is no gain using PLSR on the raw curves if data analysis is of interest. The predictions using bi-exponential fitting gave slightly better results (RMSECV ranging from 0.86 to 1.26 and correlation coefficients ranging from 0.63 to 0.87) than the predictions using the CPMG PLSR or the SLICING (RMSECV ranging from 1.05 to 1.49 and correlation coefficients ranging from 0.54 to 0.79). The predictions using the results from the distribution analysis gave varying results, and in general these results are inferior to the results from bi-exponential fitting.

Acknowledgements

Povlsen wish to thank the STVF (Danish Research Council) for financial support of the project Applied Quality Monitoring in the Food Production Chain (AQM), while Rinnan wish to thank the STVF for financial support through Project 1179. Thanks to professor Rasmus Bro (KVL) for valuable discussions, to Ole Dahl Pedersen (DJF) for performing the NMR analysis, and to the sensory panel (DJF).

References


