ELSEVIER

Trends in Food Science & Technology 13 (2002) 235–244

TRENDS IN FOOD SCIENCE & TECHNOLOGY

Review

Multivariate data analysis as a tool in advanced quality monitoring in the food production chain

Rasmus Bro*,† Frans van den Berg†, Anette Thybo‡, Charlotte M. Andersen†,§, Bo M. Jørgensen§ and Henrik Andersen¶

†Department of Food Science, Royal Veterinary and Agricultural University, DK-1958 Frederiksberg, Denmark (e-mail: rb@kvl.dk)

‡Department of Horticulture, Danish Institute of Agricultural Sciences, DK-5792 Aarslev, Denmark

§Department of Seafood Research, Danish Institute for Fisheries Research, DK-2800 Kgs.Lyngby, Denmark

¶Department of Animal Product Quality, Danish Institute of Agricultural Sciences, DK-8830 Tjele, Denmark

This paper summarizes some recent advances in mathematical modeling of relevance in advanced quality monitoring in the food production chain. Using chemometrics – multivariate data analysis – it is illustrated how to tackle problems in food science more efficiently and, moreover, solve problems that could not otherwise be handled before. The different mathematical models are all exemplified by food related subjects to underline the generic use of the models within the food chain. Applications will be given from meat storage, vegetable characterization, fish quality monitoring and industrial food processing, and will cover areas such as analysis of variance, monitoring and handling of sampling variation, calibration, exploration/data mining and hard modeling.

Introduction

There is an increased public and political focus on food production. Safety, nutritional value, eating quality, ethical, environmental, economic and social aspects are all issues that the food industry needs to be aware of and respond to. This is a result of the fact that most food industries produce low profit products, the ever rising wages in most industrialized countries, and the ever ongoing changes in life style. Increased information and purchasing power have triggered the latter, hereby making customers increasingly sophisticated, demanding and powerful. This leads to a necessity for efficient tools in monitoring, optimization, characterization, speciation and general handling of raw materials, processes intermediates and final products. Together with prediction of quality throughout the production chain, this becomes a must for the food industry to be competitive at the global niche markets of the future.

Optimal utilization of available data obtained throughout the production chain is an important aspect of developing the tools necessary to fulfill the above mentioned demands. Often, the food industry performs a number of different measurements throughout the process, typically for specific, dedicated purposes. This generates a large amount of data, which is seldom used outside its direct scope. Rather, it is used distinctly for one specific purpose. However, it can be of great interest to combine all available information in order to extract

* Corresponding author.

0924–2244/02/$ - see front matter
PII: S0924-2244(02)00138-3

even more relevant information from the collected data. Moreover, introduction of alternative measurements will also be necessary to accomplish the future demands, especially in relation to the prediction of quality. For example, on-line or in-line spectroscopy is a promising area where non-destructive and cheap measurements can be made providing *multivariate* and very general but accurate information on different chemical and physical properties of the samples measured (Archibald & Kays, 2000; Bro, 1998; Colquhoun, 1998; Engelsen, Mikkelsen, & Munck, 1998; Isaksson, 1990; Munck, Norgaard, Engelsen, Bro, & Andersson, 1998; Scotter, C., 1994; Scotter, C.N.G., 1997; Shibata, Ono, & Hirano, 2000; Simpkins & Harrison, 1995).

In this paper, examples will be given on how advanced multivariate data analysis can be helpful in analyzing complicated data sets obtained from monitoring production at different steps in the food production chain. The focus will not be on the mathematical aspects of these methods, but rather on the practical results obtained. The examples thus serve as illustrations of the benefits that can be obtained by utilizing multivariate data analysis on food related data. Emphasis is on the areas: *visualization, optimization* and *calibration*; all three are of importance in relation to developing tools ensuring optimal control in future food production.

The findings reported here were acquired in an ongoing project named Advanced Quality Monitoring involving several partners with quite distinct disciplinary backgrounds in Danish food research tradition. The paper illustrates the necessity of joining forces to solve complicated food chain related problems by modern scientific methods.

Visualization

Exploratory data analysis is an often neglected but highly useful discipline (Andersson, 2000; Munck *et al.*, 1998; Tukey, 1977; Weihs, 1993). Usually, data analysis is performed as a confirmatory exercise, where a postulated hypothesis is claimed, data generated accordingly and the data analysed in order to either verify or reject this hypothesis. In effect, no new knowledge is obtained in confirmatory analysis except the possible verification of a prior postulated hypothesis. Confirmatory analysis, though, is highly useful in many forms of traditional quality control.

In contrast, using exploratory analysis, the data are gathered in order to represent as broadly and well as possible the problem under investigation. The data are analysed and through the, often visual, inspection of the results, hypotheses are suggested *on the basis of the empirical data*. Thus, the aim in exploratory analysis is broader than in confirmatory analysis. Rather than defining the whole problem mentally and merely verifying these mental constructs through data and analysis, the empirical real-life observations of the problem are used to obtain knowledge of the underlying characteristics. This gives an increased possibility for obtaining new, different and interesting information about the problem, possibly leading to new hypotheses. Consequently, exploratory data analysis is an extraordinary tool in displaying thus far unknown information from established and potential monitoring methods, which subsequently can be used to establish solid measuring methods of importance in food quality control. Two examples will be given on how to use dedicated mathematical models combined with relevant advanced measurements to extract, display and understand important underlying causes and effects in different food problems.

Characterizing water distribution of food samples by low-field NMR

The distribution and mobility of water in complex systems such as muscle-based food are important for perceived eating quality and in many cases also for suitability for processing and storage. In the muscle, water may e.g. be bound tightly to proteins, encompassed by the muscle fibrils, in the cytoplasm or sarcoplasm, or in the extra-cellular fluid.

Low-Field Nuclear Magnetic Resonance (LF-NMR) transverse relaxation measurement is the method of choice when the task is to probe the state of water in samples from such food systems. Although it is possible to imagine a large number of possibilities (for example, water–protein interactions) and consequently an almost continuous distribution of water mobilities, it turns out in practice that a few compartments or pools of water can be identified from the LF-NMR transverse relaxation signal (Pedersen, Bro, & Engelsen, 2002). This signal is a weighted sum of mono-exponential decays, the number of which equals the number of different pools. The weights are proportional to the pool size (amount of water) and the relaxation times are a function of the characteristic water mobilities. These parameters therefore provide a useful picture of the states of water in a given sample. From the matrix of pool sizes (each row representing a sample and each column a pool) one can often predict quality-related properties by multivariate calibration methods.

Traditionally, the LF-NMR transverse relaxation signal has been treated by curve fitting methods. A set of parameters is obtained independently from each sample. The main problem is the highly correlated parameters, and if the number of components is not known, over-fitting is a realistic risk. Comparisons between samples are also impeded by the fact that the resulting relaxation times may differ profoundly among samples so that it is not obvious whether the components are comparable at all. It is therefore practical to stabilize the model fit by forcing a set of components to be common to all samples, if such a set can be assumed to exist. The number of components obviously depends

on how close the relaxation times are. The samples will thus differ in the relative amounts of the various components which are linear parameters that can be determined analytically.

The most difficult task is to find the set of common relaxation components. Here, modern developments in multi-way chemometrics have provided a powerful tool: parallel factor analysis (PARAFAC) combined with a data rearrangement that will result in tri-linear data (Pedersen *et al.*, 2002; Windig & Antalek, 1997). The PARAFAC model is a mathematical model that, under certain mild conditions, can separate measured signals into the underlying contributions (Bro, 1997; Harshman, 1970; Harshman & Lundy, 1994). This method was recently applied to the study of water states in pre-frozen cod stored in modified atmosphere at $+2^{\circ}C$ (Jensen, Guldager, & Jørgensen, in press). It was possible to identify four pools of water and to determine how the water distribution changed during frozen storage dependent on the temperature (-20 or $-30^{\circ}C$; Fig. 1). Moreover, the distribution during cold storage could be observed to change in a way only slightly dependent on the storage temperature. Combined with other methods such as microscopy and calorimetry, measuring the water distribution as described provides new insight in those changes in the muscle that lead to quality deterioration during storage.

Combining multiple data sets with multi-block methods

Many data analysis problems can be specified in terms of blocks. For example, a single block of spectra can be analyzed by Principal Component Analysis (PCA) (Wold, Esbensen, & Geladi, 1987) to find the main phenomena in the set. The same block of spectra can be used together with a block of quality values to build a predictive Partial Least Squares (PLS) regression model (Vandeginste & Massart, 1997) so that the quality can be predicted in the future directly from the easily measured spectra. In many research and process questions, different but distinct sources of information are available on the same set of objects. Examples are different analytical techniques (spectroscopy, rheology, wet-chemistry, etc.) collected on the same samples or the same parameters measured on the same samples at different stages in a production process. In recent years, methods have been developed to handle multiple blocks of this kind (Westerhuis, Kourti, & MacGregor, 1998). These modeling methods are extensions of well established so-called one- and two-block factorial models such as PCA and PLS.

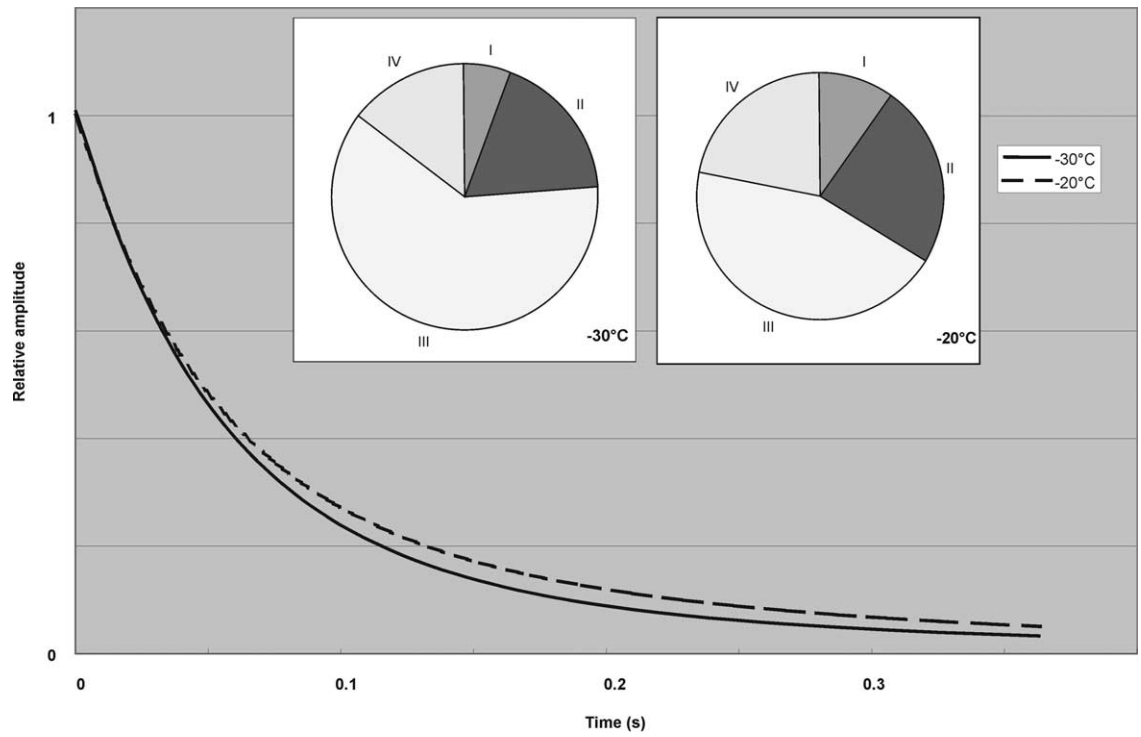

Fig. 1. ^{1}H NMR relaxation signals and water pools in samples from thawed cod. Intact cod muscle was frozen for three months at $-30^{\circ}C$ and subsequently for three months at the temperature indicated (-30 or $-20^{\circ}C$). After thawing, the muscle samples were minced and the spin-spin relaxation signals recorded. From these signals four water pools (marked I–IV) could be identified with pool I being the most mobile (free) water. The normalized relaxation curves and the relative amount of water in the four pools are shown for each of the two storage conditions. The figure was constructed from data in Jensen, Guldager, & Jørgensen (2002).

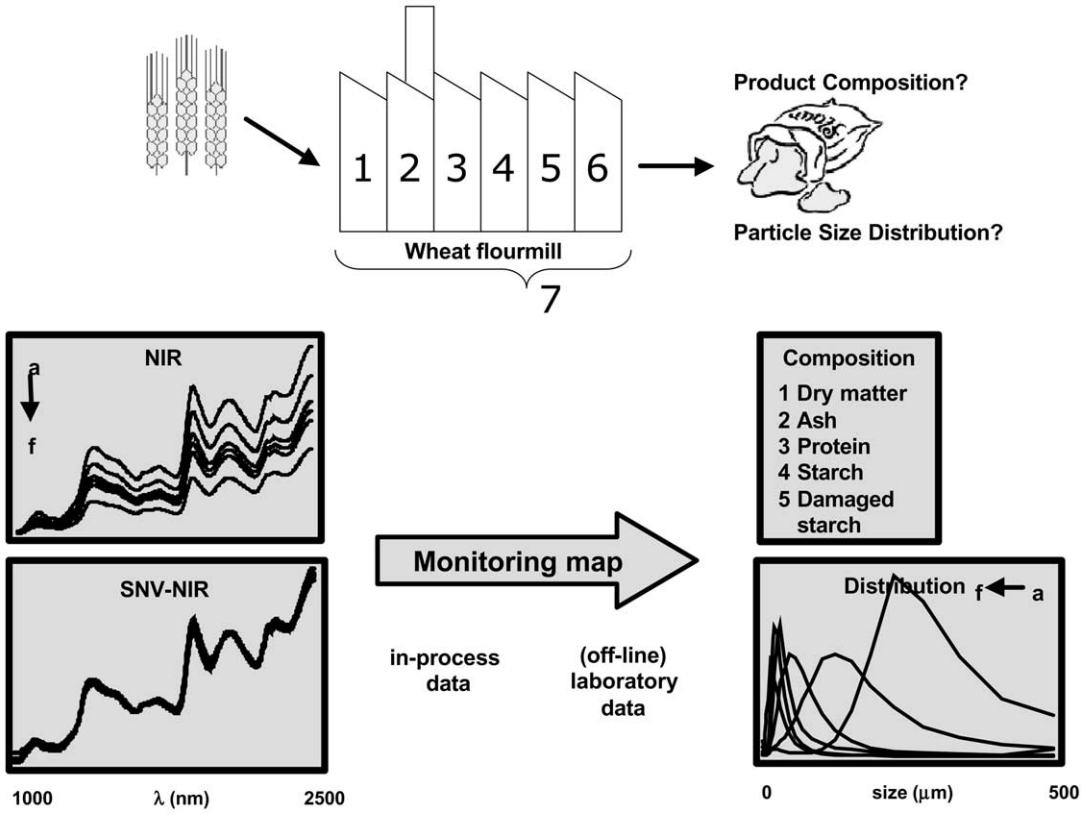

Fig. 2. Building blocks of the wheat flourmill data-set.

The concept of multi-block modeling will be illustrated using data from a wheat flourmill (Nielsen, Bertrand, Micklander, Courcoux, & Munck, 2001). At six different streams in the mill, samples were taken. A seventh sample was composed from equal proportions of the first six. The samples were separated in six size fractions labeled *a–f*, where *f* is the remaining fraction after separation (see Fig. 2; Berg, 2001). Fast (in-process) near infrared (NIR) spectral analysis and slower (laboratory) chemical composition and laser-scatter size distribution analysis are collected for all the samples. In order to emphasize chemical information in the spectra over the scatter information, a standard normal variate signal correction is applied to the NIR. After this action there are four contributors in the factor model: the two predictor blocks NIR and SNV-NIR and the two response blocks composition and distribution. The aim is to create a *monitoring-map* that will show the position of future mill samples using only their NIR-spectra. For this purpose a multi-block PLS model is constructed that seeks to predict the two blocks of chemical and distribution data from the two blocks of NIR data (see Fig. 3).

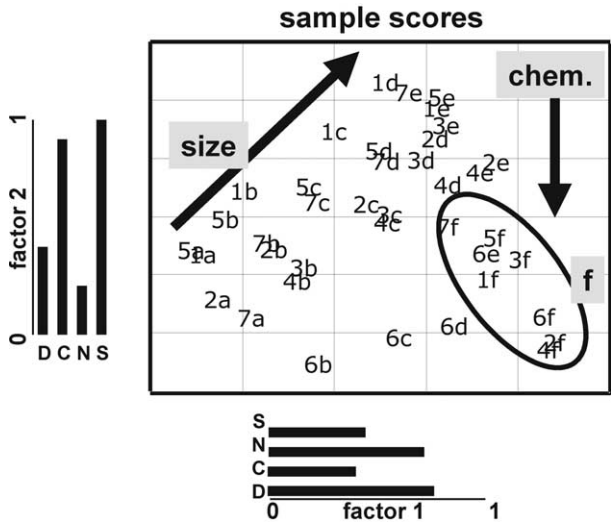

Fig. 3. Multi-block monitoring-map for the wheat flourmill data-set; 'S' is SNV-NIR, 'N' is NIR, 'C' is chemical composition and 'D' is particle size distribution.

Fig. 3 provides a visualization of the multi-block model in terms of the samples marked by their labels. It is remarkable that an accurate summary of all the individual 1437 variables organized in four different blocks can be given in such a condensed map. Samples that lie close together are similar while samples that are distant are dissimilar. It is seen that the horizontal axis primarily

explains the size distribution. Next to the axis, the relative block contribution is plotted which shows that all four blocks of data have a considerable contribution in this direction, moderately dominated by the NIR and Distribution data. Thus NIR and distribution data are the ones mainly reflecting size distribution. The second axis – dominated by the SNV-NIR and Compositional/Chemical direction – splits of the samples from location 6, which are known to be chemically different in composition (Nielsen *et al.*, 2001). The *f*-samples are seen to form a separate cluster in the map.

The essence of multi-block methods is to do data analysis thinking in terms of building blocks rather than individual variables. This significantly reduces the risk of being overwhelmed even when a lot of (different) data, related to the same set of objects, has to be analysed. Being an extension of 'conventional' factorial models – PCA and PLS – all their strong features remain valid, augmented with block-specific (preferably graphic) information and diagnostics.

Optimization

Optimization is a challenging problem in any process industry. Optimization can be used to minimize product quality variability, maximize yield, minimize energy consumption, etc. In the food industry, optimization is often based on empirical observations, which do not take into consideration all processing steps involved. In a more scientific setting, optimization is often performed in a systematic way through experimentally designed data analysed by so-called analysis of variance (Hirsh, 1977; Latorre, 1984; Massart *et al.*, 1988; Montgomery, 1991; Morgan, Burton, & Church, 1989; Stahle & Wold, 1989). Various important factors are systematically and independently varied in order to verify their influence and interactions on the property under investigation. In this section, two interesting and new applications of optimization are described.

Visualizing experimentally designed data of meat storage properties

Modified atmosphere packaging is widely used to extend the shelf life of fresh meat. One aspect of optimizing the atmosphere is to retain the red color of the meat, which is favorable for consumer preferences. Meat color was monitored in an experiment using *Longissimus dorsi* muscles of several animals. The factors varied were storage time, storage temperature, O_2 content in headspace, and amount of light exposure. Red color was measured for different settings of these factors; the settings being defined through a D-optimal design because the limited number of samples prohibited a full or fraction factorial design to be used (Bro & Jakobsen, 2002).

It is characteristic of many biological systems, that the influence of different factors on certain properties, such as the color in this case, is not simple. Often, the influence of one factor is dependent on the level of other factors. Hence, there are *interactions* between the factors. In fact, the interactions may be mainly responsible for the relation between the factors and the property. For example, the influence of soil type and fertilizer on wheat yield is mainly a function of the combination of the two factors rather than independent functions of both.

A new model called GEneralized Multiplicative ANalysis OF Variance—GEMANOVA (Bro, 1998; Bro *et al.*, 2001; Bro & Heimdal, 1996; Heimdal, Bro, Larsen, & Poll, 1997) has been suggested for analysing data where interactions are likely to be the main source of variation. The GEMANOVA model was applied to the above meat data and the result is illustrated in Fig. 4. The GEMANOVA model states that the color can be explained by two independent phenomena: the initial color of the meat and the degradation of color. The absolute degradation is thus independent of the initial level of color. The initial color of the meat is shown in the lower right part of Fig. 4 where the model estimates are compared to the actual measured color of the six different muscles used, showing excellent agreement. The degradation is given by a three-way interaction between storage time, temperature and light exposure. Oxygen is (surprisingly) found to have no significant effect in the investigated domain. For a specific storage time, temperature and light exposure, the estimate of the absolute degradation is found by reading the ordinate of the three corresponding graphs in Fig. 4 and multiplying these together. Thus at storage time zero there is no degradation as the storage effect is zero (upper left). On the other hand, going from temperature 2 to 8°C, it is seen that the temperature effect increases from 1.2 to 2.4. Therefore, the effect of this change will be that the overall decrease is twice as high at 8°C as it is at 2°C, regardless of all other factors.

As can be seen, this *multiplicative* model of the three-way interaction generates a model which is easily and intuitively understood in terms of the underlying problem. A similar understanding would not be possible with a traditional ANOVA model.

Characterizing the effect of biological inhomogeneity with analysis of variance

In many applications, instrumental measurements are performed at only one specific part of the product due to the size of the measurement device. The measurements are subsequently related to the overall properties of the sample. This introduces a random sampling error that can be reduced by measuring replicates at different parts of the sample. The number of replicates used for building the calibration model and used for future

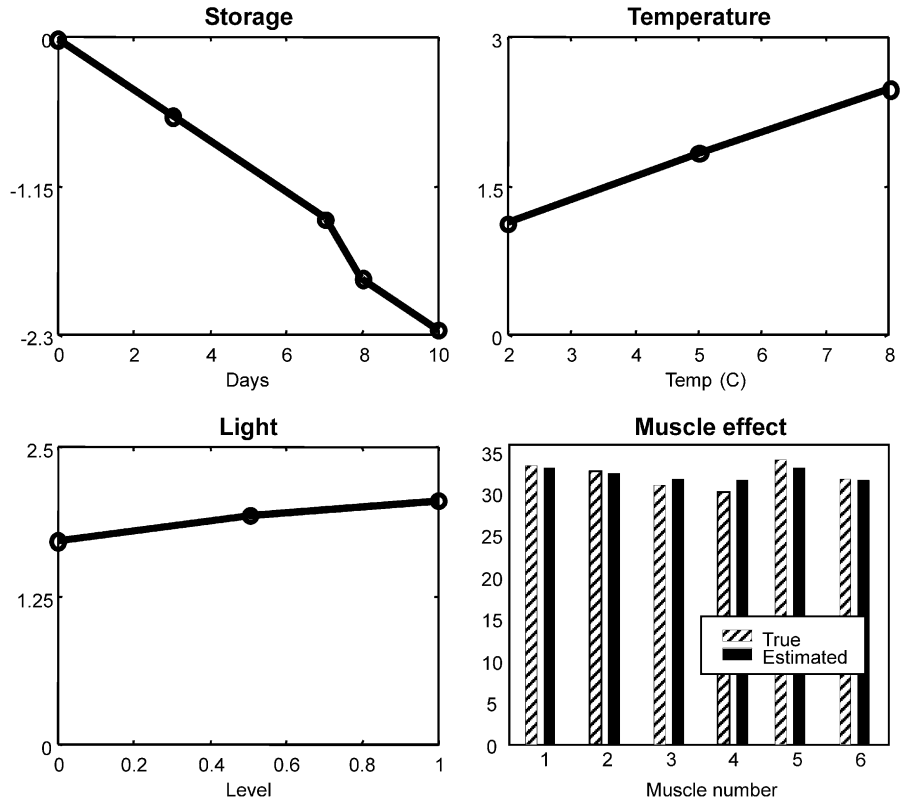

Fig. 4. Results of GEMANOVA model showing the influence of storage, temperature, light and muscle on meat color. Oxygen has no significant effect.

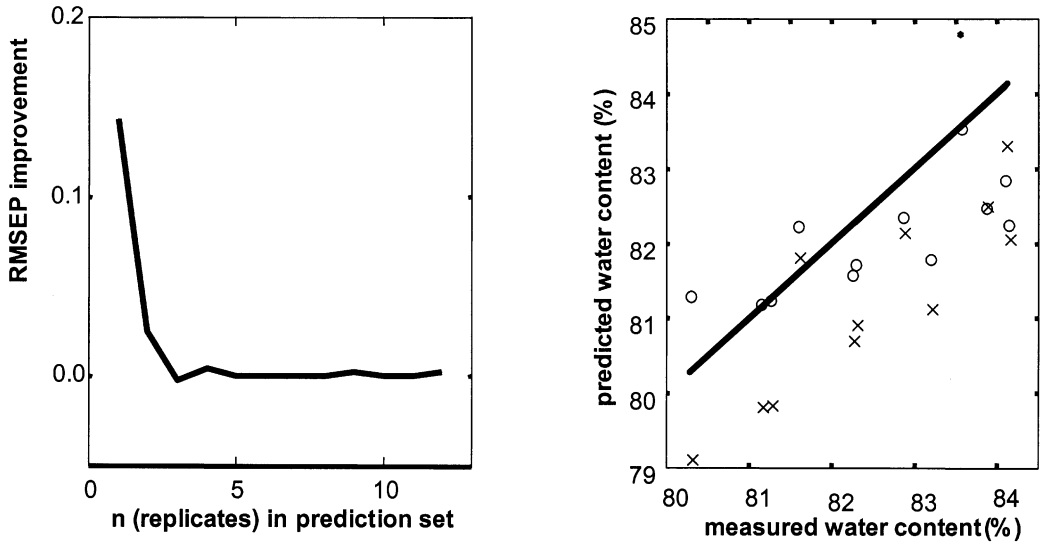

Fig. 5. (left) Difference in the average prediction error, Root Mean Squared Error of Prediction (RMSEP), between the corrected estimation and least squares estimation when six replicates are used for calibration and from one to 12 replicates are used in prediction; (right) an example of the predicted water content versus the measured water content obtained with nine replicates in the calibration set and only one measurement in prediction (crosses—least squares; circles—corrected least squares).

prediction may not be the same. In such situations, the predictions can be improved by correcting the regression coefficients so that they correspond to the least squares estimate with the number of replicates in the calibration set. The advantage of applying this correction based on the theory of the measurement error models is illustrated in Fig. 5 where prediction of water content of fish muscle from ^{1}H low-field NMR relaxations is used as an example.

The difference between the predicted error obtained using direct least squares regression and when correcting for the number of measurements in the prediction set is largest when only one measurement is used for prediction. When the measurements of the prediction samples consist of few replicates, there seems to be an advantage in correcting the regression coefficients. However, this advantage diminishes when more than approximately three measurements are used for each sample.

An example of the improved predictions obtained by the correction is shown in Fig. 5. The model is calibrated using samples with nine replicates and validated on a test set where the samples are measured only once.

Calibration

Although calibration of sensors and measurements is a universal task of process operators and laboratory technicians throughout the food industry, many novel methods of data handling are interesting supplements to current technologies. Specifically the use of multivariate calibration as propagated in the field of chemometrics can turn measurement signals with no apparent selectivity into models with good predictive performance for a wide variety of properties (Martens & Næs, 1989). Three examples will hint on the diversity of potential applications.

Predicting final product sensory quality from low field NMR measurements on the raw product

Texture is an important quality attribute for cooked potatoes. It is closely related to the dry matter content. The potato product industry is demanding methods that can predict the final texture of potatoes non-destructively. This will enable sorting of potatoes in gradings of various texture qualities with less variability, having a more appropriate quality for a given final product. This will increase the quality within products, reduce the waste, and increase the income as well as the satisfaction of the consumers.

Low field ^{1}H NMR (LF-NMR) is known to reflect compartmentalization of water phases and hence texture in many food products such as meat, bread and fish mince. For potatoes, a prediction of texture in cooked potato samples from raw sample measurement by LF-NMR have shown high correlations for many sensory texture attributes using NMR relaxation curves (Fig. 6) as well as bi-exponential fit parameters (T_2s) (Thybo, Bechmann, Martens, & Engelsen, 2000; Thygesen, Thybo, & Engelsen, 2001). In a recent experiment, 23 potato samples of different varieties and dry matter content were investigated. A PARAFAC (see above) prediction of the sensory attributes was compared with a PLS prediction using NMR relaxation curves as well as T_2s (Povlsen, Rinnan, van den Berg, Andersen, & Thybo, 2002). The results are given in Table 1 and show that most of the sensory attributes are well predicted from LF-NMR. The PLS predictions seemed to give the best predictions.

LF-NMR is a non-destructive method and hence interesting for process monitoring of texture in the food

Table 1. Prediction of sensory texture attributes using low-field NMR given by correlation coefficients between measured and predicted attributes

Sensory texture attributes	PLS on relaxation curves[a]	PLS on bi-exponential fitting parameters[b]	PARAFAC on relaxation curves[a]
Hardness	0.80	0.81	0.66
Cohesiveness	0.84	0.83	0.74
Adhesiveness	0.74	0.63	0.53
Mealiness	0.85	0.90	0.73
Graininess	0.71	0.76	0.62
Moistness	0.85	0.90	0.76

[a] Four-factor models. [b] M_2 and T_2 values used.

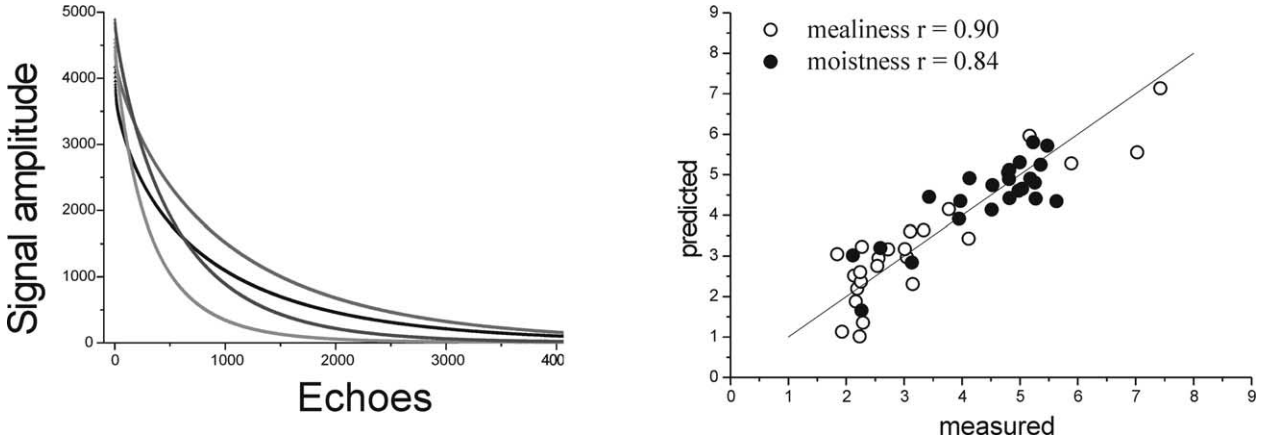

Fig. 6. (left) Low-field ^{1}H NMR relaxation curves of four examples of different potato varieties; (right) prediction curves of mealiness and moistness using NMR.

industry. Through multivariate calibration modeling, quality can be assessed directly and quickly giving more relevant and time-critical information for controlling the process.

Prediction of sensory texture attributes from full uniaxial compression curves on raw material

Uniaxial compression is a destructive instrumental approach to determining mechanical properties such as hardness, crispness or springiness. Uniaxial compression is not considered a rapid instrumental method, but can be used for (off-line) randomly sampled quality control.

In uniaxial compression a sample from a potato tuber is compressed (e.g. 75% at constant velocity). Usually, two to four parameters are extracted from the compression curves (force and deformation at fracture and moduli before fracture) and used for further interpretation and correlation with sensory texture attributes.

Using multivariate data analysis it is possible to use the full compression curve instead of only a few features extracted. Compared with the information content of the four curve features, more information may be found using the full curves. Prediction of sensory texture attributes of cooked potatoes from either full curves or from curve parameters on raw samples are given in Table 2. Most of the sensory attributes were better predicted from full curves than from curve parameters. This indicates that more information is found in the full curves and that these can replace the traditional calculation of curve parameters. This may be an

Table 2. Prediction of sensory texture attributes using uniaxial compression given by correlation coefficients between measured and predicted attributes

Sensory texture attributes	PLSR on full force-deformation curves	PLSR on 4 curve parameters
Hardness	0.89	0.81
Cohesiveness	0.84	0.82
Adhesiveness	0.78	0.63
Mealiness	0.81	0.76
Graininess	0.77	0.75
Moistness	0.79	0.72

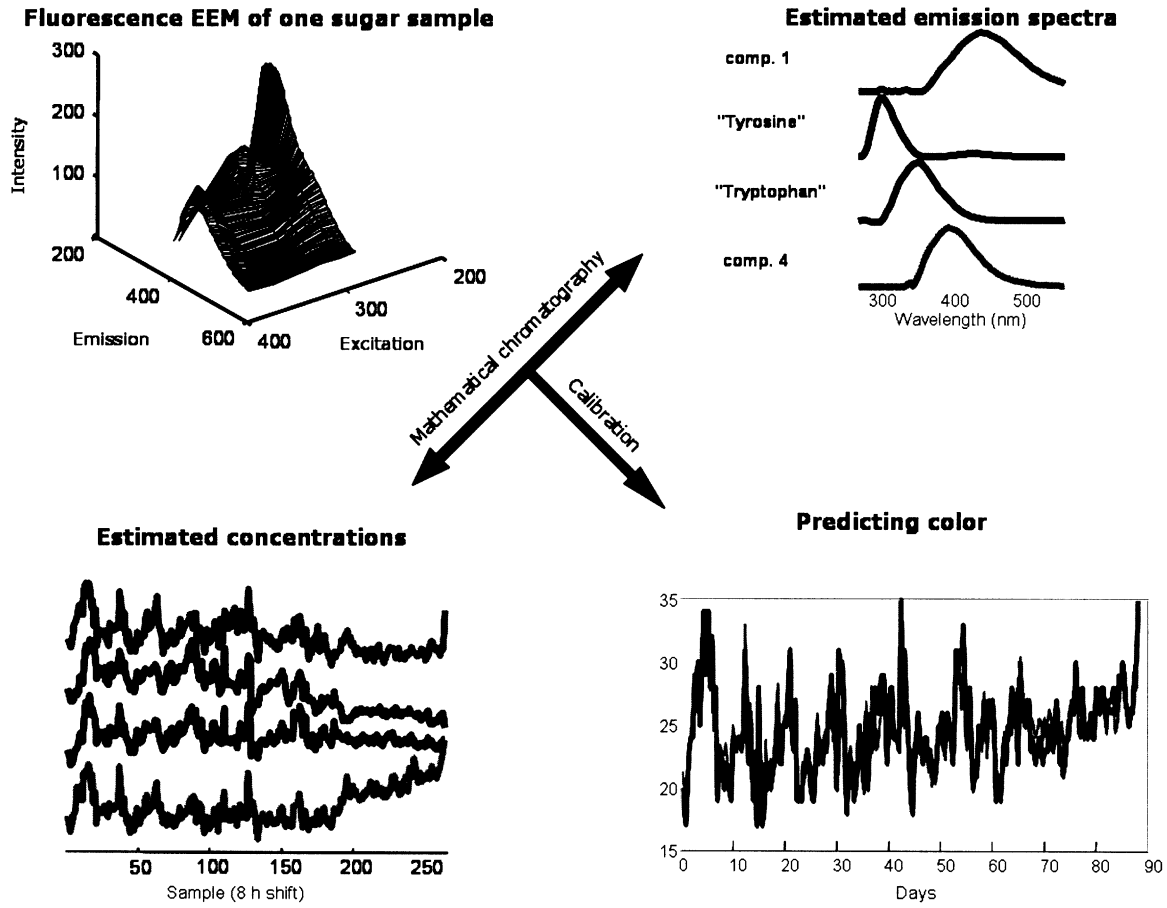

Fig. 7. Sugar samples were taken every eighth hour, dissolved in water and measured spectrofluorometrically. An example is shown upper left. PARAFAC decomposition of the data provides relative concentrations (lower left), relative excitation spectra (not shown) and relative emission spectra (upper right). The four concentrations are used for predicting color of the sugar shown lower right.

advantage as some of the curve parameters (e.g. moduli) are defined in many different ways making the comparison of results difficult (Thybo & van den Berg, 2002).

Predicting sugar quality from fluorescence-based mathematical chromatography

In the sugar industry, sugar is manufactured from either beet or cane. Through a number of unit operations, the crystalline product is obtained. The main quality parameters are color and ash content (Nørgaard, 1995b). These are used externally for reporting quality and internally for controlling the last part of the process. Measuring these parameters requires manual sampling and wet-chemical laboratory work. It is thus expensive and there is a certain lag between the actual sampling and the determination of quality. Having a cheaper, faster, on-line measurement would be highly beneficial for process control purposes.

This can be made possible by the use of fluorescence spectroscopy and indirect multivariate calibration as first described by Nørgaard (1995a, 1995b). The results have been further refined over the years. Bro (1999) showed that the fluorometric data of sugar can be decomposed into chemically meaningful components with the use of a PARAFAC model. In this case, fluorescence excitation–emission matrices (EEM) as shown in Fig. 7 (upper left) are measured on sugar samples dissolved in water. A sample is taken every eighth hour, the frequency with which the quality parameters are determined traditionally. In all, 268 samples are taken, covering the 3 months' yearly production.

A PARAFAC model of these fluorescence data directly provides relative concentrations, emission and excitation spectra of the *underlying* chemical analytes as shown in Fig. 7. Thus, the holistic fluorometric characterization of the sugar samples can be expressed by only four underlying phenomena. This means that during the three months, *any* sugar sample can be fully characterized with respect to its fluorescence fingerprint by varying amounts of four different estimated fluorophores. These four relative amounts/concentrations are shown in the lower left part of the figure. The most intriguing aspect in this model though, is that the disentangled fluorophores, determined *solely* from the direct measurements, can be identified on a chemical basis. In this case, comparing the emission and excitation spectra with known fluorophores, two of them are identified as tryptophan and tyrosine, respectively. As in ordinary chemical chromatography, these findings can be further substantiated in different ways (Baunsgaard, Andersson, Arndal, & Munck, 2000; Baunsgaard, Nørgaard, & Godshall, 2000).

By multiple linear regression it is possible to make quantitative models, e.g. predicting the color of the sugar from the four concentrations found by PARAFAC as shown in the figure. Thus, a quantitative model is obtained which is much easier to handle than the standard wet-chemical approach and which has a chemical basis making it more transparent (directly related to the four components) and descriptive than the standard methods.

Conclusion

In this paper we show that the recent advances in chemometrics, e.g. combined with the use of new on-line or at-line spectroscopic measurements provide an important and interesting direction of research. Through the use of dedicated mathematical models it is possible to overview huge data sets and complicated problems in an intuitive and straightforward manner.

Examples have shown the benefit of proper *visualization* in data analysis in e.g. details on the water distribution of products or modeling of large data-set via multi-block methods. Also shown are the promising prospects of *optimization* in e.g. the parameter estimation for storage and packaging, or the characterization of in-homogeneity in raw materials for the food industry. In the last paragraph the potential of multivariate *calibration* was demonstrated by examples on the prediction of final product properties from measurements on the raw material and the use of advanced modeling techniques for process parameter monitoring and prediction.

All these benefits arise from the proper use of advanced mathematical models and show that exploratory, multivariate data analysis guided by visualization is a fruitful complementary discipline in food research, and can generate results that either immediately or readily can be used in combination with existing food quality monitoring methods or potential measuring methods at a level which in the future can fulfill the demands that the food industry is facing with regard to totally quality control.

Acknowledgements

The authors gratefully acknowledge support provided by LMC (Center for Advanced Food Studies) and AQM (Advanced Quality Monitoring in the Food Production Chain) supported by the Danish Ministry of Research.

References

Andersson, C. A. (2000). *Exploratory multivariate data analysis with applications in food technology*. PhD thesis, Dept. Dairy & Food Science, The Royal Veterinary & Agricultural University, Denmark.

Archibald, D. D., & Kays, S. E. (2000). Determination of total dietary fiber of intact cereal food products by near-infrared reflectance. *Journal of Agricultural and Food Chemistry, 48*, 4477–4486.

Baunsgaard, D., Andersson, C. A., Arndal, A., & Munck, L. (2000). Multi-way chemometrics for mathematical separation of

fluorescent colorants and colour precursors from spectrofluorimetry of beet sugar and beet sugar thick juice as validated by HPLC analysis. *Food Chemistry*, *70*, 113–121.

Baunsgaard, D., Nørgaard, L., & Godshall, M. A. (2000). Fluorescence of raw cane sugars evaluated by chemometrics. *Journal of Agricultural and Food Chemistry*, *48*, 4955–4962.

Berg, van den F. W. J. (2001). Multi-block PLSR models in food technology. In Proceedings on 'PLS and related methods' (pp. 385–394)

Bro, R., & Jakobsen, M. (2002). Exploring complex interactions in designed data using GEMANOVA. Color changes in fresh beef during storage. *Journal of Chemometrics*, *16*, 294–304.

Bro, R. (1997). PARAFAC, Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, *38*, 149–171.

Bro, R. (1998). *Multi-way analysis in the food industry. Models, algorithms, and applications*. NL: University of Amsterdam Available from <http://www.mli.kvl.dk/staff/foodtech/brothesis.pdf>.

Bro, R. (1999). Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometrics and Intelligent Laboratory Systems*, *46*, 133–147.

Bro, R., & Heimdal, H. (1996). Enzymatic browning of vegetables, Calibration and analysis of variance by multiway methods. *Chemometrics and Intelligent Laboratory Systems*, *34*, 85–102.

Colquhoun, I. J. (1998). High resolution NMR spectroscopy in food analysis and authentication. *Spectroscopy Europe*, *10*, 8, 10, 12, 14, 16, 18, 10.

Engelsen, S. B., Mikkelsen, E., & Munck, L. (1998). New approaches to rapid spectroscopic evaluation of properties in pectic polymers. *Progress in Colloid and Polymer Science*, *108*, 166.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Harshman, R. A., & Lundy, M. E. (1994). PARAFAC: parallel factor analysis. *Computational Statistics and Data Analysis*, *18*, 39–72.

Heimdal, H., Bro, R., Larsen, L. M., & Poll, L. (1997). Prediction of polyphenol oxidase activity in model solutions containing various combinations of chlorogenic acid, (-)-epicatechin, O2, CO2, temperature and pH by multiway analysis. *Journal of Agricultural and Food Chemistry*, *45*, 2399–2406.

Hirsh, R. F. (1977). Analysis of variance in analytical chemistry. *Analytical Chemistry*, *49*, 691A–700A.

Isaksson, T. (1990). *Multivariate calibration of near infrared reflectance data from food products. Experimental design, linearization and applications*. Department of food Science, Chalmers Uni. of Technology and MATFORSK Norwegian Food Research Institute

Jensen, K.N., Guldager, H.S., & Jørgensen, B.M. Three-way modelling of NMR relaxation profiles from thawed cod muscle. *Journal of Aquatic Food Product and Technology* (in press)

Latorre, G. (1984). Analysis of variance and linear models. In B. R. Kowalski (Ed.), *Mathematics and statistics in chemistry* (pp. 377–391). Dordrecht: D: Reidel Publishing.

Martens, H., & Næs, T. (1989). *Multivariate calibration*. Chichester: Wiley & Sons.

Massart, D.-L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., Kaufman, L., Vandeginste, B. G. M., Kaufman, L. (Eds.). (1988). *Chemometrics: a textbook*. Amsterdam: Elsevier.

Montgomery, D. C. (1991). *Design and analysis of experiments*. New York: John Wiley & Sons.

Morgan, E., Burton, K. W. C., & Church, P. A. (1989). Practical exploratory experimental designs. *Chemometrics and Intelligent Laboratory Systems*, *5*, 283–302.

Munck, L., Nørgaard, L., Engelsen, S. B., Bro, R., & Andersson, C. A. (1998). Chemometrics in food science—a demonstration of the feasability of a highly exploratory, inductive evaluation strategy of fundamental scientific significance. *Chemometrics and Intelligent Laboratory Systems*, *44*(1–2), 31–60.

Nielsen, J. P., Bertrand, D., Micklander, E., Courcoux, P., & Munck, L. (2001). Study of NIR spectra, particle size distributions and chemical parameters of wheat flours: a multi-way approach. *Journal of Near-Infrared Spectroscopy*, *9*, 275–285.

Nørgaard, L. (1995a). A multivariate chemometric approach to fluorescence spectroscopy. *Talanta*, *42*, 1305–1324.

Nørgaard, L. (1995b). Classification and prediction of quality and process parameters of beet sugar and thick juice by fluorescence spectroscopy and chemometrics. *Zuckerindustrie*, *120*, 970–981.

Pedersen, H. T., Bro, R., & Engelsen, S. B. (2002). Towards rapid and unique curve resolution of low-field NMR relaxation data: Trilinear SLICING versus two-dimensional curve fitting. *Journal of Magnetic Resonance*, *157*, 141–155.

Povlsen, V. T., Rinnan, Å., van den Berg, F., Andersen, H. J., & Thybo, A. K. 2002. Direct decomposition of NMR relaxation profiles and prediction of sensory attributes of potato samples. *Food Science and Technology* (submitted for publication)

Scotter, C. (1994). Food, glorious food, a look at the menu for spectroscopists. Analysis Europa, October: 23–25

Scotter, C. N. G. (1997). Non-destructive spectroscopic techniques for the measurement of food quality. *Trends in Food Science and Technology*, *8*, 285–292.

Shibata, K., Ono, M., & Hirano, S. (2000). Quality evaluation of chocolate by using near infrared spectroscopy. *Journal of the Japanese Society for Food Science and Technology—Nippon Shokuhin Kagaku Kogaku Kaishi*, *47*, 692–699.

Simpkins, W., & Harrison, M. (1995). The state of the art in authenticity testing. *Trends in Food Science and Technology*, *6*, 321–328.

Ståhle, L., & Wold, S. (1989). Analysis of variance (ANOVA). *Chemometrics and Intelligent Laboratory Systems*, *6*, 259–272.

Thybo, A. K., Bechmann, I. E., Martens, M., & Engelsen, S. B. (2000). Prediction of sensory texture of cooked potatoes using uniaxial compression, near infrared spectroscopy and low field H-1 NMR spectroscopy. *Food Science and Technology*, *33*, 103–111.

Thybo, A. K. & van den Berg, F. (2002). Full uniaxial compression curves for predicting sensory texture quality of cooked potatoes. *Journal of Texture Studies*, 33, 119–137.

Thygesen, L. G., Thybo, A. K., & Engelsen, S. B. (2001). Prediction of sensory texture quality of boiled potatoes from low-field ^{1}H NMR of raw potatoes. *The role of chemical constituents. Food Science and Technology*, *34*, 469–477.

Tukey, J. W. (1977). *Exploratory data analysis*. USA: Addison Wesley.

Vandeginste, B. G. M., & Massart, D. L. (1997). *Handbook of chemometrics and qualimetrics B*

Weihs, C. (1993). Multivariate exploratory data analysis and graphics: a tutorial. *Journal of Chemometrics*, *7*, 305–340.

Westerhuis, J. A., Kourti, T., & MacGregor, J. F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, *12*, 301–321.

Windig, W., & Antalek, B. (1997). Direct exponential curve resolution algorithm (DECRA): a novel application of the generalized rank annihilation method for a single spectral mixture data set with exponentially decaying contribution profiles. *Chemometrics and Intelligent Laboratory Systems*, *37*, 241–254.

Wold, S., Esbensen, K. H., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, *2*, 37–52.