New aspects of chemometrics applied to spectroscopy

- examples from research in fish for production

Ph.D. thesis by

Charlotte Møller Andersen

Supervisors

Professor Rasmus Bro
Food Technology
Department of Dairy and Food Science
The Royal Veterinary and Agricultural University
Rolighedsvej 30, DK-1958 Frederiksberg C

Senior Researcher Bo M. Jørgensen
Department of Seafood Research
Danish Institute for Fisheries Research
Søltofts Plads, DTU bygn. 221
DK-2800 Kgs. Lyngby

Preface

This thesis is submitted to fulfil the requirement for obtaining the Ph.D. degree. The project is a part of the framework programme "Advanced Quality Monitoring in the Food Production Chain" supported by the Danish Research Council and involving several research groups within the food science area.

The work has been carried out at the Food Technology Group, Department of Dairy and Food Science, The Royal Veterinary and Agricultural University and Department of Seafood Research (FF), Danish Institute for Fisheries Research from September 1999 to March 2003 under supervision of professor Rasmus Bro and senior researcher Bo Jørgensen with whom I have had many inspiring and encouraging discussions, and who have given useful help and constructive comments during the work.

The project has included three months of work in Norway at MATFORSK where researcher Jens Petter Wold has been a very inspiring and enthusiastic colleague and adviser. I want to thank both him and rest of the staff at MATFORSK for making me feel comfortable and welcome.

I would like to thank all my colleagues at the Food Technology Group, and at the Raw Material and Product Technology Group at FF for their help and scientific inspiration and for providing a good working environment. I especially wish to thank Hugo Ladefoged, Claus Reesbøl, Tina Rathjen and Rie Sørensen for assistance during the experimental and practical work. Vibeke Povlsen, Frants Torp Madsen and Gilda Kischinovski are thanked for proofreading the manuscript. Finally, I would like to thank Søren T. Christensen and Birgitte A. Sobiecki for their help on providing the literature of interest.

Frederiksberg, March 2003
Charlotte Møller Andersen

Summary

The overall purpose of the present Ph.D. thesis is to show how new as well as known chemometric methods can be applied to spectroscopic data. The spectroscopic measurements are obtained as multivariate or multi-way data, which due to the complexity of the measured samples can be rather complex. Therefore, chemometric methods are necessary in order to resolve and fully understand the underlying structure reflected in the measurements.

Fluorescence spectroscopy and nuclear magnetic resonance (NMR) measured on fish muscles are used as examples to illustrate the potential use of the methods on real products. Focus has been on three main subjects, which are exemplified in the six included papers. The three subjects are:

1) Exploratory analysis of spectroscopic data
2) Validation of models
3) Quantifying and handling errors in instrumental measurements

Application of spectroscopic measurements makes it possible to measure the product non-destructively. However, such measurements can be rather complex due to the composition and heterogeneity of the food product. This is illustrated by exploratory analysis of fluorescence emissions of cod and salmon, fluorescence landscapes of fish muscle extracts and low-field NMR measured on gadoid fish muscle. Using principal component analysis (PCA) and parallel factor analysis (PARAFAC) to interpret the data, it is suggested that fluorescence spectra of fish muscle are primarily due to the connective tissue, NADH, oxidation products, lipids, pigments and amino acids. Slicing modelling of low-field NMR decays shows the distribution of water within fish fillets and the division of water into different pools. The content of water in these pools is correlated to the water holding capacity (WHC) of the fish muscle, which is an important quality parameter.

Furthermore, the exploratory analysis shows how three-way data can be decomposed and interpreted. Problems involved in PARAFAC modelling of fluorescence excitation-emission data such as missing values and scatter are discussed and examples of how these can be handled are illustrated.

Sampling is an important aspect to consider, since fish muscle as other biological material is heterogeneous. An example is given of how sampling errors in instrumental measurements can be quantified and handled when these measurements are used for estimating a reference value by univariate or multivariate regression. Furthermore, it is shown how traditional univariate statistics can be used in multivariate situations by calculating the net analyte signal.

Validation is another important factor that should be considered in all data analysis. Validation is necessary to ensure valid and reliable modelling and is included as a part of the chemometric modelling in all the papers. In the thesis it is described how the chosen validation method depends on the purpose of the data analysis and the chemometric model.

Overall, the project has illustrated that spectroscopic methods such as fluorescence and NMR in combination with chemometrics can provide information about underlying chemical and physical parameters in fish. The results are promising and give indications of considerable potential for the used methods, which encourages further research and development.

Resumé

Det overordnede formål med denne Ph.D. afhandling er at vise hvordan nye såvel som kendte kemometriske metoder kan blive anvendt på spektroskopiske data. De spektroskopiske målinger er opnået som multivariate eller multi-vejs data og kan på grund af prøvernes kompleksitet være meget komplekse. Forståelse og analyse af den underliggende struktur af målingerne kan derfor nødvendiggøre anvendelse af kemometriske metoder.

Fluorescens-spektroskopi og nuklear magnetisk resonans (NMR) målt på fiskemuskler er benyttet som eksempler til at illustrere den mulige anvendelse af metoderne på reelle produkter. Der er fokuseret på tre områder, som er eksemplificeret i de seks vedlagte artikler.

1) Eksplorativ analyse af spektroskopiske data
2) Validering af modeller
3) Kvantificering og håndtering af fejl i instrumentelle målinger

Anvendelse af spektroskopiske teknikker giver mulighed for at foretage ikke destruktive målinger direkte på produktet. Imidlertid kan sådanne målinger være meget komplekse, da produktet kan være heterogent og indeholde mange forskellige komponenter. Dette er illustreret ved eksplorativ analyse af fluorescens-emission målt på torsk og laks, fluorescenslandskaber målt på ekstrakter af fiskemuskel og lav-felts NMR målt på gadoide fisk. Ved anvendelse af principal komponent analyse (PCA) og parallel faktor analyse (PARAFAC) er det foreslået, at fluorescensspektre målt på fiskemuskler kan henføres til bl.a. bindevæv, NADH, oxidationsprodukter, lipider, pigmenter og aminosyrer. Slicing modellering af NMR relaksationer har vist fordelingen af vand i torskefileter og er benyttet til inddeling af dette vand i forskellige fraktioner. Koncentrationen af vand i disse fraktioner er korreleret til fiskemusklens vandbindingsevne, der er en vigtig kvalitetsparameter.

I forbindelse med den eksplorative analyse er det endvidere vist, hvordan tre-vejs data kan analyseres og fortolkes. Problemer involveret i udviklingen af en PARAFAC model for fluorescens emissions- og excitationsdata kan skyldes manglende værdier og spredt lys. Det er vist, hvordan sådanne faktorer kan håndteres.

Sampling er et vigtigt aspekt at overveje, da fisk som andet biologisk materiale er heterogent. Der er givet et eksempel på, hvordan samplingfejl i instrumentelle målinger kan kvantificeres og håndteres, når disse målinger anvendes til at estimere en referencemåling ved univariat eller multivariat kalibrering. Det er endvidere vist, hvordan traditionel univariat statistik kan overføres til multivariate problemstillinger via beregning af det såkaldte "net analyte signal".

Validering er en anden vigtig faktor, der skal overvejes i forbindelse med dataanalyse. Validering er nødvendig til sikring af valide og pålidelige resultater og er inkluderet som en del af den kemometriske databehandling i alle vedlagte artikler. Det er beskrevet, hvordan den valgte valideringsmetode afhænger af data, formålet med dataanalysen og den anvendte kemometriske model.

Samlet set har projektet illustreret, at spektroskopiske metoder som fluorescens og nuklear magnetisk resonans i kombination med kemometri kan give information om underliggende kemiske og fysiske parametre af fisk. Resultaterne er lovende og giver indikationer af betydeligt potentiale, som ansporer til yderligere forskning og udvikling.

List of papers and appendices

Paper I
On the fluorescence of muscle and connective tissue from cod and salmon
Charlotte M. Andersen and Jens Petter Wold. *Journal of Agricultural and Food Chemistry*, 51, 470-476 (2003)

Paper II
Practical aspects of PARAFAC modelling of fluorescence excitation-emission data
Charlotte M. Andersen and Rasmus Bro. *Journal of Chemometrics*, accepted

Paper III
Distribution of water in fresh cod
Charlotte M. Andersen and Åsmund Rinnan. *Lebensmittel Wissenschaft und Technologie*, 35, 687-696 (2002)

Paper IV
On the relation between water pools and water holding capacity in cod muscle
Charlotte M. Andersen and Bo M. Jørgensen. *Journal of Aquatic Food Product Technology*, submitted

Paper V
Effect of sampling errors on predictions using replicated measurements
Charlotte M. Andersen, Rasmus Bro and Per B. Brockhoff. *Journal of Chemometrics*, submitted

Paper VI
Quantification and handling of sampling errors in instrumental measurements: A case study
Charlotte M. Andersen and Rasmus Bro. In preparation

Appendix 1
Freshness assessment of thawed and chilled cod fillets packed in modified atmosphere using near-infrared spectroscopy
Niels Bøknæs, Kristina N. Jensen, Charlotte M. Andersen and Harald Martens. *Lebensmittel Wissenschaft und Technologie*, 35, 628-634 (2002)

Appendix 2
Theory of net analyte signal vectors in inverse regression
Rasmus Bro and Charlotte M. Andersen, submitted

Appendix 3
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, *Trends in Food Science & Technology*, 13, 235-244 (2002)

Abbreviations

CPMG	Carr-Purcell-Meiboom-Gill
DECRA	direct exponential curve resolution algorithm
EEM	excitation-emission matrix
FID	free induction decay
M	amplitude of an NMR signal
MLR	multiple linear regression
MSEC	mean squared error of calibration
NADH	nicotinamide adenine dinucleotide, reduced form
NAS	net analyte signal
NMR	nuclear magnetic resonance
OLS	ordinary least squares
PARAFAC	parallel factor analysis
PC	principal component
PCA	principal component analysis
PCR	principal component regression
PLS	partial least squares
RMSECV	root mean squared error of cross validation
RMSEP	root mean squared error of prediction
s_e	model error
s_x	variance of errors in instrumental measurements (x)
s_y	variance of errors in reference measurements (y)
T_1	longitudinal or spin-lattice relaxation time constant
T_2	transverse or spin-spin relaxation time constant
Vpe	variance of prediction error
WHC	water holding capacity
WHC_d	water holding capacity given as gram water per gram dry matter
WHC_w	water holding capacity given in per cent relative to the original water content

Content

Preface **3**
Summary **4**
Resumé **6**
List of papers and appendices **8**
Abbreviations **9**
Content **10**
1. Introduction **13**
1.1. Chemometrics applied to spectroscopy 14
1.2. Outline of the thesis 15
2. Chemometrics and spectroscopic methods **17**
2.1. Chemometrics 17
2.1.1. Data structure *17*
2.1.2. Qualitative/descriptive analysis *18*
2.1.2.1. Principal component analysis (PCA) 18
2.1.2.2. PARAFAC (PARAllel FACtor analysis) 19
2.1.2.3. Slicing 20
2.1.3. Regression methods *21*
2.2. Regression with errors in instrumental variables 22
2.2.1. Measurement error models *22*
2.2.2. Net analyte signal (NAS) *23*
2.2.3. Quantification of errors *23*
2.3. Fluorescence spectroscopy 26
2.3.1. Fluorescence *26*
2.3.2. Modelling fluorescence data *27*
2.3.3. Application of chemometrics in the interpretation of fluorescence measurements *28*
2.4. Low-field ^{1}H nuclear magnetic resonance (Low-field NMR) 29
2.4.1. Nuclear magnetic resonance *29*
2.4.2. Measuring NMR relaxations *29*
2.4.3. Application of chemometrics in the interpretation of low-field NMR decays *30*
3. Exploratory analysis of spectroscopic data **33**
3.1. Multivariate analysis of fluorescence spectroscopic measurements 33
3.2. Multi-way analysis of fluorescence excitation-emission matrices (EEM) 35
3.3. Interpretation of low-field NMR relaxations 38
3.3.1. Distribution of water in cod *39*

3.3.2. Distribution of water in four fish species *40*
3.3.3. Relationship between water distribution and water holding capacity (WHC) *41*
4. Validation **45**
4.1. Validation methods 46
4.1.1. Test set validation *46*
4.1.2. Cross validation *46*
4.1.3. Jack-knife validation *47*
4.1.4. Bootstrapping *48*
4.1.5. Split-half analysis *49*
4.1.6. Distribution of residuals *49*
5. Sampling **51**
5.1. Errors influencing predictions in multivariate regression 52
5.2. Heterogeneity of biological materials 53
5.3. Quantification of errors 54
5.4. Handling errors in instrumental variables 56
5.4.1. Correction of the regression coefficient in case of different numbers of replicates *56*
5.4.2. Application of the correction for univariate and multivariate data *57*
6. Conclusion and perspectives **61**
6.1. Exploratory analysis of complex spectroscopic data 61
6.2. Validation of models 63
6.3. Quantifying and handling sampling errors of instrumental measurements 63
7. References **65**
Paper I **77**
Paper II **87**
Paper III **119**
Paper IV **131**
Paper V **143**
Paper VI **163**
Appendix 1 ..... **181**
Appendix 2 ..... **191**
Appendix 3 ..... **203**

1. Introduction

During recent years there has been an increasing attention on food products. Eating quality, food safety, nutritional value, environmental and economical factors are some of the aspects which have received great interest from the public. This increased focus suggests a necessity for the development of new methods and techniques that can be used for controlling and monitoring a production process.

Spectroscopic techniques, such as near infrared, fourier transform infrared, fluorescence and nuclear magnetic resonance (NMR), have shown promising possibilities for providing valuable and accurate information regarding various chemical and physical factors of a product (Scotter 1997; Munck et al. 1998). The methods typically work rapidly in a non-destructive way, yielding multivariate data containing general but accurate information on the relevant properties of the samples. Development of fast non-destructive methods for measuring parameters such as freshness, lipid content and texture has been an important research area in fisheries research in recent years (Isaksson et al. 1995; 2002; Nordtvedt et al. 1998; Sigurgisladottir et al. 1999; Jepsen et al. 1999; Bechmann et al. 1999; Jensen et al. 2002; Bøknæs et al. 2002, appendix 1). The fast instrumental methods may be implemented industrially, giving fast and highly reproducible measurements suitable for controlling the end quality of the fish product and continuously monitoring the production.

It is important to ensure optimal extraction of the relevant information in data when developing new techniques such as spectroscopic measurements. Chemometrics – including multivariate and multi-way analyses - is suited for handling complex data efficiently, providing information that could not be obtained otherwise.

The objective of this thesis is to discuss aspects of chemometric modelling applied to spectroscopic measurements. Fluorescence spectroscopic data and nuclear magnetic resonance measured on fish muscles are used as examples to illustrate the potential use of the methods on real products. Focus has been on three main subjects.

1) Exploratory analysis of spectroscopic data
2) Validation of models
3) Quantifying and handling sampling errors in instrumental measurements

These three subjects constitute an extensive group of parameters which one has to be aware of in order to obtain reliable results from instrumental measurements. Knowledge about the underlying chemical phenomena in the spectroscopic measurements is necessary in order to extract information in data, which is important for the future purpose. Validation is an important factor in almost every analysis and is necessary to ensure valid and reliable modelling.

Furthermore, knowledge of how to sample can be used to minimise product variability, maximise yield etc. Thus, information regarding optimal monitoring and handling of sample variation is an extremely important step in the development of new techniques.

This project is a part of a larger framework programme, Advanced Quality Monitoring in the Food Production Chain, involving several research groups working with various subjects within food science. The overall purpose of the framework programme has been to introduce new and known on-line/at-line techniques by developing new mathematical models based on real-world problems including sensory, engineering and statistical parameters (Bro et al. 2002a, appendix 3).

1.1. Chemometrics applied to spectroscopy

Application of chemometrics enables exploratory analysis of complex data structures where no initial hypothesis is necessary prior to the analysis. By exploratory data analysis the often large multivariate data sets can be analysed in an intuitive straightforward manner, visually displaying information that can not be seen by traditional data analysis. Subsequently, these results can be used to improve the performance of existing analysis methods, establish robust and accurate measurement methods, and illustrate possibilities for developing stable non-destructive spectroscopic measurements for new applications.

Spectroscopic measurements are often performed non-destructively on complex food materials. Thus, the measurements may be rather complex themselves, increasing the necessity for data analytical methods that can handle these types of data. Furthermore, before industrial application of instrumental techniques, it is important to understand the nature of the samples, including the underlying chemical and physical phenomena. Such an understanding ensures that optimal and relevant information is extracted from the data.

Validation is an important step in the development of new alternative analysis methods based on spectroscopy and chemometrics. Validation can be seen as an assessment of the appropriateness of the developed models, which includes an evaluation of how well data are modelled and an evaluation of how well the model reflects the structure in data. It is used to prevent incorrect model estimates and invalid interpretations leading to erroneous conclusions. Other aspects of ensuring optimal model performance include evaluation of instrumental noise, detection limits, influence from the surroundings and the sampling procedure.

Typically, measurements are influenced by significant errors due to the heterogeneity of the biological material. Furthermore, instrumental measurements performed on a localised area of an object can be used to represent the whole object. Such measurements will almost always be influenced by a sampling error. Therefore, it is important to optimise the sampling procedure to

ensure that the data collected are representative for the actual purpose and contain relevant information. This is an important, but often neglected step when performing instrumental measurements. Optimisation of sampling concerns both a quantification of the various types of errors and, if possible and necessary, a correction for the sampling errors.

1.2. Outline of the thesis

The thesis is divided into four main chapters. Chapter 2 describes the chemometric and spectroscopic methods used. The next three chapters include the main subjects of the thesis.

Examples of decomposing, visualising and interpreting complex spectroscopic data are given in chapter 3 and the Papers I-IV. Papers I and II deal with the application of fluorescence spectroscopy describing (1) multivariate analysis of salmon and cod spectra especially in relation to indication of freshness and the presence of collagen and (2) how fluorescence landscapes can be decomposed by multi-way modelling. This includes how to handle problems due to missing values and scatter. Papers III and IV concern the possibilities of describing the distribution of water within the fish muscle and predicting water holding capacity (WHC) by the use of low-field NMR.

Chapter 4 gives an introduction to various validation methods including a description of the purpose of each method, how the methods work, and how the results can be interpreted and visualised. Examples taken from Papers I to IV are used to illustrate the applicability of the validation methods.

Sampling is treated in Chapter 5 and in the Papers V and VI. The contribution of different errors in multivariate regression models is illustrated together with a description of how heterogeneity of spectral measurements in fish muscles can be investigated and visualised. The main objectives of the chapter and the two papers are (1) to show how the importance of sampling errors in the spectroscopic measurements can be quantified and (2) to illustrate a method by which such types of errors can be handled.

Chapter 6 presents conclusions and perspectives, describing the findings in the present project and pointing out ideas for further work that turned up directly from the results shown in the previous chapters.

2. Chemometrics and spectroscopic methods

This chapter presents the chemometric and spectroscopic methods used in this thesis. Section 2.1 describes the chemometric methods applied. Section 2.2 describes theory of regression modelling in situations with error in the independent variables, e.g. instrumental measurements. Sections 2.3 and 2.4 introduce fluorescence spectroscopy and low-field ^{1}H NMR, respectively, which are the spectroscopic methods used.

2.1. Chemometrics

Chemometrics has been defined by the International Chemometric Society (http://www.emsl.pnl.gov:2080/docs/incinc/chemdef.html (2003-02-03)) as

> *the science of relating measurements made on a chemical system or process to the state of the system via application of mathematical or statistical methods*

The chemometric techniques make it possible to explore large data sets in an explorative and non-selective way by decomposing the complex data into simpler structures, thus improving the interpretation and understanding of the information in the available data.

Application of chemometrics has evolved during the last decades due to the development of new instruments, algorithms and faster computers. In this project, the chemometric analysis is based on spectroscopic measurements, which are obtained by fast, non-destructive instruments making the techniques suitable for application on-line or at-line.

There are two main purposes for applying chemometrics: a *qualitative* and a *quantitative*. In the qualitative analysis, the data are decomposed into a few factors, enabling the study of patterns in the data, trends, outliers, etc, whereas the purpose of the quantitative analysis is typically to predict a certain parameter using regression models. The chemometric methods are further elaborated on in the following. Validation of the chemometric models is described in Chapter 4.

2.1.1. Data structure

Classical statistics typically deals with univariate data, characterised by *zero-order* measurements. For such data, signals from interferents cannot be distinguished from the analyte signal and the decomposition and understanding of data may in some situations be difficult or even impossible. *First-order* measurements are characterised by a vector, for example spectral measurements. Combining several *first-order* measurements gives a matrix that can be analysed and decomposed with multivariate methods. Principal component analysis

(PCA) is commonly applied for the descriptive analysis, whereas regression methods, such as partial least squares regression (PLS) and principal component regression (PCR), are often used for predictions.

Second-order data are obtained when the measurements of one sample can be represented as a matrix. Combining several samples of *second-order* measurements gives a box (three-way structure) that ideally, if the data are trilinear, can be analysed with N-way methods. PARAFAC (PARAllel FACtor analysis) is well suited for decomposing the N-way structure describing the underlying phenomena in data. N-PLS is an extension of PLS to higher order situations and is applied for regression and prediction.

In addition to the N-way methods, multi-way data can be unfolded and analysed with the multivariate techniques. However, N-way methods generally give more adequate, robust and interpretable models. Furthermore, N-way methods are less sensitive to noise and give loadings that can directly be related to the chemical structure of the data. Geladi (1989), Kroonenberg (1992), Harshman and Lundy (1984; 1994), and Bro (1996; 1997; 1998) give introductions to some of the N-way methods.

2.1.2. Qualitative/descriptive analysis

2.1.2.1. Principal component analysis (PCA)

PCA is a bilinear model that extracts the systematic information in data by decomposing the signals/measurements into principal components (PCs) using a least squares approach (Equation 2.1) (Wold et al. 1987). The PCs describe independent variations in the data set.

$$x_{ij} = \sum_{f=1}^{F} t_{if} p_{jf} + e_{ij} \qquad i = 1,\ldots,I \quad j = 1,\ldots..,J \tag{2.1}$$

In Equation 2.1, F is the number of PCs chosen by the analyst, x_{ij} is the measurement of the i-th sample at the j-th variable, t_{if} is the score of the i-th sample for the f-th component, p_{jf} is the corresponding loading for the j-th variable and e_{ij} is the residual. Equation 2.1 can be expressed in a matrix notation as

$$\mathbf{X} = \mathbf{TP'} + \mathbf{E} \tag{2.2}$$

where the matrix **X** holds the measured data with I samples and J variables. The model is made by the product of the two matrices, **T** and **P'** . **T** is the score matrix containing the samples as rows and PCs as columns. It contains the information about the amount of each component in every sample. **P** is the loading matrix with the variables as rows and the PCs as columns

containing information about the contribution of the variables to each PC. **E** is a matrix of residuals, carrying the information in **X** that is not covered by the model and ideally contain no chemical information.

2.1.2.2. PARAFAC (PARAllel FACtor analysis)

PARAFAC (Harshman 1970) can be seen as an extension of the bilinear PCA into the multilinear situations. It was developed in 1970 independently by two research groups and was named PARAFAC by Harshman and CANDECOMP (CANonical DECOMPosition) by Carrol & Chang. Subsequently, PARAFAC has been developed and applied in many scientific fields such as chemistry and food science (Munck et al. 1998; Bro 1998).

The PARAFAC model (Harshman 1970) is described by Equation 2.3 and illustrated for three-way data in Figure 2.1.

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \tag{2.3}$$

PARAFAC can be explained in a similar way as PCA. x_{ijk} is the intensity of the *i*-th sample at the *j*-th variable (in the second mode) and at the *k*-th variable (in the third mode). a_{if} b_{jf} and c_{kf} are the elements in the three loading matrices, **A**, **B** and **C**, respectively, and contain information about the importance of the samples/variables to each component. As for PCA, the residuals, e_{ijk}, contain the variation not captured by the model. In Figure 2.1, the data, $\underline{\mathbf{X}}$, is decomposed into two components (*F*) represented by a set of a, b and c. Choosing the right number of components is important, as it ensures a unique solution that discovers the latent structure in the data.

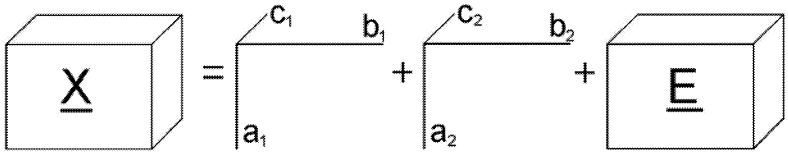


Figure 2.1 Representation of the PARAFAC model for three-way data with two components.

In order for the PARAFAC model to provide unique solutions, it is necessary that all loadings are linearly independent, meaning that no two loadings have the same shape. Ideally, PARAFAC recovers the underlying structure uniquely, if the data are approximately trilinear, the right number of components is chosen and the signal-to-noise ratio is appropriate. This means that the estimated PARAFAC model cannot be rotated without loss of fit, as opposed to two-way analysis where one may rotate scores and loadings without changing the fit of the model. Bilinear models can therefore not estimate pure spectra without external information.

By PARAFAC modelling, the true underlying phenomena in data can be identified up to a scaling factor. Therefore, PARAFAC can be used for curve resolution. This is illustrated in Paper II and Section 3.2, showing how to estimate the pure excitation and emission spectra in aqueous fish muscle extracts.

2.1.2.3. Slicing

Slicing was recently proposed by Pedersen, Bro and Engelsen (2001 and 2002). It is based on the principles of Direct Exponential Curve Resolution (DECRA) (Windig and Antalek 1997) and can be used to analyse data consisting of a sum of exponentially decaying curves. The exponential function for one sample can be described by Equation 2.4

$$\mathbf{x} = \sum_{f=1}^{F} M_f e^{\frac{t}{T_{2,f}}} \tag{2.4}$$

where **x** is a vector holding the signal, F is the number of exponential functions, M_f denotes the amplitude of the signal and T_{2f} is the characteristic relaxation time constant for factor f.

The idea is to create second order data from the first order exponential decays by splitting the data into two or more overlapping slabs yielding a three-way structure as shown in Figure 2.2. The figure shows that the three-way structure is formed by two sub-matrices (slabs) that contain the same number of variables. One of these holds the first number of variables (0 to *b*) and the other holds the last number of variables (*a* to *J*). The two sub-matrices will be of the same size and contain the same information, but shifted horizontally. The number of variables to shift between two slabs is referred to as *lag*. The transformation is possible since exponentially decaying curves retain the relaxation characteristics when shifted in the variable direction. Only the intensity of the signal changes. Slicing can be performed with more than two slabs and different lag sizes. However, for simplicity only two slabs are shown in the figure. The optimal number of slabs depends on the noise in data, the number of samples and the number of exponential components (Pedersen, Bro and Engelsen 2002).

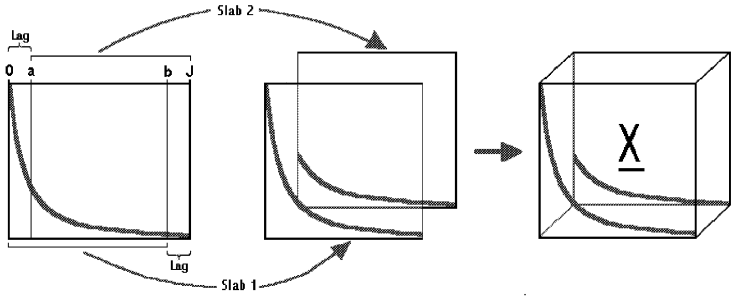


Figure 2.2 Illustration of slicing yielding three-way data from two-way data

When the splitting is performed in the same way for all samples in a data set, a three-way structure is obtained. These three-way data can be decomposed by an N-way method such as PARAFAC. This enables identification and interpretation of the underlying exponential profiles. The loadings in the variable mode will be mono-exponential and for each factor these loadings describe the exponential function of the corresponding component. Application of slicing requires that the data contains more than one sample and that all samples contain the same exponential phenomena. Decomposing NMR relaxation profiles by slicing based on PARAFAC is described in Papers III and IV and in Section 3.3.

2.1.3. Regression methods

PCR and PLS are multivariate regression methods often applied for relating multivariate data, **X**, to a reference parameter, **y**, such as the concentration of a certain analyte, sensory score value or storage condition. The overall purpose is to predict the reference value in future samples and to interpret the relationship between the two data sets. The multivariate regression models can be described as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \qquad (2.5)$$

where **b** is the regression coefficient vector obtained by the calibration model and **e** denotes the residuals containing model errors and other variation in data not captured by the model. PLS performs a simultaneous decomposition of **X** and **y** in such a way that the information in **y** is directly used as a guide for the decomposition of **X** after which regression of **y** is done (Martens and Næs 1989). As for the decomposition methods described above, PLS works by generating new latent components describing the variation in data that contains the information most important for the relation between the two data sets.

N-PLS is an extension of PLS to multi-way situations and predicts **y** in a way similar to PLS (Bro 1996). The model is not unique (loadings can be rotated) and it can, therefore, not be used for curve resolution, as is the case for PARAFAC. Compared with unfolding the data and using bilinear PLS regression, N-PLS is claimed to give more stable and parsimonious models that are easier to interpret and less influenced by noise due to the incorporation of information in all modes (Bro 1996).

In addition, both PCA and PARAFAC can be used for regression and prediction. In these cases the score values are used as independent variables and related to **y** by multiple linear regression (MLR). For PCA, this gives PCR. MLR is the classical regression method, but does not apply with covariate data (such as spectra). Furthermore, MLR does not apply when there are more variables than samples.

2.2. Regression with errors in instrumental variables

This section describes data analytical methods that in Papers V and VI and Sections 5.3 and 5.4 are used to evaluate and handle errors in instrumental measurements (**X**). Section 2.2.1 deals with the theory of measurement error models used for correcting biased regression parameters. The net analyte signal is illustrated in Section 2.2.2. It allows the calculation of figures of merits such as sensitivity and signal to noise ratio. Here, it is used to represent a multivariate signal as a univariate, which can be used for traditional statistical testing. Section 2.2.3 presents an expression of the prediction error variance that makes it possible to compare the size of the errors influencing the predictions. The three data analytical methods are only some of the methods that may apply in situations with errors in the instrumental variables, but are the methods used in this work.

2.2.1. Measurement error models

Ordinary least squares regression (OLS) is widely used for regression and prediction. However, correct use of OLS requires that errors in the instrumental variables are negligible. Otherwise, the parameters estimated by OLS will be biased. However, when the aim is to predict new samples, OLS will be adequate, giving the lowest prediction errors, if the residuals are homoscedastic and independent of the x-variables. Furthermore, the samples to be predicted should originate from the same distribution as the samples used for calibration (Fuller 1987).

When the two data sets do not come from the same distribution or when the purpose is to obtain un-biased estimates of regression coefficients, OLS may no longer be the optimal regression method. Measurement error models described by Cochran (1968) and Fuller (1987; 1995) are examples of methods that can be used to investigate the effect of measurement/sampling errors and to correct for these errors in the independent variables. The overall purpose is to provide more accurate estimates of the regression coefficients by correcting the model parameters. This is done by estimating the variance of the true x-variables and using this for regression modelling instead of estimates of the measured x-variables. The theory of measurement error models is described in Paper V and will thus not be considered further here.

Measurement error models have the drawbacks that they do not take the errors of each individual sample into account and they assume the ratio of variances for the reference and the instrumental variables to be constant. To overcome some of these problems, other methods such as weighted least squares regression may be used (Ripley and Thompson 1987; Riu and Rius 1995; 1996).

2.2.2. Net analyte signal (NAS)

The net analyte signal denotes the part of a spectrum that is unique for the analyte meaning that it is the signal important for predicting a certain parameter such as the concentration of an analyte in a sample (Lorber 1986). There are several methods for calculating this signal (Lorber et al. 1997; Xu and Schechter 1997; Faber 1998; Goicoechea and Olivieri 1999; 2001; Ferré et al. 2001). Bro and Andersen (2002, appendix 2) present a new simple way, which is expressed in Equation 2.6.

$$\mathbf{x}_{k,i}^{*} = \mathbf{b}(\mathbf{b}^{T}\mathbf{b})^{-1}\mathbf{b}^{T}\mathbf{x}_{i} \tag{2.6}$$

$\mathbf{x}_{k,i}^{*}$ denotes the net analyte signal of analyte *k* for the *i*-th sample, **b** is the regression coefficient vector obtained by PLS or PCR and $\mathbf{x}_i$ is the spectrum of the *i*-th sample. The scalar net analyte signal is usually defined as the norm of the net analyte signal vector. However, this is not valid for mean centred models and when the predicted value is below zero (Faber 1998; Bro and Andersen 2002, appendix 2). Instead, one should apply Equation 2.7, which does not have the problem of negative predictions.

$$x_{k,i}^{*} = \frac{\mathbf{x}_{k,i}^{*\,T}\mathbf{b}}{\|\mathbf{b}\|} \tag{2.7}$$

Equation 2.6 has the advantages that it does not rely on any assumptions on the structure of the data and it gives exactly the same predictions as the calibration model from which it is calculated. The scalar net analyte signal makes it possible to represent the multivariate calibration model as a univariate least squares regression model using the scalar values as independent x-variables. In addition, the net analyte signal allows the calculation of figures of merits (sensitivity, selectivity, etc.) that are used as quantitative measures of the performance of the regression model.

2.2.3. Quantification of errors

An approximation of the prediction error variance (*Vpe*) obtained from a multivariate regression model based on errors in the instrumental variables (**X**), error in reference measurements (**y**), error in the estimated calibration model (regression vector) and model error is given by Faber and Kowalski (1996; 1997). The expression makes it possible to compare the influence of the different errors and is shown in Equation 2.8. It is an approximation based on a linearization of the model. The expression assumes that the noise of all x-variables is of the same size for both calibration and test samples, errors are small and that the data are unbiased. Further, all errors are considered to be independently and identically distributed random variables with zero mean. This may not always hold for instrumental variables such as spectroscopic measurements,

where the noise may be correlated. However, as an approximation, the expression may still be valid even in such cases.

$$Vpe \approx (1/n + h_i)(s_e^2 + s_{\Delta y}^2 + \|\mathbf{b}\|^2 s_{\Delta X}^2) + s_{e_i}^2 + \|\mathbf{b}\|^2 s_{\Delta X_i}^2 \quad (2.8)$$

$s_{\Delta y}^2$ denotes the variance of errors in the reference value. $s_{\Delta X}^2$ and $s_{\Delta X_i}^2$ are the errors in the instrumental measurements for the calibration set and the test set, respectively. s_e^2 and $s_{e_i}^2$ are the model errors for calibration and validation, respectively. h_i is the leverage of the test sample calculated from the scores and n denotes the number of samples in the calibration set. Equation 2.8 estimates the prediction error variance in case of a mean-centred model. The term $1/n$ should be left out, if data is not mean-centred.

Equation 2.8 can be used for estimating the prediction error variance for new samples. In this project, the approach is used to compare the relative impact of the different errors influencing the predictions. The variance of the **X** and **y** measurements can be estimated e.g. from replicated measurements, and the regression coefficients (**b**) and the leverage (h_i) are determined from the calibration model and the scores of the test samples, respectively. The model error contribution (s_e^2) can be estimated from knowledge of the other error contributions as suggested by Faber et al. (1998). This expression is given in Equation 2.9

$$s_e^2 = \mathrm{MSEC} - s_{\Delta y}^2 - \|\mathbf{b}\|^2 s_{\Delta X}^2 \quad (2.9)$$

where MSEC denotes the mean squared error of calibration and the other factors are given as above. Equation 2.9 gives one estimate of the model error for each model. The rationale for this equation depends on the fact that MSEC is an estimate of $(s_e^2 + s_{\Delta y}^2 + \|\mathbf{b}\|^2 s_{\Delta X}^2)$ for a well specified model.

To give an impression of the validity of this way of estimating the model error variance, a small simulation study is described. The simulated data are made by randomly choosing scores, loadings and regression coefficients for two-component models. True x-variables and y-values of 50 samples are calculated from these model estimates. Random normally distributed noise with a standard deviation of 0.015 is added to each x-variable. Noisy y-values are obtained by adding random normally distributed model error and measurement error in **y** to each sample. These error contributions have a standard deviation of 0.05 or 0.1. With the given noise settings, the noise in **X** and **y** makes up less than 1% of the variation in data. Simulations are performed 1000 times and the model error calculated by Equation 2.9 using two PLS-components. The distribution of these estimates is shown in Figure 2.3, illustrating that the estimated model errors are approximately normally distributed with a mean value around 0.0026 or 0.0104. These values are in accordance with the true model errors that are 0.0025 and 0.01, respectively.

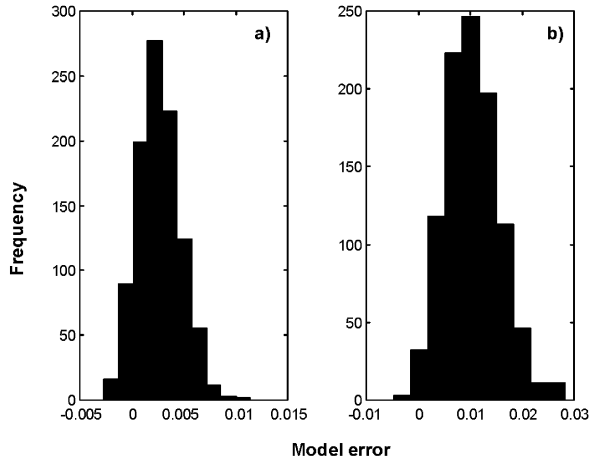


Figure 2.3 Distribution of the model error estimated using Equation 2.9 for simulated data where noise with a standard deviation of 0.015 is added to the x-variables. Error in **y** and model error are added to the y-values. The standard deviation of these is 0.05 (a) and 0.1 (b).

Equation 2.8 estimates the true prediction error variance, denoted as the *actual* prediction error variance. However, when **y** contains errors, this value can not be evaluated because the true reference values are not known. Instead, an *apparent* prediction error is obtained, which is a measure of the difference between the noisy measured y-values and the y-values predicted by the model. This estimate will always be higher than the *actual* prediction error that estimates variation between the true y-values and the y-values predicted by the model. The relationship between the *actual* and the *apparent* prediction error is given by DiFoggio (2000) and Fernández Pierna et al. (2003) as shown in Equation 2.10. The difference between the two prediction error estimates is due to the variance of errors in **y**.

$$\textit{actual}\ \text{MSEP} = \textit{apparent}\ \text{MSEP} - s_y^2 \tag{2.10}$$

From Equation 2.8 and 2.10, the *apparent* prediction error variance can be estimated by Equation 2.11 where the variance of **y** is added to the expression of the prediction error variance.

$$Vpe_{\text{apparent}} \approx (1/n + h_i)(s_e^2 + s_{\Delta y}^2 + \|\mathbf{b}\|^2 s_{\Delta X}^2) + s_{e_i}^2 + \|\mathbf{b}\|^2 s_{\Delta X_i}^2 + s_{\Delta y}^2 \tag{2.11}$$

Both Equation 2.8 and 2.11 give valid estimates of the prediction error variance. It is important to decide which of those to choose. The difference is that Equation 2.8 estimates the error in relation to the true y-values and Equation 2.11 estimates the error in relation to the measured y-values. The *apparent* prediction error variance is used in Section 5.3 and Paper VI for estimating the prediction error variance.

2.3. Fluorescence spectroscopy

Fluorescence spectroscopy is commonly used in many scientific fields such as chemistry, biochemistry and medicine. This section describes the basic theory of fluorescence including examples of applications within food science. Focus is on factors important for multivariate and multi-way analysis and for interpretation of the spectroscopic measurements. For a discussion on instrumental and more specific areas reference is made to the textbooks of Munck (1989) and Lackowicz (1999).

2.3.1. Fluorescence

Fluorescence is caused by an immediate radiation of light by molecules after excitation to an excited singlet state. Excitation is due to absorption of light with a characteristic wavelength, always in the UV or visual wavelength area. Only light in these areas possesses enough energy to excite electrons. After absorption of light, thermal vibrations cause loss of a part of the energy. Therefore, fluorescence will be of lower energy than the excitation light meaning that it will be found at longer wavelengths. This shift between excitation and emission is called the Stokes shift.

Not all molecules will exhibit fluorescence emission after excitation to the singlet state. Phosphorescence and collisions with other molecules are ways by which the excited molecule can return to the ground state without emitting fluorescence. When measuring fluorescence directly on a food or a fish sample, chemical compounds occurring naturally within the sample matrix induce fluorescence emission. This type of fluorescence is denoted as autofluorescence. The fluorophores include mainly aromatic compounds but some carbonyls and molecules with highly conjugated double bonds are also able to fluoresce. Each fluorescent molecule has a characteristic excitation and emission spectrum, which can be used to separate and identify fluorescent molecules as well as to differentiate between different substitutions and conformations of the same molecule. These characteristic excitation and emission spectra also gives a high degree of sensitivity and selectivity, which is one of the advantages of fluorescence spectroscopy.

Fluorescence is influenced by several factors such as the solvent, pH and temperature. These factors may induce a shift in the peak maxima, making fluorescence dependent on the sample matrix, instrument, and experimental conditions. Quenching of the fluorescence signal is an important factor in complex systems such as food and fish. It is caused by (1) a deactivation by other molecules such as oxygen, halogens and amines, (2) an absorption of the excitation or emitted light (concentration quenching), or (3) an absorption of fluorescence by another molecule in the same sample (Lackowicz 1999). The third possibility is discussed in Paper I where it is suggested that NADH and collagen fluorescence are absorbed by red pigments in

the salmon muscle, such as astaxanthin. The reason for this phenomenon is that the emission of NADH and collagen has peak maxima between 430 and 480 nm, which are in the same wavelength area as the absorbance of the red pigments (Britton 1995).

2.3.2. Modelling fluorescence data

If a sample is measured at several emission wavelengths for several excitation wavelengths, an excitation-emission matrix (EEM) is obtained (Figure 2.4). Ideally, such data will be trilinear and can then be analysed with multi-way chemometric models such as PARAFAC. This enables curve resolution and identification of the underlying spectral phenomena (Leurgans and Ross 1992; Ross and Leurgans 1995; Bro 1997).

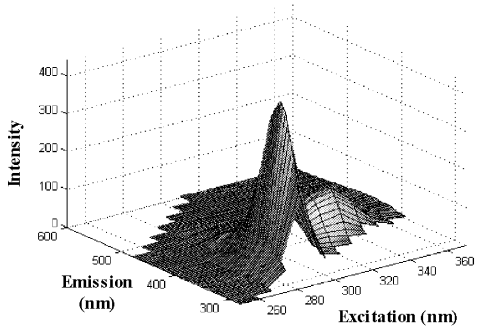


Figure 2.4 Example of a fluorescence excitation-emission matrix (EEM)

However, factors such as missing values, scatter and spectral similarities influence the fluorescence EEMs making PARAFAC modelling problematic. There is no emission below excitation, because the emission will always be of lower energy than the excitation. Hence, these data elements do not conform to the trilinear model. This has to be taken into account and is discussed in Paper II, explaining why emission below excitation cannot generally be set to zero but should, for example, be given as missing values. However, an increased amount of missing values makes parameter estimation more uncertain, as is also described in the paper.

Scatter disturbs the trilinearity of the fluorescence landscapes and may overlap with the analyte signal. It is due to scattered light caused by small particles in the samples. There are two main scatter phenomena: Rayleigh scatter and Raman scatter. Rayleigh scatter is seen at the same wavelength as the excitation wavelength and shows up as diagonal lines in the fluorescence landscape. It takes place when a molecule is excited to a higher vibrational level. There is no electronic conversion and the absorption and emission will be of the same energy. The emitted light thus has the same wavelength as the exciting light. Second order Rayleigh scatter appears at twice the excitation wavelength. For interpretation and curve resolution by PARAFAC, both

first and second order Rayleigh scatter must be handled, for example, by replacing the data elements influenced by scatter with missing values.

Raman scatter is weaker than Rayleigh scatter and may not be recognised, if the fluorescence intensity is high. It appears at longer wavelengths than the excitation wavelength with a constant distance to the Rayleigh scatter peak and is due to vibrational energy being added to, or subtracted from, the excitation photon. It has been shown that multivariate and multi-way methods are capable of handling Raman scatter present in the fluorescence spectra (Nørgaard 1995; Nikolajsen et al. 2001). Raman scatter is not treated further here.

In Paper II fluorescence EEMs measured on fish muscle extracts are analysed and some of the problems involved in PARAFAC modelling caused by missing values and scatter are illustrated. For example, it is shown how such phenomena result in unreliable parameter estimates, which should be taken into account. These inferior loadings are handled by applying constraints.

2.3.3. Application of chemometrics in the interpretation of fluorescence measurements

The introduction of fluorescence spectroscopy for fast non-destructive measurements requires that fluorescence is measured directly on the object, often a solid sample. The application of solid-sample fluorescence spectrophotometry is shown in evaluations of fat and connective tissue in beef (Wold et al. 1999; Egelandsdal et al. 2002), fish meal quality (Dahl et al. 2000), light-induced oxidation in poultry and dairy products (Wold and Mielnik 2000; Wold et al. 2002) and sugar quality (Baunsgaard et al. 2000a,b). In Paper I fluorescence emission is measured on cod and salmon fillets and the emission spectra analysed with PCA, illustrating that fluorescence emission is complex and involves several fluorophores such as collagen, NADH and oxidation products (see Section 3.1 for a description of the results). Even though solid sample spectrophotometry has advantages in industrial quality monitoring, the high sensitivity and selectivity of the fluorescence spectra is lost due to lower signal intensities. Furthermore, the complexity of the products gives rise to complex fluorescence signals. Interferences such as re-absorption, various forms of light scattering, unknown fluorophores and quenching may influence the fluorescence spectra, thereby reducing the sensitivity and selectivity.

Examples of the application of PARAFAC for resolving and describing fluorescence EEMs are given by Ross et al. (1991), Jiji et al. (1999), Bro (1999), Baunsgaard et al. (2000c), Moberg et al. (2001) and Pedersen, Munck and Engelsen (2002). Paper II also provides an example of how to use PARAFAC for curve resolution and interpretation of the underlying structure in the fluorescence landscapes. From the EEMs measured on fish muscle extracts, pure excitation and emission spectra are estimated illustrating that three fluorophores are present, one of which is suggested to be tryptophan (see Section 3.2 for more discussion of the results).

2.4. Low-field ^{1}H nuclear magnetic resonance (Low-field NMR)

Nuclear magnetic resonance is widely used as an analytical technique in many scientific fields. With the development of NMR bench-top instruments it has become possible to obtain fast non-destructive and reproducible NMR measurements for quality control within the food industry. The signals measured are functions of time and give information about relaxation processes. These relaxation phenomena can provide useful information on various aspects like moisture or fat content as well as information on the physical structure of the sample.

This section deals with the theory of low-field NMR and includes a description of how to measure and analyse the NMR decay curves. A more detailed description of the theory is given in textbooks (Deleanu and Paré 1997; Ruan and Chen 1998) as well as in scientific review articles (Rutledge 1992; Colquhoun 1993).

2.4.1. Nuclear magnetic resonance

NMR is observed for isotopes having a non-zero nuclear spin. Such nuclei contain a net magnetic moment or net spin of the whole sample. The spin of each nucleus can be described as miniature bar magnets interacting with each other and with an externally applied magnetic field. When placed in an external magnetic field, the spins will align with the applied field. Some spins are placed in parallel with the applied magnetic field and others are placed anti-parallel. The energy difference between these two situations depends on the applied external magnetic field.

The NMR signal is produced by absorption or emission of a photon caused by a radio frequency pulse whose energy is equal to the difference between the two spin states. The pulse tips the magnetisation away from the static field, inducing an NMR signal. The hydrogen atom is by far the most commonly used nucleus. This is due to a high sensitivity and a large abundance in most types of products. Other nuclei used in NMR are for example ^{13}C, ^{19}F and ^{31}P.

2.4.2. Measuring NMR relaxations

After a radio frequency pulse is applied, the spins relax to their equilibrium states at a certain rate. Relaxation time constants are used to characterise this decay rate. The constants, denoted as T_1 (spin-lattice) and T_2 (spin-spin), give information about molecular mobility and can, therefore, be used to study chemical and physical properties of the sample. T_1 is a measure of the time taken to restore the component parallel to the applied magnetic field and characterises the time needed to obtain equilibrium between the spin system and the surrounding sample. T_2 is a measure of the time taken for the spins to reach equilibrium transverse to the external field and, therefore, a measure of the duration of the NMR signal. It is due to interactions between spins, chemical exchange and molecular diffusion.

In low resolution NMR, several pulse sequences are commonly applied such as inversion recovery (measuring T_1), FID (measuring T_2*)[1] and CPMG (measuring T_2). The CPMG pulse sequence used in Papers III and IV is devised by Carr and Purcell (1954) and Meiboom and Gill (1958). It is generally preferred for estimation of T_2 values, because the whole relaxation curve is obtained in one measurement, and the contribution from inhomogeneity in the magnetic field is avoided. The sequence is illustrated in Figure 2.5. First, a 90° pulse is applied, which tips the magnetisation into a plane perpendicular to the applied magnetic field. After a short time, denoted τ, a 180° pulse is applied, then a new 180° pulse is applied after another two τ periods, and so on. The sequence produces broad peaks of NMR intensity denoted as echoes. A series of echoes with decreasing intensity is obtained, if the process is repeated several times. Typically, the maximum of every second echo is acquired to give the NMR decay curve.

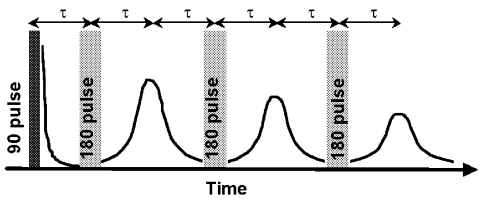


Figure 2.5 Illustration of the CPMG sequence with echoes as the solid line.

2.4.3. Application of chemometrics in the interpretation of low-field NMR decays

Low-field NMR measures the mobility of protons and is therefore suited for estimating the characteristics of water and fat in the product. For this purpose chemometric methods such as PCA and PLS have been applied in a number of studies where the whole decay curves were measured (Hernandez and Rutledge 1994; Gerbanowski et al. 1997; Rutledge and Barros 1998; Jepsen et al. 1999; Brøndum et al. 2000; Pedersen et al. 2000; Brown et al. 2000; Thybo et al. 2000). However, since the NMR decay produced from the CPMG sequence can be described as a sum of exponentially decaying curves the data can also be analysed with various exponential curve fitting procedures.

Discrete exponential curve fitting gives a few characteristic relaxation components and related magnitudes[2] for each sample and is one of the traditional ways of analysing NMR decays. It has been used to identify water pools and to interpret and evaluate the importance of these pools to various quality parameters in meat and fish (Renou et al. 1985; Fjelkner-Modig and Tornberg 1986; Lambelet et al. 1995; Steen and Lambelet 1997; Botlan et al. 1998; Jepsen et

[1] T_2* is the spin-spin relaxation time constant measured by FID but depends on the homogeneity of the magnetic field. Therefore, it is normally shorter than the real T_2 value, which is measured by the CPMG sequence.

[2] The relaxation time constants and magnitudes are given in Equation 2.4 as T_2 and M, respectively.

al. 1999). One advantage of discrete curve fitting is that the relaxation time constants and magnitudes are estimated for each sample allowing the relaxation characteristics to be different. However, knowledge of the number of relaxation components is required. Discrete exponential fitting can be made on all samples together and is then denoted as “matrixfit”. It gives a few characteristic relaxation time constants that are equal for all samples, but with magnitudes varying among the samples.

By distributed exponential fitting, the relaxation curve for each sample is fitted to a continuous distribution of characteristic relaxation times. It is claimed that in muscle foods water does not exist in distinct pools, but the surroundings vary for each proton, giving different relaxation phenomena (Brown et al. 2000; Bertram et al. 2002a). By distributed exponential fitting these variations are taken into account and the method is, therefore, typically applied for interpretation of the underlying exponential characteristics, as shown for pork meat by Bertram et al. (2001 a,b and 2002a,b) and by Brown et al. (2000). Hills and Remigereau (1997) and Lillford et al. (1980) give other examples of the use of distributed exponential fitting.

Distributed exponential fitting may give a more fully understanding of the underlying relaxation characteristic since the relaxation phenomena of each proton is recognised. However, different sample characteristics and noise levels may have an impact on the relaxation characteristics obtained by distributed exponential fitting. In addition, there are many distribution curves, which will have similar fit to a given measurement but different interpretations (Pedersen 2001). Typically, the same overall relaxation characteristics are obtained by discrete exponential fitting. Thus, discrete fitting may be preferred when the above mentioned considerations are taken into account.

Slicing (Section 2.1.1.3) is used in Papers III and IV to analyse the NMR decays measured on cod muscle (see also Section 3.3). Paper III shows how the NMR relaxation curves measured on cod muscle can be decomposed into two exponential components describing two water pools whose concentration varies within the length of the fish and between the fish. In Paper IV it is shown how the size of the exponential components obtained by slicing can be related to the WHC. This illustrates a relation between WHC and the mobility of water within the fish muscle. It should be mentioned that the results of slicing are similar to the results obtained by matrixfit. However, there are advantages of slicing such as the possibility of using simple and well known validation tools to find the number of relaxation components.

The methods discussed above for analysing low-field NMR relaxations are only some examples of how the data can be analysed. There are advantages and disadvantages with all methods and the choice of a certain analysis method may depend on the actual data and the purpose of the analysis. In the experimental work performed here, slicing is primarily used, because the purpose is to describe the water mobility with a few characteristic parameters that are common to all samples.

3. Exploratory analysis of spectroscopic data

It is important to understand the underlying chemical and physical phenomena that give rise to the instrumental signals before spectroscopic measurements can be applied for industrial quality assessment. This chapter gives examples of how data obtained by three different methods can be analysed and interpreted. The three methods include two types of fluorescence spectroscopic measurements and ^{1}H low-field NMR.

3.1. Multivariate analysis of fluorescence spectroscopic measurements

Fish muscle exhibits autofluorescence (Manohar 1969; Davis 1982), which has been related to parameters such as lipid oxidation and freshness (Munck 1989; Undeland et al. 1998). Paper I presents an explorative investigation of autofluorescence from salmon and cod muscle. An indication of the fluorophores present is obtained a visual appearance of the scores and loadings obtained by PCAs. In addition, an interpretation of changes due to storage time, site of measurements and variations between the two fish species.

Fish muscle texture and the degree of gaping are two parameters important for fish product quality. Gaping is a phenomenon where the connective tissue can no longer hold the muscle blocks together. It has been found that the two parameters depend on the content of connective tissue, in that higher collagen content is observed in fish with a firmer texture and low collagen content is found in fish with tender meat vulnerable to gaping (Sato et al. 1986; Hassan et al. 1999). Since connective tissue and thereby collagen is important for the overall impression of fish products, determination of collagen content and its degradation by a fast on-line method can be important for controlling and optimising processing and marketing. Connective tissue does emit autofluorescence, which is supposed to be due to collagen cross-links and components such as pyrodinoline (Bailey et al. 1995). In addition, autofluorescence of bovine connective tissue has been characterised (Swatland 1987a; Egelandsdal et al. 1996; 2002; Wold et al. 1999). Therefore, it was supposed that collagen content, texture and degree of gaping could be evaluated by measuring autofluorescence of fish muscles.

Collagen type I and type V are the two collagen types which occur naturally in fish muscle (Sato et al. 1988; 1989; Mizuta et al. 2002). Therefore, fish muscle autofluorescence is compared with autofluorescence of collagen type I and type V. Figure 3.1 shows the pure emission spectra of the two types of collagen and the loadings of PCAs based on fluorescence spectra of salmon and cod.

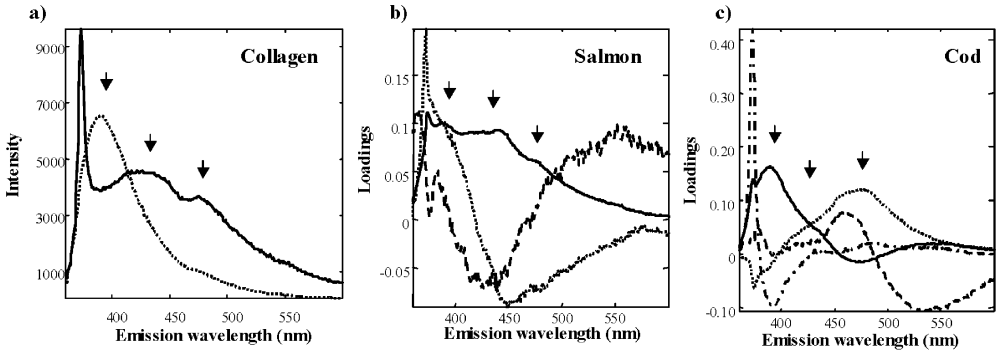


Figure 3.1 a) Autofluorescence spectra of collagen type I (solid) and collagen type V (dotted) obtained after excitation 330 nm, b) loadings of a PCA based on emission spectra of salmon obtained after excitation 330 nm, and c) loadings of a PCA based on emission spectra of cod obtained after excitation at 330 nm. In b and c the components are indicated by: PC1 (solid), PC2 (dotted), PC3 (dashed) and PC4 (dash-dotted).

The loadings obtained for salmon illustrate peaks in PC1 (shown as arrows in Figure 3.1) corresponding to those of collagen type I and type V. Thus, collagen likely contributes to the autofluorescence and to the variation in intensity. This is verified by orthogonal rotation of the loadings. The loadings obtained from a PCA based on salmon spectra and spectra of collagen types I and V are rotated orthogonally towards the loadings obtained from a PCA based only on the salmon spectra (Jennrich 1970). This rotation results in similarities between the loadings obtained from the two models (see Figure 8 in Paper I). The presence of collagen in cod muscle is not so clearly identified. Only the loadings of PC1 have a large peak below 400 nm corresponding to the peak of collagen type V. The orthogonal rotation also shows that collagen fluorescence is not as pronounced for cod as for salmon (see Figure 9 in Paper I).

The fish were stored for different periods and changes due to storage time were studied. Storage of cod for longer periods gives increased fluorescence intensity slightly above 500 nm, which is probably due to some oxidation phenomena. In addition, a decrease in the peak with position between 450 and 480 nm is possibly correlated to a decrease in the NADH content. These two phenomena can be seen in the loadings of PC2 and PC3 in Figure 3.1c. The relation between fluorescence intensity and lipid oxidation is confirmed by other studies showing increased fluorescence intensity of aqueous and organic fish muscle extracts together with increased lipid oxidation (Aubourg et al. 1998a,b; Aubourg and Medina 1999). The same was illustrated by solid sample fluorescence spectrometry on a dried fish muscle system (Hasegawa et al. 1992) and on herring filets (Undeland et al. 1998).

The actual fluorescent properties of the two fish species are probably quite similar, even though differences in fluorescence characteristics have been found. Fluorescence microscopy showed that the major source of fluorescence from adipose tissue was from the connective tissue fibres located between the adipose cells (Swatland 1987b). Therefore, due to the larger fat content, it

seems reasonable that collagen fluorescence is more pronounced for salmon. Fluorescence of salmon muscle is probably re-absorbed by red pigments in the muscle fibres such as astaxanthin. The absorption by those pigments may also be seen in PC3 (Figure 3.1b) where the negative loadings between 430 and 480 nm correspond to the wavelengths with absorption of the pigments (Britton 1995). This effect of pigments in salmon creates the visible contrast between muscle fibres and connective tissue and explains why collagen fluorescence can be visualised, whereas fluorescence of the muscles cannot. Cod muscle is white and no pigments will re-absorb the autofluorescence. Together with the lower fat content, this indicates that fluorescence of cod muscle is more intense than that of connective tissue.

In order to study the potential of using fluorescence spectroscopy for fast non-destructive measurements of texture and gaping, predictions by PLS regression were performed. As shown in Paper I, the predictions do not perform well, probably due to non-optimal sample representativity, measurement error and small sample variation. For texture predictions, the same weak relations are obtained as were found by application of visual/near-infrared reflectance spectroscopy (Isaksson et al. 2002). However, it has been illustrated that determination of beef and pork connective tissue by application of fast on-line fluorescence probes is possible (Swatland 1991; Swatland 2000).

The possibility of measuring connective tissue in mammal meat by fluorescence spectroscopy and the identification of collagen in the fluorescence spectra of salmon makes the method seem promising as a rapid tool for determining such parameters. However, the results are not unambiguous, for which reason other sampling methods and instrumentation should be tested. In addition, the possibility to measure the decrease in NADH content and increase in lipid oxidation of cod fillets indicates that fluorescence spectroscopy can be a valid method for estimating the storage time of fish muscles in industrial applications.

3.2. Multi-way analysis of fluorescence excitation-emission matrices (EEM)

Paper II describes how fluorescence spectroscopic data can be decomposed, the underlying chemical phenomena resolved and the data interpreted by PARAFAC modelling. This may be of industrial importance, since in order to use fluorescence spectroscopy for quality measurements, it is important to understand the underlying factors that govern the measured spectral data. Fluorescence excitation-emission matrices (EEMs) of aqueous cod extracts have been measured, enabling the use of PARAFAC for data analysis. The extracts were obtained from frozen-thawed modified atmosphere packed cod fillets from the Barents Sea. The cod fillets were stored at 2°C for up to 21 days. To expand the variation in the product, different frozen storage temperatures, frozen storage periods and chill storage periods were used.

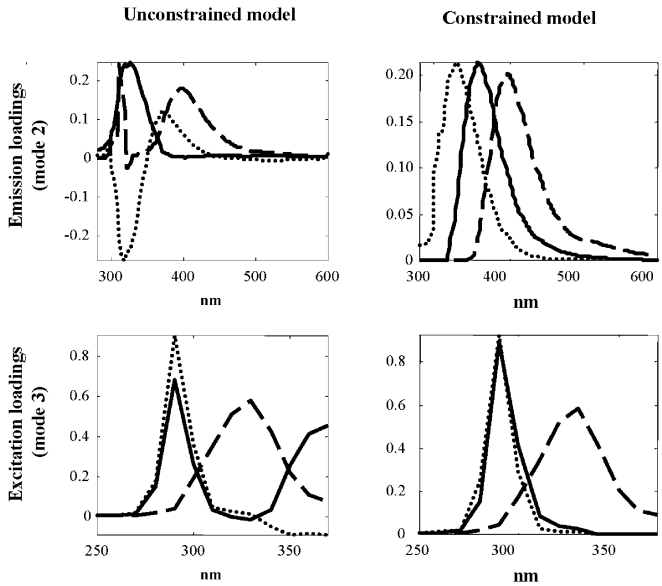


Figure 3.2 Loadings of an unconstrained three-component PARAFAC model and of a three-component PARAFAC model constrained with non-negativity in the first mode and non-negativity and uni-modality in the second and third modes.

Initial analyses indicate that three fluorophores are present. The loadings of a three-component PARAFAC model are shown in Figure 3.2 (left part). However, due to Rayleigh scatter and missing values the loadings of this model do not give a valid description of the underlying phenomena in the fluorescence landscapes. The sharp peak around 300 nm in the emission loadings and the large negative region indicate that the model is not reliable. In Paper II various methods are used to evaluate and illustrate the inappropriateness of the model such as jack-knife, visualization of residuals and appearance of the loadings.

It has been shown that PARAFAC can be accomplished even when non-trilinear parts in the EEMs disturb the assumptions of the model. By a weighted PARAFAC model, elements in the EEMs are up- or down- weighted in such a way that the fluorophore signal will contribute most to the modelling. It was applied to circumvent Rayleigh and Raman scatter overlapping with the fluorescence landscapes (Jiji and Booksh 2000; Jiji et al. 2000). A weighted least squares fitting method called maximum likelihood least squares (MILES), where the model parameters are estimated by least squares algorithms in an iterative way, was suggested by Bro et al. (2002b).

In addition, constraining the model parameters can be used to handle non-trilinear parts. Applying constraints may increase the modelling capability and ensure that the estimated spectra make sense. For example, non-negativity constraints were applied to resolve the pure spectra of three fluorescent dyes (Lee et al. 1991). Both non-negativity and uni-modality

constraints were used to handle scatter and missing values when analysing fluorescence EEMs of sugar solutions (Bro 1999; Baunsgaard et al. 2000a,c). In Paper II it is shown that constraining the parameters of the PARAFAC model with uni-modality and non-negativity improves the interpretation, curve resolution and model stability. However, before applying constraints, the appropriateness of these should be considered (Bro 1998). For the data presented in Paper II, non-negativity constraints are reasonable due to the fact that spectra of pure analytes cannot be negative and negative loadings are obtained for the unconstrained model (Figure 3.2). Furthermore, fluorescence spectra are generally characterised by one peak, which together with the uni-modal appearance (apart from an unstable part) of the loadings supports the appropriateness of uni-modality constraints. The loadings of the thus obtained three-component constrained model (Figure 3.2) illustrate that application of the chosen constraints is reasonable. Peaks are obtained at the same positions as for the unconstrained model, but by applying constraints, negative parts are removed, and the loadings become uni-modal. Indeed, for the emission loadings, the inferior peak caused by scatter and missing values has been removed.

Two of the components estimated by the constrained model have the same position of the excitation maximum, but different positions of the emission maxima that are found at 330 and 360 nm, respectively (Figure 3.2). Even though the two components have the same excitation maxima, they are not equal since the emission maxima differ. The reason for the apparent similarity of these loadings can be that they reflect the same fluorophores, but with different substitutions or they appear in slightly different environments, thus leading to slight differences in spectral characteristics. Standard addition showed that one of these components describes the presence of tryptophan. The other may also be tryptophan, but with a substitution, or it may be bounded to another molecule. There is a weak relationship between the relative concentration of these compounds and the chill storage time indicating the development of smaller protein fragments with the storage time. The fluorophore described by the third component has its excitation maximum around 330 nm and emission maximum around 400 nm (Figure 3.2). This component needs still to be identified, but may be NADH, even though the peak values vary from those given in the literature (Munck 1989). The exact position of the peaks depends on the sample matrix and the measurement conditions.

The objective of Paper II is to illustrate and discuss the steps and problems involved in finding the optimal PARAFAC model based on fluorescence data. The paper shows that fluorescence EEMs can be modelled well by PARAFAC, even when the data are influenced by missing values and scatter. In addition, the paper shows how fluorescence EEMs obtained from fish muscle extracts can be analysed. However, in order to fully use fluorescence spectroscopy and three-way modelling for fish quality measurements, more dedicated experiments are needed, including measurements on other types of extracts, whole fish muscle and application of other types of experimental design.

3.3. Interpretation of low-field NMR relaxations

Water is found in large quantities in many food products and is known to be extremely important for the overall stability and acceptability of foods. For example, sensory parameters, texture, WHC and drip loss are factors that depend on the concentration and mobility of water. In fish muscles, water can be seen as existing in different fractions depending on how tight the water is bound to the muscle structure (Ruan and Chen 1998). The distribution of water into these fractions depends on factors like storage treatment, freezing, mincing, etc. In recent studies made on porcine meat, three water fractions were identified: one trapped between the fibres, one outside the myofibrillar structure or even outside the muscle cells, and one tightly bound to the muscle proteins (Bertram et al. 2001a and 2002b). The last was found in very low concentrations that remained constant irrespective of how the meat was treated (Bertram et al. 2002b).

Low-field NMR is an ideal tool for quantification and characterisation of water in foods since it measures the mobility of protons and thus also the mobility of water (Schmidt and Lai 1991; Berendsen 1992). Several examples have illustrated the applicability of low-field NMR for studying the water distribution in foods. For example, the relation between NMR relaxations and the state of water within the food product is used to predict quality parameters such as the WHC of meat (Brøndum et al. 2000; Bertram et al. 2001b and 2002a,b) and of frozen-thawed cod (Jepsen et al. 1999). Muscle changes taking place after slaughter influence the water distribution and have been described by NMR relaxometry (Tornberg et al. 2000). Texture changes in frozen cod have been correlated to changes in NMR relaxation parameters (Steen and Lambelet 1997). Furthermore, sensory parameters such as juiciness and tenderness depend on the water availability and have been correlated to the NMR relaxation constants (Fjelkner-Modig and Tornberg 1986).

The following shows how low-field NMR can be used to describe the distribution of water within gadoid fish muscles. The NMR measurements were obtained by CPMG pulse sequences (see Section 2.4.2), giving information about the T_2 relaxation characteristics. Section 3.3.1 contains a description of the distribution of water within cod fillets (Paper III), Section 3.3.2 describes the distribution of water in four gadoid fish species and Section 3.3.3 discusses the relationship between the distribution of water into different pools and WHC (Paper IV). Accordingly, this subchapter gives a thorough discussion of the applicability of low-field NMR for studying the states of water in fish muscles and a few derived quality parameters.

3.3.1. Distribution of water in cod

Water content and NMR relaxation profiles were measured in 40 to 60 samples taken systematically from whole fillets of five cod. The NMR relaxations were analysed with the slicing method (see Section 2.1.2.3), giving mono-exponential profiles corresponding to the common characteristic water pools. Relaxation time constants (T_2 values) were calculated from these profiles.

Two exponential functions were obtained, illustrating the presence of two different water pools with the characteristic T_2 values of 50 and 94 ms. Figure 3.3 shows, for two of the five fillets, the distribution of the score values of the slow relaxing component (component two) and the distribution of water within the length of the fillet. The figure reveals that the water content as well as the score values of the slow relaxing component (component two) generally increase with the distance from the head. This indicates that the score values contain information about the total water content, which can also be verified by the correlation of 0.94 between these score values and the water content. The distribution of the water content is in agreement with the results obtained by Dambergs (1963).

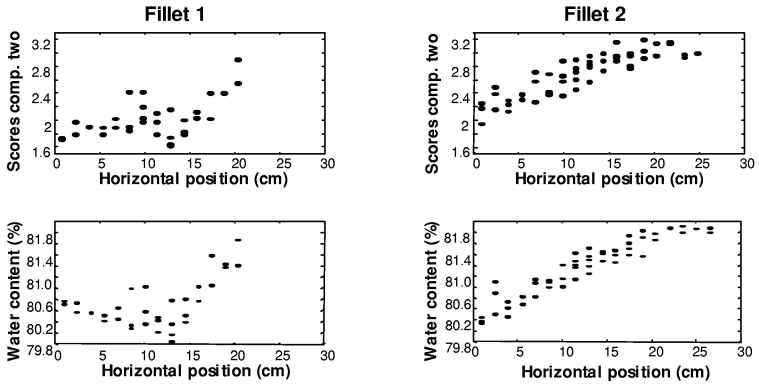


Figure 3.3 Distribution of scores of component two and water content. The position is given as the distance from the head (0 cm).

Water with the fastest relaxation is found in larger concentration near the head and water with the slowest relaxation is primarily found near the tail. Cell sizes and muscle fibres are smaller towards the tail compared with the rest of the fillet (Love 1988). Therefore, the variation in both the total water content and the distribution of the water pools can likely be explained by the size of the muscle cells or fibres meaning that water is bound more tight in the larger muscle fibres closer to the head (Love 1988), which also contain less water as it is shown in Figure 3.3.

3.3.2. Distribution of water in four fish species

Low-field NMR is used to investigate the distribution of water in four gadoid fish species (cod, haddock, whiting, and saithe). The purpose is to see if the distribution of water varies systematically between the four fish species. A three-component slicing model seems to give the best description of the data. The reliability of this model is supported by fit values, core consistency diagnostics, distributions of residuals and split-half analyses[3]. However, both distributed and discrete exponential fitting point to two components being optimal[4]. Figure 3.4 shows some results obtained by distributed fitting. To the left, the distribution of T_2 values for ten samples of each fish species is shown. All samples have a main water component with T_2 values between 50 and 100 ms. Another smaller component has T_2 values ranging from 100 to 1000 ms, but the exact values vary between the samples. The distribution of the T_2 values also differs within one fish species and between the fish species. The right part of the figure shows an enlargement of the area containing the main component. Mean relaxation time distributions of haddock and whiting are shown; illustrating that haddock typically has lower T_2 values than whiting[5]. These differences between fish species and within one fish may explain why slicing cannot model the data appropriately. Even though all samples contain two water pools, these cannot be identified by slicing because of the large variation in the T_2 values. The appropriateness of two water pools in fresh gadoid fish muscle is supported by other studies (see Paper III and Lambelet et al. 1995).

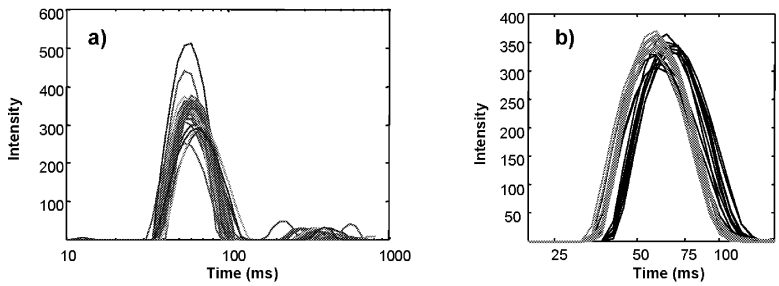


Figure 3.4 a) Distributed exponential fitting of T_2 relaxation data from 10 samples of whiting, haddock, saithe and cod and b) enlargement of the distributions between 20 and 100 ms for average distributions of whiting (black) and haddock (grey). Averages are made for each fish.

All four fish species are gadoid and it is assumed that the muscle structure of the four species is not very different. The relaxation time constants vary among the four fish species, but also within one species. This indicates a variation in the water pool distribution. However, no

[3] See paper II for a description of the validation methods that can be applied in PARAFAC. Split-half analysis and distribution of residuals are explained in Chapter 4.
[4] See section 2.1.2.3 and 2.4.3 for a discussion on the data analytic methods.
[5] Mean relaxation time distributions are the mean distribution of each fish. Eighteen measurements are performed on each fish.

systematic variation between the four fish species is found in the water pool sizes probably due to a large variation within one fish species.

Slicing is a powerful data analytical method which can be applied when all samples have the same underlying structure. However, as discussed above, slicing may give misleading interpretations when these assumptions are not met. Therefore, the appropriateness of slicing should always be considered, before using it for data analysis. This can be done by relating the slicing results to other information about the data as illustrated above.

3.3.3. Relationship between water distribution and water holding capacity (WHC)

The ability of the muscle to hold water is important for the acceptability of the product since poor WHC influences the texture, juiciness and overall impression of the fish product. An increased amount of fish is frozen before sale or is sold as a frozen product. By freezing and thawing, muscle proteins are degraded, which leads to decreased protein solubility and loss of functional properties (Shenouda 1980; Mackie 1993). This decreases the WHC and the product quality. Therefore, the development of instrumental methods, such as low-field NMR, to measure WHC is industrially important. Furthermore, the applicability of low-field NMR for studying the distribution of water into different pools enables a fundamental study of the relationship between water in the muscle and WHC. Such knowledge can be used to optimise freezing and thawing as well as the storage conditions.

Low-field NMR measured on cod fillets subjected to different freezing and storage treatments are obtained and decomposed with slicing and discrete exponential fitting. The amount of water in each pool is calculated as described in Paper IV and related to the WHC by multiple linear regression (MLR). WHC is reported as (1) the amount of water present in the sample after centrifugation relative to the original amount of water (WHC_w) or as (2) the amount of water present in the sample after centrifugation relative to the amount of dry matter (WHC_d). These two quantities are perfectly correlated as long as the dry matter content does not vary among the samples. However, when the dry matter content varies, as is the case here, the two measures will give different WHC.

Tables 3.1 and 3.2 show the performance of the predictions including also the results obtained by PLS regression using the whole relaxation curves as x-variables. Generally, large correlations (r) and low error estimates (RMSECV) are obtained; illustrating that low-field NMR is a valid method to assess WHC in fish muscles. It is interesting that when expressing WHC relative to the original amount of water (WHC_w), optimal predictions are obtained on whole and minced muscle. The poor predictions obtained from the centrifuged fillets are reasonable, since water is removed by the centrifugation and the samples contain no information about the original water content. However, the centrifuged samples give good predictions when WHC is expressed relative to the dry matter content (WHC_d). The centrifuged samples have been given

exactly the same treatment (mincing and centrifugation) as the samples used for the physical determination of WHC. Therefore, it is reasonable that these samples give more accurate estimated values and thus better predictions.

Table 3.1 Prediction of WHC_w expressed as percentage of water left after centrifugation relative to the original amount of water. Variation in WHC_w is between 68 and 88 %. RMSECV is given in per cent.

Whole fillet		Minced fillet		Centrifuged fillet	
r	RMSECV (%)	r	RMSECV (%)	R	RMSECV (%)
		PLS on whole relaxations[a]			
0.86	2.95	0.77	3.75	0.60	4.80
		M values obtained by slicing			
0.73	3.96	0.83	3.27	0.67	4.47
		M values obtained by slicing			
0.77	3.84	0.68	4.33	0.48	5.47

[a] The PLS regressions are made with three components for whole and minced fillet and one component for centrifuged fillet.

Table 3.2 Prediction of WHC_d expressed as g water left after centrifugation relative to the g dry matter. Variation in WHC_w is between 2.37 and 3.92 g/g. RMSECV is given in the original values (g/g).

Whole fillet		Minced fillet		Centrifuged fillet	
r	RMSECV (g/g)	R	RMSECV (g/g)	r	RMSECV (g/g)
		PLS on whole relaxations[a]			
0.77	0.817	0.79	0.233	0.92	0.157
		M values obtained by exponential fitting			
0.68	0.272	0.82	0.221	0.79	0.246
		M values obtained by slicing			
0.77	0.248	0.70	0.276	0.92	0.156

[a] The PLS regressions are made with three components.

In contrast to the fresh fish muscles shown in the two previous sections, three water pools were found. Presumably, the extra component corresponds to a water pool that develops during freeze storage. The appearance of this water pool is likely due to protein deterioration and leakage of water into the extracellular space and can be identified as thaw exudates (Lambelet et al. 1995).

One objective of this study has been to investigate the relationship between the concentration of water in the different pools and the WHC. This is evaluated from the multiple linear regressions by a jack-knife approach that estimates the significance of the regression coefficients (Section 4.2.3). The prediction of WHC_w of whole and minced fillet shows that both the fastest and the

slowest relaxing water are significant, indicating that a large concentration of water within the myofibrillar structure and a low concentration of very mobile water ensures high WHC. However, it may be that the concentrations of the two water pools are negatively correlated. In this situation both of these will be found to be important by the jack-knife validation even though one of them may not actually have an influence on the WHC. The slowest relaxing and loosest bound water is removed during centrifugation and is probably also removed as a function of lower WHC. Therefore, it is reasonable that this water pool is important for the predictions. Prediction of WHC_d of the centrifuged samples shows that the two fastest relaxing components are significant. By centrifugation, part of the slowest relaxing water is removed and will probably have only a limited influence on the predictions.

The above shows that low-field NMR is a valid method to measure WHC of fish muscle. One purpose has been to evaluate low-field NMR as a non-destructive method, which can be applied industrially. This makes the prediction performance of the whole fillets interesting and encourages further studies. The good predictions obtained with the centrifuged samples are interesting from a scientific point of view and can be used in research to improve the understanding of the relationship between water mobility and WHC. In the longer run such an understanding can be used to optimise storage and processing.

4. Validation

Validation is an important aspect in all data analysis. The overall purpose is to ensure that the model will work in the future for new, similar data. In addition, explorative methods such as PCA and PARAFAC should be validated to assure that the interpretation is correct and reasonable. For regression models such as PLS and PCR, validation should be more focused on obtaining as good predictions as possible, meaning that the predicted value should be close to the value of the reference measurement. Thus, validation includes, among other things, determination of the number of components, outlier detection and evaluation of the estimated model parameters (Harshman 1984).

Validation cannot be applied using a few general methods. The chosen validation method will depend on the data, the problem, and the chemometric methods. In the following, several validation methods will be described. Examples taken from the Papers I to VI are included to illustrate the applicability of the various methods. The objective is to show the applicability and interpretability of the methods and not to give the mathematical and statistical background.

The validation methods described here do not represent a full list of all possible methods. The description rather illustrates some important methods and should give an indication of how the models can be validated in many similar situations. Furthermore, it should be recognised that the results obtained by a certain validation method are not the truth, but only provide indications. The results should be related to external knowledge about the data in order to ensure that the model is valid in relation to what is already known. For example, spectral measurements such as fluorescence typically have a smooth form with positive and uni-modal peaks, as described in Paper II. This can be used to verify an obtained estimate of pure analyte spectra. Likewise, in Papers III and IV it is described that the loadings obtained from slicing modelling of NMR relaxations should be mono-exponential, as they represent the pure relaxation profiles of the different exponential components.

4.1. Validation methods

4.1.1. Test set validation

Test set validation is used to determine the number of components to use in multivariate calibration models and to give an estimate of the prediction errors. The Root Mean Squared Error of Prediction (RMSEP) is typically used to evaluate the results. It is expressed as

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{4.1}$$

where $\hat{y}_i$ is the predicted value of the i-th sample, y_i is the measured/true value of the i-th sample and n is the number of samples. When calculating RMSEP using increasing number of components, the error estimate usually decreases until it reaches a minimum. The number of components should be chosen at this minimum. With a further increase in the number of components, RMSEP increases again or flattens out.

Test set validation can be applied when two sets of samples are at hand. One of these is used for calibration. The other is used as an independent test set that should estimate the prediction performance of new samples not included in the model. The test set can also be chosen by taking samples out of one data set, if the data set contains enough samples. One part of the samples will be used to make the model; the other will be used as an independent test set to evaluate the predictions.

Test set validation requires that the two data sets are representative of future measurements, that they span the expected variation in the samples and reflect the variation arising due to sampling and measurement. Therefore, the actual measurements/samples in the test set have to be chosen based on the information it is sought to reflect. Test set validation is used in Papers V and VI to evaluate the prediction performance of unknown samples even though the data sets contain few samples.

4.1.2. Cross validation

Cross validation is typically used to estimate the number of components in a calibration model similarly to test set validation. It is performed by leaving out one sample or a group of samples and using this sample or these samples for validating a model based on all other samples. After that, another sample/group of samples is left out and used for validation. This procedure continues until all samples or segments of samples have been left out exactly one time. The

method is called full-leave-one-out cross validation when one sample is left out at a time and segmented cross validation when a group of samples is left out.

The Root Mean Squared Error of prediction as estimated by Cross Validation (RMSECV) is used to evaluate the results. RMSECV is used and interpreted in the same way as RMSEP described in the previous section. Cross validation is often chosen when the data set contains only a few samples or when a representative test set cannot be drawn from the data. It is applied in Paper I to estimate the number of components and obtain an indication of the amount of variance explained in PCAs made on fluorescence spectra of salmon and cod. In Paper IV, cross validation is used when predicting WHC of fish muscles from low-field NMR decays.

When segmented cross validation is used, it is important to consider how the segmentation of samples is made. For example, if the data set contains replicates, these should be included in the same segment, as it is done in Paper IV. If replicates are separated into different segments, the model will be validated on the same samples that are used for calibration, and the error estimates may be too optimistic. Another way to handle replicates is to average the replicates and then use full-leave-one-out cross validation. However, in this approach variation between replicates will not be detected.

4.1.3. Jack-knife validation

Jack-knife validation is used to evaluate the uncertainty of parameter estimates. It can be considered as an extension of cross validation. Leaving out samples in the same way as for cross validation creates a number of new data sets. One model is made on each of these data sets, giving a set of parameter estimates for each of the “new” data sets. The model is validated on the sample(s) left out (Efron and Tibshirani 1993). As for cross-validation, jack-knifing can be performed by leaving out segments of samples or by leaving out only one sample at a time.

The jack-knifing results can be evaluated by calculating the standard deviation of the estimated parameters, which is estimated by Wehrens et al. (2000) as

$$s_V = \sqrt{\frac{n-1}{n}\sum (V_i - \bar{V})^2} \qquad (4.2)$$

where V_i is the estimated value obtained when fitting the model on the data excluding sample *i*, $\bar{V}$ is the mean value of the estimated V_i values, and *n* is the number of new data sets.

The theory and some applications of jack-knifing are described by Efron and Tibshirani (1993) and by Martens and Martens (2000). Jack-knifing can be applied to eliminate useless variables or to find outliers. The determination of significant x-variables in PLS regressions and the

elimination of disturbing and useless ones is shown in a sensory assessment of cocoa drinks for predicting physical and chemical parameters (Martens and Martens 2000). Bøknæs et al. (2002, appendix 1) applied jack-knifing to determine significant wavelengths in the prediction of chill storage time of frozen-thawed cod from near infrared spectroscopy. Paper IV shows the use of jack-knifing for evaluating the importance of the different water pools in the prediction of WHC, where the concentration of water in each water pool is obtained from low-field NMR decays.

Riu and Bro (2002) showed how identification of outliers in a trilinear model such as PARAFAC can be performed based on jack-knifing results. This approach is used in Paper II, which also shows how the stability of the model parameters can be evaluated by jack-knife validation.

4.1.4. Bootstrapping

Bootstrapping is another resampling approach used to estimate the uncertainty of model parameters. Many new data sets, called bootstrap samples, are created from the original data set by sampling with replacement. As for jack-knifing, one model is made for each of these bootstrap samples. This gives a number of estimated parameters from which statistical estimates such as mean, standard error and confidence intervals can be obtained. The standard error is obtained as expressed by Wehrens et al. (2000) (Equation 4.3) as

$$s_z = \sqrt{\frac{1}{b-1}\sum (Z_i - \bar{Z})^2} \qquad (4.3)$$

where b is the number of bootstrap samples, Z_i the i-th bootstrap estimate and $\bar{Z}$ the mean value of the bootstrap estimates. It is important that the number of bootstrap samples is large enough to ensure stable and reliable estimates.

A thorough description of bootstrapping is given by Wehrens et al. (2000) and Efron and Tibhshirani (1993). The application of bootstrapping in PCR is described by Wehrens and van der Linden (1997). Furthermore, bootstrapping is applied in Paper III, illustrating how the uncertainty of PARAFAC loadings can be used for choosing the correct model.

In jack-knifing all samples contribute equally to the model parameter estimation. When resampling is performed with replacement as in bootstrapping, the estimated model parameters depend on the actual samples selected. Furthermore, bootstrapping renders more difficult to recognise phenomena occurring in very few samples. Therefore, jack-knifing should be preferred for identifying outliers and detecting insignificant variables. In addition, jack-knifing is easily performed using the cross validation schemes, whereas bootstrapping requires additional calculations. However, bootstrapping may be preferred, when statistical uncertainty of parameter estimates is of primary importance.

4.1.5. Split-half analysis

Split-half analysis is used to estimate the number of components to use in models with a unique solution such as PARAFAC. In the split-half analysis, the data is divided into two parts that are analysed independently (Harshman 1984; Harshman and de Sarbo 1984). When the model is fitted on the two parts, equal model parameters are obtained in the non-splitted directions (modes) if the model is valid. Split-half analysis is a strong method to validate the model dimensionality since the model parameters will only be equal if the correct number of components is chosen.

It is important to consider how data is divided as both sets should contain all underlying phenomena in the data. Furthermore, all split-half data sets should contain a sufficient number of independent samples/variables. If the data set is too small, modelling each split-half data set may not be able to reveal the important variation.

If one phenomenon is only present in one of the split-half data sets, the parameter estimates may not be stable. This induces a risk of rejecting the model even though it may be true and would have been accepted if the splitting had been performed in another way. In order to avoid unfortunate splitting, where the split halves are actually different, it is suggested to divide the data into four parts. The four parts are then combined two and two and the split-half analysis is performed on different combinations of the four parts (Harshman 1984). Split-half analysis is applied in Paper II to validate fluorescence EEMs modelled with PARAFAC and in Paper III and IV to estimate the correct number of components to use in slicing models of low-field NMR relaxation data.

4.1.6. Distribution of residuals

Residuals should ideally be randomly distributed, meaning that the remaining unexplained variations would only be caused by random noise. A systematic distribution of the residuals is an indication that the data contains systematic variation not included in the model and more components have to be used. Therefore, the residuals should be investigated in order to find any systematic variation.

In addition, sample specific residuals may be used to identify outliers, trends, etc. The identification of outliers by investigating the distribution of sample residuals is illustrated in Paper II, which also shows that the sum of squared residuals of each sample can be used as a measure of outlier identification.

5. Sampling

Instrumental measurements are of importance in industry, for example, for quality control or in quality sorting in production. The instrumental measurement is meant to provide information about the whole object, even though it only measures at a localised area. In these cases large errors could arise if sampling is not considered sufficiently, since the expectation of the measurement taken will not necessarily be representative of the average of the whole object.

Sampling involves selecting a part of a population for estimating a certain parameter of the whole population. Measuring only a part of a population will always introduce an error, as the part of the population measured will not be exactly the same as the rest of the population. This causes a sampling error, which is defined as the difference between the measurement taken and the measurement that would be obtained if the whole object was measured (Cochran 1977). Such errors decrease the applicability of the results, and it is important to find out whether they are significant.

An important factor in the optimisation of sampling is that the sample should be as representative as possible of other and future objects from the same population. Gy (1998) defines a representative sample as a sample that is accurate, reproducible and precise. Besides reducing the sampling error by choosing a representative sample, the number of samples needed to obtain a reliable estimate is important. It is necessary to know the minimum number of samples that should be measured in order to reduce the effect of sampling error, to be able to detect significant variations between different groups and to avoid unnecessary measurements.

This chapter considers errors in instrumental measurements when these are used to predict some parameter by univariate or multivariate regression. Section 5.1 describes the errors influencing the predictions. In Section 5.2 it is illustrated how sample heterogeneity contributes to the error in the instrumental measurements. Section 5.3 shows how errors can be quantified and Section 5.4 contains an example of how errors in instrumental measurements can be handled.

5.1. Errors influencing predictions in multivariate regression

Various types of errors influence the predictions when relating multivariate instrumental measurements to a certain quality parameter, for example, by PLS or PCR. The decomposition of the errors contributing to the total prediction error of a measured sample containing errors in the measurement of the reference value (y-error) is shown in Figure 5.1. When considering errors in **y**, the total prediction error is denoted as the *apparent* prediction error. If the prediction error is based on true reference values there will be no y-error and it will then be denoted as an *actual* prediction error[1]. The total error is the error expected when predicting a parameter of a new object which is not a part of the model estimation.

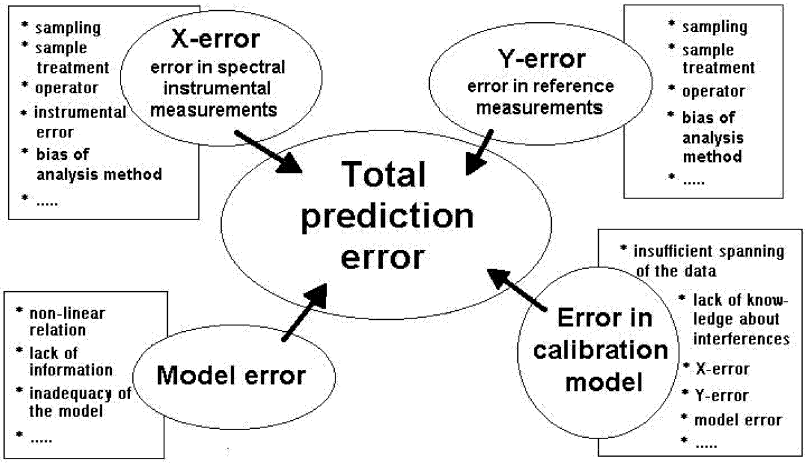


Figure 5.1 Errors influencing multivariate regression and prediction.

The sampling error in both the instrumental (x-error) and the reference measurements (y-error) may contribute significantly to the *apparent* prediction error. These errors are for example due to variation in chemical and physical composition within the same object. The instrumental noise is generally not very large due to high accuracy and reproducibility of the instruments considered here. But together with instrumental drift it may contribute somewhat to the total prediction error. Furthermore, the sample treatment introduces an error in both **X** and **y** due to weighing, extraction, pre-concentration, contamination, storage conditions etc. Effects of the operator and a bias due to the choice of a non-optimal analysis method may also be important.

[1] See section 2.2.3 for a discussion on the difference between *apparent* and *actual* prediction error.

The model error and the error in the calibration model reflect the part of the total error that arises from the model or from estimating the model parameters. The model error is the inadequacy of the model even for data without noise. It can be due to a non-linear relation between the instrumental and the reference values, a lack of information in the instrumental measurements for predicting the reference data or an inadequacy of the linear model used. The error in the calibration model is a function of the three other types of errors. Furthermore, it may be due to an insufficient spanning of the variation in data and lack of knowledge about interferences.

The variance of the individual contributions can be added, if independence among the different types of errors is assumed. Thus, it is possible to identify and combine all the individual errors and from these to estimate the total prediction error. However, the error in the calibration model is a function of the three other types of errors, which makes it more difficult to decompose the total prediction error into the individual error components.

5.2. Heterogeneity of biological materials

The heterogeneity of the object or population sampled has an important impact on the overall sampling error (Gy 1998). When the object/population is homogeneous, the samples can be taken out arbitrarily. However, for biological materials, homogeneity is almost never the case and the samples must therefore be taken out in some kind of rational systematic way. Studies have shown a large variation in chemical composition even between fish within the same batch, where a batch is a group of fish caught at the same time or reared together by one farmer (Nordtvedt et al. 1998 Refsgaard et al. 1998; Isaksson 1995).

The heterogeneity of instrumental measurements within a cod fillet is illustrated in Figure 5.2. Low-field NMR relaxations are measured on samples taken at every 1.5 cm^2 and with the depth corresponding to the thickness of the fillet. Data are analysed with multivariate image analysis (Geladi et al. 1989; Esbensen and Geladi 1989) by which a PCA on all NMR relaxations measured is made and the score values of PC1 plotted on a figure of the fillet.

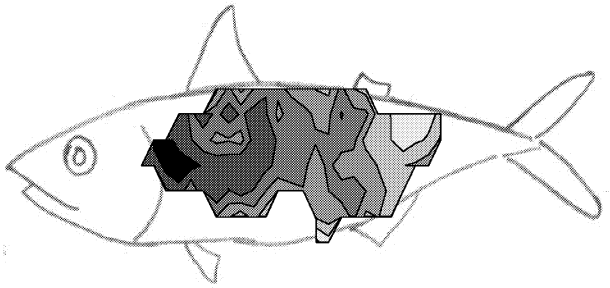


Figure 5.2 Visualisation of the variation in low-field NMR measurements within cod fillets. A PCA is made on low-field NMR relaxations and the score values of the first PC plotted. Different grey-scale levels indicate the different levels of the score values.

The low-field NMR relaxations show a systematic variation between the head and the tail part, corresponding to the distribution of water and protein (Paper III; Dambergs 1963). Thus, there is a variation in the instrumental measurements, depending on which part of the fillet that is measured. The variation within one fish can make the experiments quite complex because of the variation in muscle fibre structure, connective tissue, chemical composition, etc. Therefore, the heterogeneity of fish is important when performing instrumental measurements on a localised area only. This has been demonstrated for low-field NMR in Paper III and in studies using near infrared spectroscopy on a number of different fish species (Sigernes et al. 1998; Nordtvedt et al 1998; Downey 1992; Isaksson 1995). However, none of these studies go into detail with optimisation of sampling in order to obtain optimal predictions. The sampling of instrumental texture measurements made on raw salmon fillets is discussed in Paper I and has been investigated by Sigurgisladottir et al. (1999) and Isaksson et al. (2002), illustrating that the texture varies depending on where on the fillet measurements are taken.

5.3. Quantification of errors

Errors (e.g. sampling) in instrumental variables (**X**) may be important for multivariate regression and prediction. Knowledge about the importance of these errors can reveal how errors influence the prediction results and can be used for sampling optimisation. This section illustrates how errors in instrumental variables can be quantified. Prediction of water content of fish muscles from low-field NMR relaxations is used as an example in the following and in Paper VI. Paper VI also shows an example of predicting analyte concentrations from fluorescence spectroscopic measurements.

The approaches used in this and the next section apply for random errors. As shown in Figure 5.2 the variation in instrumental measurements within a fish seems to be systematic. However, if one measurement is taken at a random position, the error arising from the specific location can be considered random. Furthermore, the approaches assume that no bias is present, since such errors will not be identified, even when they are of a considerable size.

An approach suggested by Faber and Kowalski (1996) is applied for quantification of errors influencing the predictions (see Section 2.2.3 for a thorough description of the approach). Low-field NMR relaxations are measured at 18 different positions within each fish. Predictions are made using six randomly chosen NMR measurements from each fish in a calibration set and six other NMR measurements in a test set. The six measurements are used directly without calculating the average. The water content (reference value) is the same for the two data sets. Thus, data from all fish are included in both the calibration set and the test set, but the x-measurements in the two data sets differ. The mean variance of prediction error expressed as

the mean of all samples is illustrated in Table 5.1. Table 5.1 also shows how much each error term contributes to the prediction error variance.[7]

Table 5.1 Mean variance of prediction error and errors contributing to the mean variance of prediction error expressed in per cent.

	Mean prediction error variance	Error in calibration model (%)	Model error (%)	Contribution of errors in **X** (%)	Contribution of errors in **y** (%)
Cod	2.1	13.3	85.3	0.02	1.4
Saithe	1.9	7.6	91.2	0.001	1.2
Whiting	8.5	8.0	91.2	0.007	0.8
Haddock	1.4	3.6	93.5	0.007	2.9

The mean variance of prediction error is of the same order of magnitude for the four fish species, but somewhat larger for whiting. For all four fish species, the model error is most important, but also the error in the calibration model and the contribution of errors in the reference measurements seem to have some influence on the predictions. The errors in **X** only make up a small portion of the prediction error variance. This contribution is somewhat larger for cod than for the three other fish species. Ideally, in PLS regression only the structural part of the data is included in the model. Noise is left in the residuals. Therefore, it seems reasonable that errors in **X** only make up a very small part of the total prediction error.

It can be assumed that when the error in **X** only constitutes a very small part of the prediction error variance, it will not have a significant influence on the predictions. However, as shown in Figure 5.3, the contribution of errors in **X** is important for cod but not for haddock. Improved predictions will be obtained when using more replicates, if errors are important. For cod, lower RMSEP is obtained when the number of replicates increases. This illustrates that errors in **X** are important. For haddock, RMSEP does not change by increasing the number of replicates probably because errors in **X** are not important. The number of replicates has to be considered. Ideally, it should be as small as possible (to reduce the amount of experimental work) and chosen as the number at which the error estimate (here RMSEP) no longer decreases significantly by increasing the number of replicates.

[7] Mean prediction error variance and contribution of each error term is estimated by Equation 2.11.

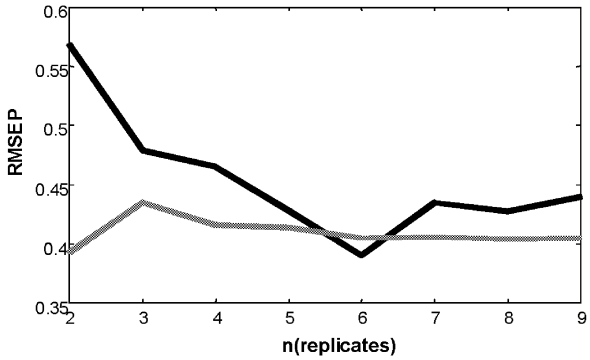


Figure 5.3 RMSEP as a function of the number of replicates in both calibration and validation for cod (black line) and haddock (grey line).

5.4. Handling errors in instrumental variables

Knowledge of the contribution of instrumental error to the total prediction error makes it possible to optimise sampling and calibration. For example, if the random spectral error is large, the total prediction error can, as mentioned earlier, be reduced using replicates. Least squares regression gives optimal predictions, when the assumptions underlying least squares estimation are met. If not, it is important to consider if least squares regression still gives adequate predictions or if one should find another method or modify the least squares expression. This section shows an example of such a modification, which is based on the measurement error theory and applies when the number of replicates in calibration and validation differs.

5.4.1. Correction of the regression coefficient in case of different numbers of replicates

The error in the x-variables varies depending on the number of replicates due to an averaging effect. If the validation set is based on an average of another number of replicates than the calibration set, the predictions obtained directly by least squares estimation may no longer be optimal. This may be the case, even though all other conditions required for least squares estimation to be optimal are fulfilled. In Paper V a correction of the regression coefficient is proposed based on the measurement error theory. It is expressed in Equation 5.1 and applies for univariate data[8]. The approach is an attempt to improve the prediction errors by obtaining a least squares estimate corresponding to the number of replicates in the validation set. This is in contrast to the general purpose of applying measurement error models where non-biased parameter estimates are the primary aim.

[8] The derivation of Equation 5.1 based on least squares regression and the measurement error theory are described in paper 5.

$$b_{\text{lsval}} = b_{\text{lscal}} \frac{s_x^2 + s_{\Delta x,\text{cal}}^2}{s_x^2 + s_{\Delta x,\text{val}}^2} \tag{5.1}$$

In Equation 5.1, b_{lscal} denotes the least squares estimated regression coefficient for the calibration set. b_{lsval} is a calculated regression coefficient that is equal to the regression coefficient of the calibration set but corresponds to the number of replicates in the validation set. Furthermore, s_x^2, $s_{\Delta x,\text{val}}^2$ and $s_{\Delta x,\text{cal}}^2$ are the variances of the true x-values and of the x-error in the validation and the calibration set, respectively. When different numbers of replicates are used in the validation and the calibration set, the variances of the two error estimates will be different and the corrected regression coefficient will differ form the least squares estimate.

The application of Equation 5.1 is illustrated in the following and in Papers V and VI. Water content of fish muscles is predicted from low-field NMR relaxations after which RMSEP is used to evaluate the results.

5.4.2. Application of the correction for univariate and multivariate data

The application of Equation 5.1 requires univariate data. Therefore, the applicability of the equation is illustrated choosing one variable on the NMR relaxation curve. Figure 5.4 shows the prediction error estimates obtained for the four fish species. The left plot shows RMSEP obtained from the least squares estimated regression coefficient as a function of the number of replicates in the validation set. The middle plot shows the same, but RMSEP is obtained using the corrected regression coefficient. These plots show how RMSEP changes with an increase in the number of replicates. It was expected that the prediction errors would decrease by increasing the number of replicates due to more accurate measurements. However, this is only clearly seen for the whiting. For saithe, RMSEP even increases with an increasing number of replicates in the validation set. Probably, this is due to other factors than the sampling error such as an increased variability caused by the uncertainty in the estimated correction factors.

The right plot in Figure 5.4 shows the relative difference in RMSEP obtained by least squares regression and when correcting for the number of measurements in the validation set. A positive value indicates that the correction performs best. The difference is largest when only one measurement is used for prediction. When the validation set is made of up to a few measurements, there seems to be an advantage of correcting the regression coefficients. However, this advantage diminishes when more than approximately three measurements are used.

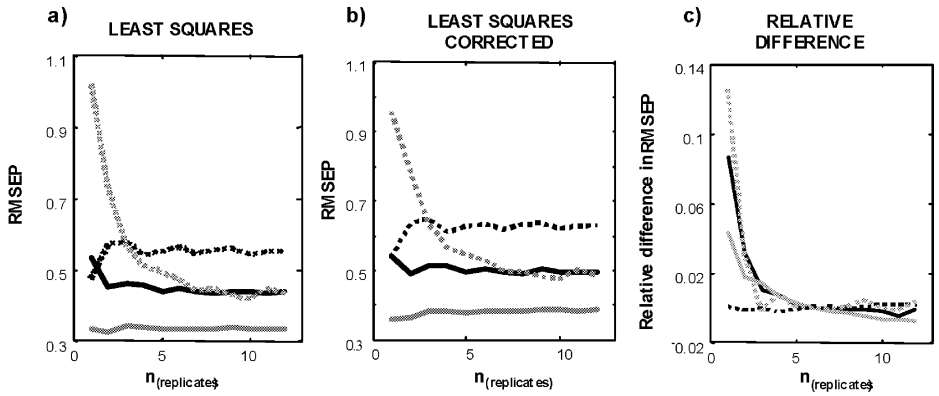


Figure 5.4 Prediction of water content from ^{1}H low-field NMR relaxations. One variable from the NMR relaxation curve is used as x-variable. Models are validated with full cross validation and include the average of six replicates in the calibration set. a) RMSEP as a function of the number of replicates in the validation set where the regression coefficient is estimated by least squares regression, b) RMSEP as a function of the number of replicates in the validation set where the regression coefficient is estimated by correcting the least squares regression and c) relative difference in RMSEP expressed as RMSEP obtained by least squares regression minus the RMSEP obtained from Equation 5.1 relative to the least squares estimate. The lines indicate: cod (solid black), haddock (solid grey), saithe (dotted black) and whiting (dotted grey).

Paper VI shows how Equation 5.1 can be applied on multivariate data using the scalar net analyte signals as univariate independent variables. Figure 5.5 shows RMSEP as a function of the number of replicates in the validation set using the norm of the PLS regression vector as regression coefficient (a) and the norm of the PLS regression vector corrected according to Equation 5.1 (b). The right plot (c) shows the relative difference in RMSEP obtained by the two methods.

Figure 5.5 shows that for cod and haddock the correction gives better predictions when very few replicates are used in the validation set. Saithe does not show a clear difference between the RMSEP obtained by the two methods probably because the errors in **X** are not important for the prediction of the water content. With many replicates in the validation set neither cod nor haddock show an effect of applying the correction. This is due to non-important errors in the NMR measurements for these situations. By increasing the number of replicates, errors in the instrumental measurements become less important and there is no effect of correcting for such errors.

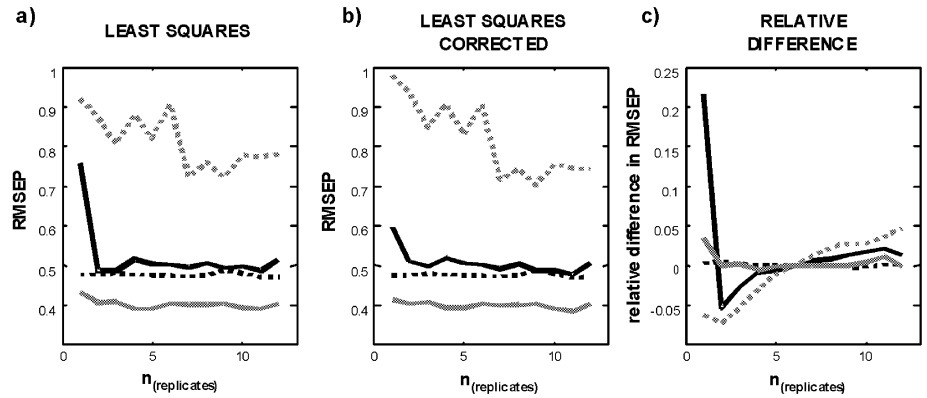


Figure 5.5 Prediction of water content from ^{1}H low-field NMR relaxations. Scalar net analyte signals are used as x-variable. Models are validated with full cross validation and include the average of six replicates in the calibration set. a) RMSEP as a function of the number of replicates in the validation set where the norm of the PLS-regression coefficient vector is used, b) RMSEP as a function of the number of replicates in the validation set where the correction is applied on the norm of the PLS-regression coefficient vector and c) relative difference in RMSEP expressed as the difference in RMSEP relative to RMSEP obtained using the norm of the regression coefficient vector. The lines indicate: cod (solid black), haddock (solid grey), saithe (dotted black) and whiting (dotted grey).

The results obtained for whiting seem more peculiar, in that the application of the correction gives poorer predictions with few replicates in the test set and better predictions with many replicates. These results may be due to the larger RMSEP obtained for whiting, which indicates that these models do not perform well. This can also be seen from Table 5.1 where the prediction error variance is larger for whiting than for the other fish species.

As shown above, there is a difference between the results obtained for real univariate data taking out one x-variable and when using the scalar net analyte signal as x-variable. The variable taken out to make the univariate data may not contain all information in the NMR relaxation profiles that are related to the water content. Furthermore, the univariate data may be influenced by interferents, which are removed when calculating the net analyte signal (see appendix 2). These factors entail that the two types of x-variables will give different models and thereby different predictions.

The above shows that the correction is useful when applied on real data. If errors in the instrumental variables are important, the regression coefficient should be corrected, so it corresponds to the number of measurements in the validation set. This is typically the situation with very few replicates. However, the application of the correction requires further studies including other types of data. Such studies will lead to a thorough understanding of the application of the correction on multivariate data and may illustrate where the advantages of the method are most pronounced.

6. Conclusion and perspectives

This project has elaborated on the application of known and new chemometric methods to decompose, interpret and optimise spectroscopic measurements. Examples of using fluorescence spectroscopy and NMR on fish muscles are illustrated. The focus has been on three areas:

1) Exploratory analysis of complex spectroscopic measurements
2) Validation of models
3) Quantifying and handling sampling errors of instrumental measurements

6.1. Exploratory analysis of complex spectroscopic data

Applying fast instrumental methods, such as fluorescence spectroscopy and NMR, has in this Ph.D. project been found to give informative and reliable data on several aspects concerning fish quality. When performing non-destructive measurements, the data obtained may be rather complex due to the large complexity of the food materials. This complexity is illustrated for fluorescence emission of cod and salmon.

The fluorescence spectra are shown to be influenced by factors such as connective tissue, NADH, oxidation products, lipids and pigments. This possibility of identifying various fluorophores present in the fish muscle encourages further studies on the relationship between the development or disappearance of certain fluorophores and the storage conditions or some quality parameter. One aim would be to develop on-line techniques for quality control. For example, the detection of NADH and oxidation products illustrates the possibility of using fluorescence for determination of chill storage time of cod. However, more experiments are needed before the potential of fluorescence spectroscopy for industrial determination of chill storage time can be fully elucidated. Such experiments may include fish caught at different seasons or places and fish treated with other storage conditions.

The applicability of fluorescence to measure the degradation of connective tissue has been investigated by PCAs and by relating the spectra to texture and degree of gaping. Rather poor predictions were obtained probably due to non-optimal sampling and small sample variation. However, fluorescence spectroscopy may still be applicable as a fast non-destructive method for quantification of collagen and may also be of importance within research investigating the influence of collagen on fish quality. Other instrumentation or different sample preparation should be tested in order to obtain a fundamental understanding of the fluorescence properties of connective tissue. This may induce further knowledge of the potential for on-line

implementation of fluorescence spectroscopy for measuring quality parameters related to the connective tissue.

Fluorescence EEMs were obtained for fish muscle extracts and decomposed by PARAFAC, illustrating how underlying spectral and chemical information can be extracted. PARAFAC enables identification of pure excitation and emission spectra of the individual fluorophores. However, modelling may be troublesome and also time consuming. For example, missing values and scatter can cause problems which can be handled, for example, by constraining the loadings with non-negativity and unimodality.

An initial investigation of the fluorophore composition in fish muscles is shown. However, further experiments involving other types of extracts or whole fish muscle are needed to fully understand the underlying chemical properties of the fish muscles.

The steps and problems involved in PARAFAC modelling are illustrated showing how fluorescence spectroscopic data can be decomposed and interpreted and illustrates the potential of three-way modelling of fluorescence excitation-emission data for resolving the underlying chemical structure in the measured spectra. This may give an incentive to apply PARAFAC modelling to fluorescence EEMs of other products.

Water content and the mobility of water are important for fish products, as they influence quality parameters such as texture and WHC. Thus, knowledge of the distribution of water can be important for product optimisation. Low-field NMR has been applied to study how water is distributed in gadoid fish muscles. There is a variation in the total amount of water within one fillet as well as in the content of the individual water pools. This variation likely relates to the variation in cell and muscle fibre size and illustrates variations in the measurements, depending on which part of the fillet that is used for the analysis. Further studies could be of interest investigating the distribution and state of water within the fish muscle. This may contribute to a more detailed knowledge of the composition of the muscle structure and the differences within one fish and between different fish species.

WHC of frozen-thawed fish muscle has been predicted from low-field NMR relaxations. Low error estimates and high correlations between the measured and the predicted WHC were obtained, showing that NMR is a valid method for predicting WHC. Furthermore, the relationship between WHC and the concentration of the individual water pools estimated from the NMR relaxations encourages further studies that can illustrate how processing and storage influence muscle structure and WHC. In the longer run, such studies can be used to optimise storage and processing.

6.2. Validation of models

Validation is an important aspect of data analysis. It is necessary to ensure valid and reliable modelling and can be seen as an assessment of the appropriateness of the developed models. In this project, experiments have not been made with the purpose of investigating validation. Rather, validation has been an integrated part of the data analysis and has been included in all the experimental work.

Various validation methods have been presented and applied in this thesis and in the included papers. The thesis thereby illustrates how the choice of validation method depends on the actual situation such as the type of data, models applied and purpose of the data analysis. For example, it is shown that test set or cross validation should be used to evaluate the prediction performance and can, together with split-half analysis, be used to estimate the number of components in chemometric models. Jack-knifing and bootstrapping estimate the uncertainty of the model parameters and are suitable for variable selection, outlier detection and evaluation of the model reliability.

6.3. Quantifying and handling sampling errors of instrumental measurements

The homogeneity of instrumental measurements within cod fillets has been investigated illustrating that the measurements depend on which part of the fillet that is used for the analysis. Therefore, sampling may be important when measurements are performed on a part of the fillet and is necessary to consider in further experimental work.

Knowledge about the influence of errors (e.g. sampling) on predictions is important. The quantification of errors can be done by an approach which gives the prediction error variance based on factors such as the contribution of errors in the instrumental measurements, errors in the reference measurements, error in calibration model and model error. By multivariate regression, the largest part of the noise is left unmodelled in the residuals. Therefore, the contribution of errors in the instrumental measurements typically only makes up a very small part of the total prediction error. However, these errors may still be important. The approach is a useful method for comparing the size of the errors influencing the predictions in multivariate regressions and can be a valid method for optimising sampling and experimental conditions. This should be investigated further using other types of data.

Least squares estimation gives optimal predictions when residuals are homoscedastic and there is no model error. This is the case even with large errors in the instrumental measurements. However, applying a correction of the regression coefficient for univariate data improves the predictions when the instrumental error is considerable and the number of replicates used for

calibration and validation differs. This correction is also applicable for multivariate data using the scalar net analyte signal instead of the multivariate signals. However, the results are not unambiguous and may be influenced by various types of errors. Therefore, in order to fully understand when and how the correction can be applied more studies are required involving other types of samples and other instrumental measurements.

7. References

Aubourg, S.P. and Medina, I. Influence of storage time and temperature on lipid deterioration during cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) frozen storage. *Journal of the Science of Food and Agriculture*, **79**, 1943-1948 (1999).

Aubourg, S.P., Medina, I. and Gallardo, J.M. Quality assessment of blue whiting (*Micromesistus poutassou*) during chilled storage by monitoring lipid damages. *Journal of Agricultural and Food Chemistry*, **46**, 3662-3666 (1998a).

Aubourg, S.P., Soleto, C.G. and Perez-Martin, R. Assessment of quality changes in frozen sardine (*Sardine pilchardus*) by fluorescence detection. *Journal of the American Oil Chemists Society*, **75**, 575-580 (1998b).

Bailey, A.J., Sims, T.J., Avery, N.C. and Halligan, E.P. Nonenzymatic glycation of fibrous collagen – reaction-products of glucose and ribose. *Biochemical Journal*, **305**, 385-390 (1995).

Baunsgaard, D., Munck, L. and Nørgaard, L. Evaluation of the quality of solid sugar samples by fluorescence spectroscopy and chemometrics. *Applied Spectroscopy*, **54**, 438-444 (2000a).

Baunsgaard, D., Munck, L. and Nørgaard, L. Analysis of the effect of crystal size and color distribution on fluorescence measurements of solid sugar using chemometrics. *Applied Spectroscopy*, **54**, 1684-1689 (2000b).

Baunsgaard, D., Andersson, C.A., Arndal, A. and Munck, L. 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 (2000c).

Bechmann, I.E., Pedersen, H.T., Nørgaard, L. and Engelsen, S.B. Comparative chemometric analysis of transverse low-field NMR relaxation data. In: Advantages in Magnetic Resonance in Food Science (Belton, P.S., Hills, B.P. and Webb, G.A., eds), pp 217-225, The Royal Society of Chemistry, Cambridge, 1999.

Berendsen, H.J.C. Rationale for using NMR to study water relations in foods and biological tissues. *Trends in Food Science & Techology*, **31**, 202-205 (1992).

Bertram, H.C., Dønstrup, S., Karlsson, A.H. and Andersen, H.J. Continuous distribution analysis of T_2 relaxation in meat – an approach in the determination of water holding capacity. *Meat Science*, **60**, 279-285 (2002a).

Bertram, H.C., Purslow, P.P. and Andersen, H.J. Relationship between meat structure, water mobility and distribution: A low-field nuclear magnetic resonance study. *Journal of Agricultural and Food Chemistry*, **50**, 824-829 (2002b).

Bertram, H.C., Karlsson, A.H., Rasmussen, M., Dønstrup, S., Petersen, O.D. and Andersen, H.J. The origin of T_2 relaxation in muscle myowater. *Journal of Agricultural and Food Chemistry*, **49**, 3092-3100 (2001a).

Bertram, H.C., Karlsson, A.H. and Andersen, H.J. Comparative study of low-field NMR relaxation measurements and two traditional methods in the determination of water holding capacity of pork. *Meat Science*, **57**, 125-132 (2001b).

Botlan, D.L., Rugraff, Y., Martin, C. and Colonna, P. Quantitative determination of bound water in wheat starch by time domain NMR spectroscopy. *Carbohydrate Research*, **308**, 29-36 (1998).

Britton, G. UV/Visible spectroscopy. In: UV/Visible spectroscopy in carotenoids. (Britton, G., Liaaen-Jensen, S. and Pfander, H., eds), chapter 2, Birkhäuser, Basel, 1995.

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

Bro, R. Multi-way analysis in the food industry. Models, algorithms and applications. Doctoral thesis, 1998.

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

Bro, R. Multiway calibration. Multilinear PLS. *Journal of Chemometrics*, **10**, 47-61 (1996).

Bro, R., van den Berg, F., Thybo, A., Andersen, C.M., Jørgensen, B. and Andersen, H. Multivariate data analysis as a tool in advanced quality monitoring in the food production chain. *Trends in Food Science & Technology*, **13**, 235-244 (2002a) – appendix 3.

Bro, R., Sidiropoulos, N.D. and Smilde, A.K. Maximum likelihood fitting using ordinary least squares algorithms. *Journal of Chemometrics*, **16**, 387-400 (2002b).

Bro, R. and Andersen, C.M. Theory of net analyte signal vectors in inverse regression. submitted (2002) – appendix 2.

Brown, R.J.S., Capozzi, F., Cavani, C., Cremonini, M.A., Petracci, M. and Placucci, G. Relationships between ^{1}H NMR relaxation data and some technological parameters of meat: A chemometric approach. *Journal of Magnetic Resonance*, **147**, 89-94 (2000).

Brøndum, J., Munck, L., Henckel, P., Karlsson A. ,Tornberg, E. and Engelsen, S.B. Prediction of water holding capacity and composition of porcine meat with comparative spectroscopy. *Meat Science*, **55**, 177-185 (2000).

Bøknæs, N., Jensen, K.N., Andersen, C.M. and Martens. H. Freshness assessment of thawed and chilled cod fillets packed in modified atmosphere using near infrared spectroscopy. *Lebensmittel Wissenschaft und Technologie*, **35**, 628-634 (2002) – appendix 1.

Carr, H.Y. and Purcell, E.M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, **94**, 630-638 (1954).

Cochran, W.G. Sampling techniques. John Wiley and Sons, New York, 1977

Cochran, W.G. Errors of measurements in statistics. *Technometrics*, **10**, 637-666 (1968).

Colquhoun, I.J. NMR spectroscopy in food science. *Nutrition & Food Science*, **1**, 8-12 (1993).

Dahl, P.L., Christensen, B.M., Munck, L., Larsen, E.P. and Engelsen, S.B. Can spectroscopy in combination with chemometrics replace mink digestibility tests. *Journal of the Science of Food and Agriculture*, **80**, 365-374 (2000).

Dambergs, N. Extractives of fish muscle. 3. Amounts, sectional distribution and variances of fat, water-solubles, protein and moisture in cod (*Gadus morhua* L.) fillets. *Journal of the Fisheries Research Board of Canada*, **20**, 909-918 (1963).

Davis, H.K. Fluorescence of fish muscle: Description and measurement of changes occurring during frozen storage. *Journal of the Science of Food and Agriculture*, **33**, 1135-1142 (1982).

Deleanu, C. and Paré, J.R.J. Nuclear magnetic resonance spectroscopy (NMR): principles and applications. In: Instrumental methods in food analysis (Paré, J.R.J. and Bélanger, J.M.R., eds), pp 179-236, Elsevier Science, Amsterdam, 1997.

DiFoggio, R. Guidelines for applying chemometrics to spectra: feasibility and error propagation. *Applied* Spectroscopy, **54**, 94A-113A (2000).

Downey, G. Non-invasive and non-destructive percutaneous analysis of farmed salmon flesh by near infra-red spectroscopy. *Food Chemistry*, **55**, 305-311 (1992).

Efron, B. and Tibshirani, R. An introduction to the bootstrap. Chapman and Hall, New York, 1993.

Egelandsdal, B., Wold, J.P., Sponnich, A., Neegård, S. and Hildrum, K.I. On the attempts to measure the tenderness of *Longissimus Dorsi* muscles using fluorescence emission spectra. *Meat Science*, **60**, 187-202 (2002).

Egelandsdal, B., Kvaal, K. and Isaksson, T. Autofluorescence spectra as related to tensile properties from bovine masseter. *Journal of Food Science*, **61**, 342-347 (1996).

Esbensen, K. and Geladi, P. Strategy of multivariate image analysis (MIA). *Chemometrics and Intelligent Laboratory Systems*, **7**, 67-86 (1989).

Faber, N.M. Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Analytical Chemistry*, **70**, 5108-5110 (1998).

Faber, N.M., Duewer, D.L., Choquette, S.J., Green, T.L. and Chesler, S.N. Characterizing the uncertainty in near-infrared spectroscopic prediction of mixed-oxygenate concentrations in gasoline: sample-specific prediction intervals. *Analytical Chemistry*, **70**, 2972-2982 (1998).

Faber, K. and Kowalski, B.R. Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares. *Journal of Chemometrics*, **11**, 181-238 (1997).

Faber, K. and Kowalski, B.R. Prediction error in least squares regression: further critique on the deviation used in The Unscrambler. *Chemometrics and Intelligent Laboratory Systems*, **34**, 283-292 (1996).

Fernández Pierna, J.A., Jin, L., Wahl, F., Faber, N.M. and Massart, D.L. Estimation of partial least squares regression prediction uncertainty when the reference values carry a sizeable measurement error. *Chemometrics and Intelligent Laboratory Systems*, **65**, 281-291 (2003).

Ferré, J., Brown, S.D. and Rius, F.X. Improved calculation of the net analyte signal in inverse multivariate calibration. *Journal of Chemometrics*, **15**, 537-553 (2001).

Fjelkner-Modig, S. and Tornberg, E. Water distribution in porcine *M. longissimus dorsi* in relation to sensory properties. *Meat Science*, **17**, 213-231 (1986).

Fuller, W.A. Estimation in the presence of measurement error. *International Statistical Review*, **63**, 121-147 (1995).

Fuller, W.A. Measurement error models. John Wiley, New York, 1987

Geladi, P. Analysis of multi-way (multi-mode) data. *Chemometrics and Intelligent Laboratory Systems*, **7**, 11-30 (1989).

Geladi, P., Isaksson, H., Lindqvist, L., Wold, S. and Esbensen, K. Principal component analysis of multivariate images. *Chemometrics and Intelligent Laboratory Systems*, **5**, 209-220 (1989).

Gerbanowski, A., Rutledge, D.N., Feinberg, M.H. and Ducauze, C.J. Multivariate regression applied to time domain nuclear magnetic resonance signals: Determination of moisture in meat products. *Sciences des Aliments*, **17**, 309-323 (1997).

Goicoechea, H.C. and Olivieri, A.C. A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemometrics and Intelligent Laboratory Systems*, **56**, 73-81 (2001).

Goicoechea, H.C. and Olivieri, A.C. Enhanced Synchronous spectrofluorometric determination of tetracycline in blood serum by chemometric analysis. Comparison of partial least-squares and hybrid linear analysis calibrations. *Analytical Chemistry*, **71**, 4361-4368 (1999).

Gy, P. Sampling for analytical purposes. John Wiley and Sons, England,1998.

Harshman, R. “how can I know if it’s real?” A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling. In: Resarch methods for multimode data analysis (Law, H.G., Snyder, C.W., Hattie J.A. and McDonald R.P., eds), pp 566-591, Praeger, New York, 1984.

Harshman, R.A. Foundations of the PARAFAC procedure: model and conditions for an “explanatory” multi-mode factor analysis. *UCLA Working Papers in phonetics*, **16**, 1-84 (1970).

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

Harshman, R.A. and Lundy, M.E. The PARAFAC model for three-way factor analysis and multidimensional scaling. In: Resarch methods for multimode data analysis (Law, H.G., Snyder, C.W., Hattie J.A. and McDonald R.P., eds), pp 122-215, Praeger, New York, 1984.

Harshman, R.A. and de Sarbo, W.S. An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In: Resarch methods for multimode data analysis (Law, H.G., Snyder, C.W., Hattie J.A. and McDonald R.P., eds), pp 602-643, Praeger, New York, 1984.

Hasegawa, K., Endo, Y. and Fujimoto, K. Oxidative deterioration in dried fish model systems assessed by solid sample fluorescence spectrophotometry. *Journal of Food Science*, **57**, 1123-1126 (1992).

Hassan, F., Goerge, F., Mukundan, M.K. and Sherief, P.M. Role of collagen in gaping of fish fillets. *Fisheries Technologie,* **36**, 40-42 (1999).

Hemminga, M.A. Introduction to NMR. *Trends in Food Science & Technology*, **3**, 179-186 (1992).

Hernandez, C.V. and Rutledge, D.N. Characterization of cocoa masses: low resolution pulse NMR study of the effect of geographical origin and roasting on fluidification. *Food Chemistry*, **49**, 83-93 (1994).

Hills, B.P. and Remigereau, B. NMR studies of changes in subcellular water compartmentation in parenchyma apple tissue during drying and freezing. *International Journal of Food Science and Technology*, **32**, 51-61 (1997).

Isaksson, T., Swensen, L.P., Taylor, R.G., Fjæra, S.O. and Skjervold, P.O. Non-destructive texture analysis of farmed Atlantic salmon using visual/near-infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, **82**, 53-60 (2002).

Isaksson, T., Tøgersen, G., Iversen, A. and Hildrum, K.I. Non-destructive determination of fat, moisture and protein in salmon fillets by use of near-infrared diffuse spectroscopy. *Journal of the Science of Food and Agriculture*, **69**, 95-100 (1995).

Jensen, K.N., Guldager, H.S. and Jørgensen, B.M. Three-way modelling of NMR relaxation profiles from thawed cod muscle. *Journal of Aquatic Food Product Technology*, **11**, 201-214 (2002).

Jennrich, R.I. Orthogonal rotational algorithms. *Psychometrika*, **35**, 229-235 (1970).

Jepsen, S.M., Pedersen, H.T. and Engelsen, S.B. Application of chemometrics to low-field ^{1}H NMR relaxation data of intact fish flesh. *Journal of the Science of Food and Agriculture*, **79**, 1793-1802 (1999).

Jiji, R.D., Andersson, G.G. and Booksh, K.S. Application of PARAFAC for calibration with excitation-emission matrix fluorescence spectra of three classes of environmental pollutants. *Journal of Chemometrics*, **14**, 171-185 (2000).

Jiji, R.D. and Booksh, K.S. Mitigation of Rayleigh and Raman spectral interferences in multi-way calibration of excitation-emission matrix fluorescence spectra. *Analytical Chemistry*, **72**, 718-725 (2000).

Jiji, R.D., Cooper, G.A. and Booksh, K.N. Excitation-emission matrix fluorescence based determination of carbamate pesticides and polycyclic aromatic hydrocarbons. *Analytica Chimica Acta*, **397**, 61-72 (1999).

Kroonenberg, P.M. Three-mode component models. A survey of the literature. *Statistica Applicata*, **4**, 619-633 (1992).

Lackowicz, J.R. Principles of fluorescence spectroscopy. 2nd edition, Kluwer Academic/Plenum Publishers, New York, 1999.

Lambelet, P., Renevey, F., Kaabi, C. and Raemy, A. Low-field nuclear magnetic resonance relaxation study of stored or processed cod. *Journal of Agricultural and Food Chemistry*, **43**, 1462-1466 (1995).

Leurgans, S. and Ross, R.T. Multilinear models: Applications in spectroscopy. *Statistical Science*, **7**, 289-319 (1992).

Lee, C.-H., Kim, K. and Ross, R.T. Trilinear analysis for the resolution of overlapping fluorescence spectra. *Korean Biochemical Journal*, **24**, 374-379 (1991).

Lillford, P.J., Jones, D.V. and Rodger, G.W. Water in fish. Water in fish tissue – a proton relaxation study of *post rigor* minced cod. In: Advances in fish science and technology (Connell, J.J. and staff of Torry Research, eds), pp 495-497, Fishing News Books Ltd., England, 1980.

Lorber, A. Error propagation and figures of merit for quantification by solving matrix equations. *Analytical Chemistry*, **58**, 1167-1172 (1986).

Lorber, A., Faber, K. and Kowalski, B.R. Net analyte signal correction in multivariate calibration. *Analytical Chemistry*, **69**, 1620-1626 (1997).

Love, R.M. The food fishes; their intrinsic variation and practical implications. Farrand Press, London, 1988.

Mackie, I.M. The effects of freezing of flesh proteins. *Food Reviews International*, **9**, 575-610 (1993).

Manohar, S.V. Characteristics of white muscle fluorescence in pre-rigor fish. In: Fish inspection and quality control (Kreuzer, R., eds), pp 211-215, Fishing News Books Ltd., England, 1971.

Martens, H. and Martens, M. Modified jack-knife estimation of parameter uncertainty in bilinear modeling by partial least squares regression. *Food Quality and Preference*, **11**, 5-16 (2000).

Martens, H. and Næs, T. Multivariate calibration. John Wiley & Sons, England, 1989.

Meiboom, S. and Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *The Review of Scientific Instruments*, **29**, 688-691 (1958).

Mizuta, S., Hwang, J. and Yoshinaka, R. Molecular species of collagen from wing muscle of scate (*Raja kenojei*). *Food Chemistry*, **76**, 53-58 (2002).

Moberg, L., Robertsson, G. and Karlberg, B. Spectrofluorimetric determination of chlorophylls and pheopigments using parallel factor analysis. *Talanta*, **54**, 161-170 (2001).

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

Munck, L. Fluorescence analysis in foods. Longman Singapore Publishers Ltd., Singapore, 1989.

Nikolajsen, R., Hansen, Å.M. and Bro, R. Attempt to separate the fluorescence spectra of adrenaline and noradrenaline using chemometrics. *Luminescence*, **16**, 91-101 (2001).

Nordtvedt, R., Torissen, O.J. and Tuene, S. Application of near-infrared transmittance spectroscopy in the determination of fat, protein and dry matter in Atlantic halibut fillet. *Chemometrics and Intelligent Laboratory Systems*, **42**, 199-207 (1998).

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

Pedersen, H.T. Low-field nuclear magnetic resonance and chemometrics applied in food science. Ph.D. Dissertation, 2001.

Pedersen, D.K., Munck, L. and Engelsen, S.B. Screening for dioxin contamination in fish oil by PARAFAC and N-PLSR analysis of fluorescence landscapes. *Journal of Chemometrics*, **16**, 451-460 (2002).

Pedersen, H.T., Bro, R. and Engelsen, S.B. 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 (2002).

Pedersen, H.T., Bro, R. and Engelsen, S.B. SLICING – A novel approach for unique deconvolution of NMR relaxation decays. In: Magnetic resonance in food science: a view to the future (Webb, G.A., Belton, P.S., Gill, A.M. and Delgadillo, I., eds), pp 202-209, The Royal Society of Chemistry, Cambridge, 2001.

Pedersen, H.T., Munck, L. and Engelsen, S.B. Low-field ^{1}H nuclear magnetic resonance and chemometrics combined for simultaneous determination of water, oil, and protein contents in oilseeds. *Journal of the American Oil Chemists Society*, **77**, 1069-1076 (2000).

Refsgaard, H.H., Brockhoff, P.B. and Jensen, B. Biological variation of lipid constituents and distribution of tocopherols and astaxanthin in farmed Atlantic salmon. *Journal of Agricultural and Food Chemistry*, **46**, 808-812 (1998).

Renou, J.P., Monin, P. and Sellier, P. Nuclear magnetic resonance measurements on pork of various qualitites. *Meat Science*, **15**, 225-233 (1985).

Ripley, B.D. and Thompson, M. Regression techniques for the detection of analytical bias. *Analyst*, **112**, 377-383 (1987).

Riu, J. and Bro, R. Jack-knife for estimation of standard errors and outlier detection in parafac models. *Chemometrics and Intelligent Laboratory Systems*, **65**, 35-49 (2002).

Riu, J. and Rius, F.X. Assessing the accuracy of analytical methods using linear regression with errors in both axes. *Analytical Chemistry*, **68**, 1851-1857 (1996).

Riu, J. and Rius, F.X. Univariate regression models with errors in both axes. *Journal of Chemometrics*, **9**, 343-362 (1995).

Ross, R.T. and Leurgans, S. Component resolution using multilinear models. *Methods in Enzymology*, **246**, 679-700 (1995).

Ross, R.T., Lee, C.-H., Davis, C.M., Ezzedine, B.M. Fayyad, E.A. and Leurgans, S.E. Resolution of fluorescence spectra of plant-pigment complexes using trilinear models. *Biochimica et Biophysica Acta*, **1056**, 317-320 (1991).

Ruan, R.R. and Chen, P.L. Water in foods and biological materials. A nuclear magnetic approach. Technomic Publishing Company, Lancaster, 1998.

Rutledge, D.N. Low resolution pulse nuclear magnetic resonance in the agro-food industry. *Journal of Chemical Physics*, **89**, 273-285 (1992).

Rutledge, D.N. and Barros, A.S. Method for detecting information in signals: application to two-dimensional time domain NMR data. *Analyst*, **123**, 551-559 (1998).

Sato, K., Yoshinaka, R., Itoh, Y. and Sato, M. Molecular species of collagen in the intramuscular connective tissue of fish. *Comparative Biochemistry and Physiology B*, **92**, 87-91 (1989).

Sato, K., Yoshinaka, R., Sato, M., Itoh, Y. and Shimizu, Y. Isolation of types-I and types-V collagen from carp muscle. *Comparative Biochemistry and Physiology B*, **90**, 155-158 (1988).

Sato, K., Yoshinaka, R., Sato, M. and Shimizu, Y. Collagen content in the muscle of fishes in association with their swimming movement and meat texture. *Bulletin of the Japanese Society of the Science of Fisheries*, **52**, 1595-1600 (1986).

Schmidt, S.J. and Lai, H.-M. Use of NMR and MRI to study water relations in foods. In: Water relations in food (Levine, H. and Slade, L., eds), pp 405-452, Plenum Press, New York, 1991.

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

Shenouda, S.Y.K. Theories of protein denaturation during frozen storage of fish flesh. *Advances in Food Research*, **26**, 275-311 (1980).

Sigernes, F., Esaiassen, K., Heia, K., Wold, J.P., Eilertsen, G. and Sørensen, N.K. Assessment of fish (cod) freshness by VIS/NIR spectroscopy. In: Methods to determine the freshness of fish in research and industry (Olafsdóttir, G., Luten, J.P., Dalgaard, P., Careche, M., Verrez-Bagnis, V., Martinsdóttir, E. and Heia, K., eds), pp 369-375, International Institute of Refrigeration, Paris, 1998.

Sigurgisladottir, S., Hafsteinsson, H., Jonsson, A., Lie, Ø., Nordtvedt, R., Thomassen, M. and Torrissen, O. Textural properties of raw salmon fillets as related to sampling method. *Journal of Food Science*, **64**, 99-104 (1999).

Steen, C. and Lambelet, P. Texture changes in frozen cod mince measured by low-field nuclear magnetic resonance spectroscopy. *Journal of the Science of Food and Agriculture*, **75**, 268-272 (1997).

Swatland, H.J. Connective and adipose tissue detection by simultaneous fluorescence and reflectance measurements with an on-line meat probe. *Food Research International*, **33**, 749-757 (2000).

Swatland, H.J. Evaluation of probe designs to measure connective tissue fluorescence in carcasses. *Journal of Animal Science*, **69**, 1983-1988 (1991).

Swatland, H.J. Effect of excitation wavelength on the separation of types I and III collagen by fiber optic fluorometry. *Journal of Food Science*, **52**, 865-868 (1987a).

Swatland, H.J. Autofluorescence of adipose-tissue measured with fiber optics. *Meat Science*, **19**, 277-284 (1987b).

Thybo, A.K., Bechmann, I.E., Martens, M. and Engelsen, S.B. Prediction of sensory texture of cooked potatoes using uniaxial compression, near infrared spectroscopy and low field ^{1}H NMR spectroscopy. *Lebensmittel Wissenschaft und Technologie*, **33**, 103-111 (2000).

Tornberg, E., Wahlgren, M., Brøndum, J. and Engelsen, S.B. Pre-rigor conditions in beef under varying temperature and pH-falls studied with rigometer, NMR and NIR. *Food Chemistry*, **69**, 407-418 (2000).

Undeland, I., Ekstrand, B. and Lingert, H. Lipid oxidation in herring (*Clupea harengus*) light muscle, dark muscle, and skin, stored separately or as intact fillets. *Journal of American Oil Chemists Society*, **75**, 581-590 (1998).

Wehrens, R., Putter, H. and Buydens, L.M.C. The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, **54**, 35-52 (2000).

Wehrens, R. and van der Linden, W.E. Bootstrapping principal component regression models. *Journal of Chemometrics*, **11**, 157-171 (1997).

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

Wold, J.P., Jørgensen, K. and Lundby, F. Non-destructive measurement of light-induced oxidation in dairy products by fluorescence spectroscopy and imaging. *Journal of Dairy Science*, **85**, 1693-1704 (2002).

Wold, J.P. and Mielnik, M. Non-destructive assessment of lipid oxidation in minced poultry meat by autofluorescence spectroscopy. *Journal of Food Science*, **65**, 87-95 (2000).

Wold, J.P., Lundby, F. and Egelandsdal, B. Quantification of connective tissue (hydroxyproline) in ground beef by autofluorescence spectroscopy. *Journal of Food Science*, **64**, 377-383 (1999).

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

Xu, L. and Schechter, I. A calibration method free of optimum factor number selection for automated multivariate analysis. Experimental and theoretical study. *Analytical Chemistry*, **69**, 3722-3730 (1997).

Paper I

On the fluorescence of muscle and connective tissue from cod and salmon

Charlotte M. Andersen and Jens Petter Wold

470 J. Agric. Food Chem. 2003, 51, 470−476

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Fluorescence of Muscle and Connective Tissue from Cod and Salmon

CHARLOTTE M. ANDERSEN*,†,‡ AND JENS PETTER WOLD§

Danish Institute for Fisheries Research, Department of Seafood Reseach, Technical University of Denmark, Søltofts Plads, Building 221, DK-2800 Kgs. Lyngby, Denmark, MATFORSK, Osloveien 1, N-1438 Aas, Denmark, and Department of Dairy and Food Science, Food Technology, The Royal Veterinary and Agricultural University, Rolighedsved 30, DK-1958 Frederiksberg C, Denmark

Autofluorescence of salmon and cod muscle was measured and compared with autofluorescence of collagen type I and type V. Similarities between fluorescence of fish muscle and collagen were found in that the same peaks were obtained around 390, 430, and 480 nm. These similarities are supported by principal component analyses. Texture and gaping score were predicted from the fluorescence spectra by partial least-squares regression. However, the predictions did not perform well. Relating fluorescence to the gaping score gave a prediction error of 0.91 and a correlation of 0.43 when measuring gaping on a scale from 0 to 5. There was no relation between texture and fluorescence spectra. Fluorescence of fish muscle could be related to the storage time. However, this relation seemed not to be induced by changes in collagen.

KEYWORDS: Salmon; cod; collagen; connective tissue; fluorescence

INTRODUCTION

The collagen content of fish muscle varies considerably from species to species. Contents from approximately 3 to 20 g/kg of fish muscle have been found in a number of fish species (*1*). These values include the connective tissue of myocommata, perimysium, and endomysium. Collagen type I and type V have been identified, and it is stated that these two types of collagen make up the major part of the intramuscular connective tissue in fish muscle. Collagen type I is higher in content than collagen type V (*2*−*5*).

It has been shown that the texture of fish muscles and the degree of gaping depend on the content of connective tissue in that a higher collagen content was observed in situations with a firmer texture of the raw meat and a low collagen content was found in tender meat vulnerable to gaping (*1*, *6*). Since the connective tissue and thereby collagen are important for the overall quality of fish products, determination of the collagen content and its degradation by a fast on-line method, such as fluorescence spectroscopy, can be important for controlling and optimizing processing and marketing.

Fluorescence spectroscopy has been used to study connective tissue of ground bovine meat (*7*). The pure spectra of connective tissue was estimated, and it was shown that excitation at 332 nm gave emission with high intensities between 400 and 520 nm. Egelandsdal et al. (*8*) studied the fluorescence of connective tissue from bovine masseter muscle and found emission maxima at 380 and 465 nm after excitation at 335 nm. The sources of fluorescence from connective tissue are not fully understood at the molecular level yet, but it is supposed that different collagen cross-links and components such as pyridinoline are contributors (*9*, *10*).

It has been known for long time that fish muscle is able to exhibit autofluorescence (*11*, *12*), and it has been shown that autofluorescence changes during frozen storage (*13*). Previous experiments measuring fluorescence of fish products have primarily focused on the ability of fluorescence spectroscopy as a method to follow lipid oxidation. Increases in fluorescence of aqueous and organic extracts of cod, haddock, and sardine during freeze storage (*14*, *15*) and of blue whiting during chill storage (*16*) were correlated to some measures of lipid damage. Hasegawa et al. (*17*) used solid sample fluorescence spectrometry on a dried fish model system and found a correlation to the oxidative deterioration. Another application of fluorescence spectroscopy was the development of a method that can detect bones in fish fillets (*18*). Other sources contributing to the autofluorescence of fish muscle may be proteins, amino acids, NADH, FAD, pigments, and some nucleic acids. All these factors will make fluorescence of heterogeneous materials such as fish muscles very complex.

This paper presents an investigation of autofluorescence from fish muscle especially in relation to the occurrence of collagen type I and type V. By use of multivariate methods, the aim is to indicate the presence of these fluorophores within cod and salmon muscle. Changes in autofluorescence characteristics during storage are investigated as well as variations within the

* Corresponding author. E-mail: chm@dfu.min.dk. Fax: +45 45884774.
† Danish Institute for Fisheries Research.
‡ The Royal Veterinary and Agricultural University.
§ MATFORSK.

10.1021/jf020524d CCC: $25.00
Published on Web 12/12/2002

fillets. A second objective is to relate gaping and texture to the autofluorescence spectra of salmon. This relation is expected due to the influence of collagen on quality parameters such as texture and gaping (*1*, *6*) and the theory that collagen makes up an important part of the autofluorescence of fish muscle.

EXPERIMENTAL PROCEDURES

Materials and Sample Preparation. *Cod.* Four gutted and bled cod were obtained from Frederikstad, Norway, the day after catch. One fillet from each fish was measured the day after slaughtering. The other fillet was stored at 2 °C until the fifth day after slaughtering and then measured. Head and bones were kept on during storage. Another five fillets were bought from the fish auction hall in Oslo. The exact day of catch of these fish from the auction hall was not known. However, judged upon the odor, these fish were supposed to be older than the fish mentioned previously.

Fluorescence was measured on three samples from each cod fillet. Two samples were taken from the loin, one as close to the head as possible and the other from the middle of the fillet between the tail and the head. The third sample was taken as close to the tail as possible. This gave a total of 39 samples, 12 from storage day one, 12 from storage day five, and 15 obtained from the fish auction hall fillets. Skin was removed from the fillets before cutting out the samples. The samples were placed into specially designed sample cuvettes, which exposed a flat circular surface with a diameter of 5 cm for measurement. Samples were placed in the cuvettes such that the interior of the fillets was exposed for autofluorescence measurements. Thus, only fluorescence of the white muscle was measured. Before measurement, the samples were conditioned at 18 °C for at least 1 h.

Salmon. Salmon were raised, slaughtered, and filleted at the Institute of Aquaculture Research (Akvaforsk) in Norway. The fish were stored with ice at 2 °C for 1, 2, or 3 weeks before measuring. After 1 week of storage 60 salmon were analyzed, after 2 weeks 60 new ones were alalyzed, and after 3 weeks 30 additional salmon were analyzed. Texture and gaping were measured at Akvaforsk, after which the fillets were brought to the Norwegian Food Research Institute (Matforsk) for fluorescence measurements. These were measured 1 day after the texture and gaping score determinations. Samples of salmon were taken from the loin and were prepared as described for the cod. In addition, samples as close to the tail as possible were taken from the fish stored for 1 week.

Collagen. Fluorescence was measured on collagen type I purified from calf skin and collagen type V purified from bovine achilles tendon (Sigma Aldrich Sweden AB, Stockholm).

Autofluorescence Measurements. Fluorescence emission spectra were measured directly on the fish muscle. An optical bench system optimized for measuring rather large sample surfaces (up to diameter ≈ 6 cm) was used. The excitation light was generated by a 300 W xenon light source (Oriel 6258, Oriel Corp., Stratford, CT). The light was directed onto the samples at an angle of 45°. The spectra were collected by an imaging spectrograph (Acton SP-150, Acton Research Corp., Acton, MA) connected to a sensitive charge coupled device (CCD-camera) (Princeton TEA/CCD-512-TKBM1, Princeton Instruments Inc., Trenton, NJ). A cutoff filter was positioned in front of the spectrograph slit to suppress excitation light reflected from the samples. The samples were excitated with a 332 nm (Oricl, bandwidth 10 nm) interference filter where after the emission was measured from 360 to 600 nm. The exposure time was 4 s, and the temperature of the samples ranged from 16 to 19 °C. For all samples, two spectra were collected and the average was used for further analysis. The sample illumination was not perfectly homogeneous, so samples were rotated 90−180° between the two exposures. Spectrograph and detector were controlled by the software WinSpec Ver. 1.4.3.4 (Princeton Instruments Inc., Trenton, NJ). Spectra were digitized to a resolution of 16 bit and the length reduced from 512 to 256 points by averaging two adjacent readings, resulting in one reading for approximately every 1 nm.

Images were collected with the same system. A Nikon 102 mm photographic lens was mounted on the imaging spectrograph, the spectrograph slit was removed, and the grating was exchanged with a mirror. Spectral images were created by placing a 40 nm bandwidth interference filter with center wavelength 450 nm in front of the lens. Samples were illuminated with 332 nm light, and fluorescence images were measured. The spatial resolution of the images were 512 pixels/35 mm. Exposure time for all images was 30 s. For both fish species, a few samples from each storage period were used for collecting the autofluorescence images.

Fluorescence emission from collagen type I and type V was measured with the same optical setup as for the fish muscles.

Gaping and Texture. Gaping was evaluated according to the procedure described by Andersen et al. (*19*). Each fillet was given a score from zero to five. Zero was given for a fillet without gaping, and five was given for a fillet with severe gaping.

Fillet texture was evaluated instrumentally by compression using a Texture Analyzer TA.XT2 (Stable Micro Systems, Surrey, England) equipped with a cylindrical plunger (12.5 mm diam). The plunger was pressed into the fillets at a constant speed of 1 mm s^{-1} until it reached 90% of the sample height. The resistance force was recorded at the maximum force obtained during compression. Two measurements were performed on each fillet, and the mean value was used for the analysis. The measurements were made in front of the dorsal fin, about 1.5 cm above the lateral line.

Data Analysis. The fluorescence spectra were analyzed using the multivariate methods principal component analysis (PCA) (*20*) and partial least-squares regression (PLSR) (*21*). By principal component analysis (PCA), the main systematic variation in data is extracted by resolving the information into principal components (PCs). The data (fluorescence spectra) are contained in a matrix **X** and decomposed into a score matrix **T** and a loading matrix **P**. The residuals, which are the variations that are not described by the scores and loadings, are collected in a residual matrix **E**.

The general purpose of PLSR is to find a mathematical relation between two data sets, **X** (fluorescence spectra) and **Y** (gaping score or texture). PLSR performs a simultaneous decomposition of the **X** and **Y** matrices in such a way that the information in the **Y** matrix is directly used as a guide for the decomposition of **X** after which regression of **Y** is done. The regression models were evaluated by the correlation between the measured and the predicted values and by the root-mean-squared error of prediction (RMSEP), which is defined by

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}}$$

where y_i is the measured value of sample i and $\hat{y}_i$ is the predicted value of the same sample. All models were validated with full cross validation. The data analysis was performed using The Unscrambler version 7.6, CAMO A/S, Trondheim, Norway.

RESULTS AND DISCUSSION

Spectral Characteristics of Fish Muscle and Collagen. Fluorescence emission spectra of purified collagen show that collagen type I has peaks around 430 and 480 nm irrespective of the excitation wavelength used (**Figure 1**). Collagen type V has a peak just below 400 nm and a shoulder around 480 nm.

The autofluorescence spectra of the two fish species (**Figure 2**) are somewhat different but have common features that are similar to those of collagen type I and type V. Especially for salmon, the peaks and shoulders obtained at 390, 430, and 480 nm agree with the three peaks of collagen. The figure also shows a difference in intensity between the salmon sample taken from the loin and the sample taken from the tail. A difference between the two cod samples is also seen, but it is not as clear as for the salmon.

The gray scale images visualize the difference in fluorescence measured around 450 nm between cod and salmon (**Figure 3**).

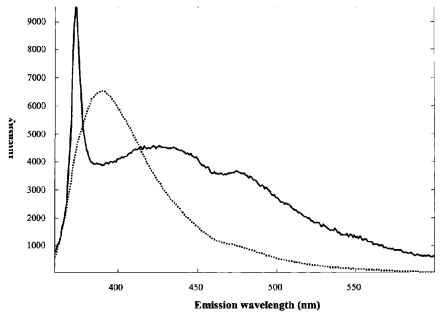


Figure 1. Autofluorescence spectra of calf skin collagen type I (solid) and bovine collagen type V (dotted) obtained after excitation at 330 nm.

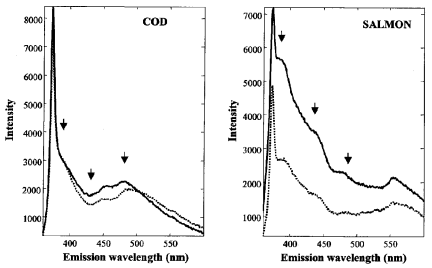


Figure 2. Autofluorescence of cod and salmon after excitation at 332 nm. Cod samples were stored for 5 days (left panel), from the middle section (solid) and froim the tail part (dotted) of the same fillet. Salmon samples (right panel) were from the middle section (solid) and the tail part (dotted) of the same fillet. The arrows indicate the peak of collagen as seen in **Figure 1**.

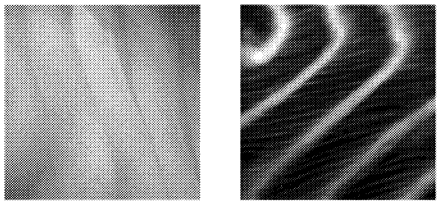


Figure 3. Autofluorescence images of cod stored for 1 day (left panel) and of salmon (right panel). The excitation wavelength was 332 nm, and the intensity of emitted light recorded at 450 nm.

The myocommata of salmon emits a bright white fluorescence, whereas the muscle fibrils are more reddish/pink. The thin fibrils of perimysial and endomysial connective tissue can be seen as white stripes connecting two thick stripes of myocommata. In contrast, the myocommata of cod emits slightly weaker fluorescence than the muscle fibrils. The whole fillet emits a blue/white fluorescence.

The actual fluorescence of the two species is probably quite similar. In salmon, however, fluorescence from collagen and other sources is most likely reabsorbed by the red pigments in the muscle fibrils. These pigments, astaxanthin and cantaxanthin, absorb strongly in the region 350−520 nm (*22*). This phenomenon creates the visible contrast between muscle fibers and connective tissue. Cod muscle is white, and no pigments will reabsorb the autofluorescence. Therefore, fluorescence of the muscle in cod is as pronounced as that of myocommata, and for the particular sample shown here, the fluorescence of the muscle fibers seems to be somewhat more intense. The peak at 430 nm of collagen type I can hardly be seen in the fluorescence spectra of cod, which might be due to the larger fluorescence of the muscle fibers, which do not allow this collagen peak to be detected. However, this phenomenon needs to be studied further.

Above, similarities between fluorescence of fish muscle and mammalian collagen type I and type V are illustrated. The collagen structure of fish muscle may not be exactly similar to the structure of collagen isolated from a mammalian source, and possibly the fluorescence of collagen from fish differs from that measured on mammalian collagen. However, it has been found that the structure and amino acid composition of collagen type I and type V from fish muscle is quite similar to that of mammalian and avian muscle (*2*, *3*). Hence, it is expected that the very small differences will not be detected by the method used here. Even though we have not found any publications stating that other types of collagen are present in fish muscles, it might be possible. However, collagen type I and V are the types found in largest concentration. Therefore, considering the low concentration of other collagen types and the complexity of the fish muscle, it is not expected that other types of collagen will be detected by the method used here. Elastin is another possible fluorophore but is only to be found in very low concentration and is shown to have an emission spectrum different from that of collagen (*23*).

Explorative Data Analysis. *Salmon.* Two PCAs based on emission spectra of salmon taken from the loin were made. One of the models included the samples of collagen, and the other was made only with the salmon samples. In this way 150 samples were used in the model without collagen and 152 samples were used in the model including collagen. Before modeling, the spectra of collagen were normalized to be of the same size as the salmon spectra. Three-component models explained 96% and 97% of the variation in data for the models with and without collagen, respectively. The observation that the addition of spectral features from collagen did not lead to an extra dimension in the models and that the two models explain approximately the same amount of variation suggests that collagen exhibits a major part of the autofluorescence of salmon.

The loadings of the three components obtained by the two models show similarities supporting the resemblance between the fluorescence of salmon and collagen (**Figure 4**). The PC1 loadings describe an average intensity in the emission spectra and illustrate peaks corresponding to those of collagen type I and type V. Thus, collagen may be a great contributor to the variation in intensity and, probably, the samples had different amounts of collagen exposed for illumination.

The PC2 loadings of the model including collagen have a clear peak around 390 nm corresponding to the peak of collagen type V. The significance of this peak is illustrated by the score values (**Figure 5**). All salmon samples have PC2 score values around zero whereas collagen type V has a high PC2 score value. This large difference between collagen type V and salmon indicates that phenomena other than collagen contribute to the autofluorescence of salmon. Both collagen type I and type V vary from the salmon in PC3. Thus, PC3 may explain some structure of collagen. Furthermore, in the model made without

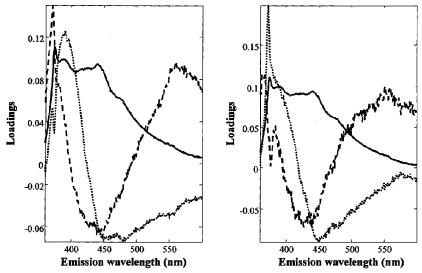


Figure 4. Loadings of PCAs based on emission spectra: (left) model including samples of salmon and collagen type I and type V, PC1 (86%) (solid), PC2 (8%) (dotted), and PC3 (3%) (dashed); (right) model including only salmon samples, PC1 (78%) (solid), PC2 (13%) (dotted), and PC3 (5%) (dashed).

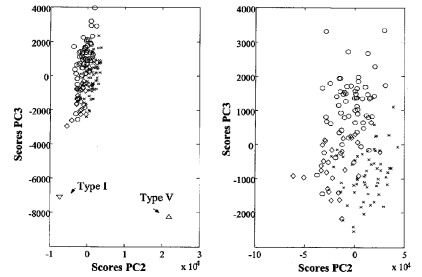


Figure 5. PC2 score versus PC3 score for PCAs based on emission spectra after excitation at 332 nm: (left) model including samples of salmon and collagen type I and V; (right) model including only salmon samples. Key: storage time 1 (○); storage time 2 (×); storage time 3 (◇).

collagen there is a separation of the salmon samples as well. This may be due to a variation in the amount of pigment exposed for illumination. The large negative loadings of PC3 between 430 and 480 nm corresponds to the wavelengths with absorption of the pigments (*22*). Even though salmon may contain both tocopherol and vitamin A, these fluorophores do not seem to influence the salmon spectra due to low concentrations, low quantum yields, and excitation/emission maxima at other wavelengths (*24*).

A separation of data depending on the storage time is obtained both when modeling only salmon and when modeling collagen together with the salmon (**Figure 5**). The possibility of separating the samples according to the storage time illustrates the advantage of multivariate methods, since it was not possible to detect any variation of the storage time by studying the spectra visually. For the model made without collagen, PC2 separates samples stored for 2 or 3 weeks. PC3 separates samples stored 1 week from samples stored 2 or 3 weeks. There is no systematic separation of the sample groups in PC1 (not illustrated). As mentioned, this component mainly describes a variation in the overall fluorescence intensity.

Cod. Two PCAs based on autofluorescence spectra of cod were made. One included the two samples of collagen type I and type V, and the other was made only on cod samples. Thus, the two models were made using 41 and 39 samples, respectively. As for the models based on salmon, the collagen spectra were normalized to have approximately the same intensity as

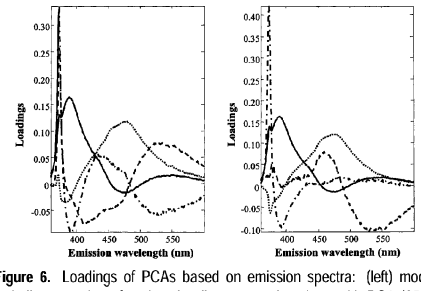


Figure 6. Loadings of PCAs based on emission spectra: (left) model including samples of cod and collagen type I and type V, PC1 (65%) (solid), PC2 (26%) (dotted), PC3 (6%) (dashed), and PC4 (2%) (dash–dotted); (right) model including only cod samples, PC1 (72%) (solid), PC2 (24%) (dotted), PC3 (2%) (dashed), and PC4 (2%) (dash–dotted).

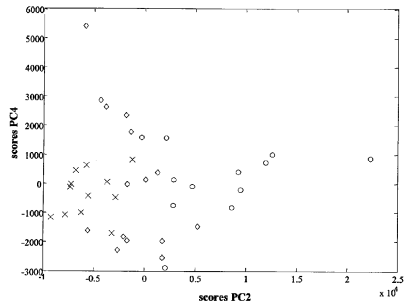


Figure 7. Scores of PC2 versus scores of PC4 from a PCA of the autofluorescence spectra after excitation at 332 nm. Storage time: 1 day (○); 5 days (×); old cod (◇). The PCA is made only on cod samples.

the cod samples. Four-component models explained 99% and 100% of the variation in data for the model with collagen and the model without collagen, respectively.

There are similarities in the PC1 and PC2 loadings of the two models (**Figure 6**). PC1 has a large peak below 400 nm corresponding to the peak of collagen type V shown in **Figure 1**. PC4 explains 2% of the variation when including collagen in the model and seems to describe the autofluorescence of both collagen types since the loadings have a negative peak below 400 nm and two positive peaks around 430 and 480 nm corresponding to the peaks obtained from pure collagen (**Figure 1**). When collagen is omitted from the model, there are still peaks in the loadings of PC4 corresponding to peaks in the spectra of the two types of collagen. Possibly, the fourth component describes the variation in collagen also in the model made on the basis of the data without collagen. However, there is no apparent systematic variation in the score values of PC4 (**Figure 7**) in any of the models.

A systematic variation of the storage time is seen in PC2 (**Figure 7**). Samples stored 1 day have higher score values corresponding to a larger peak between 450 and 480 nm. Samples from the fish auction hall have PC2 score values that are between the samples stored 1 day and the samples stored for 5 days. Conclusions based on samples with only three different storage times might be very weak. However, the separation in the score values may be due to an increase in the overall fluorescence intensity after longer storage. A broad peak

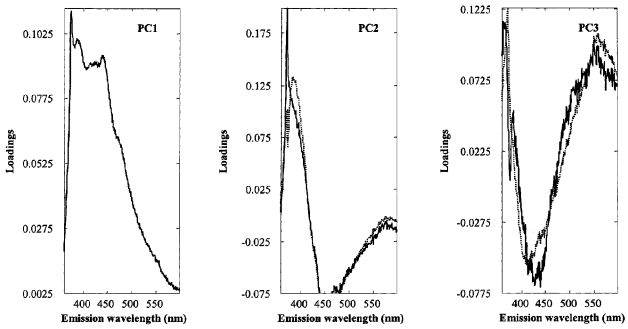


Figure 8. Loadings of a PCA based on emission spectra of salmon excluding samples of collagen (solid) and orthogonal rotated loadings of a PCA based on emission spectra of salmon and collagen (dotted).

above 500 nm developed after long storage may be so strong that it increases the intensity of the whole emission spectrum. This broad peak corresponds to the peak in the loadings of PC3 around 525 nm. As for the visual assessment (**Figure 2**) no clear variation according to the sampling site was seen in any of the components. Variation due to sampling site was expected since a variation in collagen content between head and tail has been shown for other fish species (*25*, *26*).

The development of the peak slightly above 500 nm during storage may be due to some oxidation phenomena (*16*, *17*, *27*). Adding a small amount of malondialdehyde directly to fresh cod muscle resulted in a strong increase in fluorescence intensity similar to the peak just above 500 nm (not illustrated). Wold et al. (*28*) used the same approach. They showed that fluorescence from poultry meat with added aldehydes exhibited fluorescence similar to that of freeze stored poultry. It was also shown that the emission from one sample can be caused by several oxidation products (*28*, *29*). Hasegawa et al. (*17*) described that emission developed just above 500 nm could originate from protein related oxidation. Due to the low fat content of cod of approximately 0.5%, it seems reasonable that the fluorescence developed by oxidation is related to the proteins.

The importance of the wavelength area around 450 nm in the PC2 loadings may be due to a decrease in the NADH content. NADH may be used as an indicator of freshness and makes up a large part of the fluorescence from fresh muscle. The content and thereby the intensity decreases during storage (*30*), which is in agreement with a decrease in the peak between 450 and 480 nm.

The fluorescence intensity obtained by the oxidation products and by NADH seems to be so strong that only weak indications of peak from collagen can be detected in the fluorescence spectra of cod. This may also explain why the peak at 430 nm of collagen type I is not seen in the fluorescence spectra. Probably, measuring only the connective tissue would reveal all the collagen peaks since NADH and oxidation products are found in the muscle fibers.

Collagen and Fish Muscle Fluorescence. The multivariate analysis indicated the presence of collagen type I and type V in the two fish species. Peaks possibly originating from fluorescence of collagen were clearly seen for salmon, but only collagen type V was easily detected in the spectra of cod. The identification of collagen can be further verified by rotating the loadings obtained from the PCAs. With application of orthogonal rotation, the loadings obtained from the two data sets (with and without collagen) should be approximately similar if the two sets span the same variation (*31*). By the rotation, the loadings obtained from PCAs on the basis of data containing collagen are regressed orthogonally on the loadings obtained from PCAs on the basis of data without collagen. The results thus obtained are illustrated in the **Figures 8** and **9**. For salmon, the rotated loadings corresponds well with the loadings of the model made on data without collagen. This similarity is not as clear for cod, where only the loadings of the two first components seem to be similar. These loadings may describe collagen type V, NADH, and lipid oxidation products. Thus, rotating the loadings by orthogonal rotation supports the results already indicated from the visual investigation. Another way to identify collagen could be to relate fluorescence spectra to a measured collagen content by partial least-squares regression.

The collagen content and its distribution depend on the swimming activity of the fish (*32*). Furthermore, higher collagen content in the tail part than near the head is found in flounder, hake, and trout (*25*, *26*). Therefore, it was expected that similar variations could be detected within the fillets. For salmon, there was an intensity difference between samples taken from the loin and samples taken from the tail. The myocommata is thicker in the middle section of the fillet compared to the tail part. Therefore, this variation in intensity may be explained by how much of the myocommata that was exposed for measurement. Furthermore, for ground beef it was shown that the connective and adipose tissue had almost the same emission spectra (*7*). Fluorescence microscopy showed that the major source of fluorescence from adipose tissue was from the connective tissue fibers located between the adipose cells (*33*). Therefore, the higher lipid content in the loin may add to the variation in fluorescence intensity between the two parts of the fillet. In addition, the larger amount of lipid and connective tissue in the loin reduces the relative amount of pigmented muscle and thus reduces the reabsorption by astaxanthin.

Cod did not show the same variations within one fillet. Probably this is due to the relatively low contribution of collagen to the fluorescence spectra, the low fat content of cod, and the fact that other fluorophores inside the muscle flesh are clearly more pronounced such as NADH and lipid oxidation products as mentioned above. Even though there is a variation in collagen content within the cod muscle, the content and the variation may be so low that it cannot be easily detected with the fluorescence techniques used here.

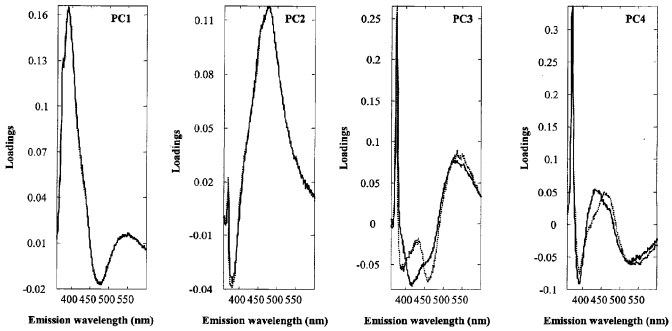


Figure 9. Loadings of a PCA based on emission spectra of cod excluding samples of collagen (solid) and orthogonal rotated loadings of a PCA based on emission spectra of cod and collagen (dotted).

Relating Fluorescence of Salmon to the Gaping Score and Texture. Autofluorescence spectra of salmon were correlated to the gaping score and to the texture by partial least-squares regression (PLSR). Above it was suggested that collagen type I and type V cause a major part of the structure in the autofluorescence spectra of salmon. Furthermore, texture and gaping are supposed to be influenced by the collagen content as well as its degradation during storage. Therefore, it was expected that the autofluorescence spectra could be correlated to either the texture, the degree of gaping, or both.

There seemed to be some relation between the autofluorescence spectra and the gaping score. A five-component PLS model gave a correlation of 0.43 between measured and predicted values and a RMSEP of 0.91 when the gaping scores varied between 0 and 3. The gaping score was evaluated subjectively and could be influenced by errors. Furthermore, the fluorescence intensity expressed how much collagen that was exposed at the surface of the sample and not the total amount of collagen within the fillet. Thus, the sample measured may not be representative for the whole fillet, which may induce errors in the predictions. It could be relevant to study the relation between gaping score and fluorescence in more detail using other instrumentation or sample preparation to develop on-line/at-line techniques for quality sorting. Since gaping occurs along myocommata, measurements can for instance be done in these areas.

The correlation between the measured texture, which varied between 5.5 and 14.1 N, and the texture predicted by the fluorescence measurements was close to zero. This low correlation may be explained by a variation due to the sampling technique, a very low variation in collagen content or texture, or that there is in fact no relation between texture and fluorescence.

As denoted by Sigurgisladottir et al. (*34*), optimizing the sampling technique is important for texture determinations. A relatively small area of the muscle is sampled by the cylindrical probe used here. Fluorescence is measured on another small part of the fillet. Thus, it may be that measurements performed on these small samples are not representative for the whole fillet and the two samples may not correspond exactly to each other. Furthermore, it has been shown that Kramer shear force measurements gave better texture determinations than the texture profile analysis used here (*35*) and should be used when more experiments are to be made.

Collagen contents of 2.9 g/kg (*5*) and 6.6 g/kg (*36*) have been found in salmon. Supposedly, the variation in collagen in the salmon muscle is so small that the correlation to texture and partly to the gaping score cannot be found. The relation between collagen content and texture or gaping found in other studies has been shown for different fish species with a much larger range in collagen content (*1*, *6*).

A relation to texture might be found if other instrumentation or a different sample preparation is used. One could either homogenize the samples or measure only myocommata.

Furthermore, concerning collagen degradation, we cannot exclude the possibility that even though collagen is degraded, the collagenous fluorophores may still remain active for which reason the degradation will not be measured with the technique used here.

In conclusion, collagen type I and type V seem to be major contributors to the structure of the fluorescence spectra. Although low correlations were obtained between fluorescence and texture and gaping on salmon, the method is still promising as a rapid tool for determining such quality parameters. Other sampling methods and instrumentation should be tested in this effort. The study shows that fluorescence from fish muscle is complex and depends on factors such as freshness, oxidation, and site of measurement.

ACKNOWLEDGMENT

The authors are grateful for financial support through the AQM (Advanced Quality Monitoring) supported by the Danish Ministries of Research and Industry. Mia Bencze Rørå at Akvaforsk, Norway, is gratefully acknowledged for providing the salmon and the texture and gaping measurements. Frank Lundby is thanked for technical assistance with the fluorescence measurements. Rasmus Bro and Bo Jørgensen are gratefully acknowledged for valuable advice during the data analysis and preparation of the manuscript.

LITERATURE CITED

(1) Sato, K.; Yoshinaka, R.; Sato, M.; Shimizu, Y. Collagen content in the muscle of fishes in association with their swimming movement and meat texture. *B. Jpn. Soc. Sci. Fish.* **1986**, *52*, 1595−1600.

(2) Sato, K.; Yoshinaka, R.; Sato, M.; Itoh, Y.; Shimizu, Y. Isolation of types-I and types-V collagen from carp muscle. *Comp. Biochem. Physiol., B* **1988**, *90*, 155−158.

(3) Sato, K.; Yoshinaka, R.; Itoh, Y.; Sato. M. Molecular-species of collagen in the intramuscular connective- tissue of fish. *Comp. Biochem. Physiol., B* **1989**, *92*, 87−91.
(4) Mizuta, S.; Hwang, J.; Yoshinaka, R. Molecular species of collagen from wing muscle of skate (*Raja kenojei*). *Food Chem.* **2002**, *76*, 53−58.
(5) Aidos, I.; Lie, Ø.; Espe, M. Collagen content in farmed Atlantic salmon (*Salmo salar L.*). *J. Agric. Food Chem.* **1999**, *47*, 1440−1444.
(6) Hassan, F.; George, S.; Mukundan, M. K.; Sherief, P. M. Role of collagen in gaping of fish fillets. *Fish. Technol.* **1999**, *36*, 40−42.
(7) Wold, J. P.; Lundby, F.; Egelandsdal, B. Quantification of connective tissue (hydroxyproline) in ground beef by autofluorescence spectroscopy. *J. Food Sci.* **1999**, *64*, 377−383.
(8) Egelandsdal, B.; Kvaal, K.; Isaksson, T. Autofluorescence spectra as related to tensile properties for perimysium from bovine masseter. *J. Food Sci.* **1996**, *61*, 342−347.
(9) Bailey, A. J.; Sims, T. J.; Avery, N. C.; Halligan, E. P. Nonenzymatic glycation of fibrous collagen−reaction-products of glucose and ribose. *Biochem. J.* **1995**, *305*, 385−390.
(10) Bailey, A. J.: Light, N. D. *Connective tissue in meat and meat products*; Elsevier Science Publishers Ltd.: London, 1989.
(11) Davis, H. K. Fluorescence of fish muscle: Description and measurement of changes occurring during frozen storage. *J. Sci. Food Agric.* **1982**, *33*, 1135−1142.
(12) Manohar, S. V. Characteristics of white muscle fluorescence in pre-rigor fish. In *Fish inspection and quality control*; Kreuzer, R., Ed.; Fishing News (Books) Ltd.: London, 1971; pp 211−215.
(13) Davis, H. K.; Reece, P. Fluorescence of fish muscle: causes of change occurring during frozen storage. *J. Sci. Food Agric.* **1982**, *33*, 1143−1151.
(14) Aubourg, S. P.; Medina. I. Influence of storage time and temperature on lipid deterioration during cod (Gadus morhua) and haddock (Melanogrammus aeglefinus) frozen storage. *J. Sci. Food Agric.* **1999**, *79*, 1943−1948.
(15) Aubourg, S. P.; Sotelo, C. G.; Perez-Martin, R. Assessment of quality changes in frozen sardine (Sardina pilchardus) by fluorescence detection. *J. Am. Oil Chem. Soc.* **1998**, *75*, 575−580.
(16) Aubourg, S. P.; Medina, I.; Gallardo, J. M. Quality assessment of blue whiting (Micromesistius poutassou) during chilled storage by monitoring lipid damages. *J. Agric. Food Chem.* **1998**, *46*, 3662−3666.
(17) Hasegawa, K.; Endo, Y.; Fujimoto, K. Oxidative deterioration in dried fish model systems assessed by solid sample fluorescence spectrophotometry. *J. Food Sci.* **1992**, *57*, 1123−1126.
(18) Huss, H. H.; Sigsgaard, P.; Jensen, S. A. Fluorescence of fish bones. *J. Food Prot.* **1985**, *48*, 393−396.
(19) Andersen, U. B.; Strømsnæs, A. N.; Thomassen, M. S.; Steinsholt, K. Fillet gaping in farmed Atlantic salmon. *Norw. J. Agric. Sci.* **1994**, *8*, 165−179.
(20) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37−52.
(21) Martens, H.; Næs. T. *Multivariate calibration*; John Wiley & Sons: Chichester, England, 1989.
(22) Britton, G. UV/Visible Spectroscopy. In *UV/Visible spectroscopy in carotenoids*; Britton, G., Liaaen-Jensen, S., Pfander, H., Eds.; Birkhäuser: Basel, 1995; Vol. 1B, Chapter 2.
(23) Wold, J. P. Rapid quality assessment of meat and fish by using near-infrared spectroscopy autofluorescence spectroscopy and image analysis. Ph.D. Dissertation, Agricultural University of Norway, 2000.
(24) Duggan, D. E.; Bowman, R. L.; Brodie, B. B.; Udenfriend, S. A spectrophotofluorometric study of compounds of biological interest. *Arch. Biochem. Biophys.* **1957**, *68*, 1−14.
(25) Feinstein, G. R.; Buck, E. M. Relationship of texture to pH and collagen content of yellowtail flounder and cusk. *J. Food Sci.* **1984**, *49*, 298−299.
(26) Montero, P.; Borderias, J. Distribution and hardness of muscle connective-tissue in hake (*Merliccius-merluccius L*) and trout (*Salmo-irideus gibb*). *Z. Lebensm.-Unters.-Forsch.* **1989**, *189*, 530−533.
(27) Undeland, I.; Ekstrand, B.; Lingnert, H. Lipid oxidation in herring (Clupea harengus) light muscle, dark muscle, and skin, stored separately or as intact fillets. *JAOCS* **1998**, *75*, 581−590.
(28) Wold, J. P.; Mielnik, M.; Aaby, K.; Pettersen, M. K.; Baardseth, P. Rapid assessment of rancidity in complex meat products by front face fluorescence. *J. Food Sci.* **2002**, in press.
(29) Wold, J. P.; Mielnik, M. Nondestructive assessment of lipid oxidation in minced poultry meat by autofluorescence spectroscopy. *J. Food Sci.* **2000**, *65*, 87−95.
(30) Munck, L. *Fluorescence analysis in foods*; Longman Scientific and Technical: Harlow, England, 1989.
(31) Jennrich, R. I. Orthogonal rotational algorithms. *Psychometrika* **1970**, *35*, 229−235.
(32) Yoshinaka, R.; Sato, K.; Anbe, H.; Sato, M.; Shimizu, Y. Distribution of collagen in body muscle of fishes with different swimming modes. *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* **1988**, *89*, 147−151.
(33) Swatland, H. J. Autofluorescence of adipose-tissue measured with fiber optics. *Meat Sci.* **1987**, *19*, 277−284.
(34) Sigurgisladottir, S.; Hafsteinsson, H.; Jonsson, A.; Lie, Ø.; Nordtvedt, R.; Thomassen, M.; Torrissen, O. Textural properties of raw salmon fillet as related to sampling method. *J. Food Sci.* **1999**, *64*, 99−104.
(35) Isaksson, T.; Swensen, L. P.; Taylor, R. G.; Fjæra, S. O.; Skjervold, P. O. Nondestructive texture analysis of farmed Atlantic salmon using visual/near-infrared reflectance spectroscopy. *J. Sci. Food Agric.* **2002**, *82*, 53−60.
(36) Eckhoff, K. M.; Aidos, I.; Hemre, G.-I.; Lie, Ø. Collagen content in farmed Atlantic salmon (*Salmo salar*, L.) and subsequent changes in solubility during storage on ice. *Food Chem.* **1998**, *62*, 197−200.

Received for review May 7, 2002. Revised manuscript received September 17, 2002. Accepted October 3, 2002.

JF020524D

Paper II

Practical aspects of PARAFAC modelling of fluorescence excitation-emission data

Charlotte M. Andersen and Rasmus Bro

Summary

This paper presents a dedicated investigation and practical description on how to apply PARAFAC modeling to complicated fluorescence excitation-emission measurements. The steps involved in finding the optimal PARAFAC model are described in detail based on the characteristics of fluorescence data. These steps include choosing the right number of components, handling problems with missing values and scatter, detecting variables influenced by noise and identifying outliers. Various validation methods are applied in order to ensure that the optimal model has been found and several common data-specific problems as well as their solutions are explained. At last, interpretations of the specific models are given. The paper can be used as a tutorial for investigating fluorescence landscapes with multi-way analysis.

KEYWORDS: validation, scatter, missing values, outliers

Introduction

Fluorescence spectroscopy has been used in many scientific fields, such as chemistry, medicine, environmental and food science. However, fluorescence signals can be rather complex and, therefore, the analysis might become complicated due to interferences, scatter, overlapping signals, etc.

When autofluorescence of every sample is measured at several emission wavelengths for several excitation wavelengths, the interpretation can be facilitated by the use of multi-way models such as PARAFAC [1], Tucker, and N-PLS [2-4]. If the data are approximately trilinear, curve resolution is possible by the use of PARAFAC, possibly providing estimates of the spectra and concentration profiles of the underlying chemical analytes, thus relating the measured autofluorescence of mixtures to chemical knowledge about the material [2,5]. When modeling fluorescence data, interference from scattered light (primarily Rayleigh and Raman scattering) can give a slight model inadequacy influencing the estimated model parameters. Furthermore, other factors such as quenching and instrumental noise can cause problems. Therefore, considering the various influential contributions is important when analyzing fluorescence excitation-emission data and when interpreting the derived models.

Several studies have explored and described the underlying chemical phenomena in fluorescence spectral data by applying multi-way modeling such as PARAFAC. For example, Ross et al. [6] have described the underlying structure of fluorescence spectra obtained from pigment complexes in pea thylakoids. In other studies, PARAFAC models based on fluorescence landscapes of sugar solutions were used to obtain information regarding both sugar quality and process parameters [7,8]. It was shown that tryptophan and tyrosine as well as high molecular weight Maillard reaction polymers and polyphenolic compounds were important fluorophores. In an investigation of the interactions between a non-fluorescent DDT-

type pesticide and a fluorescent dye, PARAFAC was employed to find the fluorescence profiles of complexed dye states [9]. Furthermore, the ability to quantify trace pesticides and polycyclic aromatic hydrocarbons by fluorescence spectroscopy and PARAFAC modeling was investigated by Jiji et al. [10]. They showed a possibility for resolving the analyte spectra from overlapping fluorescence signals, scatter and instrumental background. PARAFAC was also applied to describe and predict the amount of dissolved chlorophyll and pheophytin pigments [11]. Dioxin content of fish oils were estimated by PARAFAC modeling of fluorescence landscapes [12]. It was suggested that the relations obtained were due to quenching effects or other complex chemical factors in the fish oils.

There are several references describing the theory and giving examples of PARAFAC modeling [1,3,4,13,14]. Furthermore, the applications given above illustrate the advantages of using PARAFAC for analyzing fluorescence spectroscopic data and show how PARAFAC can be applied to interpret such data. However, no articles discuss in depth the practical aspects of using PARAFAC for decomposing fluorescence excitation-emission matrices (EEMs). Important practical issues such as how to determine the optimal number of components, how to handle disturbing scatter signals and how to deal with missing values, find outliers and validate the model are mostly just described superficially.

This paper demonstrates in detail how the use of a suitable mathematical model such as PARAFAC can help understanding the underlying spectral and chemical phenomena in a complex system. In particular it provides guidelines and tools for proper handling of many of the typical problems that arise in the modeling of fluorescence excitation-emission data. Thus, the paper can be used as a tutorial for investigating fluorescence landscapes with multi-way analysis. Measurements on fish muscle extracts are used as an example and illustrate a possibility of using fluorescence spectroscopy for on-line quality monitoring in the fish industry. Additionally, fluorescence excitation-emission matrices of solutions with known concentrations of four fluorophores are analyzed. Including both data sets in the presentation provides an illustration of how both well-characterized (and well-behaved) data as well as less well-characterized data can be analyzed.

Materials and methods

Fish data

A factorial design with two frozen storage temperatures (-20°C and -30°C), four frozen storage periods (3, 6, 9 or 12 months) and five chill storage periods (0, 3, 7, 14 or 21 days at 2°C) was used. Cod (*Gadus morhua*) from a single catch were caught in February 1999 in the Barents Sea. Experiments after 3 months of frozen storage at -20°C were left out for practical reasons giving a total of 35 storage conditions. The chill storage was made in modified atmosphere (40%CO_2/40%N_2/20%O_2). Three packs with each storage treatment were analyzed giving a total of 105 samples.

Aqueous extracts were made by homogenizing 25 g fish muscle with 75 ml water. The pH was reduced to 5.2 with 2 M HCl and the mixture was heated to 70°C, cooled to room temperature and filtered to remove precipitated proteins. This treatment provides a clear solution containing a number of fluorophores but of lower chemical complexity than if the raw fish muscles were measured. The extracts were stored at -30°C.

The extracts were measured spectrofluorometrically at 22°C in a 10 by 10 mm thermostated quartz cuvette on a Perkin Elmer LS50 B spectrofluorometer. Raw non-smoothed data was recorded. For every sample, an excitation-emission matrix (EEM) was obtained by measuring the emission spectra from 270 to 600 nm in 2 nm intervals with excitation at every 10 nm from 250 to 370 nm. These wavelengths were chosen from a preliminary experiment based on which areas that showed the highest variability (not shown). The measurements started with the highest excitation wavelength and ended with the lowest in order to minimize photodecomposition of the sample [15]. Based on the visual appearance of the preliminary experiments excitation and emission slit widths were both set to 7 nm and the scan speed was 750 nm/s.

All measurements were made within a few days to minimize the effect of instrumental drift and change in lamp intensity. Such changes could have required suitable standardization [16]. A solution of 15 g ordinary white sugar per 100 ml double deionized water was used as standard and measured three times a day to verify that instrumental changes did not influence the results [16]. No considerable variations among the sugar samples were found. The fish muscle extracts were measured in random order.

Data with known fluorophores

This data set contains fluorescence landscapes of 27 samples containing different concentrations of four fluorophores with rather similar spectral properties [17]. The four compounds are: Phenylanaline, 3,4-dihydroxyphenylalanine (DOPA), 1,4-dihydroxybenzene and tryptophan.

The measurements were performed on the same Perkin Elmer LS50 B fluorescence spectrofluorometer as the fish data with excitation wavelengths ranging between 200 and 315 nm (5 nm interval) and emission wavelengths ranging from 250 to 459 nm (1 nm intervals). Both excitation and emission slit widths were set to 5 nm and the scan speed was 1500 nm/min.

Multi-way data analysis

Ideally, the data arranged in an $I \times J \times K$ three-way array will be trilinear [2]. The first index (I) refers to the samples, the second (J) to the emission wavelengths, and the third (K) to the excitation wavelengths (Figure 1).

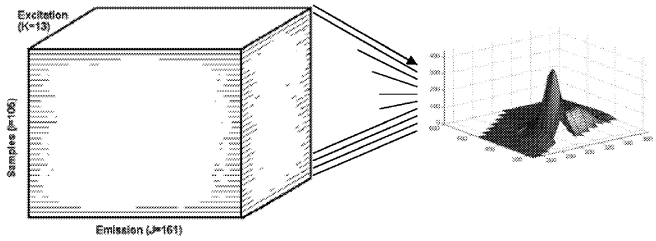


Figure 1 Arrangement of the data in a three-way structure

PARAFAC [1,13] was used to model the data. The PARAFAC model can be written as

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{jf} c_{kf} + e_{ijk} \qquad i = 1,..,I;\ j = 1,..,J;\ k = 1,..,K$$

where x_{ijk} is the intensity of the i-th sample at the j-th variable (emission mode) and at the k-th variable (excitation mode). a_{if}, b_{jf} and c_{kf} are parameters describing the importance of the samples/variables to each component. The residuals, e_{ijk}, contain the variation not captured by the model.

The PARAFAC components will be estimates of the signals from the individual fluorophores if the data are approximately low-rank trilinear and when the correct number of components is used. In that case, the scores in a_{if} may be interpreted as the relative concentration of analyte f in sample i. The J-vector $\mathbf{b}_f$ with elements b_{jf} (j=1,..,J) is the estimated emission spectrum of this analyte and likewise $\mathbf{c}_f$ is the estimated excitation spectrum. In order for the decomposition to be unique and provide meaningful estimates, some mathematical conditions must be fulfilled. For example, no two spectra may be identical. In fact, if all spectra and concentration profiles are linearly independent, the decomposition is unique. More relaxed conditions can also be given [18]. The model states that for each analyte, the contribution to the measured excitation-emission matrix is $a_{if}\mathbf{b}_f\mathbf{c}_f^T$, which is the landscape of a pure analyte scaled by the concentration. This reflects that additivity and linearity of the signal is assumed to be valid. A change in concentration only changes the magnitude of the contribution (a_{if}) not the actual shape ($\mathbf{b}_f\mathbf{c}_f^T$). Both concentrations and spectra are only determined up to a scaling because e.g. the vector $\mathbf{b}_f$ may be exchanged with $½\mathbf{b}_f$ as long as for example $\mathbf{c}_f$ is exchanged with $2\mathbf{c}_f$. Such a change will not change the contribution to the model ($a_{if}\mathbf{b}_f\mathbf{c}_f^T = a_{if}½\mathbf{b}_f 2\mathbf{c}_f^T$).

Inner-filter effects caused by high concentrations, scattering and quenching can disturb the trilinearity of the data. Furthermore, abundance of missing data (see later) and spectral

similarities can lead to uncertain estimates. Therefore, the chemical interpretation should be exercised with care. In the following it is shown how a PARAFAC model can be validated using fit values, visual assessment of the loadings, residual analysis, core consistency diagnostic, jack-knifing and by a split-half analysis [19].

For further in-depth descriptions of PARAFAC theory and algorithms we refer to other publications [1,2,4,5,13,14]. All analyses were performed with the N-way toolbox (www.models.kvl.dk) and Matlab ver. 5.3 (The MathWorks Inc., Natick, MA).

Data-pretreatment

For the fish data set, emissions from 270 to 280 nm were removed from the data to reduce the amount of Rayleigh scatter and missing values. This gives a three-way array of the size 105 × 161 × 13. For the data set with the four known fluorophores, the size of the array is 27 × 210 × 24.

There is no emission below the excitation wavelength as this would correspond to higher energy being emitted than the energy causing the emission. Therefore, below the excitation wavelength, the emission is zero (or some noise equivalent to zero) regardless of the chemistry. Hence, for a complete EEM landscape a triangular part will be trivially zero which has to be respected by the subsequent model. These zero-values do not conform to the trilinear model which is valid for the remaining part of the data. If this is not respected, misleading results can be obtained. The most common way to handle this problem is to set these values to missing in the analysis.

It is worth considering in more detail why emission below excitation cannot generally be simply set to or kept at zero in PARAFAC modeling. Emission wavelengths below the excitation wavelength do not exhibit any fluorescence and the measured intensity is therefore zero up to noise. However, a part of an estimated excitation spectrum may have non-zero values at wavelengths higher than the corresponding estimated emission spectrum. An example of this is given for tyrosine in Figure 2 (upper part). The PARAFAC model of a measured excitation-emission landscape is found as the outer product of the estimated excitation and emission spectrum. Hence, for emission below excitation, the model will have non-zero values as shown for tyrosine. However, this is physically impossible in practice. No fluorescence will occur and only zero signal (plus possible scatter) will be measured. Thus, the tri-linear PARAFAC model will not be valid for this part.

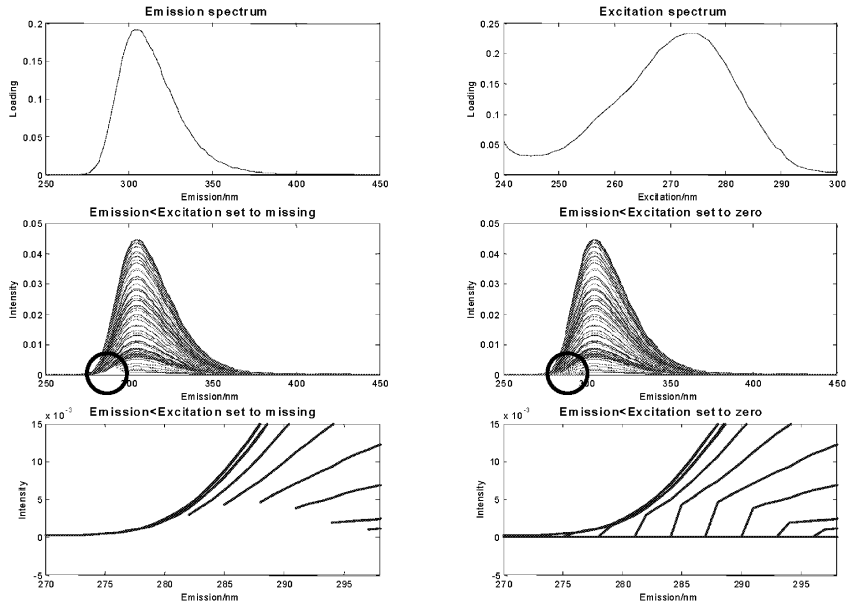


Figure 2 Illustration of the inadequate use of PARAFAC for fluorescence emission below the excitation wavelength. Top plots show the emission and excitation spectrum of tyrosine. The left middle plot show the emission spectra measured at several excitation wavelengths with emission below excitation set to missing. A smaller fraction indicated by the circle is shown blown up below. The right middle plot and lower plot shows the same but with emission below excitation set to zero. As can be seen lower-right, a rank-one approximation is not valid when zeros are present because the different emission spectra no longer have the same shape due to the incorporated zeros.

The middle-left-hand plot in Figure 2 shows that similar spectra are obtained for different excitation wavelengths when the model of the data is set to missing for emission below excitation. The plot also shows that a one-component bilinear model can model the data adequately because all spectra have the same shape regardless of excitation. The relevant parts of the plot are shown in detail in the lower left part of the figure (indicated by a circle in the middle plot). When the emission is set to zero as in the middle-right-hand plot, an interesting phenomenon appears. Every emission spectrum is now different in shape because the zeros appear at different places for different excitations. Clearly, such a data set can not be modeled by a single bilinear component as is the model in PARAFAC. In fact, many components are needed in order to handle the fact that every emission spectrum is different. This is contradictory to the underlying assumptions and hence the incorporation of zeros hinders adequate analysis of data.

In the above example, the problematic area is only the emission wavelengths slightly below the excitation wavelengths. The problem is that the chemical model, which is only valid for emission above excitation, is extended below for mathematical reasons. Emissions *far* below the excitation wavelength, though, are mostly zero both physically and mathematically. Hence, in

these areas it is less problematic to introduce zeros. However, other problems than the above-mentioned can cause zeros to be invalid. For example, for a double-fluorophoric molecule, a double-peak might occur in both the excitation and the emission spectra of the molecule and then this can lead to ghost peaks in the emission below excitation area. These will then be incorrectly forced to zero if zeros are maintained in this area. Before discussing the possible ways to handle emission below excitation another problematic feature of EEM data is described.

Scattering (or reflection) is brought about by small particles in the samples, and causes the light to deviate from its original path and spread in all directions. The scattering does not contain any information on the fluorescence properties of the sample. Even non-fluorescent samples give rise to several peaks in a fluorescence spectrum. These peaks arise from elastic Rayleigh scatter, inelastic Raman scatter and complex, but predictable, harmonic order reflections of these same scatter peaks caused by the diffraction grating employed in the monochromator of the instrument (0th, 1st etc. order scattering). In fluorescence EEMs, the scatter effects show up as diagonal lines across the landscapes. For elastic scattering there is no energy loss, so the scattered emission wavelength is identical to that of the exciting light. For inelastic scattering there is an energy loss associated with the scattering light, i.e. the scattered emission is shifted to longer wavelengths compared to the incident light. Both Rayleigh and Raman scattering intensities will vary with solvent type and the quantity of dissolved particles in solution but Rayleigh scatter is often the most troublesome to handle [15]. Due to Rayleigh scatter, emission in a window around the excitation wavelength does not conform to a trilinear model or rather, the emission above the excitation has a contribution from the trilinear fluorescence signal as well as from the non-trilinear scatter signal.

There are several ways to deal with the non-chemical areas (the zeros and the scatter). In this case, the emission measurements up to slightly above the excitations are simply set to missing. Emissions obtained at the wavelengths around twice the excitation wavelength will be influenced by second order Rayleigh scatter and were also replaced with missing values. The algorithms used to handle the missing data are described by Bro [4]. In the literature other approaches have also been adopted [14,20-24], such as using zeros or down-weighting elements where emission is lower than excitation. The purposes of using these alternatives are basically the same (minimize the influence of non-trilinear parts) and no systematic investigations have yet been made to compare them. The current experience is that as long as the influence of the most significant Rayleigh scatter containing parts are minimized, the results are often similar.

Results and discussion

The steps and problems involved in finding a valid PARAFAC model for describing fluorescence excitation-emission data are given below. The appropriate number of components is determined based on several different criteria. For example the visual appearance of the loadings is a useful diagnostic because fluorescence spectra of liquids are typically characterized by broad and often uni-modal peaks. Furthermore, the variance explained by the model, the number of iterations and the so-called core consistency diagnostic [4] are used. An important discussion on the influence of scatter and missing values is included in the following. How to correct for possible artifacts from these by constraining the model is described next. Only the determination of the number of components for the fish data set is illustrated in detail, since the procedure is similar for the other data set. An evaluation of the reliability of the score values by a jack-knife based approach is used together with the leverage and the distribution of sample residuals for identifying outlying samples. After removing grossly disturbing outliers, the models are validated by a split-half analysis. All diagnostic tools used to assess the models have pitfalls and only by using several quantitative as well as qualitative tools, a thorough conclusion can be reached.

Building the PARAFAC model

Fish data

Initial PARAFAC modeling

Unconstrained PARAFAC models are calculated using from one to five components. It is expected that a two-component model is optimal since the fluorescence landscapes contain two visually recognizable peaks (Figure 3). However, peaks of fluorophores with a low quantum yield or low concentration may not be seen because of the dominant response of other fluorophores. This is especially so, when their excitation and emission maxima are in the same wavelength areas as the dominant ones. Therefore, it is possible that more than two fluorophores contribute to the fluorescence spectra obtained here.

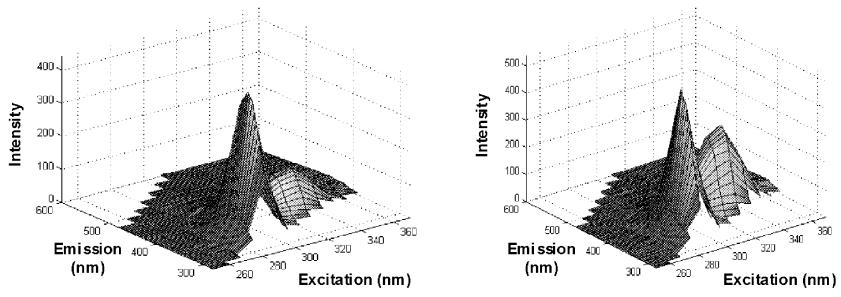


Figure 3 Examples of fluorescence excitation-emission landscapes. Emission spectra were collected from 280 to 600 nm after excitation from 250 to 370 nm. Note that triangular parts of the landscapes have been set to missing.

The explained variance indicates that three components are optimal because the increase obtained with more than three components is small relative to the increase in explained variance obtained using up to three components (Table 1). Additionally, the percentage of variance explained using three components is adequate for this type of data. This is known from experience but can also be quantitatively evaluated based on the pooled standard deviation of replicate measurements, which is found to vary between 0.9 and 9.3. Comparing this with the pooled residuals of the three-component model of 4.3 and 6.5 for the two-component model shows that, for both the two-component and the three-component model, the amount of un-modeled variation in the data agrees with the variation between replicates and, thus, the un-modeled part of the data can largely be attributed to expected residual variation. The pooled residual variation is found as the average standard deviation for all data elements.

Table 1 Explained variance and core consistencies in percentages vs. the number of components for PARAFAC models of the fluorescence data with from 1 to 5 components

Number of components	1	2	3	4	5
Explained variance (%)	84.7	98.1	99.4	99.6	99.8
Core Consistency (%)	100	100	37	31	10

The core consistency diagnostic is an approach suggested for finding the number of components to use in multi-way models such as PARAFAC [4]. A so-called Tucker3-like core array is calculated from the data and the PARAFAC loadings. The relative sum-of-squared difference between this core and a superdiagonal core of ones is called the core consistency. It is usually expressed in percentages. A PARAFAC model can be represented as a constrained Tucker3 model where the core has been forced to a superdiagonal of ones (trilinearity) and the consistency provides a quantitative measure of how well the loadings represent variation in the data consistent with this assumption. If the core consistency is not close to 100%, the model does not give an appropriate description of the data and a lower number of components should be chosen (see Bro [4] for more details). Core consistencies for the five models are shown in Table 1. A two-component model gives a core consistency of 100 %. The core consistency of the three-component model of 37 % indicates that this model might not be stable and hence core consistency points to two components being optimal. However, as noted above the percentage of variation explained for these data points to three components possibly being adequate. It is not uncommon that different diagnostics point to different number of components, especially in this early stage of an analysis. Further elaboration will help to clear up such controversies and no final conclusions are drawn at this stage.

When fitting a model several times from random starting positions, the same solution is typically obtained for all or most models if the right number of components is chosen. When too many components are extracted, the number of local minima (different model fit values when fitting from different starting points) often increases. In the actual situation, the three-component model repeatedly converged to the same solution. This does not provide conclusive evidence, but together with the conclusions from looking at the fit-values, it does indicate that three components may be feasible.

A so-called two-factor degeneracy is a situation in which two of the components are virtually identical but of opposite sign [25]. This typically occurs when too many components are extracted. Here, no degeneracy is found for the three-component solution. Although absence of local minima and degeneracies are only to be used as very ad hoc diagnostics, these signs together with the fit values do indicate that three components might be an option upon further analysis.

Visual appearance of the loadings

When the fluorescence data follow a trilinear model, the correctly validated PARAFAC emission and excitation loadings are estimates of the pure analyte fluorescence spectra when the correct number of components is used. Figure 4 shows the loadings of the two- and three-component models. The emission spectra in both models have characteristics that are not consistent with the expectations. The sharp peaks around 300 nm and the large negative regions indicate that the models are not correctly identifying spectra of pure chemical components.

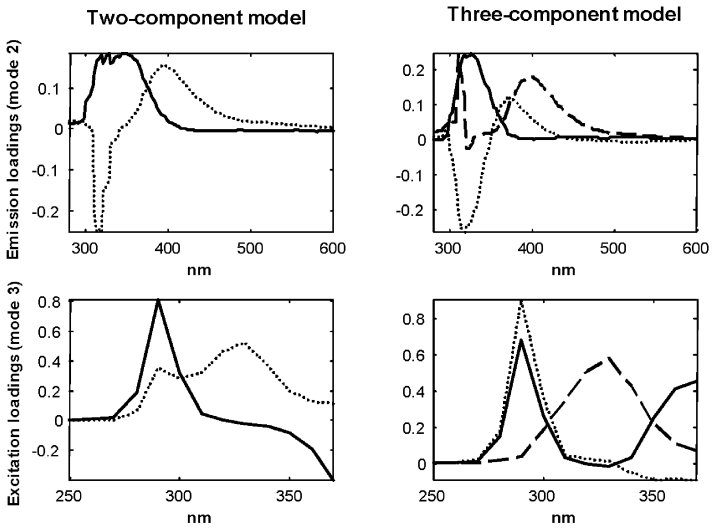


Figure 4 Emission and excitation loadings of two- and three-component PARAFAC models. The lines indicate: component one (solid), component two (dotted) and component three (dashed).

Probably, the high and narrow peak in the estimated emissions of component three (Figure 4) is an artifact caused by a combination of the large amount of low-emission missing data (Figure 3) and the presence of small amounts of scatter remaining in the specific area. This will be elaborated on below.

Jack-knife validation of the loadings

One way to evaluate the stability of the model is by an exploratory use of the jack-knife method. Jack-knifing is a resampling method that can be used for assessing the uncertainty of the model parameter estimates such as scores and loadings [26]. Multiple models are fitted by leaving out one sample at a time. For these data, then, 105 versions of the loadings are obtained, which can be used for evaluating the model stability and for outlier detection [27]. Figure 5 shows the standard errors of the 105 estimated emissions for the three-component model. Large standard errors are obtained for component three in the wavelength area with the high and narrow peak (Figure 4). Also for component two a part of the wavelengths corresponding to areas with supposedly unreliable loadings have high standard errors. This component shows emission loadings with both a negative and a positive peak.

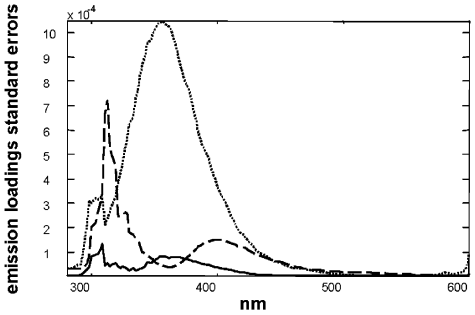


Figure 5 Standard errors of the loadings of the three-component PARAFAC model (Figure 4) obtained by Jack-knife validation. Jack-knife is performed by leaving one sample out at a time. The lines indicate: component one (solid), component two (dotted) and component three (dashed).

Explaining artifactual loadings

By jack-knifing, a quantitative measure of the uncertainty is obtained, but no explanation for the problematic loading shapes can be obtained from these results directly. The sharp low wavelength peak is, in fact, a spurious result of the pattern of missing values. In order to see why the peak appears, the structure described by component one and two of the three-component model are subtracted from the original data. Thus, only noise and the structure described by the third component remains in these residuals and hence represents the variation that component three is seeking to explain. Visualizing this part of the data, can help understanding why the estimated emission profile is peculiar. Figure 6.a shows an example of these residuals illustrating the peak caused by a chemical phenomenon corresponding to the broad peak of component three. The high and narrow peak around emissions at 300 nm is not visually detectable. In addition, the outer product of the emission and excitation loadings of

component three is given (Figure 6.b). The outer product provides a full landscape but if the missing elements in the data are set to missing in the component three landscape, Figure 6.c is obtained. The high and narrow peak in the emission loadings causes a narrow peak in the full landscape positioned in a part where there are no measurements. The peak arises because it is describing a small remaining scatter peak in the original landscape in the low excitation area (circle in Figure 6.a). Due to the pattern of missing values, the size of this small peak relative to the large chemical peak is largely *unrelated*. Thus, the magnitude of the two peaks in component three cannot be compared directly. The apparent large sharp peak is an artifact that appears due to the combined effect of the small amount of Rayleigh scatter left and the missing data. As such the peak is correct but its relatively high magnitude is disturbing from an exploratory point of view. A similar explanation of the inferior appearance of the loadings of component two can be given even though the appearance of component two differs from component three.

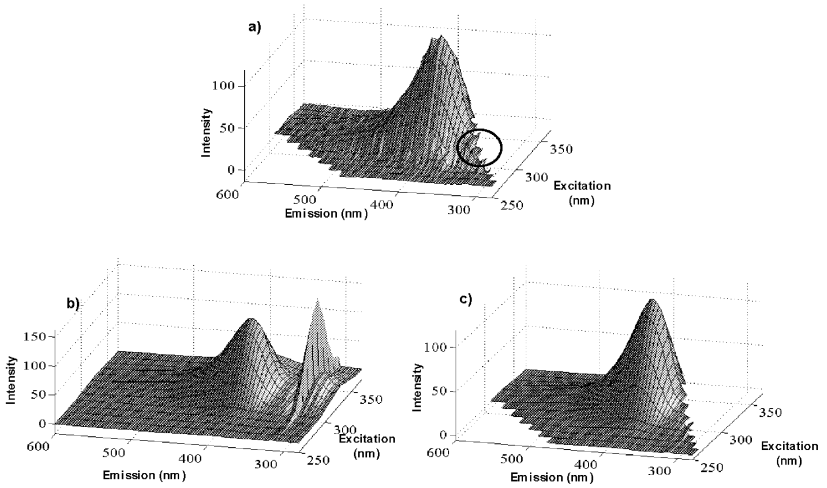


Figure 6 a) Example of an excitation-emission matrix after removing component one and two obtained from the three-component PARAFAC model, b) outer products of emission and excitation loadings of component three of the three-component PARAFAC model and c) as in b) with the wavelength areas corresponding to missing values in the original data set to missing.

The above indicates that more data elements need to be set to missing because there are still parts influenced by scatter. However, setting more elements to missing would mean that parts of the interesting chemical variation are also eliminated which leads to a lower information level, which is not feasible. There are several alternative ways of dealing with this problem [4,24,28]. Below it will be shown how the application of constraints on the parameters can be helpful to this end. It is emphasized, though, that the apparently artifactual peaks are not incorrect numerically but merely annoying in visual interpretation of the parameters. Before applying

constraints an initial look at the currently obtained score values is performed to check for extreme outliers.

Initial evaluation of score values

Extreme score values in the sample mode indicate possible outliers. Figure 7 shows scatter plots of the scores obtained from the three-component PARAFAC model (loadings are shown in Figure 4). All scores are of similar magnitude and no extreme outliers are indicated on the basis of the score values. However, there may be more subtle outliers, which can not be identified only by studying the score values. More dedicated diagnostics will be needed for identification of such outliers but is postponed as the immediate check for gross outliers suffices at this stage of the analysis.

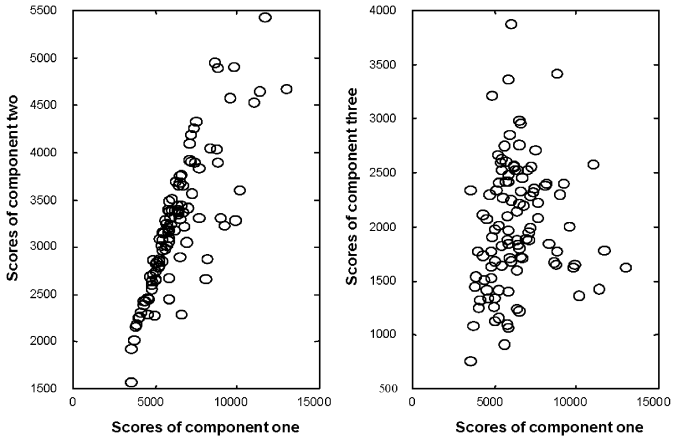


Figure 7 Scatter plot of the score values obtained for the three-component PARAFAC model.

The score plot shows an unexpected partly collinear behavior of the scores for two of the components. While this may be due to chemical properties of the system, it is disturbing that it apparently only occurs for a (major) part of the data. If this is reflecting a real underlying phenomenon, then likely an unfortunate sampling has been used. However, comparing with Figure 4, the corresponding emission spectra seem to be partly confounded having a similarly shaped low-emission peak. Most likely these two components are hence partly describing the same underlying phenomena, possibly caused by scattering. In the actual model described later, the same phenomenon is not observed, hinting at the appropriateness of the constraints suggested next.

Applying Constraints

Constraining the parameters of the PARAFAC model can be helpful in terms of interpretability, curve resolution and stability of the solution. It may be argued that constraints should not be necessary if the data behave according to the model. However, as exemplified above, minor

details in the data can impose large changes in the estimated parameters that disturb proper interpretation. Constraints can help to remedy such problems. Caution is needed though, to make sure that the constraints are not otherwise disturbing the appearance of important phenomena.

The loadings shown in Figure 4 are uni-modal apart from the unstable part. It is expected that that the real spectra are non-negative and thus, constraining with both non-negativity and uni-modality could possibly help in modeling the problematic part of the fluorescence landscapes. Various combinations of non-negativity and uni-modality constraints in the three modes were tried and evaluated by explained variance, core consistency and visual appearance of the loadings. It was found that applying non-negativity in the sample mode and both uni-modality and non-negativity in the two other modes (excitation and emission) performed well. The loadings thus obtained for two- and three-component models are illustrated in Figure 8. The visual appearance of the loadings indicates that the constrained models give a more appropriate description of the underlying spectral phenomena than the un-constrained models, not being excessively influenced by the minor remaining scatter variation. The second excitation component of the two-component constrained model has a shoulder indicating that the data contain more than two components.

In order to validate the appropriateness of constraints, the changes caused by these must be explainable. For example, for non-negative constraints the loadings should show peaks at reasonable positions compared to the raw data but with negative parts removed. The changes can be seen by comparing Figure 8 with Figure 4. The loadings have been forced to be uni-modal whereby inferior peaks caused by scatter and missing values have been removed. The danger of applying constraints is that the parameters are forced to correspond to the a priori knowledge without necessarily being more adequate. However, by comparing the solutions carefully, the risk of misleading models can easily be avoided. In this case, the spectral profiles are similar to the unconstrained profiles except that understood low-variance artifacts are removed.

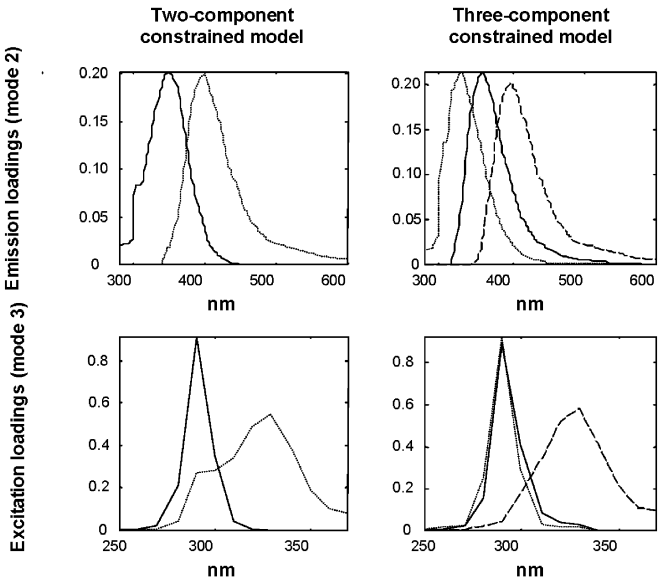


Figure 8 Emission and excitation loadings of two- and three-component PARAFAC models constrained with non-negativity in mode one and non-negativity and uni-modality in mode two and three. The lines indicate: component one (solid), component two (dotted) and component three (dashed).

A core consistency of 59 % for the three-component model shows that systematic variation is present in all three factors. Compared with the core consistency of the three-component un-constrained model of 37 % (Table 1) this shows that applying constraints increases the model validity although the core consistency is still lower than hoped for. This, probably, reflects that the three components are somewhat difficult to estimate. The explained variance increases from 97.6 to 99.1 by increasing the number of components from two to three. The last value should be compared with the explained variance of the three-component unconstrained model of 99.4% (Table 1). The fit of a constrained model will be lower than for an un-constrained model per definition, but the similarity of fits shows that the constrained model is mainly filtering off insignificant noise and not systematic variation. Hence, the constraints seem valid also from this perspective.

From the discussion above, the non-negativity and uni-modality constrained three-component model is found to give a valid description of the excitation-emission data.

Data with known fluorophores

Since the data set contains four analytical compounds, it is expected that four PARAFAC components are appropriate and the reliability of four components will be verified below. Initial modeling indicates the presence of four or five components as evaluated from the explained variance and the number of iterations (Figure 9). When too many components are included, the

number of iterations will typically increase dramatically as seen for the six-component model. However, the number of iterations can only be used as an indication since e.g. highly correlated components will often require many iterations and a model with too many factors can sometimes be easy to fit. Furthermore, an unlucky initialization may lead to many iterations even though the model is perfectly sound. Still, using ad hoc measures such as the number of iterations, number of local minima etc. is useful in practice.

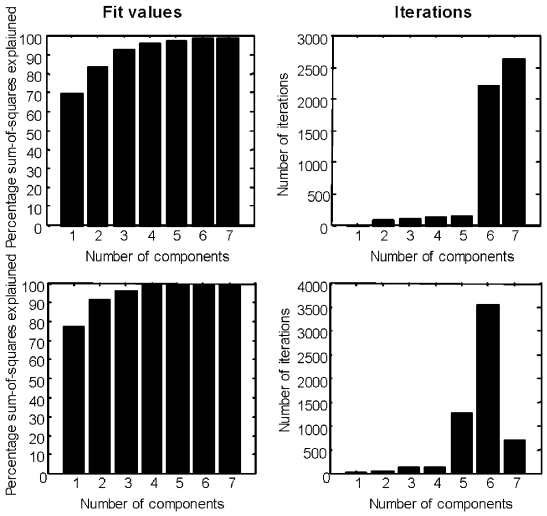


Figure 9 Fit values and number of iterations as a function of the number of components, all data included (top) and low excitation and emission removed (bottom).

The five-component model indicates that the low excitations are very noisy. This is seen in Figure 10, which shows an example of the residuals of sample one. The reason for this noise is to be found in the properties of the instrument. In order to reduce the effect of the noise, excitations below 230 nm were excluded as were also emissions below 260 nm.

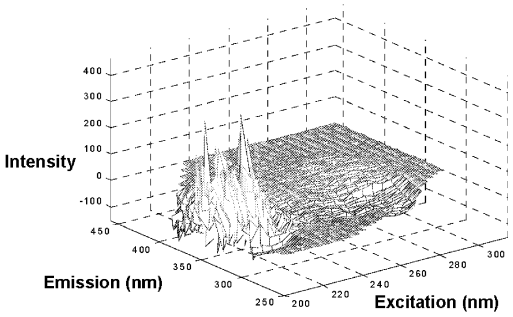


Figure 10 Residuals of sample one from a five-component PARAFAC model.

After excluding the noisy part it becomes clear that four components provide a valid PARAFAC model. The four-component model explains approximately 100 % of the variation in the data and the number of iterations increases considerably when fitting more components. The reliability is further verified by the excitation and emission loadings, which have shapes resembling pure spectra (Figure 11).

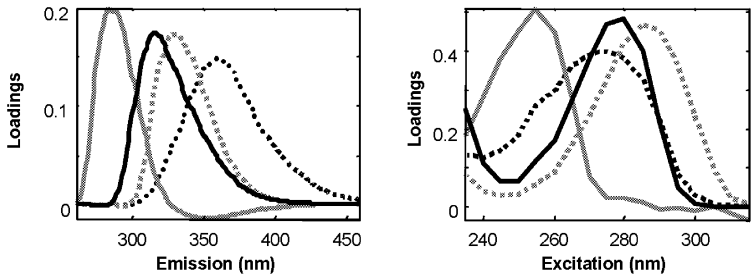


Figure 11 A four-component PARAFAC model on all calibration samples with low excitations removed, emission mode loadings (left) and excitation mode loadings (right). The lines indicate: component one (black solid), component two (black dotted), component three (gray solid) and component four (gray dotted).

Finding outliers

Fish data

The three-component model constrained with non-negativity in the first mode and non-negativity and uni-modality in both the second and the third mode is chosen for further detailed analysis and validation. After the initial analysis, the possible presence of more subtle outliers should be monitored for and possibly removed. Here, jack-knife resampling and an investigation of the residuals is used for the identification of outliers.

An identity match plot obtained from the jack-knife procedure is suggested by Riu and Bro [27] for identification of outliers. In the plot, the predicted score of the sample left out is plotted against the score obtained by the overall model. Potential outliers will be placed away from the ideal identity line, showing that the quantitative information of the samples differs markedly whether or not the sample is included in the analysis. The plot is made for each of the three components separately (Figure 12, upper plots) showing that sample number 42 is a clear outlier. Its score values differ markedly depending on whether it is included in the model or not. Hence, its spectral characteristics are not covered by those of the remaining samples which almost per definition makes it an outlier.

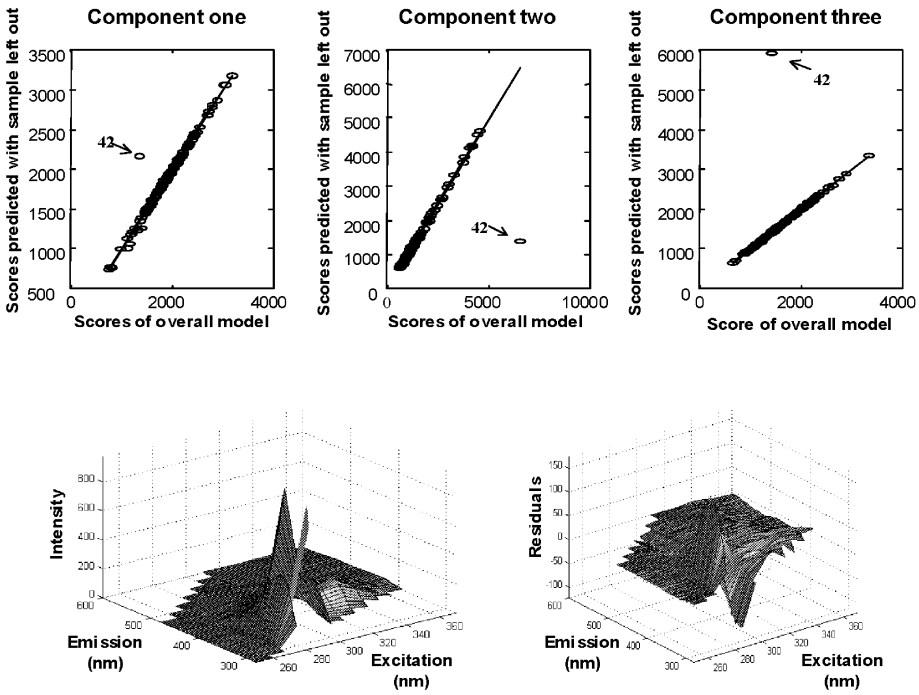


Figure 12 Upper plots: Identity match plots of the scores from the constrained three-component PARAFAC model, lower-left plot: fluorescence excitation-matrix of sample **42** and lower-right plot: residuals of sample **42** for a three-component PARAFAC model.

The fact that sample 42 is an outlier is supported by the fluorescence landscape and the residuals of this sample. The lower-left plot in Figure 12 shows the fluorescence landscape. One of the peaks contains elements with an intensity larger than what can be measured with the given instrument settings (out-of-range). These elements are given as missing values in the data set. Even though this peak has a similar shape as the same peak in all the other samples, it is possible that the larger amount of missing values makes the modeling of the sample difficult or uncertain. This is supported by the systematic variation left in the residuals shown in the plot to the right. Thus, there is significant structure in the fluorescence landscape of sample 42, which is not modeled.

Other diagnostics than the identity match plot can be used for detecting outliers by the jack-knife technique. For example, an outlying sample may have a large uncertainty in the score values as estimated from the resampled models. Furthermore, an indication of an outlier could be a sample where the estimated loadings obtained when leaving out this sample differ from the corresponding loadings obtained from all samples. These approaches were also applied here, but did not indicate additional outliers.

All samples should have low residuals compared to the measured spectra. Furthermore, residuals should ideally be randomly distributed, meaning that the remaining unexplained variations would only be caused by random noise. Sometimes though, systematic variation can remain in the residuals reflecting unmodeled scatter effects of chemical interactions. The model may still be valid, as long the estimated parameters can be properly validated.

A constrained three-component PARAFAC model with sample 42 left out shows that three other samples may be considered outliers as judged from their residuals. The upper-left plot in Figure 13 shows the sum of squared residuals of each sample. Fairly large values are observed for the samples 14, 37 and 82, which also have systematically distributed residuals (Figure 13). Therefore, new models are evaluated with those samples removed.

An explained variation of 99.1 % and a core consistency of around 50 % is obtained irrespective of the removal of outliers or not. Figure 14 shows the loadings of PARAFAC models with sample 42 left out and with the samples 14, 37, 42 and 82 left out. The loadings of the two models are similar indicating that the three samples (14, 37 and 82) are of similar type as the remaining and hence help in supporting the model. Therefore, these three samples are kept in the model.

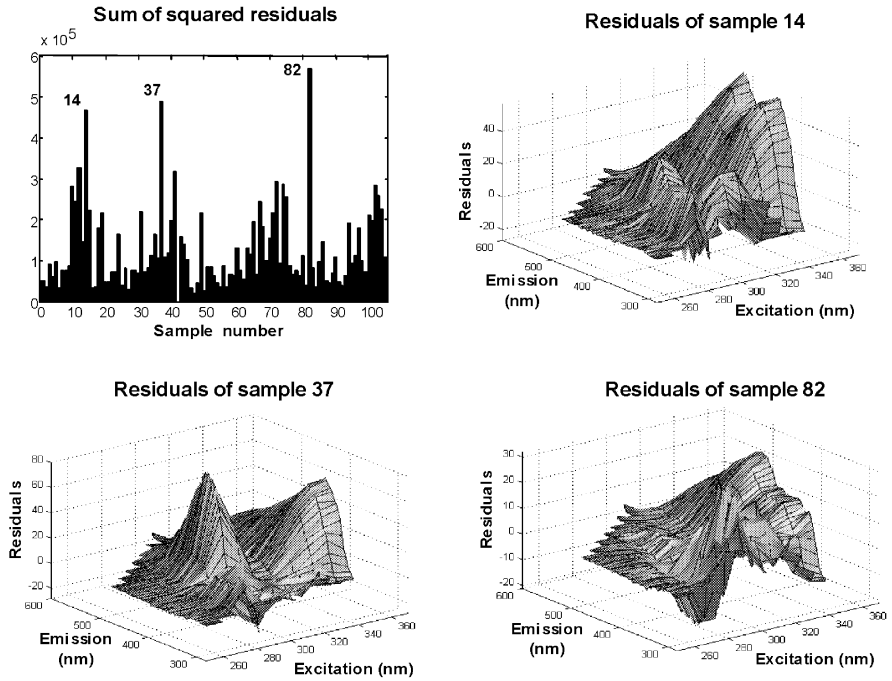


Figure 13 Plots illustrating possible outliers in a constrained three-component PARAFAC model where sample number 42 is removed. Sum of squared residuals for all sample and distribution of residuals for sample 14, 37 and 82 are shown.

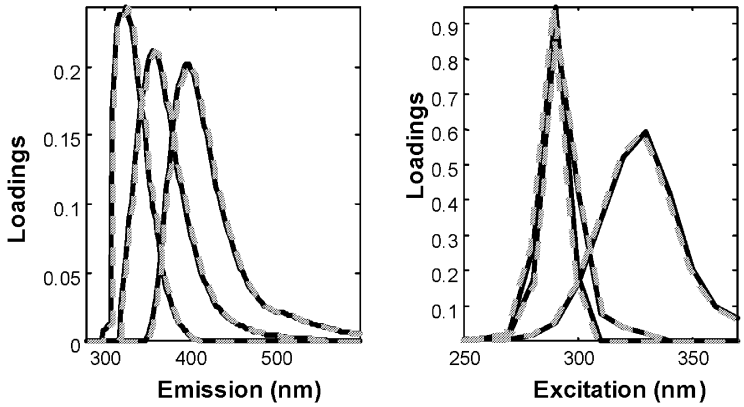


Figure 14 Emission (left) and excitation (right) loadings for constrained three-component PARAFAC models with sample 42 left out (black solid) and samples 14, 37, 42 and 82 left out (gray dashed).

Data with known fluorophores

Another way to illustrate the presence of outliers is by a residual and influence analysis. This is illustrated for the four-component PARAFAC model made on the fluorescence landscapes of the four analytes. Figure 15 shows that no samples have very large residuals but the samples 2, 3 and 4 seem to have somewhat large leverages. As for the other data set, the influence of the high-leverage samples on the model parameter estimates should be evaluated. It was found that the loadings did not change by removing the possible outliers and the samples should, therefore, be kept in the model.

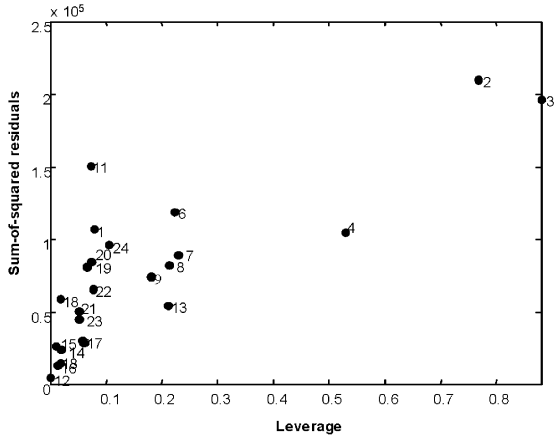


Figure 15 Influence plot of four-component PARAFAC model. The sum-of-squared residuals summed within the sample mode are plotted against the leverage of the first mode components. Each sample is labeled by number.

The samples 2, 3, and 4 have high concentrations of hydroquinone, tryptophan or DOPA. Thus, the high residuals may be due the high concentrations leading to slightly changed line-shapes of the spectra e.g. due to inner-filter effects.

Split-half validations

In order to finally validate the appropriateness of the models build on data with the chosen outliers removed, split-half analysis [19] is applied. In the split-half analysis, different subsets of data are analyzed independently. Due to the uniqueness of the PARAFAC model, the same spectral loadings will be obtained from different samples if the samples reflect the same fluorophores, when the correct number of components is chosen and enough data are available in each subset.

Fish data

A model constrained with non-negativity in the first mode and uni-modality together with non-negativity in the second and third mode is used. In addition, one sample (42) has been excluded as an outlier. The samples are divided according to storage temperature. Thereby all variation in the fluorescence spectra will be found in both data sets. The loadings obtained from two split-half data sets are illustrated in Figure 16. The differences between the two data sets can hardly be discerned and the chosen model, therefore, seems to give a reliable and adequate description of underlying fundamental phenomena.

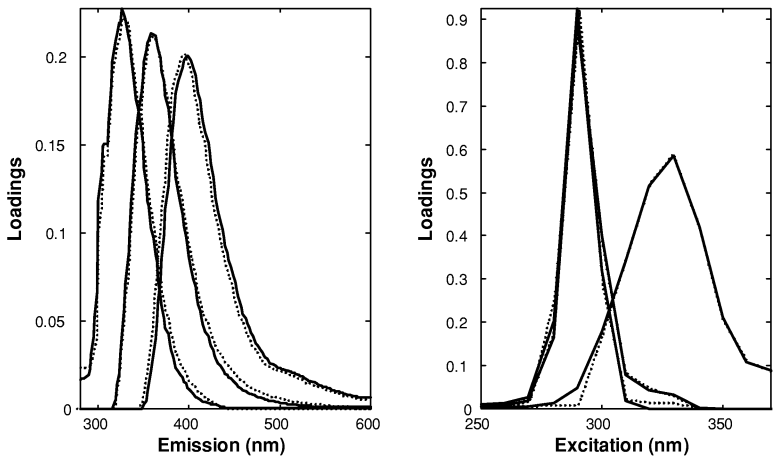


Figure 16 Results from a split-half analysis. The split-half is performed by dividing the samples into groups depending on the storage temperature. Solid lines show samples stored at -30°C and dotted lines show samples stored at -20°C.

Data with known fluorophores

The split-half analysis was performed by dividing the samples into two subsets where both subsets contain information on all four fluorophores. The splitting was made in two ways giving four different subsets that were pair-wise independent. The graphical representation in Figure 17 illustrates the reliability of four components.

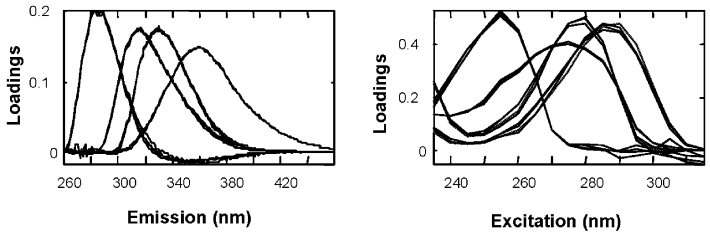


Figure 17 Results from split-half analysis, emission mode loadings (left) and excitation mode loadings (right).

Interpretation

Interpretation of scores and loadings of the fish data

The estimated excitation spectra have maxima at 290 nm and 330 nm (Figure 8). This corresponds to the peaks in the raw fluorescence landscapes (Figure 3). Component one and two have the same position of the excitation maximum but different positions of the emission maxima that are found at 360 and 330 nm for the two components, respectively. In fact, the excitation loadings of those two components are almost identical. The reason for the similarity of these loadings can be that they reflect the same fluorophores but with different substitutions, one or both are bounded to another molecule or that they are identical fluorophores appearing in slightly different electronic environments. This leads to slight differences in spectral characteristics. This was also illustrated by Ross and Leurgans [5] who showed that the estimated excitation spectrum of tyrosine dissolved in water was almost similar to the estimated excitation spectrum of tyrosine bound to phosphate. However, the minor differences in the loadings were sufficient for estimating them separately. The possibility of separating fluorophores with similar excitation maxima shows the value of using PARAFAC for curve resolution and for describing the underlying spectral phenomena. The fluorophore described by component three has excitation maximum around 330 nm and emission maximum around 400 nm.

Investigating press juices and aqueous extracts of fish muscles, a fluorophore associated with dissolved muscle proteins was earlier shown to have excitation/emission maxima at 292/340 nm [29]. Thus, the two fluorophores obtained in the present study with excitation maxima around 290 nm might be related to the presence of muscle proteins. Tryptophan, tyrosine and phenylalanine are the amino acids responsible for the fluorescence of proteins.

Adding tryptophan in various concentrations to the cod extracts and comparing the thus obtained fluorescence with the fluorescence measured on pure extracts indicates the presence of tryptophan. It is found that loadings of a PARAFAC model made on fluorescence spectra of cod extracts with added tryptophan are similar to the loadings of the PARAFAC model made on fluorescence spectra of the pure cod extracts (Figure 18.a-b). The presence of tryptophan in the cod extracts is further substantiated by a linear relation between the score values of component one and the concentration of tryptophan added to the extracts (Figure 18.c). The fluorophore of component two, which has the same excitation spectrum as the tryptophan component but different emission spectrum might also be tryptophan with a substitution or bounded to another molecule. There is no linear relation between the scores of this component and the tryptophan concentration.

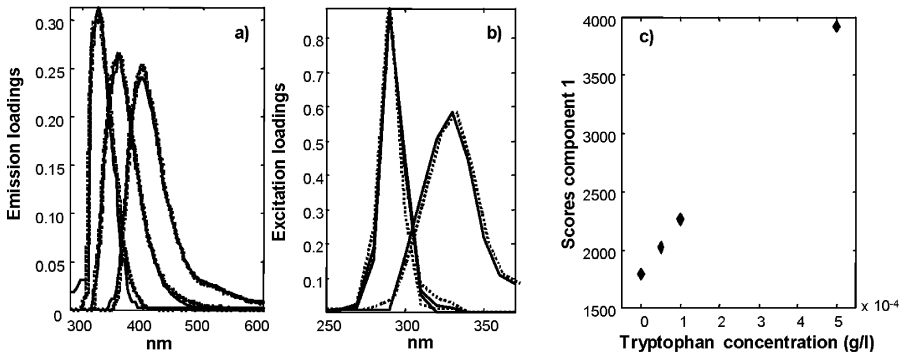


Figure 18 Emission (a) and excitation (b) loadings for a three-component PARAFAC model made on fluorescence spectra of cod extracts added with tryptophan (solid) and excitation and emission loadings for a three-component PARAFAC model made on fluorescence spectra of the pure cod extracts (dotted). Relation between the score values of component 1 and the concentration of added tryptophan (c).

The scores of the model are estimates of the relative concentrations of the fluorophores identified by the loadings and are therefore measures of the amount of the fluorophores present. For component two and perhaps also component one, there are weak correlations to the chill storage time (Figure 19). These relations could not be seen when comparing visually the intensity of the two peaks with the chill storage time. The change in fluorescence intensity with storage time has been shown in other studies [30,31]. However, the present study illustrates an explorative way of identifying chemically meaningful quality parameters that can possibly be monitored industrially.

Another study using the same fish material showed a correlation between the chill storage time and drip loss of water and between chill storage time and the sensory parameter juiciness [32]. When proteins degrade, smaller peptides are produced. This reduces the ability of the muscle to hold water and results in an increased drip loss related to the development of smaller protein

fragments. Thus, the increase in the score values may reflect an increase in the content of small peptides produced during chill storage. This is also verified by the positions of peak maxima in the excitation and emission loadings of component one and two.

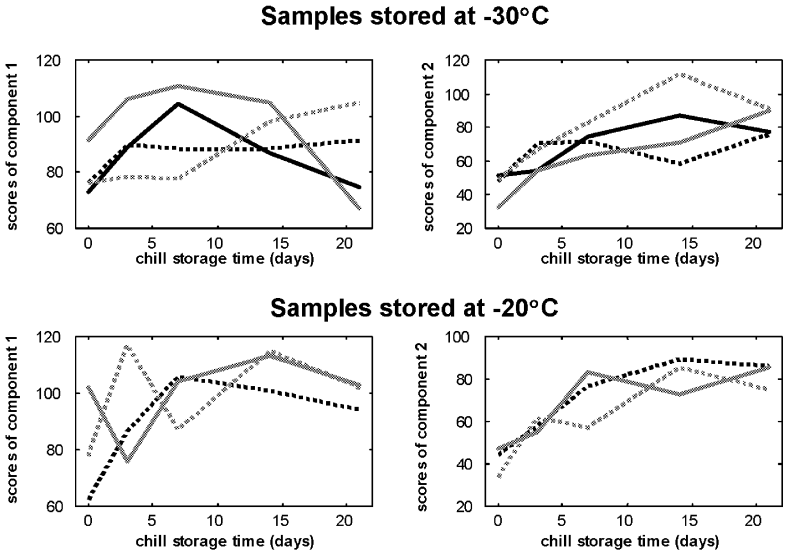


Figure 19 Relation between chill storage time and scores of the constrained three-component PARAFAC model. Scores are corrected for the different weight of fish muscle used for making the extracts. The lines indicate different freeze storage time: 3 months (solid black), 6 months (dotted black), 9 months (solid gray) and 12 months (dotted gray).

No relation was found between the score values and the freeze storage temperature nor between the score values and the freeze storage time. This suggests that these two storage conditions do not have an effect on the development or disappearance of fluorescent chemical compounds in the fish muscle extracts.

Fluorescence correlated to lipid oxidation was found to have excitation maximum at 360-370 nm [33,34]. In the present study, fluorescence with excitation larger than 370 nm was not measured because no emission was obtained for excitation larger than 370 nm. Therefore, lipid oxidation is probably not detected. The NADH does have an important impact on autofluorescence originating from fresh fish muscle [35], but it has been found that the NADH content of fish muscle has almost disappeared after freezing and thawing [29] and probably NADH will have no or only very little influence on the fluorescence spectra obtained in this study. However, component three with excitation/emission maxima around 330/400 nm may conform to the excitation and emission maxima of NADH that are found at 340 and 450 nm, respectively [35]. The difference in peak maxima between the component obtained here by PARAFAC and the value for NADH given in the literature may be an effect of the sample matrix in that the exact

position of the peaks depends on the sample composition and the measurement conditions. However, if component three describes NADH, an effect of the chill storage time is expected, which was not seen. Yet, the relation between fluorescence of fish muscle and NADH or lipid oxidation should be investigated more thoroughly because the NADH content decreases and lipid oxidation develops during storage. Therefore, prediction of these parameters by fluorescence spectroscopy could be an indicator of fish freshness.

It is not possible to identify fully the underlying spectral phenomena that PARAFAC found to be contributing to the fluorescence of aqueous cod extracts. In order to do so, several approaches can be feasible. Chromatographic separation before fluorescence measurements as was shown by Baunsgaard et al. [7] may improve the identification of the fluorophores. In addition, specific peptide sequences can be separated by capillary electrophoresis. This may be used for identifying the component with excitation maximum at 290 nm, which was not tryptophan. Furthermore, standard addition can be used for verifying the presence of certain fluorophores as shown for tryptophan.

Using the model on new data

Below it is shown how the parameters of the PARAFAC model can be used on new fluorescence spectroscopic data e.g. for the determination of analyte concentrations. In the data set with the four fluorophores, the concentration of the four analytes in each sample is known. Therefore, it is possible to make a calibration model from the relation between the score values and concentration of one of the analytes. If the PARAFAC model is made only on a part of the samples, the scores of the samples left out can be estimated by fitting a new PARAFAC model with the spectral loadings fixed to those of the original solution. From the estimated scores, the concentration of the analyte can be estimated from the calibration model.

The loadings of component four resemble the pure spectra of tryptophan. Therefore, score values of this component and the corresponding analyte concentrations are used for making a univariate calibration model for tryptophan. Three samples were not used in building the PARAFAC model and the score values of these samples as well as the analyte concentration can be estimated from the model. Figure 20 shows the tryptophan concentration predicted from the calibration model versus the known concentrations illustrating that the calibration model performs well. The concentration of tryptophan in the samples left out (test samples) is obtained from the estimated score values and the regression coefficient of the calibration model. In the predicted versus measured plot it is seen that sample 5 is not predicted well, whereas good predictions are obtained for sample 10 and 15. A high residual and large leverage was obtained in the estimation of the score values of sample 5. This reveals that it is an extreme sample and also explains why the tryptophan concentration is not predicted well.

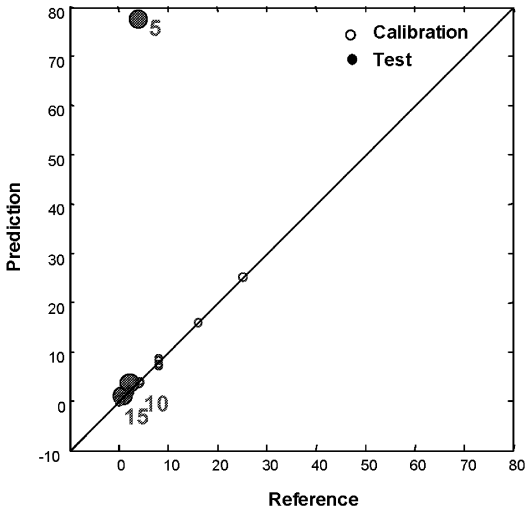


Figure 20 Predicted versus reference concentration of tryptophan. Test samples are shown with solid circles.

Conclusion

This paper has given a presentation of the steps involved in finding an optimal PARAFAC model. Focus has been on the practical application of PARAFAC modeling of complex data. Problems with e.g. scattering and missing values have been given special attention. It has been shown when and how to apply constraints, identify outliers and validate the model. Finally, model interpretation and the possibility to use the results of the PARAFAC model on new samples are illustrated.

Acknowledgements

The authors thank for financial support through LMC (Center for Advanced Food Studies), and AQM (Advanced Quality Monitoring) supported by the Ministries of Research and Industry. R. Bro is grateful for support from EU-project Project GRD1-10337, NWAYQUAL. Niels Bøknæs is gratefully acknowledged for providing the materials.

References

1. Harshman R A. Foundations on the PARAFAC procedure: model and conditions for an "explanatory" multi-mode factors analysis. *UCLA Working Papers in phonetics* 1970; **16**:1-84.

2. Leurgans S, Ross R T. Multilinear models: Applications in spectroscopy. *Statist. Sci.* 1992; **7**:289-319.
3. Kroonenberg P M. Three-mode component models. A survey of the literature. *Statist. Appl.* 1992; **4**:619-633.
4. Bro R. Multi-way analysis in the food industry, Theory algorithms and applications, Doctoral dissertation, University of Amsterdam 1998.
5. Roos R T, Leurgans S. Component resolution using multilinear models. *Methods Enzymol.* 1995; **246**:679-700.
6. Ross R T, Lee C-H, Davis C M, Ezzeddine B M, Fayyad E A, Leurgans S E. Resolution of fluorescence spectra of plant-pigment complexes using trilinear models. *Biochim. Biophys. Acta* 1991; **1056**:317-320.
7. Baunsgaard D, Andersson C A , Arndal A, Munck L. 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 Chem.* 2000; **70**:113-121.
8. Bro R. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometr. Intell. Lab. Syst.* 1999; **46**:133-147.
9. Jiji R D, Andersson G G, Booksh K S. Application of PARAFAC for calibration with excitation-emission matrix fluorescence spectra of three classes of environmental pollutants. *J. Chemometr.* 2000; **14**:170-185.
10. Jiji R D, Cooper G A, Booksh K S. Excitation-emission matrix fluorescence based determination of carbamate pesticides and polycyclic aromatic hydrocarbons. *Anal. Chim. Acta* 1999; **397**:61-72.
11. Moberg L, Robertsson G, Karlberg B. Spectrofluorimetric determination of chlorophylls and pheopigments using parallel factor analysis. *Talanta* 2001; **54**:161-170.
12. Pedersen D K, Munck L, Engelsen S B. Screening for dioxin in fish oil by PARAFAC and N-PLSR analysis of fluorescence landscapes. *J. Chemometr.* 2002; **16**:451-460.
13. Harshman R A, Lundy M E. PARAFAC: Parallel factor analysis. *Comp. Stat. Data. Anal.* 1994; **18**:39-72.
14. Bro R. Parafac: Tutorial and applications. *Chemometr. Intel. Lab. Syst.* 1997; **38**:149-171.
15. Ewing G W. *Instrumental methods of chemical analysis*, McGraw-Hill Book Company, Singapore, 1985.
16. Nørgaard L. Direct standardisation in multi-wavelength fluorescence spectroscopy. *Cemometr. Intell. Lab. Syst.* 1995; **29**:283-293.
17. Baunsgaard D. Factors affecting 3-way modelling (PARAFAC) of fluorescence landscapes. *The Royal Veterinary and Agricultural University*, 1999.
18. Sidiropoulos, N.D., Bro, R. On the uniqueness of multilinear decomposition of N-way arrays. *J. Chemometr.* 2000; **4**:229-239
19. Harshman R A, de Sarbo W S. An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic

techniques. In: H.G. Law, C.W. Snyder, J.A. Hattie and R.P. McDonald (Eds.), *Research methods for multimode data analyis*, Praeger, New York, 1984, 602-642.

20. Heimdal H, Bro R, Larsen L M, Poll L. Prediction of polyphenol oxidase activity in model solutions containing various combinations of chlorogenic acid, (-)-epicatechin, O_2, CO_2, temperature, and pH by multiway data analysis. *J. Agric. Food Chem.* 1997; **45**:2399-2406.
21. Bro R, Heimdal H. Enzymatic browning of vegetables. Calibration and analysis of variance by multi-way methods. *Chemometr. Intel. Lab. Syst.* 1996; **34**:85-102.
22. Jiji R D, Booksh K S. Mitigation of Rayleigh and Raman spectral interferences in multi-way calibration of excitation-emission matrix fluorescence data. *Anal. Chem.* 2000; **72**:718-725.
23. Wentzell P, Nair S S, Guy R D. Three-way analysis of fluorescence spectra of polycyclic aromatic hydrocarbons with quenching by nitromethane. *Anal. Chem.* 2001; **73**:1408-1415.
24. Bro R, Sidiropoulos N D, Smilde A K. Maximum likelihood fitting using ordinary least squares algorithms *J. Chemometr.* 2002; **16**:387-400.
25. Kruskal J B, Harshman R A, Lundy M E. How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In: R. Coppi and S. Bolasco (Eds.), *Multiway Data Analysis*, Elsevier, Amsterdam, 1989, 115-122
26. Efron B, Tibshirani R J. An introduction to the bootstrap, Chapman & Hall, New York, 1993.
27. Riu J, Bro R. Jack-knife for estimation of standard errors and outlier detection in parafac models. *Chemometr. Intel. Lab. Syst.* 2002; **65**:35-49.
28. Bro R, Sidiropoulos N D. Least squares algorithms under unimodality and non-negativity constraints. *J. Chemom.* 1998; **12**:223-247.
29. Davis H K. Fluorescence of fish muscle: Description and measurements of changes occurring during frozen storage. *J. Sci. Food Agric.* 1982; **33**:1138-1142.
30. Aubourg, S P, Medina I, Gallardo J M. Quality assessment of blue whiting (*Micromesistus poutassou*) during chilled storage by monitoring lipid damages. *J. Agric. Food Chem.* 1998; **46**: 3662-3666.
31. Andersen C M, Wold J P. Fluorescence from muscle and connective tissue from cod and salmon. *J. Agric. Food Chem.* 2003; **51**:470-476.
32. Bøknæs N, Jensen K N, Guldager H S, Østerberg C, Nielsen J, Dalgaard P. Thawed chilled Barents sea cod fillets in modified atmosphere packaging-application of multivariate data analysis to select key parameters in good manufacturing practice. *Lebensm. Wiss. Technol.* 2002; **35**:436-443.
33. Undeland I, Ekstrand B, Lingert H. Lipid oxidation in herring (*Clupea harengus)* light muscle, dark muscle, and skin, stored separately or as intact fillets. *J. Am Oil Chem. Soc.* 1998; **75**:581-590.

34. Hasegawa K, Endo Y, Fujimoto K. Oxidative deterioration in dried fish model systems assessed by solid sample fluorescence spectrophotometry. *J. Food Sci*. 1992; **57**:1123-1126.
35. Munck L. *Principles of fluorescence analysis in foods*, Longman Group, UK, 1989.

Paper III

Distribution of water in fresh cod

Charlotte M. Andersen and Åsmund Rinnan

Lebensm.-Wiss. u.-Technol., 35, 687–696 (2002)

Distribution of Water in Fresh Cod

C. M. Andersen and Å. Rinnan*

C. M. Andersen, Å. Rinnan: The Royal Veterinary and Agricultural University, Department of Food Science, Food Technology, Rolighedsvej 30, DK-1958 Frederiksberg C (Denmark)
C. M. Andersen: Danish Institute of Fisheries Research, Department of Seafood Research, DTU Building 221, Søltofts Plads, DK-2800 Lyngby (Denmark)
(Received February 4, 2002; accepted July 11, 2002)

Low-field ^{1}H nuclear magnetic resonance (NMR) transverse relaxation was used to measure water mobility and distribution of water in fresh cod fillets. The NMR relaxations were analysed with the so-called SLICING method giving uni-exponential profiles from which the transverse relaxation time (T_2-values) and the relative sizes of the water populations were calculated. Two water populations with the T_2-values of 50 and 94 ms were obtained. The shortest relaxation time was primarily found near the head, and water with the longest relaxation time was primarily found near the tail. This variation can be explained by the smaller muscle cells and muscle fibers in the tail, which may influence the distributions of water into the different pools. The amount of one of the water populations was correlated to the overall water content with a correlation coefficient of −0.94.

Keywords: water distribution; water populations; NMR; SLICING

Introduction

Water is one of the most important components for the quality of food matrices including fish muscle. Water influences quality attributes such as appearance, texture and storage stability. The water content of fish muscles can be separated into different populations according to the mobility and how tight the water molecules are bound to the muscle structure. It is not only the total amount of water that is important for the overall quality of the product, but also its state and mobility (Ruan and Chen, 1998).

The heterogeneity within biological materials has implications on the results of most types of analyses made on that material. Therefore, it is important to know about this variation before analysing muscle-based food. Only very few studies have tried to characterize the heterogeneous distribution of the water content in lean gadoid fish species. Dambergs (1963) showed an increase in the overall water content of cod from the head-end to the tail-end. There was an inverse relationship between water and protein content within the fish muscle basically because these are the only significant components in the muscle.

This paper describes the distribution of water within cod fillets. The variation in water content and the variation in the populations of water within cod fillets will be investigated by measuring the water content physically and by measuring NMR relaxations. Low-field ^{1}H nuclear magnetic resonance (NMR) measures the mobility of protons and is therefore a direct technique for investigating both the total quantity of water and the state of water within the fish muscle. Low-field ^{1}H NMR has been used to measure and describe changes in properties of fish muscle occurring during frozen storage and processing (Lambelet *et al.*, 1995; Steen and Lambelet, 1997) and for investigating quality attributes of pork meat (Larsson and Tornberg, 1988; Brøndum *et al.*, 2000; Bertram *et al.*, 2001).

The use of a fast bench-top NMR hardware makes it possible and easy to acquire entire relaxation curves. In this paper, these will be analysed with a multivariate technique—SLICING—enabling interpretation of the underlying phenomena of the measurements. SLICING is a chemometric tool that resolves multivariate data, where the signals measured are an additive sum of exponentials. SLICING has been used for describing NMR relaxations of frozen and chill stored cod fillets (Jensen *et al.*, 2002), and minced and processed meat (Pedersen *et al.*, 2001).

*To whom correspondence should be addressed.
E-mail: aar@kvl.dk

0023-6438/02/$35.00

doi:10.1006/fstl.2002.0924

lwt/vol. 35 (2002) No. 8

Materials and Methods

Materials

Five cod were caught in Øresund, Denmark, in August 2000. They were immediately brought to the laboratory where they were stored on ice for 5–7 days such that all fish were in a post-rigor state when they were measured. The five fish were coded A–E and were weighed before filleting. The weights were 1.0, 0.7, 0.6, 1.3 and 1.5 kg for the fish, A, B, C, D and E, respectively. Only one fillet from each cod was used for the analyses, since it was supposed that there was no difference between the two sides of the fish.

NMR measurements

The cod fillets were divided in squares of 1.5 cm^2 retaining information on the anatomical position of the square. The samples were taken from each of the squares such that one sample is measured for every 1.5 cm both in the horizontal and the vertical direction of the fillet. The number of samples measured was 59, 37, 38, 58 and 62 for the five fish, respectively, giving 254 samples in total. Each of the samples were weighted and prepared for the measurements. Blood, bone and red muscle, though, were manually excluded from the samples.

The measurements were performed on a Maran Benchtop Pulsed NMR analyser (Resonance Instruments, Witney, U.K.) operating at 23.2 MHz and equipped with an 18 mm variable temperature probe head. The receiver was adjusted to 3% and the receiver delay was set to 6 s. Previous experiments showed that a receiver delay of 6 s was enough for recovering of the magnet. Transverse relaxations were measured using the Carr–Purcell–Meiboom–Gill (CPMG) sequence (Carr and Purcell, 1954; Meiboom and Gill, 1958). For each measurement eight scans were performed with 1024 echoes and Tau at 500 μs. Tau is the 90–180° interpulse spacing in the CPMG sequence. Only even echoes were recorded, which gives 512 echoes measured for each sample.

All measurements were performed at 4 °C. Before measuring, the samples were equilibrated for 30 min at the chosen temperature. The fish were introduced into the NMR probe by placing samples of 2–4 g into glass tubes that matched the inner diameter of the 18 mm NMR sample tubes. The sample preparation was performed as carefully as possible in order not to alter the muscle structure by the manual treatment.

Water content

The water content was determined on exactly the same samples as were used for the NMR measurements, after the NMR relaxations were measured. The fish samples were kept in the small glass tubes and dried overnight at 110 °C. They were weighed before and after drying.

Data analysis

The measurements of six of the squares taken from the fillet were considered as outliers. Three of these were removed because of extreme water content, due to measurement errors. The three remaining outliers were removed due to high modeling residuals in X and Y, caused by errors in the NMR-measurements. The decay of mobile protons is measured by ^1H low-field NMR relaxations. In cod, the protons that can be measured with low-field NMR will almost exclusively be found in the water molecules. Thus, the amplitude of the signal will depend on the amount of water in the sample and thus the weight of the sample since all the samples contain approximately the same concentration of water. In order to properly handle the fact that the measured samples are of different weight, the measurements are normalized using maximum normalization. By this method, the

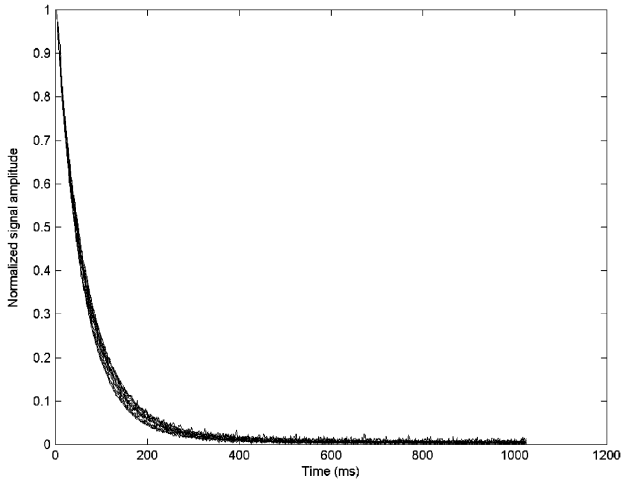


Fig. 1 Normalized CMPG curves. Each line represents the relaxation curve of one sample. Only some of the samples are shown

lwt/vol. 35 (2002) No. 8

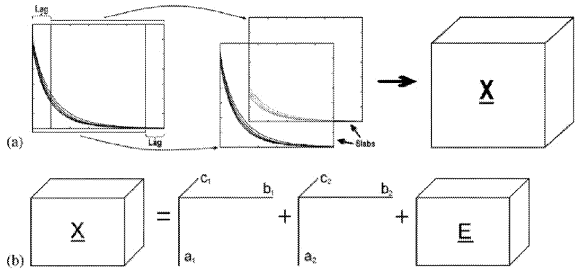


Fig. 2 (a) The SLICING method yielding three-way data from two-way data. (b) A visualization of the PARAFAC model. (b) is made on the three-way data from (a)

maximum of each relaxation curve is set to one and all other elements in the profile scaled accordingly.

Theory of SLICING. The normalized CMPG relaxation curves, illustrated in **Fig. 1**, are sums of exponentially decaying curves.

$$A(t) = \sum_{n=1}^{N} M_n e^{-t/T_{2,n}} \qquad \text{Eqn [1]}$$

Where $A(t)$ is the NMR signal, N is the number of exponential components, and M_n is the amplitude of the nth exponential and thus a measure of the relative concentration. $T_{2,n}$ is the corresponding spin–spin relaxation time constant and t is the acquisition time. Due to the normalization, the sum of the amplitudes (M_n) is one. Equation [1] represents the ideal case of the NMR signal. In real samples there will always be some noise, and a noise factor should then be included in the equation.

SLICING is a method based on the principles of direct exponential curve resolution algorithm (DECRA) (Windig and Antalek, 1997). The idea is to split the spectrum 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 relaxation curve is present in both slabs. This operation is illustrated in **Fig. 2a**. The dimensionality of the matrix will then increase from I (samles) $\times$ L (measurement points) to $I \times$ (L-maximum *lag*) $\times$ number of slabs (K). Below, this will simply be shortened to $I \times J \times K$.

The three-dimensional array (**X**) has the size $I \times J \times K$ and contains the elements x_{ijk}, where the first index (i) refers to the samples, the second (j) refers to the time and the third (k) refers to the slab number. It can be shown that the rearranged three-way data follow a so-called PARAFAC model (Eqn [2]) (Harshmann, 1970; Bro, 1997) when the original data is of the form described by Eqn [1]

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

$$(i = 1, \ldots, I; j = 1, \ldots, J; k = 1, \ldots, K) \qquad \text{Eqn [2]}$$

The element x_{ijk} is the original value in the position (i, j, k) of the data cube **X**. $a_{\bullet f}$ is the object score (magnitude) for factor f (first mode), $b_{\bullet f}$ is the estimated decay curve for the pure component f (second mode), and loading $c_{\bullet f}$ gives the ratio between the different slabs (third mode). The term e_{ijk} contains residual variation not captured by the model. Another way of presenting the PARAFAC model is shown in **Fig. 2b**. The **X**-cube represents the sliced data array. In the figure, the array is decomposed into two sets of factors, where each of these sets is called a triad. The factors (triads) are found simultaneously via an alternating least-squares algorithm (Bro, 1997) and are represented by a set of a–c. The **E**-cube to the right represents the noise that is left unmodeled. Noise is here defined as the variation in the data that contains no chemical information. Ideally, the estimated decay curves (b) are uni-exponential and by fitting one exponential to these, the corresponding T_2-value can be found.

If the residuals show random behavior and no systematic trend, it can be presumed that 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 second mode loading should be exponential because the PARAFAC model can be shown to uniquely recover the underlying model when correctly specified (Windig and Antalek, 1997). If too many components are extracted, the curves will reflect this, one or more being nonexponential. The appearance of the spectral loadings, bootstrapping (Wehrens *et al.*, 2000) by the use of split-halves (Harshmann and de Sarbo, 1994) and the distribution of the residuals will be used to estimate the right number of components.

Results and Discussion

According to Love (1970) there are no systematic differences between the chemical composition of the left and the right fillets of cod muscles. Thus, analyses performed on one of the fillets (left or right side) are used in this study to determine the general distribution of water within a fish.

Distribution of the water content

Figure 3 visualizes the distribution of the water content over the entire fillet. The figure shows the water content

lwt/vol. 35 (2002) No. 8

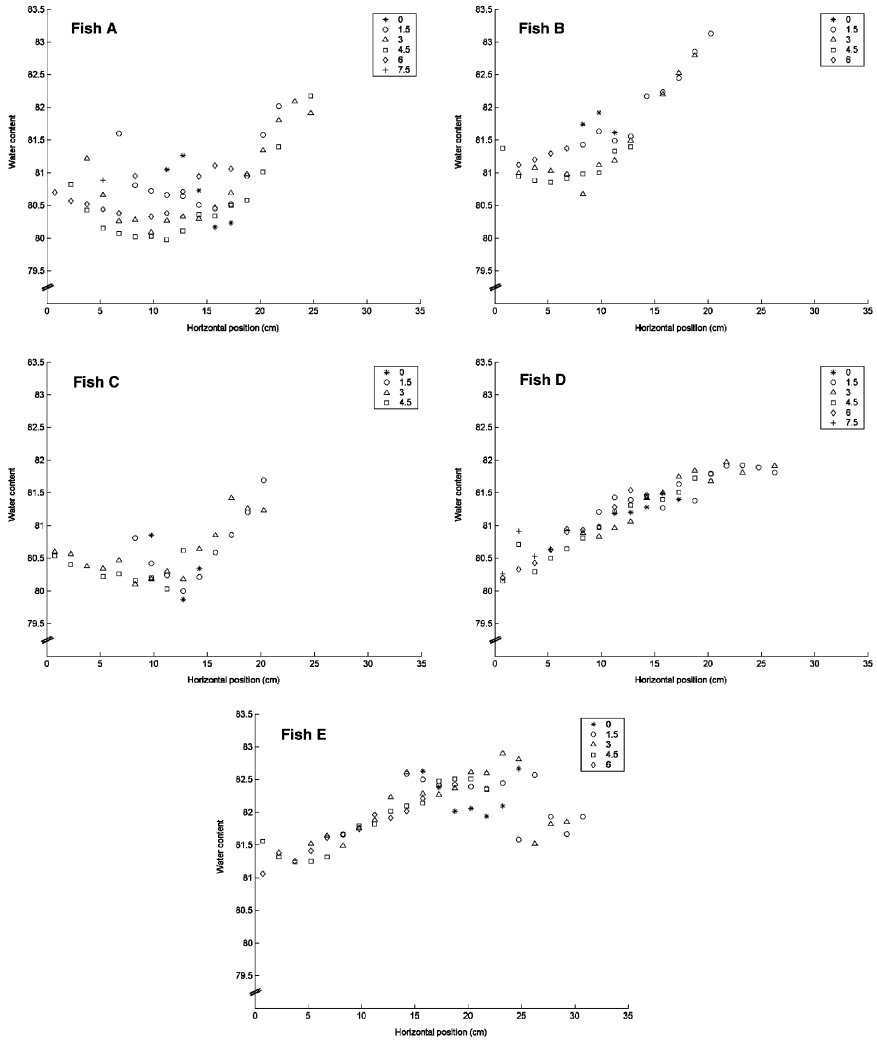


Fig. 3 Distribution of the water content within the fillets of the five fish. The water content is denoted in g water per g fish muscle and is shown as a function of the distance from the head (position 0). Differences across the fillet are shown with different marks: (*) is the sample closest to the belly, (○) is 1.5 cm from the belly, (△) is 3 cm from belly, (□) is 4.5 cm from belly, (◇) is 6 cm from belly and (+) is 7.5 cm from belly

as a function of the distance from the head. Generally, there is an increase in the water content going from the head to the tail. It seems that the smaller fish A–C have a similar water content distribution. Going from the head to the tail, the water content is getting a little lower towards the middle of the fish. Beyond that the water content increases with the highest content found at the tail-end of the fish. Fish D and E do not have the same distribution. The water content increases constantly from the head towards the tail. However, at some point the water content stops increasing, and in the last points it even decreases, especially for fish E. It is probable that this is reflecting a fundamental difference between smaller and bigger fish, but this cannot be investigated further with the present data.

The results correspond to the results obtained by Dambergs (1963). There, cod fillets were divided in three parts (head, middle/belly and tail) and showed an increase in water content going from the head to the tail. The results obtained by Dambergs (1963) were found as

lwt/vol. 35 (2002) No. 8

an average over the whole year and it is likely that the same variation would have been obtained in this study if more experiments were performed at other seasons. The opposite variation in water content was found in catfish and herring (Brandes and Dietrich, 1953; Jafri, 1973), which both have a higher fat content than cod.

Cell sizes and muscle fibers are smaller towards the tail compared with the rest of the fillet. This may have an influence on the distribution of the water content within the fillet corresponding to the results shown in the figure. Furthermore, the lower water content around 10 cm from the head seen for fish A–C can be explained by the fact that the largest muscle cells and muscle fibers are found around myotome number 12 (Love, 1988). An effect of sampling may explain some of the unexpected variations such as the decrease in water content in the outermost end of the tail. In some cases especially in the tail where very small samples are obtained, it is possible that a little part of red muscle is included in the samples because of the difficulty in separating red and white muscle in those small pieces of muscle flesh. Since red muscle has lower water content (Mannan *et al.*, 1961), it can explain why the water content decreases at the tail end.

SLICING modeling

The two-dimensional data were sliced using a lag of 1, and with two slabs, increasing the dimensionality of the matrix from 248 × 512 to 248 × 511 × 2.

Choosing the optimal number of factors. **Figure 4** illustrates the results obtained from slicing models with two or three factors. From the loading plot (**Fig. 4a**) it is clear that the two first factors are both exponential, indicating that at least two populations of water are present. It is imperative that the right number of components are chosen in the model, because all components change with the total number of components. For example, the faster relaxing factor has a T_2-value of 49.8 and 49.5 ms for the two- and three-factor models, respectively. The second T_2-values are 104.3 and 81.5 ms. The third factor for the three-factor model is probably not exponential, since it does not seem to decrease towards zero, but rather to a limit of approximately 0.03. There might be two reasons for this behavior: either the model is overfitted, or there is some kind of offset in the data. The offset may be due to the way data are collected in that all data were constrained to have positive values (H.T. Pedersen,

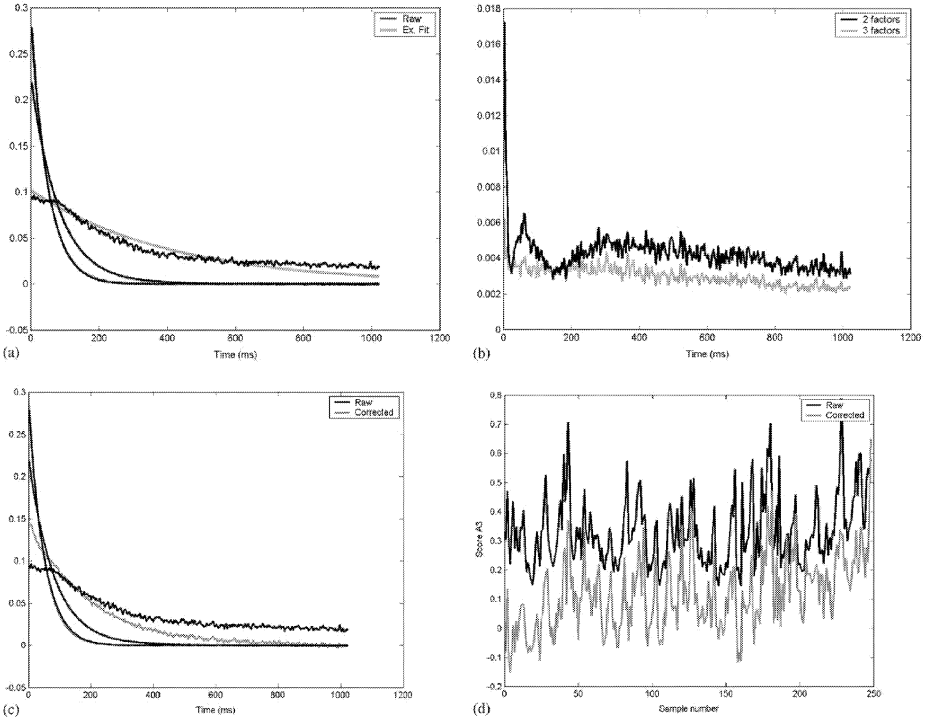


Fig. 4 (a) Second mode loadings (black) and exponential fit (gray) after three components. (b) Squared residuals summed over relaxation time from PARAFAC models with two (black) and three (gray) factors. (c) The B-loadings from three-factor models on the raw data (black) and the corrected data (gray). (d) The A-scores for the third factor. Black is the model on the raw data and the gray is the model on the corrected data

lwt/vol. 35 (2002) No. 8

pers. commun.). **Figure 4b** shows the residuals along the time axis for the two- and three-factor models. From this figure, it is clear that some information is gained by going from two to three factors. Thus, application of a three-factor model could be an advantage. Furthermore, when increasing the model complexity from two to three factors, the residuals become less structural as can be seen in **Fig. 4b**.

Corrections of the data. The possible presence of a minor third factor could be caused by an offset effect (**Fig. 4a**), as noted above. In order to investigate if the model could be improved by correcting for this artifact, the mean of the last 50 measuring points are subtracted from the whole spectra. The correction is done sample wise and is based on the assumption that the decays are measured sufficiently long so that the signal should effectively be zero unless there are offsets. As can be seen from **Fig. 4c**, the second mode loadings for the third factor has become exponential by the correction indicating the presence of a third minor factor. However, upon studying the scores (**Fig. 4d**) of the third factor, it becomes clear that some of these are negative, which gives no chemical meaning. Thus, only the two main factors give strictly positive score values.

Bootstrapping. To support the conclusion of a two-factor model, and further investigate the effect of correcting the data, bootstrapping (Wehrens *et al.*, 2000) is performed. Two- and three-factor SLICING models are made on 100 split-halves (Harshmann and de Sarbo, 1994) of the dataset with raw data and of the corrected data, giving rise to 4×200 models in total.
For comparing the results from the bootstrapping, uni-exponential fits to the second mode loadings are made. Mean and standard errors of the uni-exponential fits are calculated and are given in **Table 1**. Since the sample sets contain the same water populations, the standard error of the T_2-values should be as small as possible. Large standard errors indicate an unstable model, hence too many factors. By investigating the results in **Table 1**, it becomes clear that for both datasets the standard error of the third factor is very large. Furthermore, it was found that only 45% of the three-factor models made with corrected data have three exponential second mode loadings. The correction seems appropriate for the two-factor model since the standard error decreases going from the noncorrected to the corrected data. The bootstrapping thus supports the conclusion that the two-factor model is the optimal, but that the corrected data should be used.
For the two-factor model made on all data corrected for the offset, the fast relaxing factor has a T_2-value of approximately 50 ms and the slower relaxing factor has a T_2-value of 94 ms. Thus, one of the water populations is relatively tightly bound to the muscle structure and the other is less tightly bound.
Lambelet *et al.* (1995) obtained T_2-values of 1 and 65 ms by exponential fitting. The low T_2-value was only obtained when tau was as low as 12 μs. In the present experiment a tau of 500 μs was used and it therefore seems reasonable that a component with a T_2-value of 1 ms could not be detected. Modeling only one component gave a T_2-value of approximately 66 ms, which agrees with the slower relaxing component obtained by Lambelet *et al.* (1995). It seems that SLICING gives a possibility for identifying two water populations with great similarities, which could not be identified with the bi-exponential fitting as was used by Lambelet *et al.* (1995). Another study made on frozen–thawed cod gave three exponential components (Jepsen *et al.*, 1999). These had T_2-values of 34, 62 and 526 ms. Four populations of water with T_2-values of 37, 56, 126 and 361 ms were obtained by the SLICING technique applied on NMR-relaxations of minced frozen–thawed cod (Jensen *et al.*, 2002). The two components with the T_2-values of 56 and 126 could correspond to the water populations obtained in the present study. Thus, it seems as two more water populations are developed during processing such as mincing or freezing. Furthermore, the variation in T_2-values between this study and other studies may be due to variations in the sample preparation, season, places of catch, etc. Differences in instrumentation, parameters used and the ways of analysing the data may also cause a variation in the experimentally determined T_2-values.
For whole, minced and homogenized pork three exponential components were obtained by distributed exponential fitting (Fjelkner-Modig and Tornberg, 1986; Larsson *et al.*, 1988; Bertram *et al.*, 2001). As mentioned, only two components were obtained in the present study. It might be that fresh fish muscle in contrast to pig muscle contains only two populations of water. Furthermore, the variation and/or the content of the third water population may be so small that it cannot be described by the method used here.

Table 1 Mean relaxation times (in ms) and the standard error of second mode loadings obtained from two and three factor bootstrapping on 100 split-half models performed on both noncorrected and corrected data.

	Uncorrected data				Corrected data			
	Two factors		Three factors		Two factors		Three factors	
	Mean	s.e.	Mean	s.e.	Mean	s.e.	Mean	s.e.
$T_{2,1}$	49.8	0.6	49.5	0.6	49.6	0.4	46.5	1.8
$T_{2,2}$	104.3	2.8	81.5	1.7	93.7	1.2	82.6	5.7
$T_{2,3}$			467.7	100.8			203.1	73.6

The number of samples for these tests is 248.

lwt/vol. 35 (2002) No. 8

Interpretation of the SLICING model

Distribution of the different types of water between the cod fillets. The score values (first mode) describe how much each water population contributes to the NMR relaxation phenomena and may thus be used as a measure of the relative concentrations of the populations of water. **Figure 5** shows a score plot of factor one (relating to $T_{2,1}$) *vs.* factor two (relating to $T_{2,2}$) of all the fish. The plot is coded according to each of the fish (A–E). The ellipses are made for helping in the interpretation of the plot. The ellipses are centered in the mean of each fish along each of the two axes. The range in each direction is the standard deviation for that fish's scores in the specific direction of variation.

By comparing **Fig. 5** with **Fig. 3**, it might seem odd that fish B and D look so similar in **Fig. 5**. However, by investigating **Fig. 3** further, and taking into account that the NMR-signal is mostly due to the water in the fish, it comes as no surprise that fish B and D overlap in **Fig. 5**. The range of the water content for fish B is higher than for fish D, but the range of the water content of fish D is inside the range of fish B. Therefore, it is expected that the ellipse for fish D should be inside the ellipse for fish B, which also is the case in **Fig. 5**.

From **Fig. 5,** it is seen that the samples from the five fish are grouped together illustrating some variation among the fish. The grouping shown in **Fig. 5** can be related to the average water content of the five fish. Fish E has the highest water content of 82.02/100 g, fish B and D have the next highest of 81.53/100 g and 81.20/100 g, respectively, whereas fish A and C have the lowest water content of 80.74/100 g and 80.53/100 g, respectively. There is an inverse relation between these two types of water. The reason for this is that by normalizing the spectra to one, the scores will approximately add up to a constant value. Thus, when there are only two factors in the model and the noise level of the raw data is low, the score values will be highly correlated.

Comparing heterogeneity within the fish. **Figure 6** illustrates the scores of the first factor *vs.* the relative position of each sample. The scores are obtained from the two-factor SLICING model made on all fish with correction for offset. The results will be more or less the same if the figures are made using the results obtained from PARAFAC models made on each fish (figure not shown). Further, if the scores from the second factor were used instead of the scores from the first, the same results would be obtained. This, however, is of no surprise since the two factors are highly correlated.

There is a clear relation between the scores obtained from the PARAFAC model and the position of the sample. The black line in each figure is the 'best line' fitted to minimize the distance from every sample to the line (corresponding to the first component in a PCA). The correlation coefficients illustrate that fish A, C and E are fairly homogenous, whereas the largest correlation between the score values and the horizontal position is found for fish B (**Table 2** and **Fig. 6**). However, the same results are not obtained by looking at the range of the values, where the greatest variation is found for fish E. The low correlation and large range obtained for fish E seems to be due to a curvature in the score-values, and is consistent with what is seen in **Fig. 3**. Fish A has two samples close to the head, which have fairly low score-values compared to the others. By removing these two samples, the correlation coefficient increases to −0.66. It

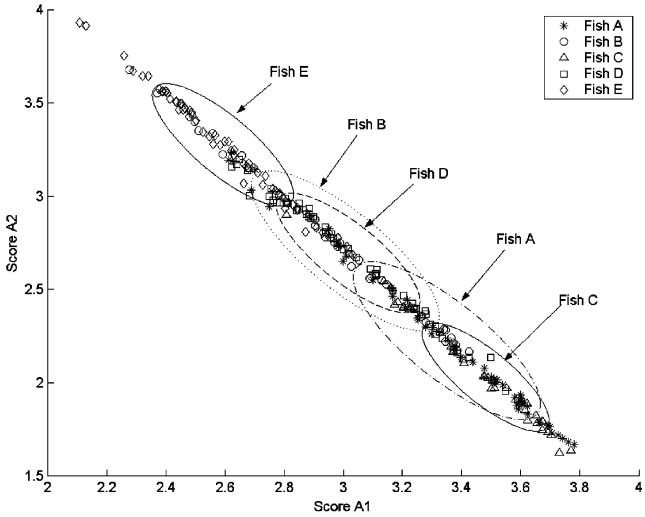


Fig. 5 Scatter plot of the scores of factor 1 *vs.* the scores of factor 2 obtained from a two-factor PARAFAC model on the raw data. Samples are coded according to fish: fish A (*), fish B (○), fish C (△), fish D (□) and fish E (◇)

lwt/vol. 35 (2002) No. 8

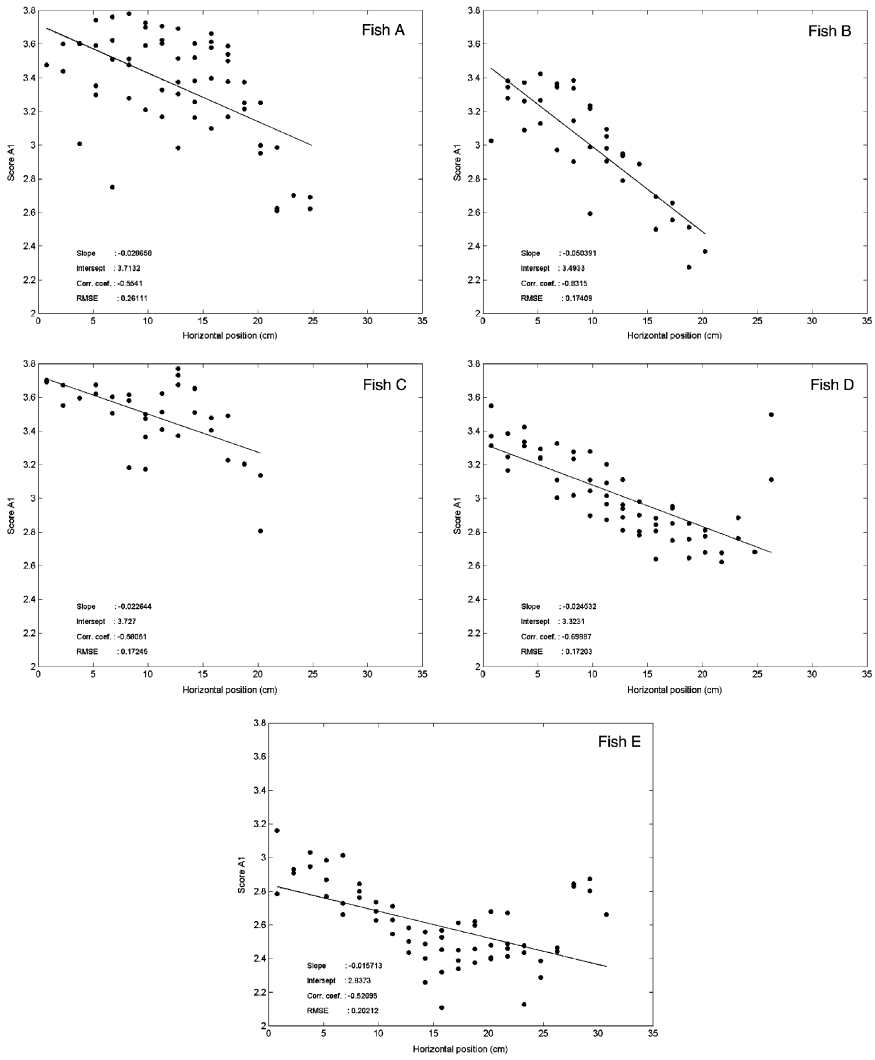


Fig. 6 Plot of the horizontal position of each sample *vs.* the A-scores of the first factor obtained from a two-factor PARAFAC model on the raw data

Table 2 Correlation between the first score and the horizontal position (r), and the range in the first score. Both are done for each fish separately

	A	B	C	D	E
No. of samples	59	36	34	57	62
r	−0.57	−0.85	−0.69	−0.74	−0.75
Range	1.20	1.05	0.88	1.06	1.09

becomes clear that fish A is homogenous up to the tail-end where the score values start to drop. Fish C seems to be homogenous in the middle of the fish, while to both sides of the middle there is a steady decrease of the score values. The reason why fish A and C in some part of the fillet are more homogenous than the other fish is not known.

There are some samples near the tail of fish D, which clearly do not follow the trend of the other samples. This might be due to some red muscle being included in the samples, and as mentioned earlier, red muscle contains less water than white muscle and will thus give a different signal. If these few samples at the tail-end of the fish are removed, the results for fish D improve,

lwt/vol. 35 (2002) No. 8

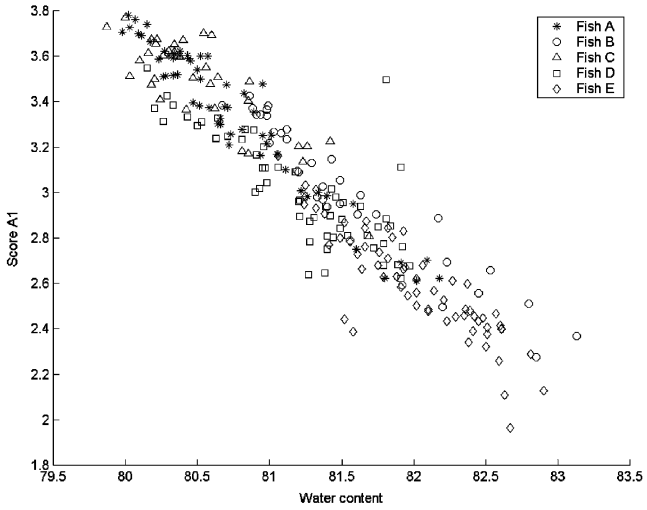


Fig. 7 Relation between the water content and the score values of the first factor

giving a correlation coefficient of −0.88. Thereby, the relation between the score values and the sample position increases resulting in a larger correlation coefficient than is obtained for fish B.

The negative correlation between the score values of the two factors means that samples with a high content of the water population described by factor one will have a low content of the water population described by factor two. Thus, samples from the head part contain more of the fast relaxing component, whereas samples from the tail part contain more of the slow relaxing component. This is consistent with theory. The muscle fibers are smaller towards the tail where the cell sizes also get smaller (Love, 1970). It is suggested that a separation of the water into two populations is due to an anatomical compartmentalization. The water with the shortest relaxation time is likely intracellular water and the water with the longest relaxation time is extracellular water (Cole *et al.*, 1993). In that case the intracellular water that is found in a higher amount near the head is bound more firmly to the muscle fibers than the extracellular water. The more loosely bound extracellular water is found in a larger amount near the tail. The relative size of intra- and extracellular water is in agreement with the variation in cell and muscle fiber sizes. Furthermore, the distribution of the score values is in accordance with the theory of the separation of water into the intracellular and the extracellular parts.

Above, it was shown that the score values of the first factor decrease when samples are taken from a position closer to the tail. This was also the case for the overall water content (**Fig. 3**). Therefore, the relative amount of each water population may correlate to the total amount of water. This is illustrated in **Fig. 7** where the water content is plotted against the score values of the first factor. It is illustrated that there is a linear relation between the two parameters. Furthermore, a correlation coefficient of −0.94 is found. This variation in water content and content of the different water populations illustrates the importance of how sampling is performed when instrumental measurements are performed on a small part of the fish.

As mentioned, other studies have already shown the multi-exponential distribution of water within muscle tissue. However, low-field NMR combined with SLICING is a new way of analysing the states of water in various products. Compared with exponential fitting, SLICING is a fast, easy and stable method giving the underlying exponential decays of the NMR relaxations. When analysing the NMR relaxations of fresh cod fillets, it has been shown that two populations of water can be identified. The concentration of the populations varies within the length of the cod fillet and seems to be related to the overall water content. Thus, these results illustrate the possibility for studying the distribution of the populations of water between and within other fish species.

Acknowledgements

Andersen thanks for financial support through Advanced Quality Monitoring (AQM), while Rinnan wish to thank the STVF (Danish Research Council) for financial support through project 1179. Rasmus Bro, Frans van den Berg and Bo Jørgensen are gratefully acknowledged for valuable advice during the data analysis and preparation of the manuscript.

References

Bertram, H. C., Karlsson, A. H., Rasmussen, M., Pedersen, O. D. Dønstrup, S. and Andersen, H. J. Origin of

lwt/vol. 35 (2002) No. 8

multiexponential T_2 relaxation in muscle myowater. *Journal of Agricultural and Food Chemistry*, **49**, 3092–3100 (2001)

Brandes, C. H. and Dietrich, R. Die Fettverteilung im Koerper des Herring. *Veroffentliches Institut für Meeresforshung in Bremerhaven*, **2**, 109–121 (1953)

Bro, R. Parafac: tutorials and applications. *Chemometrics and Intelligent Laboratory Systems*, **38**, 149–171 (1997)

Brøndum, J., Munck, L., Henckel, P., Karlsson, A., Tornberg, E. and Engelsen, S. B. Prediction of water-holding capacity and composition of porcine meat by comparative spectroscopy. *Meat Science*, **55**, 177–185 (2000)

Carr, H. Y. and Purcell, E. M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, **94**, 630–638 (1954)

Cole, W. C., Leblanc, A. D. and Jhingran, S. G. The origin of biexponential T2 relaxation in muscle water. *Magnetic Resonance in Medicine*, **29**(1), 19–24 (1993)

Dambergs, N. Extractives of fish muscle. 3. Amounts, sectional distribution, and variations of fat, water-solubles, proteins and moisture in cod (*Gadus morhua* L.) fillets. *Journal of Fisheries Research Board, Canada*, **20**, 909–918 (1963)

Fjelkner-Modig, S. and Tornberg, E. Water distribution in porcine *M. longissimus dorsi* in relation to sensory properties. *Meat Science*, **17**, 213–231 (1986)

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

Harshmann, R. A. and De Sarbo, W. S. An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints and split-half diagnostic techniques. In: Law, H. G., Snyder, C. W., Hattie, J. A. and McDonald, R. P. (Eds), *Research Methods for Multimode Data analysis*. New York: Praeger (1994)

Jafri, A. K. Fat and water distribution patterns in the flesh of the common cat-fish *Wallago attu* (bl. & schn.). *Fishery Technology*, **10**, 138–141 (1973)

Jensen, K. N., Guldager, H. S. and Jørgensen, B. M. Three-way modelling of NMR relaxation profiles from thawed cod muscle. *Journal of Aquatic Food Product Technology* (2002) in press

Jepsen, S. M., Pedersen, H. T. and Engelsen, S. B. Application of chemometrics to low-field ^{1}H NMR relaxation data of intact fish flesh. *Journal of the Science of Food and Agriculture*, **79**, 1793–1802 (1999)

Lambelet, P., Renevey, F., Kaabi, C. and Raemy, A. Low-field nuclear magnetic resonance relaxation study of stored or processed cod. *Journal of Agricultural and Food Chemistry*, **43**, 1462–1466 (1995)

Larsson, G. and Tornberg, E. An attempt to relate meat quality og pork (*M. longissimus*) to meat structure. *34th International Congress of Meat Science and Technology. Congress Proceedings Part B*, pp. 588–591 (1988)

Love, R. M. *The Chemical Biology of Fishes*. New York: Academic Press Inc. (1970)

Love, R. M. *The food fishes, their intrinsic variation and practical implications*. London: Farrand Press, (1988)

Mannan, A., Fraser, D. I. and Dyer, W. J. Proximate composition of Canadian Atlantic fish 1. Variation in composition of different section of the flesh of atlantic halibut (*Hippoglossus hipoglossus*). *Journal of Fisheries Research Board, Canada*, **18**(4), 483–493 (1961)

Meiboom, S. and Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, **29**, 688–691 (1958)

Pedersen, H. T., Bro, R. and Engelsen, S. B. A novel approach for unique deconvolution of NMR relaxation decays. In: Webb, G. A., Belton, P. S., Gill, A. M. and Delgadillo, I. (Eds), *Magnetic Resonance in Food Science: A View to the Future*. Cambridge, UK. The Royal Society of Chemistry, pp. 202–209 (2001)

Ruan, R. R. and Chen, P. L. *Water in Foods and Biological Materials. A Nuclear Magnetic Resonance Approach*. Lancaster: Technomic Publishing Co., Inc (1998)

Steen, C. and Lambelet, P. Texture changes in frozen cod mince measured by low-field nuclear magnetic resonance spectroscopy. *Journal of the Science of Food and Agriculture*, **75**, 268–272 (1997)

Wehrens, R., Putter, H. and Buydens, L. M. C. The bootstrap: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, **54**, 35–52 (2000)

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

Paper IV

On the relation between water pools and water holding capacity in cod muscle

Charlotte M. Andersen and Bo M. Jørgensen

Abstract

Low-field ^{1}H nuclear magnetic resonance (NMR) relaxations were measured on muscle, minced muscle and centrifuged mince from cod that had been treated under various freeze and chill storage conditions. Using multi-way chemometrics, uni-exponential profiles were obtained, from which the transverse relaxation times (T_2-values) and the relative water pool sizes (M-values) were determined. Three pools of water were identified with the relaxation times and M-values in the centrifuged samples reflecting the removal of loosely bound water. The M-values and the full NMR-signal decays were correlated to two measures of water holding capacity (WHC) in a way that WHC related to the original water content could be predicted well from the whole and the minced muscle. The centrifuged samples gave optimal predictions of WHC related to the dry matter content, probably because the centrifuged samples are similar to the samples used for measuring WHC. The two WHC quantities were also correlated to the water distribution itself as measured by the relative pool sizes.

Keywords: fish, water mobility, NMR relaxation, PARAFAC

Introduction

Water is the main constituent of muscle and has as such an important influence on quality-related parameters like appearance, texture and storage stability of muscle-based food. However, it is not merely the total water content that governs stability and quality. Rather, the cellular distribution, binding state and mobility of the water is important (*1*).

Muscle tissue water can be considered as distributed into different “pools” according to its mobility, which depends on how tight the water molecules are bound to sub-cellular structures, e.g. myofibrils. In porcine meat, three pools of water have been identified: one trapped between the fibres, one outside the myofibril structure or even outside the muscle cells, and one tightly bound to the muscle proteins (*2*). The distribution of water between these three pools depends on factors like storage time, freezing, mincing etc. Changes in water distribution are mainly a result of damage to proteins leading to denaturation and aggregation. This results in a decreased protein solubility and loss of functional properties such as the ability to hold water. The water holding capacity is an important quality parameter correlating with sensory attributes like texture and juiciness (*3*).

Time-domain ^{1}H low-field nuclear magnetic resonance (NMR) spectroscopy is a technique that measures nuclear spin relaxation of hydrogen atoms. The relaxation rate depends on the mobility of the nucleus and thus the water mobility in the sample of question. Transverse spin relaxations in heterogeneous material normally follow a multi-exponential decay, indicating that there are several groups of water molecules with different mobility. The number of exponential components, characterized by their relaxation time constants (T_2-values), depends on the

physical and chemical conditions of the system and corresponds to the water pools existing in the muscle tissue. It has been claimed that water does not exist in distinct pools, but that each proton has its own environment and thereby its own T_2-value (*4*,*5*). From this it follows that a distribution of T_2-values rather than a few discrete values should be sought for. However, it is in fact possible to resolve the NMR relaxation curves into a few mono-exponential components by the use of chemometric methods (*6*,*7*), possibly because exchange between slightly differently positioned water molecules within a certain micro-environment is faster than the resolution of the NMR instrument.

The present paper describes the distribution of water in cod muscle determined from low-field NMR relaxation measurements by chemometrics. The relation between the water distribution and the water holding capacity is shown for fish muscle treated under various processing and storage conditions. Additionally, the ability to predict water holding capacity from low-field NMR relaxation curves by the classical bilinear method, partial least squares (PLS) regression, is demonstrated.

Materials and methods

Sample preparation

Cod were caught in the Barents Sea or the Baltic Sea and underwent twelve different treatments, varying in freeze storage conditions and chill storage time (Table 1). The cod were filleted, de-skinned and vacuum packed during the chill storage period. Two fillets were measured in each code.

Samples of fish muscle were taken from the loin part at the inner side of the fillet and placed carefully in the NMR glass tubes in order not to alter the muscle structure by the manual treatment. The rest of the loin part of the fillet was minced and used for determination of water holding capacity and dry matter content as well as for NMR measurements on the minced samples. A third type of sample was used for the NMR measurements, namely the centrifuged mince from the analysis of water holding capacity (see below). All samples were of approximately 3 g. Measurements were performed in triplicate.

Table 1 Treatments and storage conditions of the samples.

	Place of catch	Treatment	Freeze storage temperature (°C)	Freeze storage time	2 weeks of chill storage
1	Barents Sea		-24	24 months	yes
2	Barents Sea		-24	24 months	no
3	Barents Sea	glaced	-24	24 months	yes
4	Barents Sea	glaced	-24	24 months	no
5	Barents Sea	salt glaced	-24	24 months	yes
6	Barents Sea	salt glaced	-24	24 months	no
7	Baltic Sea		-24	2 months	yes
8	Baltic Sea		-24	2 months	no
9	Baltic Sea		-20/-30[a]	14 weeks	yes
10	Baltic Sea		-20/-30[a]	14 weeks	no
11	Baltic Sea		-30	2 weeks	yes
12	Baltic Sea			none	yes

[a]The cod were stored at fluctuating temperature.

NMR measurements

The measurements were performed with a Maran Benchtop Pulsed NMR analyser (Resonance Instruments, Witney, UK) operating at 23.2 MHz and equipped with an 18 mm variable-temperature probe head. Transverse spin-relaxation was measured using the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (*8*,*9*). Only even echoes were sampled, spaced by 1.2 ms, and the number of sampling points was 512 for each measurement. The probe temperature was 30 °C, and samples were thermo-equilibrated for 30 min before measurements.

Water holding capacity

Water holding capacity was determined on approximately 2 g of minced fillet taken from the loin piece. The samples were weighed, placed in plastic tubes with a special filter bottom (pore size 0.1 mm), centrifuged (1500*g*, 10 °C, 5 min) and weighed again (*10*). Each analysis was made in quadruplicate. Water holding capacity is defined in two ways: as the amount of water left in the mince after centrifugation relative either to the amount of water originally present in the samples (WHC_w) or to the dry matter content (WHC_d).

Data Analysis

The NMR relaxation curves were divided by the sample weight and the number of variables reduced from 512 to 256 by averaging each second sample point. The transverse spin-relaxation curve is a sum of exponential decays:

$$A_i(t) = \sum_{f=1}^{F} m_{if} e^{\frac{t}{T_{2f}}} + noise \quad (1)$$

where $A_i(t)$ is a vector containing the NMR signal from sample i at time t, F is the number of exponential components corresponding to the number of water pools, m_{if} is the amount of the f-th component in sample i and T_{2f} is the spin-spin relaxation time of the water pool described by component f. A low amount of noise is also present as indicated.

The F components were extracted from the measured relaxation curves by tri-linearization and parallel factor analysis (PARAFAC; *11-13*) as previously described (*6*,*7*,*14*). This procedure is based on the principles of the Direct Exponential Curve Resolution Algorithm (*15*). The PARAFAC model recovers uniquely the underlying structure of tri-linear data as so-called loadings. Thus, from the NMR curves, mono-exponential loadings are obtained, and from the logarithm of these, the transverse spin-spin relaxation times (T_2) can be determined by weighted linear regression. Subsequently, the exponentials can be generated and the size of the water pools, the m-values, calculated from Equation (1) by multiple linear regression. The size of the water pool is given as the absolute value (m-value) or as a value (M-value) relative to the total amount of water.

Multivariate calibrations for predicting the measured water holding capacities from the low-field NMR relaxation curves themselves or from the determined m-values were made by the use of PLS regression and multiple linear regression (MLR), respectively. The regression models were evaluated by the correlation between the measured and the predicted values and by the root mean squared error of prediction as estimated by cross-validation (RMSECV). A jack-knife approach was used for evaluating the importance of the m-values as measures of the water holding capacity. An excellent review of these methods is given in the monograph by Martens and Martens (*16*).

The data analysis was performed using Matlab® Ver. 5.3 for Windows (Math Works, USA) and The Unscrambler® Ver. 7.6 (CAMO, Norway).

Results and discussion

Identifying the water pools

The first and important step in describing the water pools is to determine the number of pools. This is because the PARAFAC solution, based on minimising the sum of squared residuals, is dependent on *F* (in Equation 1). In the present case, split-half analysis (data split between replicates) (*17*), principal component analysis, and examination of how well the loadings obtained were true single exponentials, all pointed to the existence of three components in intact cod muscle tissue as well as in the minced muscle. A more thorough description on how to estimate *F* from this type of data is found in Jensen et al. (*7*). This reference reported on a fourth component in cod mince. The discrepancy is probably due to differences in the storage conditions of the fish. In another study, low-field NMR relaxation measurements on fresh cod muscle showed two exponential components only (*18*). This indicates that a third water pool develops during frozen storage, which was also shown by Lambelet et al. (*19*). Possibly, the development of a third exponential component during freeze storage is due to protein denaturation (*19*).

The estimated T_2-values (Table 2) show that the centrifuged samples had faster relaxing components, i.e. less mobile water, except for T_{21}. The relaxation parameters measured in whole and minced samples were similar, showing that the change in structure caused by the mincing process did not influence the mobility of water. The structure partly collapses by the mincing. This apparently results in a lower relative amount of tightly bound water (cf. the *M*-values in Table 2) and is possibly due to a larger protein-water interface area. Water is removed by the centrifugation leading to a low amount of the loosest bound water. A similar result was reported by Bertram et al. (*20*) who showed that there was also a slight increase in the fastest relaxing water fraction in centrifuged samples of minced pig muscle.

Table 2 T_2 and *M*-values of the three water pools in intact muscle, mince and centrifuged mince from cod treated as described in the present paper. The *M*-values are the relative concentration of water in each pool and are given in percent as the average of all samples.

	Intact muscle	Mince	Centrifuged mince
T_{21} (ms)	35	39	41
T_{22} (ms)	83	87	66
T_{23} (ms)	467	394	230
M_1 (%)	71	60	59
M_2 (%)	10	10	35
M_3 (%)	19	30	6

The water pool sizes, as expressed by the *m*-values (Equation 1), apparently did not vary in a systematic way governed by the treatment of the fish. One reason may be the existence of a large biological variation between fish within the same code, blurring a smaller effect of the different treatments in the experimental design. It was expected that samples from batch 12, which had not been frozen, would exhibit a unique water distribution, since freezing changes the structure of the muscle proteins and thereby affects the binding of water molecules (*21*,*22*). However, this was not apparent from the present results.

Predicting the water holding capacity

The water holding capacity was determined by centrifuging the samples and measuring how much water that disappeared during the centrifugation step. It is reported in two ways: as the amount of water left in the sample relative to the original amount of water (WHC_w), or as the amount of water left in the sample relative to the amount of dry matter (WHC_d). These two quantities are perfectly correlated as long as the dry matter content does not vary among samples. In the present case, the variation in dry matter content was substantial, causing an observable difference in the relation between the NMR relaxation and the two measures of WHC. Furthermore, both WHC and dry matter content varied between samples to a degree high enough that usable regression models could be made.

WHC_w and WHC_d were predicted by PLS regression using as *x*-variables the whole relaxation curves, and by MLR using as *x*-variables the *m*-values obtained from Equation 1. The predictions were validated by cross-validation, leaving replicates in the same segment, and evaluated by the correlation between the measured and the predicted water holding capacity and by RMSECV. The results are shown in Table 3 and 4. The predictions of the water holding capacity performed well, illustrating that low-field NMR is a valuable method for assessing water holding capacity in fish muscle, as was also indicated by Jepsen et al. (*23*). It is interesting that when expressing the water holding capacity relative to the original amount of water (WHC_w), optimal predictions were obtained on whole and minced muscle. The poor predictions obtained from the centrifuged fillets are reasonable since water is removed by the centrifugation and the samples contain no information about the original water content.

Conversely, the centrifuged samples gave the best predictions when expressing water holding capacity relative to the dry matter content. The centrifuged samples were given exactly the same treatment (mincing and centrifugation) as the samples used for estimating the water holding capacity. Therefore, it is reasonable that these samples give a more accurate estimate of WHC_d and thus better predictions.

Table 3 Prediction of percent water left after centrifugation relative to the original amount of water (WHC_w) from low-field NMR relaxations, *m*-values obtained by slicing. Variation in WHC_w is between 68 and 88 %. RMSECV is given in per cent.

Whole fillet		Minced fillet		Centrifuged fillet	
R	RMSECV	r	RMSECV	r	RMSECV
PLS on whole relaxations[a]					
0.86	2.95	0.77	3.75	0.60	4.80
MLR on m-values					
0.77	3.84	0.68	4.33	0.48	5.47

[a]The PLS regressions are made with three components for whole and minced fillet and one component for centrifuged fillet.

Table 4 Prediction of percent water left after centrifugation relative to the amount of dry matter (WHC_d) from low-field NMR relaxations, *m*-values obtained by slicing. Variation in WHC_w is between 2.37 and 3.92 g/g. RMSECV is given in the original values (g/g).

Whole fillet		Minced fillet		Centrifuged fillet	
R	RMSECV	r	RMSECV	R	RMSECV
PLS on whole relaxations[a]					
0.77	0.817	0.79	0.233	0.92	0.157
MLR on m-values					
0.77	0.248	0.70	0.276	0.92	0.156

[a]The PLS regressions are made with three components.

A main objective of this study has been to determine the relation between the different water pools and the water holding capacity. This was evaluated from the multiple linear regressions by a jack-knife approach estimating the uncertainty of the regression coefficients (16). Predicting WHC_w of whole and minced fillet from the *m*-values showed that both the fastest relaxing water and the slowest relaxing water are significant. This indicates that a large amount of water tightly bound to the muscle structure and a low amount of water loosely bound ensures a high water holding capacity. The slowest relaxing, i.e. loosest bound water is removed during centrifugation as seen in Table 2. Probably, part of this water is also lost for samples with decreased water holding capacity. It is therefore reasonable that this water pool is important for the quantification.

When predicting WHC_d, the centrifuged samples gave the best predictions, from which the two fastest relaxing components were found to be the most important. By centrifugation, the loosely bound water is largely removed. Only a very low concentration of the slow relaxing water is left in the samples having but a small influence on the regression model.

The results show that low-field NMR is a valuable tool for improved understanding of the relation between water mobility and of fish muscle. Furthermore, due to its potential as a non-destructive method for measuring water holding capacity, it could be of use in the industry for at-line or even on-line applications.

Acknowledgements

This work was supported through the AQM (Advanced Quality Monitoring) framework programme financed by the Danish National Research Council.
The excellent experimental work by Ms. Ulla Scherfig is gratefully acknowledged.

References

(1) Ruan, R. R; Chen, P. L. *Water in foods and biological materials. A nuclear magnetic resonance approach*; Technomic Publising Company, Lancaster, 1998.

(2) Bertram, H.C.; Karlsson, A.H.; Rasmussen, M.; Pedersen, O.D.; Dønstrup, S.; Andersen, H.J. Origin of multiexponential T_2 relaxation in muscle myowater. *J. Agric. Food Chem.* **2001,** *49*, 3092-3100.

(3) Jensen, H.S.; Jørgensen, B.M A sensometric approach to cod-quality measurement. *Food Qual. Prefer.* **1997**, *8*, 403-407.

(4) Brown, B. J. S.; Capozzi, F.; Cavani, C.; Cremonini, M. A.; Petracci, M.; Placucci, G. Relationship between ^{1}H NMR relaxation data and some technological parameters of meat: a chemometric approach. *J. Mag. Reson.* **2000**, *147*, 89-94.

(5) Bertram, H.C.; Karlsson, A.H.; Andersen, H.J. Comparative study of low-field NMR relaxation measurements and two traditional methods in the determination of water holding capacity of pork. *Meat Sci.* **2001**, *57*, 125-132.

(6) Pedersen, H.T.; Bro, R.; Engelsen, S.B. Towards rapid and unique curve resolution of low-field NMR relaxation data: trilinear slicing versus two-dimensional curve fitting. *J. Magn. Reson.* **2002**, *157*, 141-155.

(7) Jensen, K.N., Guldager, H.S., Jørgensen, B.M. (2002): Three-way modelling of NMR relaxation profiles from thawed cod muscle. *J. Aquat. Food Prod. Technol.*, **2002**, 11, 201-214.

(8) Carr, H. Y.; Purcell, E. M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Am. J. Physiol.* **1954**, *94*, 630-638.

(9) Meiboom, S.; Gill, D. Modified spin-echo method for measuring nuclear times. *Rev. Sci. Instr.* **1958**, *29*, 688-691.

(10) Eide, O.; Børresen, T.; Strøm, O. Minced fish production from capelin (*Mallotus villosus*). A new method for gutting, skinning and removal of fat from fatty species. *J. Food Sci.* **1982**, *47,* 347-349, 354.

(11) Harshman, R.A. Foundations of the PARAFAC procedure: model and conditions for an "explanatory" multi-mode factor analysis. *UCLA Working Papers in phonetics*. **1970**, *16*, 1-84.

(12) Harshman, R. A.; Lundy, M. E. PARAFAC: Parallel factor analysis. *Comput. Stats. Data Anal.* **1994**, *18*, 39-72.

(13) Bro, R. PARAFAC. Tutorial and applications. *Chemometr. Intell. Lab. Syst.* **1997**, *38*, 149-171.

(14) Jørgensen, B.M. Multivariate spectrometric methods for determining quality attributes. In *Safety and quality issues in fish processing*; Bremner, H. A. Eds.; Woodhead Publ. Ltd., Cambridge, UK, 2002; pp. 475-494.

(15) Windig W.; Antalek, B. Direct exponential curve resolution algorithm (DECRA): a novel application of the generalised rank annihilation method for a single spectral mixture data set with exponentally decaying contribution profiles. *Chemom. Intell. Lab. Syst.* **1997**, *37*, 241-254.

(16) Martens, H.; Martens, M. *Multivariate Analysis of Quality. An Introduction.* J. Wiley & Sons Ltd., Chichester, UK, 2001.

(17) Harshman, R.A. "How can I know if it's real?" A catalog of diagnostics for use with three-mode factor analysis and multidimensional scaling. In: *Research methods for multimode data analysis*; Law, H.G., Snyder, C.W, Hattie J.A., McDonald R.P., Eds.; Praeger, New York, 1984; pp 566-591.

(18) Andersen, C.M.; Rinnan, Å. Distribution of water in fresh cod. *Lebenm. Wiss. U. Technol.* **2002**, *35*, 687-696.

(19) Lambelet, P.; Renevey, F.; Kaabi, C.; Raemy, A. Low-field nuclear magnetic resonance relaxation study of stored or processed cod. *J. Agric. Food Chem.*, **1995**, *43*, 1462-1466.

(20) Bertram, H.C.; Purslow, P.P.; Andersen, H.J. Relationship between meat structure, water mobility and distribution: A low-field nuclear magnetic resonance study. *J. Agric. Food Chem.* **2002**, *50*, 824-829.

(21) Mackie, I.M. The effects of freezing of flesh proteins. *Food Rev. Int.* **1993**, *9*, 575-610.

(22) Shenouda, S.Y.K. Theories of protein denaturation during frozen storage of fish flesh. *Adv. Food Res.* **1980**, *26*, 275-311.

(23) Jepsen, S.M.; Pedersen, H.T.; Engelsen, S.B. Application of chemometrics to low-field ^{1}H NMR relaxation data of intact fish flesh. *J. Sci. Food Agric.* **1999**, *79*, 1793-1802.

Paper V

Effect of sampling errors in predictions using replicated measurements

Charlotte M. Andersen, Rasmus Bro and Per B. Brockhoff

Summary

Measurement error modelling is used for investigating the influence of measurement/sampling error on univariate predictions of water content and water holding capacity (reference measurement) from nuclear magnetic resonance (NMR) relaxations (instrumental) measured on four gadoid fish species. Reliability ratios illustrate that the four fish species are influenced differently by the error. However, the error seems to influence the predictions of the two reference measures in the same way. The effect of using replicated x-measurements is illustrated by simulated data and by NMR relaxations measured several times on each fish. The standard error of the physical determination of the reference values is lower than the standard error of the NMR measurements. In this case lower prediction error is obtained by replicating the instrumental measurements. A general formula is given for how to correct the least squares regression coefficient when another number of replicated x-measurements is used for validation than for calibration. It is shown that the correction should be applied when the number of replicates in validation is less than the number in calibration.

Keywords: measurement error model, replicates, predictions

Introduction

Instrumental methods, such as nuclear magnetic resonance (NMR) and near infrared spectroscopy, are of significant interest in many applications as they provide means for obtaining rapid on-line/at-line measurements enabling better control of production and quality. Regression techniques are commonly applied when fast instrumental measurements are used for estimating a parameter, for which alternative more direct determinations would be time consuming, expensive etc.

Least squares regression is strictly applicable only when the independent variables are measured without error. This is almost never the case since the very act of measurement implies the presence of error. Error is seldom due to one simple and specific cause but is the result of several contributing causes. For example, measurement and sampling error will have significant contributions to the overall error of the independent variable.

It is important to evaluate whether errors in the instrumental variable are important and, if so, develop methods that correct for the influence of these errors on the regression model. Measurement error models described by Fuller [1,2], Stefanski [3] and Jonsson [4] are examples of techniques that can be used to investigate the effect of measurement/sampling error and to correct for these errors in the independent variables. In practical sampling, the measurement error theory has been applied to various subjects such as an investigation of the correlations between dust lead and children's blood lead [5], a prediction of the amount of uranium enrichment as a function of the gamma count rate [6], predicting defoliation in a forest

stand by measuring the gypsy moth egg mass density [7] and in the sampling of insects from Russian wheat aphid [8].

It is claimed that if the aim is to predict new samples, ordinary least squares estimation will be adequate giving the lowest prediction errors if the samples that are to be predicted come from the same distribution as the samples used for calibration [1]. However, in other situations measurement error modelling may improve the predictions [4,7].

This paper provides guidelines on how to analyse data with error in the independent variable. The case with only one independent variable is described and analysed. The focus is on prediction; which influence does the error have on the predictions and what is the importance of replicated measurements. This includes investigating the effect of changing the number of replicates and how to find the optimal number of replicates to use in future predictions.

The theory of the measurement error models is described in the next section. After that the advantage of using a correction based on the measurement error theory to simulated data is discussed and at last the theory is applied on real data. Low-field ^{1}H NMR relaxations measured on fish muscle are used as examples. These data have shown to be correlated to the water content and the water holding capacity of fish muscle [9,10].

Theory

"Measurements" and "samples" are used for describing the data structure. Since the problem concerns how to sample and measure analytically, a single x-value is defined as a measurement. In the models, averages over a finite number of measurements are used. These averages are denoted as samples.

Regression models

The classical linear regression model with one independent variable is defined by

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1,2,\ldots,n \qquad \text{(eq. 1)}$$

where ε_i is called the model errors in the following and is assumed to be independent and normally distributed $N(0, \sigma_\varepsilon^2)$. The vector $(x_1, x_2, \ldots, x_n)$ is either fixed by design or given by a random distribution $N(\mu_x, \sigma_x^2)$. Furthermore, it is assumed that the vector $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is independent of the vector $(x_1, x_2, \ldots, x_n)$. The model error reflects the intrinsic error in the linear relationship between **y** and **x**. It may be due to an error in **y**, a non-linear relation between **x** and **y** or a lack of information in **x** for predicting **y**. There are two types of errors in **y**; one due to the natural variation in **y** and the other due to measurement/sampling error in the determination of the actual y-values used in the modelling.

Least squares parameter estimation

The parameters of the regression model can be estimated with least squares estimation where the slope is estimated by Equation 2.

$$b_{1,ls} = \frac{s_{xy}^2}{s_x^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \text{(eq. 2)}$$

$b_{1,ls}$ is the maximum likelihood estimator and is an unbiased estimator for β_1, s_{xy}^2 is the covariance between the x- and the y-samples and s_x^2 is the variance of the x-samples. The intercept (β_0) is estimated using the estimator of the slope and the average values of the independent and dependent variables (Equation 3).

$$b_{0,ls} = \bar{y} - b_{1,ls}\bar{x} \qquad \text{(eq. 3)}$$

In general terms, least squares estimation is aimed at minimising the sum of squared deviations of the observed values for the dependent variable from those predicted by the model (Equation 4).

$$Q(b_0, b_1) = \sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)^2 \qquad \text{(eq. 4)}$$

Measurement error model

Measurement error may have a great influence on the results depending on the purpose of the study. If measurement error is present, the measured x-values (x_i) used in the linear regression model can be written as the true value ($x_{i,true}$) plus a contribution from the measurement error (u_i).

$$x_i = x_{i,true} + u_i \qquad \text{(eq. 5)}$$

The measurement error (u_i) is a random error and is assumed normally distributed $N(0, \sigma_u^2)$ and independent of the x-values. Furthermore, the measurement error is independent of the model error given in Equation 1. The relation between the variances of the measured x-values, the true x-values and the measurement error can be described by

$$s_x^2 = s_{x,true}^2 + s_u^2 \qquad \text{(eq. 6)}$$

where $s_{x,true}^2$ denotes the variance of the true x-values and s_u^2 denotes the variance of the measurement errors.

Estimating the regression coefficients be the measurement error model

Six parameters have to be estimated in the measurement error model ($\mu_x, \sigma_\varepsilon^2, \sigma_x^2, \sigma_u^2, \beta_0, \beta_1$). However, only five parameters can be estimated from the vectors **x** and **y** [1]. Thus, an infinite number of models will fit when applying the measurement error theory to any type of data. Therefore, it is necessary to obtain additional information of one of the parameters in order to properly estimate the linear regression model. Here, it is assumed that the measurement error is known or that it can be estimated from repeated measurements. By correcting the variance of the x-values in Equation 2 for the measurement error, the true parameters of the linear regression model can be estimated without bias as shown in Equation 7 and 8. In ordinary least squares regression, the slope is estimated by dividing the covariance between **x** and **y** with the variance of **x** (Equation 2). When correcting for the measurement errors, the covariance between **x** and **y** is not divided by the variance of **x** but by the variance of **x** minus the variance of the measurement error (Equation 7). Thus, the estimated slope from the measurement error model will always be numerically larger than the biased least squares estimated slope. The intercept is estimated from the mean of **x** and **y** and from the estimated slope (Equation 8).

$$b_{1,\text{measerror}} = \frac{s_{xy}^2}{s_x^2 - s_u^2} \qquad \text{(eq. 7)}$$

$$b_{0,\text{measerror}} = \bar{y} - b_{1,\text{measerror}}\,\bar{x} \qquad \text{(eq. 8)}$$

Buonaccorsi [7] suggests a modified version of Equation 7, when the error variances and covariances vary from unit to unit. However, this modification will not be pursued here.

The ratio between the least squares estimated regression coefficient ($b_{1,\text{ls}}$) and the regression coefficient estimated with the measurement error model ($b_{1,\text{measerror}}$) is called the reliability ratio (r) (Equation 9).

$$r = \frac{s_{x,\text{true}}^2}{s_x^2} = \frac{s_{x,\text{true}}^2}{s_{x,\text{true}}^2 + s_u^2} = \frac{b_{1,\text{ls}}}{b_{1,\text{measerror}}} \qquad \text{(eq. 9)}$$

The reliability ratio is a measure of how much the regression coefficient is attenuated by the measurement error and is thus, indirectly, a measure of the size of the measurement error. As the reliability ratio approaches one, the measurement error will have less influence on the estimates.

Prediction

The correction of the regression coefficients described above is meant to provide more accurate estimates of the coefficients. However, even though these estimates are more accurate, this does not imply that the predictions obtained with them are more accurate. In fact, prediction by least squares estimation is normally preferred even when there are measurement errors. This holds theoretically if (y_i, x_i) is distributed as a bivariate normal random variable and if the samples to be predicted are a random selection from the same population that generated the samples used for calibration. It is claimed that correction for measurement error should be applied when an experimental design is used giving either fixed predictors in the calibration or in the test set [7]. However, applying the measurement error theory to simulated data using fixed x-variables could not substantiate this claim (results not shown). This is supported by Johnsson [4] who found that for fixed x in the test set, ordinary least squares regression gave the lowest prediction errors unless the mean x-values of the samples to be predicted lies far away from the mean x-values used for calibration. An advantage of measurement error models with respect to predictions can be obtained in situations with unequal sampling density, another subsampling method used for the samples in the test set, varying means for the true values in the random case or when the error variance in **x** is non-constant [4,7]. Jonsson [4] provides theory for estimating regression coefficients that give optimal prediction in cases where the measurement error variance of the samples in the test set varies from the mean error variance in calibration. In addition, Ganse et al. [11] used the theory of the measurement error models for predicting new samples with mean values that vary from the mean values of the original samples. In the following, the situation where samples are based on different numbers of measurements during calibration and validation will be elaborated on, explaining how to apply the least squares regression model and the measurement error theory to yield proper predictions.

Prediction in case of replicated measurements

Predictions obtained directly by least squares estimation may no longer be optimal if the x-values in the test set are based on an average over another number of measurements than is used for calibration. This may be the case even though all other conditions required for least squares estimation to be optimal are fulfilled. The reason is that the measurement error will be different when the number of measurements differs due to the averaging effect. Instead, a new regression coefficient that corresponds to the least squares estimate using the new number of measurements in the samples of the test set should be calculated, thus decreasing the prediction errors [1].

Equation 9 showed a relation between the least squares estimated regression coefficient, the corrected regression coefficient and the variances of the x-variable corresponding to the two regression coefficients. Using this relation, the least squares regression coefficient corresponding to the number of measurements in the samples of the test set can be calculated

from the least squares estimated regression coefficient and the estimated variances of the true x-values and the measurement error (Equation 10).

$$b_{1,\text{lsval}} = b_{1,\text{lscal}} \frac{s^2_{\text{x,true}} + s^2_{\text{u,cal}}}{s^2_{\text{x,true}} + s^2_{\text{u,val}}} \qquad \text{(eq. 10)}$$

$s^2_{\text{u,cal}}$ and $s^2_{\text{u,val}}$ are the variances of the measurement error for calibration and validation, respectively. In the following, the variance of the measurement error for the validation set is determined from the variance of the measurement error for the calibration set but is calculated in accordance with the number of measurements in the validation set. When the two data sets are based on different numbers of measurements, these variances will be different. $b_{1,\text{lscal}}$ is the least squares estimated regression coefficient of the calibration model. $b_{1,\text{lsval}}$ is the least squares estimator of the regression coefficient corresponding to the validation set obtained by correcting the least squares estimated coefficient from the calibration set according to Equation 10. Using this when predicting the samples in the validation set, lower prediction errors may be obtained than if the least squares estimated regression coefficient based on the calibration set is used. It is worth noting that the correction just discussed is not related to the correction in Equation 7, which has the purpose of improving the estimate of the regression coefficient itself. The correction in Equation 10 is merely an attempt to correct the coefficient to be an estimate of the least squares coefficient for a different number of measurements.

Simulated data

Data sets are made using a number of true x-values (samples). These are drawn from a N(0,1) distribution. Slope and intercept in the regression model are set to 20 and 80, respectively, and the true y-values are calculated using the randomly chosen x-values. When calculating the y-values, a model error is added, so the data follow the normal linear regression model (Equation 1). This error is drawn from a N(0,0.04) distribution. Measurement error is added to the x-values by adding a randomly normally distributed error N(0,1) to each x giving a number of noisy replicated measurements. Modelling is performed by regressing the y-values on the average of the replicated x-values and the predictions are evaluated by comparing the root mean squared error of prediction (RMSEP). RMSEP is defined as

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad \text{(eq. 11)}$$

where $\hat{y}_i$ is the predicted value of the *i*-th sample, y_i is the measured value of the *i*-th sample and *n* is the number of samples. Even though data are simulated with known measurement errors,

the variance of this error is estimated from the actual data included in the modelling as shown in Equation 12

$$s_u^2 = n^{-1}\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i.})^2 / a(n-1) \qquad \text{(eq. 12)}$$

a = (1,2....i) and n = (1,2....j)

where a is the number of samples, n is the number of measurements within each sample, $\bar{x}_{i.}$ is the average of the measurements of the i-th sample and x_{ij} is the measured value for the j-th measurement of the i-th sample.

Predictions are made using ten samples in the calibration set each being an average of ten measurements. The models are validated with test sets using 10,000 samples in the test set. These samples are made by taking the average of between one and 20 measurements. The large number of samples in the test set ensures that the predictions obtained are perfectly accurate. The predictions are performed in two ways (i) using ordinary least squares estimates and (ii) using corrected regression coefficients (Equation 10) in order to obtain the least squares estimates corresponding to the number of measurements in the test set. The modelling is repeated 1000 times and the average RMSEP is calculated.

Figure 1 shows that increasing the number of measurements in the test set reduces the prediction error. With up to around five measurements in the test set there seems to be a significant advantage in correcting the regression coefficients. This is not obtained when the number of measurements used for validating the model is larger than the number of measurements in the calibration set. As is also given by Equation 10, the two methods give exactly the same results when the same number of measurements is used in both the calibration and the test samples. It seems as the results get noisier when the number of measurements in the test set gets large in that there is an increase in the relative difference of the prediction errors.

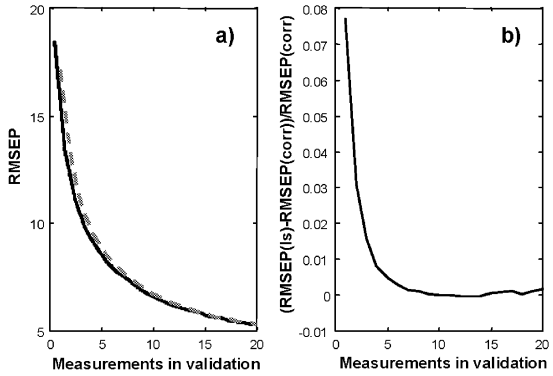


Figure 1 Prediction error as a function of the number of measurements in the test set. a) RMSEP(corr) obtained by correcting the regression coefficients in accordance with Equation 10 (grey dashed) and RMSEP(ls) obtained by least squares estimation (black solid) and b) Relative difference in RMSEP obtained by the two methods. The values stated are the average of 1000 simulations validated with test set validation. Ten samples that are the average of ten measurements are used for calibration and 10000 samples (average of the number of measurements denoted) are used for validation.

In real applications there are usually not as many samples available as is used above and the results will depend more on the actual samples in the data set. There may be errors due to variation in the material, time effects, pre-treatment, a variation in repeatability, etc. Probably, the results will not follow exactly what is expected from the theory and what is illustrated above. Therefore, predictions are also performed using only ten samples for both calibration and test set validation in order to see how well a potential gain in predictive ability can be verified with smaller data sets. As above the calibration sets are made by averaging over ten measurements. Figure 2 illustrates the RMSEP obtained using different numbers of measurements in the test set.

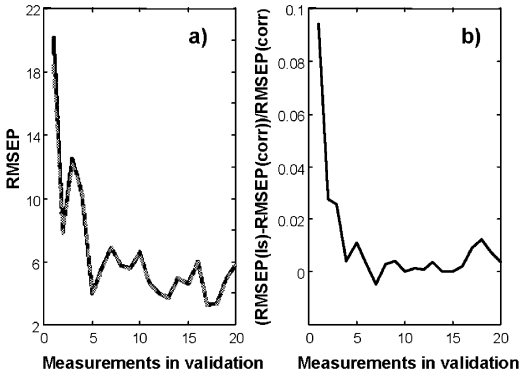


Figure 2 RMSEP as a function of the number of measurements in the test set. a) RMSEP(corr) obtained by correcting the regression coefficients in accordance with Equation 10 (grey dashed) and RMSEP(ls) obtained by least squares estimation (black solid) and b) Relative difference in RMSEP obtained by the two methods. The values stated are obtained using ten samples for both calibration and validation. The models are validated with test set validation using the average of ten measurements in the calibration set and the number of measurements denoted in the test set.

Although the results are noisier, the same tendencies are seen as was obtained when there were many samples in the test set. RMSEP decreases as the number of measurements used in the test set increases. Furthermore, there is a small difference in RMSEP obtained by the two methods. The relative difference is largest when very few measurements are used for making the test set. As above, it seems as correcting the regression coefficients is an advantage when five or less measurements are used for validation.

To conclude, the simulations indicate that corrections are feasible in situations where the predictions of the test set are made using fewer replicates than are used in calibration. It is further indicated that the improvement can be reasonably reliably estimated even for small sample sets.

Fish data

Prediction of water content and water holding capacity from ^{1}H low-field nuclear magnetic resonance (NMR) relaxations are used as examples of applying the measurement error theory to real data. The overall purpose is to investigate how to sample instrumental measurements taken at a localized area and from these predicting a parameter of the whole fish.

The data

Fish of four species were analyzed; including 12 whiting, 12 haddock, 12 saithe and 15 cod. All fish were caught in the North Sea and landed on the West coast of Jutland, Denmark. Whiting and saithe were caught in May 2001, whereas haddock and cod were caught in August 2001. In order to estimate the measurement error and the effect of replicated measurements, the NMR relaxations were measured 18 times at points spread throughout the whole fish (Figure 3).

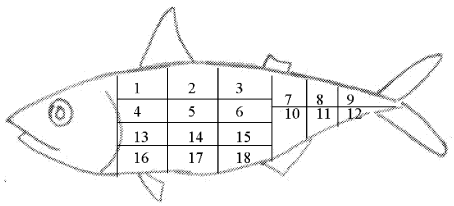


Figure 3 Sampling of NMR measurements.

Low-field NMR measurements were performed at 4°C on a Maran Benchtop Pulsed NMR analyser (Resonance Instruments, Witney, UK) operating at 23.2 MHz and equipped with an 18 mm variable-temperature probe head. The receiver gain was adjusted to 3% and the receiver delay was set to six seconds. Transverse relaxation was measured using the Carr-Purcell-

Meiboom-Gill (CPMG) sequence [12,13]. For each measurement 8 scans were performed with a 500 μs 90°-180° interpulse spacing (τ).The NMR relaxations consist of a whole relaxation curve meaning that each measurement contains approximately 500 variables. One of these variables was used as the x-values in the regressions (measurement no 88 obtained after approximately 0.2 s of relaxation).

The samples were introduced into the NMR probe by placing pieces of 2-4 g intact fish muscle into a glass tube that matched the inner diameter of the 18 mm NMR sample tube after which the samples were thermo-equilibrated for 30 min before measurement.

Water content was determined on exactly the same samples as were used for the NMR measurements, after the NMR relaxations were measured. The fish samples were kept in the small glass tubes and dried overnight at 110°C. They were weighed before and after drying. The total water content of each fish was found as the average of all water content determinations from the same fish.

Water holding capacity was determined on approximately 2 g of minced fillet using the part of the fillets that was not used for NMR measurements and water content determinations. The samples were weighed, placed in plastic tubes with a special filter bottom (pore size 0.1 mm), centrifuged (1500*g*, 10 °C, 5 min) and weighed again [14]. Each analysis was made in quadruplicate. Water holding capacity is defined as the amount of water left in the mince after centrifugation relative to the dry matter content.

Errors in the independent and dependent variables

The error in the NMR data can be divided into the error due to instrumental noise and the sampling error. NMR relaxation curves are supposed to be very accurate. Thus, the sampling error is assumed to make up almost all error in the instrumental data and is due to the variation in chemical and physical composition within the same fish. This includes the variation in muscle fibre structure and muscle cell size.

The samples are weighed before and after centrifugation or drying for the determination of water holding capacity and water content, respectively. Even though weighing is performed carefully, it will always contribute with some uncertainties to the estimates. An error caused by diffusion of water to the sample surface either during drying or centrifugation may be important in case the method is not fully optimised. This gives a variation between replicates when estimating the amount of water removed during the centrifugation step or during drying. The samples are homogenised before determining the water holding capacity. However, there will always be some heterogeneity within the sample. Thus, the reference measurements will be influenced by some sampling error as is the case for the NMR measurements.

The standard errors of the instrumental and physical determinations (Table 1) show that the water content is measured with the largest precision. Furthermore, the precision in the determination of the water holding capacity is larger than the precision of the NMR relaxations. The variation between fish (s.d.) is lowest for the determination of the water content but is of the same size for the NMR data and the water holding capacity (Table 1). From the above it is supposed that the variation in the NMR relaxations has a larger influence when predicting the water content than when predicting the water holding capacity.

Table 1 Mean values, standard error (s.e.), relative standard error (r.s.e.) and standard deviation (s.d.) for the NMR measurements, water holding capacity (WHC) and water content. Mean values are the average value obtained for each fish species and given as an intensity value for the NMR measurements, g water/g dry matter for the WHC and % (w/w) for the water content. The standard error of the measurements are calculated for each fish from replicated measurements. The average of these errors are used as the standard error for each fish species. Relative standard error is the standard error relative to the mean value. The standard deviation denotes the standard deviation of the mean values obtained for each fish species.

		Measurements			**Fish**
		Mean	s.e.	r.s.e.	s.d.
NMR	Whiting	0.094	3×10^{-3}	0.04	0.14
	Haddock	0.073	3×10^{-3}	0.04	0.07
	Saithe	0.069	3×10^{-3}	0.04	0.13
	Cod	0.082	3×10^{-3}	0.04	0.05
WHC	Whiting	4.10	0.03	6×10^{-3}	0.12
	Haddock	3.83	0.03	7×10^{-3}	0.04
	Saithe	3.86	0.02	6×10^{-3}	0.05
	Cod	3.85	0.04	10×10^{-3}	0.06
Water content	Whiting	82.16	0.25	3×10^{-3}	15×10^{-3}
	Haddock	81.24	0.20	2×10^{-3}	5×10^{-3}
	Saithe	80.67	0.15	2×10^{-3}	9×10^{-3}
	Cod	81.40	0.17	2×10^{-3}	7×10^{-3}

Regression coefficients and reliability ratios

For each fish species separately, least squares regression coefficients are calculated from Equation 2 together with the regression coefficients corrected for the measurement error (Equation 7). The models are validated by full-leave-one-fish out cross validation. The average of the 18 NMR measurements taken on each fish is used as the x-values. Since the samples consist of an average of 18 measurements, it is possible to estimate the variance of the sampling error from Equation 12 where n is the number of measurements for each fish and a is the number of fish. The variance of the sampling error will be an estimated value and it may not correspond exactly to the true variance. This adds to the uncertainty of the models and may

increase the prediction error and induce deviations from the theory. The reliability of this estimate is supported by the small confidence intervals of the estimated measurement error variance that are found to be of the size 2×10^{-6} to 5×10^{-6}. Furthermore, it was found that the measurement error follow the assumption of a normal distribution with a mean value around zero. The estimated regression coefficients and reliability ratios are shown in Table 2.

The regression coefficients become numerically larger when correcting for the sampling error as it is expected from Equation 7. By studying the reliability ratios it is seen that the models made on whiting and saithe are less influenced by the sampling error than the models obtained from haddock and cod. Whiting and saithe are the fish species that shows the greatest variations in the NMR data (Table 1). Therefore, it can be supposed that a larger variation in the data gives more stable models and the sampling error will have a lower impact on the regressions. This is supported by the fact that the standard error of the NMR data is of the same size for the four fish species.

Table 2 Regression coefficients (slope) obtained by cross validation of linear regressions (Equation 2) and regression coefficients obtained from the measurement error models (Equation 7). Models are made on each fish species and validated with full leave-one out cross validation (leaving one fish out at a time).

	Least squares estimation	**Correcting for the sampling error**	**Reliability ratio**
	$b_{1,ls}$	$b_{1,measerror}$	$b_{1,ls}/b_{1,measerror}$
	Water content		
Whiting	89.50	95.99	0.93
Haddock	6.46	8.78	0.74
Saithe	38.85	42.14	0.92
Cod	50.33	70.66	0.71
	Water holding capacity		
Whiting	22.90	24.56	0.93
Haddock	-3.11	-4.24	0.73
Saithe	1.88	1.92	0.98
Cod	11.59	16.27	0.71

In some situations, especially when predicting the water holding capacity, the regression coefficients are very low indicating that the relation between the dependent and the independent variables is not large. The poor relations between the NMR relaxations and the water holding capacity may be due to a large error in the physical determination of the water holding capacity, a low variation in the physically determined values or that the relation between the two values is in fact very low.

Prediction with replicated measurements

Predictions are made using data sets obtained by averaging over the same number of measurements in the calibration and the test set. This is done in order to evaluate the effect of replicates on the predictive ability. The regression models are validated with full-leave-one-fish-out cross validation and the average prediction error of six independent models is calculated. The actual measurements used in each of the independent models are chosen systematically; when the number of measurements increases by one, the same measurements as was used in the previous model together with one new is applied. The results are shown in the Figures 4 and 5. The prediction error expressed in percentage is approximately ten times lower for the water content compared with the water holding capacity. This is in agreement with the difference in the errors of the reference measurements (Table 1).

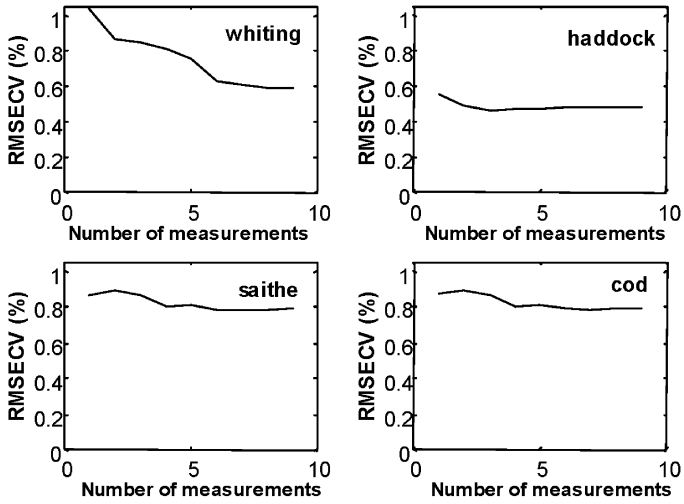


Figure 4 Predicting the water content. Prediction errors are given in percentage (relative to the total water content) as a function of the number of replicates used in both the calibration and the validation set. The results are the average of six models validated with full-leave-one-fish out cross validation.

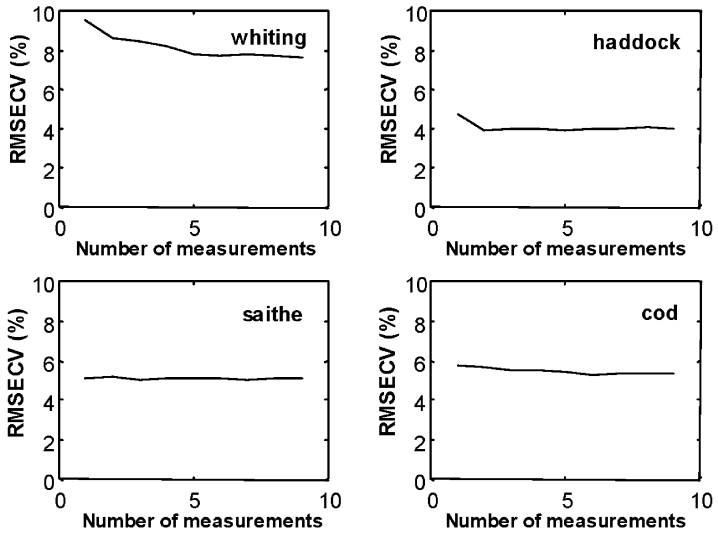


Figure 5 Prediction of water holding capacity. Prediction errors are given in percentage (relative to the water holding capacity of each fish) as a function of the number of replicates used in both the calibration and the validation set. The results are the average of six models validated with full-leave-one-fish out cross validation.

Generally, there is no large improvement of increasing the number of replicates. Only for whiting, the relative prediction error decreases steadily with the number of replicates. Whiting is the fish species that varies the most (Table 1), which as mentioned may give a more valid and stable model. When predicting the water content of the other fish species, there seems to be an improvement using three or four replicates. For the water holding capacity, the prediction error does not decrease as much with increasing number of replicates. As mentioned, the error in the physically determination of the water holding capacity is larger than the standard error obtained for the water content, which may give less improvement of an increased number of measurements compared with the prediction of the water content. Furthermore, other factors may influence the regressions, decreasing the predictive ability and thereby give results that are not expected. These factors may be that there is not a strict linear relationship between NMR data and the water holding capacity or that the NMR data does not contain enough information that relates to the water holding capacity.

The above indicates that when the standard error of the reference measurements (and the model error) is lower than the standard error of the instrumental measurements, more than one instrumental measurement is needed in order to obtain the lowest prediction error. In this case three or four measurements are needed to obtain optimal predictions of the water content. However, the exact number of measurements to take may vary depending on the material, experimental conditions, actual measurements etc.

Prediction in case of different conditions in the calibration and prediction set

In the following, it is investigated how averaging over different numbers of measurements in the test set influences the prediction error. Predictions are made with averages of six measurements in the calibration set and averages of one to 12 measurements in the validation set. The models are validated with full-leave-one-fish-out cross validation. For each number of measurements used in the validation set, the average prediction errors of six models are calculated. The actual measurements used in each model is chosen so that; when the number of measurements increases by one, the same measurements as was used in the previous model together with one new is used.

Figure 6 illustrates the results obtained when predicting the water content. Especially, whiting and cod follow what is expected from the theory and the simulations. RMSEP decreases as the number of measurements used for prediction increases. The relative difference between the prediction error obtained by least squares regression and when correcting for the number of measurements in the test set is shown in the right plot. It is largest when only one measurement is used for validation. When the test set is made of up to a few measurements, there seems to be an advantage of correcting the regression coefficients as was also seen for the simulated data. However, this advantage diminishes when more than approximately three measurements are used.

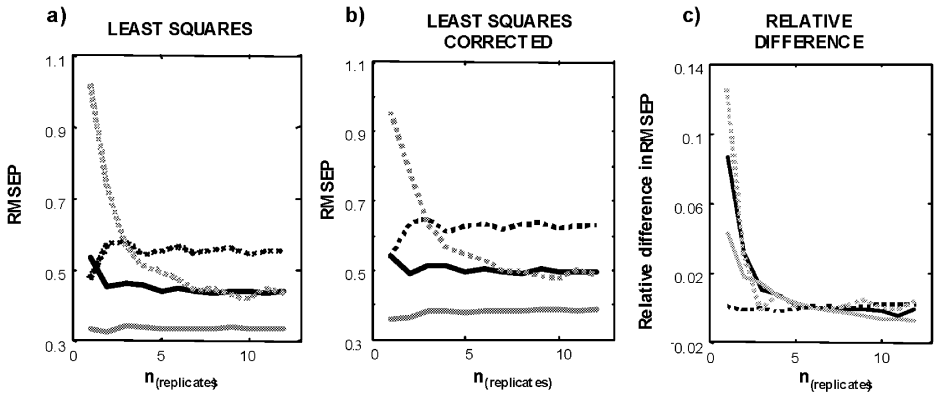


Figure 6 Prediction of water content from ^{1}H low-field NMR relaxations. One variable from the NMR relaxation curve is used as x-variable. Models are validated with full cross validation and include the average of six replicates in the calibration set. a) RMSEP as a function of the number of replicates in the validation set where the regression coefficient is estimated by least squares regression, b) RMSEP as a function of the number of replicates in the validation set where the regression coefficient is estimated by correcting the least squares regression and c) relative difference in RMSEP given as RMSEP by obtained by least squares regression minus the RMSEP obtained from Equation 10 relative to the least squares estimate. The lines indicate: cod (solid black), haddock (solid grey), saithe (dotted black) and whiting (dotted grey).

The results obtained for saithe do not show the same tendencies as are seen for the other fish species and the simulations. RMSEP increases with the increase in the number of measurements used for validation. Also, the difference between the two error estimates increases with the number of measurements. Remarkably, saithe shows inferior results compared to cod and haddock. Together with whiting, saithe is the fish species that is least influenced by the sampling error (Table 2) and the inferior results may thus be due to other factors than the sampling error such as an increased variability caused by the uncertainty in the estimated correction factors. The results obtained when predicting the water holding capacity show similar tendencies as the results obtained when predicting the water content and are not shown here.

The above illustrates that the theory of the measurement error models is useful when applied on real data. However, the results are also influenced by various types of errors that are not necessarily covered by the measurement error model and may sometimes be more influential, such as seen for the saithe. As was obtained for the simulations, the regression coefficients should be corrected so they correspond to the number of measurements in the test set if the number of measurements in the test set is much lower than the number of measurements in the calibration set. There does not seem to be an advantage of correcting the regression coefficients when the models are validated on samples consisting of many repeated measurements.

Conclusion

A simple general methodology to investigate the effect of the different types of errors of data in a specific situation has been given. Sampling error in the independent variables induces an effect on the estimated regression model. In case of lower standard error in the reference values, improved predictions can be obtained by replicating the measurements of the independent variables.

The paper gives guidelines on how to correct the estimated regression coefficients in order to obtain optimal predictions in the case where the number of measurements varies between the calibration samples and the samples to be predicted.

References

1. Fuller W A, Estimation in the presence of measurement error. *Int. Stat. Rev.* 1995; **63**:121-147.

2. Fuller WA, *Measurement error models*. John Wiley: New York, 1987.

3. Stefanski L A, Measurement error models. *Am. Stat. Assoc.* 2000; **95**:1353-1358.

4. Jonsson B, Prediction with a linear regression model and errors in a regressor. *Int. J. Forecasting* 1994; **10**:549-555.

5. Edmond M J, Lanphear B P, Watts A, Eberly S, Measurement error and its impact on the estimated relationship between dust lead and children's blood lead. *Environ. Res.* 1997; **72**:82-92.

6. Burr T L, Knepper P L, A study of the effect of measurement error in predictor variables in non-destructive assay. *Appl. Radiates. Isotopes* 2000; **53**:547-555.

7. Buonaccorsi J P, Prediction in the presence of measurement error: general discussion and an example predicting defoliation. *Biometrics* 1995; **51**:1562-1569.

8. Schaalje G B, Butts RA, Some effects of ignoring correlated measurement errors in straight line regression and prediction. *Biometrics* 1993; **49**:1262-1267.

9. Jepsen S M, Pedersen H T, Engelsen S B, Application of chemometrics to low-field ^{1}H NMR relaxation data of intact fish flesh. *J. Sci. Food Agric.* 1999; **79**:1793-1802.

10. Andersen C M, Rinnan Å, Distribution of water in fresh cod. *Lebensm.- Wiss. u. Technol.* 2002; **35**: 687-696.

11. Ganse R A, Anemiya Y, Fuller W A, Prediction when both variables are subject to error, with application to earthquake magnitudes. *J. Am. Stat. Assoc.* 1983; **78**:761-765.

12. Carr H Y, Purcell E M, Effects of diffusion on free precession in nuclear magnetic resonance, *Am. J. Physiol.* 1954; **94**: 630-638.

13. Meiboom S, Gill D, Modified spin-echo method for measuring nuclear times. *Rev. Sci. Instr.* 1958; **29**: 688-691.

14. Eide O, Børresen T, Strøm O, Minced fish production from capelin (*Mallotus villosus*). A new method for gutting, skinning and removal of fat from fatty species. *J. Food Sci.* 1982; **47**: 347-349, 354.

Paper VI

Quantification and handling of sampling errors in instrumental measurements: A case study

Charlotte M. Andersen and Rasmus Bro

Abstract

Instrumental measurements are often used to represent a whole object even though only a small part of the object is actually measured by the instrument. This can introduce an error due to the inhomogeneity of the product. Together with other errors resulting from the measuring process such errors may have a serious impact on the results when the instrumental measurements are used for multivariate regression and prediction. This paper gives examples of how errors influencing the predictions obtained by a multivariate regression model can be quantified and handled. The relevant errors contributing to the prediction error are: error in instrumental measurements, error in reference measurements, error in the estimated calibration model and model error. The size of the four error contributions are estimated by an errors-in-variables approach and the importance of the different errors is evaluated. An approach based on the theory of the measurement error models is used to show how the influence of errors in the instrumental measurements can be partly adjusted for. The approach can be used when the number of replicates used for calibration and validation differs. With few replicates in validation and when the error in the instrumental measurements has some impact on the predictions, the approach seems to provide more accurate predictions than the naive approach. Prediction of water content of cod fillets from low-field NMR relaxations and prediction of analyte concentration in mixed solutions from fluorescence emission spectra are used as examples to show the applicability of the methods.

Keywords: errors, replicates, predictions

Introduction

Instrumental methods, such as nuclear magnetic resonance (NMR) and fluorescence spectroscopy, are of significant interest in many applications as they provide means for obtaining rapid on-line/at-line measurements enabling better process and quality monitoring. Such instrumental measurements are generally characterized by high signal-to-noise ratios. However, the adequacy of the measurement as a representation of the sample may be hampered by sampling error. Such errors may in combination with measurement errors lead to biased regression parameter estimates, which can affect the predictions. Therefore, it is important to evaluate whether such errors are important and, if so, develop methods that take these errors into account.

This paper presents an investigation of the errors influencing predictions obtained by multivariate regression. This is illustrated by two examples. The first concerns the prediction of water content from low-field NMR relaxations. In the other example, the concentration of three analytes in solutions is predicted from fluorescence spectra. The evaluation of errors includes 1) quantification and visualization of the different error contributions that influence the model

parameters and the predictions and 2) how errors in the instrumental measurements can be handled.

The quantification is illustrated by an errors-in-variables approach suggested by Faber and Kowalski (1996). They presented a method to estimate the prediction error variance for each sample based on errors that are important for the parameter estimation and the predictions.

Andersen et al. (2002) proposed a correction of the regression coefficient that can be used to handle errors in instrumental measurements and in certain situations improve the predictions. The correction applies in situations where the number of replicates in calibration and validation differs and gives the least squares estimate of the regression coefficient corresponding to the number of replicates in the validation set. It is based on the measurement error theory (Fuller 1987) and requires univariate data.

Univariate data is obtained from the multivariate measurements by calculating the scalar net analyte signal. The net analyte signal was defined by Lorber (1986) as the part of a signal that is unique for a certain analyte. Later, it has been shown that the scalar net analyte signal gives an exact univariate representation of the multivariate regression model (Faber 2000; Bro and Andersen 2002).

Theory

The multivariate regression model can be expressed as

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \tag{1}$$

where $\mathbf{y}$ ($I\times1$) is a vector containing the measured reference values, $\mathbf{X}$ ($I\times J$) is a matrix holding the instrumental measurements containing measurement/sampling errors, $\mathbf{b}$ ($J\times1$) denotes the estimated regression coefficient vector, $\mathbf{e}$ ($I\times1$) is a vector holding the residuals and I and J are the number of samples and variables in $\mathbf{X}$.

Multivariate regression models, such as partial least squares (PLS) and principal component regression (PCR), are commonly applied for relating multivariate instrumental measurements to a reference value as for example the concentration of a certain analyte. By extracting a number of latent factors containing the systematic variation in data, the influence of errors in the instrumental measurements is ideally minimized. However, modeling may not be perfect and part of the error originally present in data is affecting the model parameter estimates.

The model performance is typically validated by the root mean squared error of calibration (RMSEC), which is defined as

$$\mathrm{RMSEC} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{2}$$

where $\hat{y}_i$ is the predicted value of the *i*-th calibration sample, y_i is the measured value of the *i*-th calibration sample and *n* is the number of samples. The ability of the model to predict new samples is estimated by the root mean squared error of prediction (RMSEP), which is given as

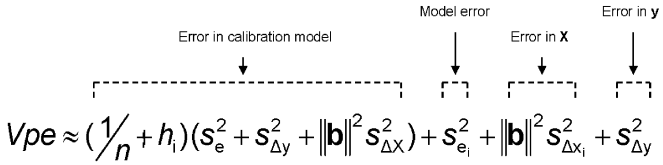
$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{3}$$

where $\hat{y}_i$ is the predicted value and y_i is the measured value of the *i*-th sample in the test set. Sometimes cross validation is used as an alternative to test set validation (Martens and Martens 2001).

Quantification of errors

An expression of the prediction error variance based on the errors that influence modeling and prediction is applied for the quantification of errors. The expression, which is shown in Equation 4, gives a sample-specific variance of prediction error and is based on a suggestion of Faber and Kowalski (1996; 1997). The expression includes a contribution to the prediction error from the errors in the instrumental variables (**X**), errors in the reference measurements (**y**), errors in the estimated calibration model (regression vector) and the model error.

When the predictions are compared in practice, the true reference values are not known. Thus, to obtain an estimate of the *apparent* variance obtained from the noisy y-measurements, an *apparent* prediction error variance can be defined. It is noted though that this error contribution is not the real in the sense that it is not reflecting an actual error in the predictions but merely reflecting the imperfect data that are used in the evaluation of the predictions. The difference between the *actual* and the *apparent* prediction error is the variance of the error in **y**, which is also the difference between the expression given by Faber and Kowalski and the expression used here.



$$Vpe \approx (1/n + h_i)(s_e^2 + s_{\Delta y}^2 + \|\mathbf{b}\|^2 s_{\Delta X}^2) + s_{e_i}^2 + \|\mathbf{b}\|^2 s_{\Delta x_i}^2 + s_{\Delta y}^2 \tag{4}$$

Equation 4 is an approximation based on linearization of the model and allows a comparison of the different error contributions. Assumptions are that the noise of the independent variables has similar size for the calibration and the test samples, errors are small and that data are unbiased meaning that all errors are considered to be independently and identically distributed random variables with zero mean. This may not always hold for instrumental variables e.g. spectroscopic measurements, where the noise is often correlated. However, as an approximation, the expression may still be valid even in such cases.

The regression coefficients (**b**) are incorporated in Equation 4 by their Euclidian norm. $s^2_{\Delta y}$ denotes the variance of errors in the reference value. $s^2_{\Delta X}$ and $s^2_{\Delta x_i}$ are the errors in the instrumental measurements for the calibration set and for the test set, respectively. s^2_e and $s^2_{e_i}$ are the model errors for the calibration model and for the test sample, respectively. h_i is the leverage of the test sample calculated from the scores and n denotes the number of samples in the calibration set. Equation 4 estimates the variance of the prediction error in case of a mean centered model. The term $1/n$ should be left out if data is not mean centered.

The influence of the four error contributions on the prediction error variance can be evaluated by comparing the size of these four error terms. The error in **X** is estimated by

$$s^2_X = n^{-1}\sum_{i=1}^{a}\sum_{j=1}^{n}(x_{ij} - \bar{x}_{i.})^2 / a(n-1) \tag{5}$$

where a=1,2....i denotes the number of samples, n=1,2....j denotes the number of replicates within each sample, $\bar{x}_{i.}$ is the average of the replicates of the i-th sample and x_{ij} is the measured value of the j-th replicate and the i-th sample. s^2_X is obtained for each variable in **X** and the average of all variables is used in the calculations. The error in **y** is estimated from replicated measurements. Furthermore, the regression coefficient (**b**) and the leverage (h_i) are estimated from the calibration model and the scores of the test samples, respectively.

The model error is estimated by Equation 6 as suggested by Faber et al. (1998).

$$s^2_e = \mathrm{MSEC} - s^2_{\Delta y} - \|\mathbf{b}\|^2 s^2_{\Delta X} \tag{6}$$

MSEC denotes the mean squared error of calibration and the other factors are given as in Equation 4. Equation 6 gives one overall estimate of the model error for each model and is valid when errors in **X** and **y** are not too large.

Net analyte signal formulation of regression model

Several methods have been published showing how to calculate the net analyte signal (Lorber et al. 1997; Xu and Schechter 1997; Faber 1998a; Goicoechea and Olivieri 1999 and 2001; Ferré et al. 2001). Here a method proposed by Bro and Andersen (2002) is used, which only relies on the specific regression model and not on any additional assumptions on the structure of the data. The net analyte signal ($\mathbf{x}^{*}_{k,i}$) for analyte *k* and sample *i* is given as

$$\mathbf{x}^{*}_{k,i} = \mathbf{b}\mathbf{b}^{+}\mathbf{x}_i = \mathbf{b}(\mathbf{b}^T\mathbf{b})^{-1}\mathbf{b}^T\mathbf{x}_i \tag{7}$$

where **b** is the regression coefficient vector estimated e.g. from a mean centered PLS or PCR calibration model and $\mathbf{x}_i$ is a vector holding the signal of the *i*-th sample. Having estimated the net analyte signal vector, the scalar net analyte signal is given as

$$x^{*}_{k,i} = x^{*}_{k,i} \times \frac{\mathbf{b}}{\|\mathbf{b}\|} \tag{8}$$

where $\|\bullet\|$ denotes the Euclidian norm. It has been stated that the scalar net analyte signal is the norm of the multivariate net analyte signal (Faber et al. 1997). However, for mean centered models and when the predicted value is below zero this is not valid and Equation 8 should be applied (Faber 1998b; Bro and Andersen 2002).

Using the scalar net analyte signal, the multivariate calibration model can be represented as a univariate least squares regression model as given in Equation 9 (Faber 2000).

$$y_i = bx^{*}_{k,i} + e_i \tag{9}$$

where *b* is a scalar regression coefficient and e_i denotes the residuals. Predictions obtained using the scalar net analyte signal as x-variable will be identical to the predictions obtained directly by the multivariate regression if the norm of the regression coefficient vector obtained by PLS or PCR is used as *b* (Ferré et al. 2001).

Plotting the concentration of the reference values versus the scalar net analyte signal illustrates the relation between the two sets of variables where the contribution of error in the instrumental measurements and the prediction errors can be shown as the deviation from the regression line in the horizontal and vertical direction, respectively. However, the visualization does not show which of the two error terms that has the largest influence on the predictions.

Correction of the regression coefficient

Andersen et al. (2002) proposed a correction of the regression coefficient based on the univariate least squares regression model (Equation 10). The correction may improve the predictions in situations where the assumptions underlying least squares regression are not met.

$$\mathbf{y} = b_{0,\text{ls}} + b_{1,\text{ls}}\mathbf{x} + \mathbf{e} \tag{10}$$

In Equation 10 the vector **x** holds the independent measured variable, the vector **y** holds the measured reference measurements and **e** denotes the errors, which are assumed to be independent of the x-variables and normally distributed. Least squares regression gives optimal predictions if the underlying assumptions are met but may give biased parameter estimates if these are not met.

If errors are present in the independent data, the x-value measured can be given as the true value ($x_{\text{i,true}}$) plus a contribution from the error (Δx_i) (Equation 11).

$$x_\text{i} = x_{\text{i,true}} + \Delta x_\text{i} \tag{11}$$

Assuming that the error (Δx_i) is random normally distributed and independent of x_i and e_i, the relation between the variances can be written as

$$s_\text{x}^2 = s_{\text{x,true}}^2 + s_{\Delta\text{x}}^2 \tag{12}$$

From the theory of the measurement error models (Fuller 1987), the regression coefficient corrected for bias ($b_{1,\text{measerror}}$) can be estimated from Equation 13 where the variance of the x-variables are corrected for the measurement error. The variance of **X** and the covariance between **X** and **y** (s_{xy}^2) can be found directly from the measurements. Furthermore, the variance of errors in **X** can be estimated from Equation 5 if replicates are measured. In ordinary least squares regression, the regression coefficient is estimated by dividing the covariance between **X** and **y** with the variance of **X**. When correcting for the measurement errors, the covariance between **X** and **y** is not divided by the variance of **X** but by the variance of **X** minus the variance of the measurement error.

$$b_{1,\text{measerror}} = \frac{s_{\text{xy}}^2}{s_\text{x}^2 - s_{\Delta\text{x}}^2} \tag{13}$$

The ratio between the least squares regression coefficient ($b_{1,ls}$) and the regression coefficient estimated from the measurement error model is called the reliability ratio (r). It can be estimated from Equation 14, which also shows the relationship between the errors.

$$r = \frac{s^2_{x,true}}{s^2_x} = \frac{s^2_{x,true}}{s^2_{x,true} + s^2_{\Delta x}} = \frac{b_{1,ls}}{b_{1,measerror}} \tag{14}$$

Predictions obtained directly by least squares estimation may be biased if the measurement error variance is changed compared to the measurement error variance in the samples used for estimating the regression coefficient.

For example, the predictions will not be optimal if the x-values are made by averaging over a number of replicates and this number differs between calibration and validation. The reason is that the measurement error will be different when the number of measurements differs due to the averaging effect. Instead, a new regression coefficient that corresponds to the least squares estimate of the number of replicates in the test set is estimated. This correction of the regression coefficient is given in Equation 15, which is based on the relations shown in Equation 14.

$$b_{1,lsval} = b_{1,lscal} \frac{s^2_x + s^2_{\Delta x,cal}}{s^2_x + s^2_{\Delta x,val}} \tag{15}$$

In Equation 15, $s^2_{\Delta x,cal}$ and $s^2_{\Delta x,val}$ are the variances of the errors in the instrumental measurements for the calibration and the validation set, respectively. When the two data sets are based on different numbers of replicates, these variances will be different. $b_{1,lscal}$ denotes the least squares estimated regression coefficient for the calibration set. $b_{1ls,val}$ is a calculated regression coefficient that is equal to the regression coefficient of the calibration set but corresponding to the number of replicates in the validation set. It is different from the regression coefficient of the calibration model when the number of replicates in calibration and validation differs. Application of this corrected regression coefficient may give better predictions than if the least squares estimated regression coefficient based on the calibration set ($b_{1,lscal}$) is used (Andersen et al. 2002).

The expression suggested in Equation 15 applies for univariate data. However, in the present study the NMR relaxation curves and the fluorescence emission spectra are obtained as multivariate signals. Equation 7, 8 and 9 showed that the scalar net analyte signal can be used to represent a multivariate regression model as a univariate regression model. This approach is used here in order to apply Equation 15 on the multivariate signals. The norm of the PLS or PCR regression vector is used as the regression coefficient of the calibration model ($b_{1,lscal}$) and the variances of the errors in the instrumental measurements of the calibration and the

validation set are estimated by Equation 5. After correcting the regression coefficient by Equation 15 the scalar net analyte signals are used for the prediction of the y-values.

Experimental

Data set 1

The total water content of fish is an important quality parameter influencing e.g. the sensory impression and texture. Therefore, it is relevant to predict the water content by a fast non-destructive method, such as low-field NMR. However, a practical problem is that only a minor fraction of the fish can be analyzed. This gives a sampling variation when the sample measured is not representative of the whole fish. Therefore, knowledge about the importance of sampling errors is important. The variation in instrumental measurements within a fish may be systematic if a certain part of the fillet is chosen for measurement each time. However, when the measured part of the fillet is chosen randomly as is done here, the sampling error will be a random error, and the theory described above, which applies for random errors, can be used.

Fish of four species are analyzed; including 12 whiting, 12 haddock, 12 saithe and 15 cod. All fish were caught in the North Sea and landed on the West coast of Jutland, Denmark. Whiting and saithe were caught in May 2001, whereas haddock and cod were caught in August 2001. In order to estimate the sampling error and the effect of replicated measurements, the NMR relaxations were measured 18 times at points spread throughout the whole fish (Figure 1).

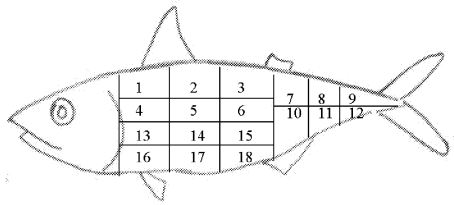


Figure 1 Sampling of NMR measurements.

Low-field NMR measurements were performed at 4°C on a Maran Benchtop Pulsed NMR analyser (Resonance Instruments, Witney, UK) operating at 23.2 MHz and equipped with an 18 mm variable-temperature probe head. The receiver gain was adjusted to 3% and the receiver delay was set to six seconds. Transverse relaxation was measured using the Carr-Purcell-Meiboom-Gill (CPMG) sequence (Carr and Purcell 1954; Meiboom and Gill 1958). For each measurement 8 scans were performed with a 500 μs 90°-180° interpulse spacing (τ). Only even echoes were recorded, of which only every second was used. This gives 256 echoes measured for each sample.

The fish were introduced into the NMR probe by placing pieces of 2-4 g intact fish muscle into a glass tube that matched the inner diameter of the 18 mm NMR sample tube after which the samples were thermo-equilibrated for 30 min before measurement. The sample preparation was performed as carefully as possible in order not to alter the muscle structure by the manual treatment.

The water content was determined on exactly the same samples as were used for the NMR measurements, after the NMR relaxations were measured. The fish samples were kept in the small glass tubes and dried overnight at 110°C. They were weighed before and after drying. The total water content of each fish is used in the data analysis. It is found as the average of all measurements from each fish.

Data set 2
The concentration of three analytes; catechol, hydroquinone, and indole, in mixed solutions is predicted from fluorescence emission spectra. The data set contains 48 samples with different concentrations of the three analytes.

Fluorescence emission spectra measured with a Perkin Elmer LS50 B spectrofluorometer after excitation at 275 nm are used for predicting the concentration of catechol and indole, whereas emission spectra obtained after excitation at 290 nm are used for predicting the concentration of hydroquinone. These excitation wavelengths are chosen because excitation maximum was found at these wavelength positions.

All samples were measured in five replicates by scanning the sample five times and leaving the sample in the instrument. Thus, there is no variation in the fluorescence spectra due to sample inhomogeneity and only instrumental noise contributes to the errors in the instrumental measurements. The y-values are given in M as the value, which it is aimed at reaching. The concentration of the analytes in the samples may not be exactly similar to the stated value due to errors in making the samples. However, with the given data the error in the reference values can not be estimated. Therefore, an error of zero in **y** is assumed. The mean concentrations are 2.7×10^{-5}, 7.0×10^{-6} and 1.7×10^{-6} M for catechol, hydroquinone and indole, respectively.

Results and discussion

Multivariate regressions are performed by regressing the water content or analyte concentration (y-variable) on the NMR relaxations or fluorescence emission spectra (x-variables) using PLS regression. For data set 1, two components are used in the PLS models since initial analysis showed that this was reasonable. Furthermore, PLS models are made on each fish species separately due to large differences between the four fish species. For data set 2, it was found that one- or two-component models gave the best predictions.

Quantification and visualization of errors

Data set 1

Predictions of water content are made using six randomly chosen NMR measurements of each fish, in a calibration set and six other NMR measurements in a test set. Average values are not used in the modeling but each measurement is used in the model one by one. The calibration set and the test set will contain different x-measurements but the same y-values. The mean variance of prediction error is obtained from Equation 4 and shown in Table 1. One value is obtained for each sample, but the average of all samples for each fish species is given in the table. Table 1 also shows how much each of the four error terms contributes to the total prediction error variance.

The mean variance of prediction error is of the same magnitude for the four fish species with whiting showing somewhat larger values. This indicates that prediction of water content of whiting is influenced by larger errors. For all four fish species, the errors in the NMR relaxations only make up a little part of the prediction error variance. However, for cod this contribution is larger than for the other three fish species. As it will be shown below this variation illustrates a difference in the importance of errors in the instrumental measurements. The model error is the error term that contributes most to the total prediction error variance, indicating that substantial improvements must come from including new types of information in the model. Furthermore, the error in the calibration model and the errors in the reference measurements seem to have some, although not a large contribution to the total prediction error variance. The size of these two error terms is approximately similar for the four fish species with the error in the calibration model being somewhat larger for cod and the error in **y** being somewhat larger for haddock.

Table 1 Mean variance of prediction error and errors contributing to the mean variance of prediction error given in per cent.

	Mean variance of prediction error	Error in calibration model (%)	Model error (%)	Contribution of errors in **X** (%)	Contribution of errors in **y** (%)
Cod	2.1	13.3	85.3	0.02	1.4
Saithe	1.9	7.6	91.2	0.001	1.2
Whiting	8.5	8.0	91.2	0.007	0.8
Haddock	1.4	3.6	93.5	0.007	2.9

The contribution of errors in **X** is made up of the squared norm of the regression coefficient vector and the estimated errors in **X**. These values are shown in Table 2. The regression coefficients (inverse sensitivities) are larger for cod and whiting indicating that the water content of these samples is predicted with larger uncertainty. Thus, even though the estimated error in **X** for cod is lower than for the other fish species, the contribution of these errors in **X** will be larger.

Table 2 Regression coefficients and errors in **X** for the four fish species

	$\|\mathbf{b}\|$	s_x^2
Cod	3×10^{-3}	142
Saithe	4×10^{-4}	240
Whiting	2×10^{-3}	1047
Haddock	4×10^{-4}	920

Replicates are typically used for minimizing the effect of large measurement errors. This is illustrated in Figure 2, which shows RMSEP as a function of the number of NMR relaxation replicates, where average values of the replicates are used. The larger importance of errors in **X** for cod is shown by the decrease in RMSEP with increasing number of replicates. Thus, even though the errors in the NMR relaxations of cod only makes up a very little part of the total prediction error variance, these are still important when predicting the water content. For haddock, there is no change in RMSEP with the change in the number of replicates showing that in this situation, errors in **X** may be negligible.

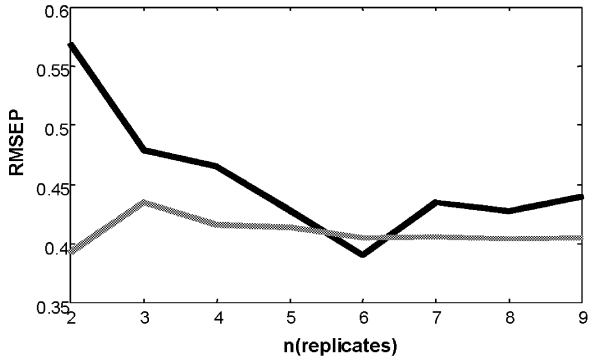


Figure 2 RMSEP as a function of the number of replicates in both calibration and validation for cod (black line) and haddock (grey line).

Figure 3 shows, for cod and haddock, the water content versus the scalar net analyte signal. In this figure, the average of six replicates are shown in contrast to the results given in Table 1, which shows results obtained using the samples individually. The samples of cod are placed closer to the regression line than the samples of haddock. This can be explained by the larger estimated errors in the NMR relaxations obtained from haddock. As seen in Table 2 these estimated errors are approximately seven times larger for haddock than for cod.

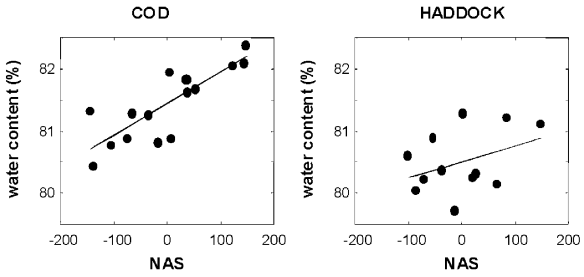


Figure 3 Plots of water content in per cent (g water/g fish muscle) versus the scalar net analyte signal (NAS) of the average of the samples in the calibration set.

The above illustrates well that both the error in **X** and the size of the regression coefficient are included in the estimated contribution of errors in **X** to the prediction error. Even though the error in the NMR relaxations of cod is very low, the contribution of these errors is large due to the low sensitivity. Furthermore, the error in the NMR relaxations of haddock is large, but due to the large sensitivity, the contribution of these errors will not be large.

Data set 2

Table 3 shows the contribution of errors in the calibration model, model error and errors in the fluorescence spectra when predicting the concentration of the three analytes. The mean variance of prediction error given as the average value of all samples is shown as well. Regression models are made with two components for catechol and indole and one component for hydroquinone using the five replicates for 24 of the samples in the calibration set and the five replicates of the other 24 samples in the test set. As for data set 1 average values of the replicates are not used.

Table 3 Mean variance of prediction error and errors contributing to the mean variance of prediction error given in per cent.

	Mean variance of prediction error	Error in calibration model (%)	Model error (%)	Contribution of errors in **X** (%)
Catechol	8×10^{-10}	5.4	94.6	5×10^{-4}
Hydroquinone	1×10^{-11}	5.5	94.5	2×10^{-5}
Indole	1×10^{-12}	5.4	94.6	1×10^{-4}

For all three analytes the prediction error variance is very small (10^{-10} to 10^{-12}) relative to the analyte concentrations (10^{-5} to 10^{-6}). The model error makes up the largest amount of the prediction error variance. Also the error in the calibration model seems to be important. The contribution of instrumental noise (errors in **X)** is very small. In order to investigate the importance of instrumental noise further, predictions were made with different number of replicates where averages over replicates were used. However, for all three analytes, no

change in RMSEP is obtained by changing the number of replicates. This further substantiates that errors in **X** are not important.

Figure 4 shows a plot of the analyte concentration versus the scalar net analyte signals for catechol and indole illustrating that the predictions perform well. The samples of indole are placed very close to the regression line whereas the catechol samples are more spread out. Both the contribution of errors in **X** and the prediction error variance (Table 3) are larger for catechol than for indole, which may explain the differences seen in Figure 4. The absolute values of the contributions of errors in **X** are 3.9×10^{-15} for catechol and 1.3×10^{-16} for indole.

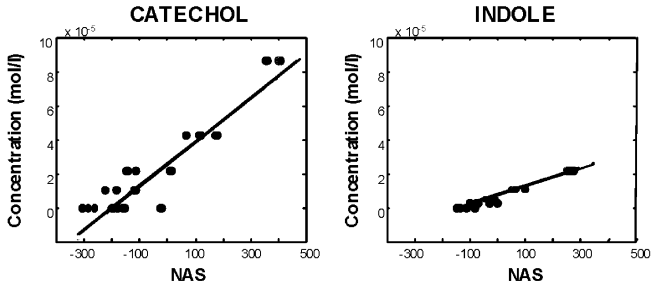


Figure 4 Plots of analyte concentration versus the scalar net analyte signal (NAS) with all samples illustrated individually.

Application of the correction based on the net analyte signal

Below, the applicability of the correction (Equation 15) is shown for the four fish species (data set 1). No effect of applying the correction was found when predicting the analyte concentrations (data set 2) from fluorescence spectra due to the very low prediction error variance and the very low influence of errors in the fluorescence spectra. Therefore, these results are not discussed further. The scalar net analyte signals are used as univariate x-values and the norm of the regression coefficient vector obtained from PLS modeling is used as the regression coefficient. Thus, the predictions will be similar to the predictions obtained directly from multivariate modeling. Regressions are made with averages of six replicates in the calibration set and averages of one to 12 replicates in the validation set. The models are validated with full-leave-one-fish-out cross validation. For each number of replicates used in the validation set, the average RMSEP of ten models is calculated. In the previous sections only results for one model were used. The actual NMR relaxation curves used in each model is chosen randomly.

Figure 5 shows RMSEP obtained directly from the PLS regression model (Figure 5.a) and by correcting the norm of the PLS regression vector (Figure 5.b). For cod and whiting RMSEP

decreases with increasing number of replicates. This is in agreement with the previous results showing that these two fish species are more influenced by errors in the NMR relaxations.

The relative difference in RMSEP between the two methods is illustrated in Figure 5.c. A positive value indicates that the application of the correction performs best. The figure shows that application of the correction may be an advantage for cod and haddock, when averages over few replicates are used in the test set. For these situations a positive relative difference in RMSEP is obtained. With more replicates in the test set of cod and haddock, the two methods give almost the same prediction errors. Saithe does not show any difference between the two methods and the relative difference in RMSEP is approximately zero irrespective of the number of replicates in the test set.

The above indicates that application of the correction may be preferred when errors in the instrumental variables have an influence on the predictions. When these errors are not important there will be no advantage of using the correction and the RMSEP obtained by the two methods will be approximately similar. This is seen for cod and haddock with more than about three replicates in the test set. For saithe this is seen for all numbers of replicates in the test set.

The results obtained for whiting seems more peculiar in that the application of the correction gives poorer predictions with few replicates in the test set and better predictions with many replicates. This may be due to the larger RMSEP and mean variance of prediction error obtained for whiting, which indicates that the regression models may not perform well.

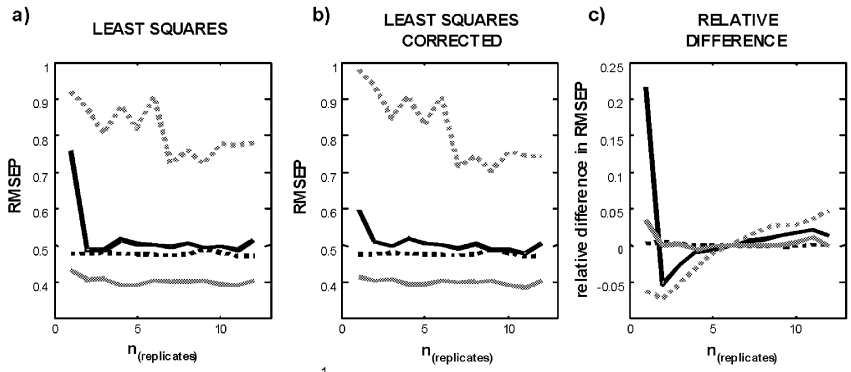


Figure 5 Prediction of water content from ^{1}H low-field NMR relaxations. Scalar net analyte signals are used as x-variable and the norm of the regression vector obtained by PLS used as the regression coefficient. Models are validated with full cross validation and include six replicates in the calibration set. a) RMSEP as a function of the number of replicates in the validation set, b) RMSEP as a function of the number of replicates in the validation set where the regression coefficient is corrected as given in Equation 15 and c) relative difference in RMSEP given as RMSEP by obtained directly by the norm of the regression vector minus the RMSEP obtained using the correction. The lines indicate: cod (solid black), haddock (solid grey), saithe (dotted black) and whiting (dotted grey).

The results for cod, saithe and haddock show that the application of the correction is an advantage when there are few replicates in the validation set and when the errors in **X** are important for the predictions. However, the inferior results obtained for whiting indicates that the correction only applies when the regression model is valid. More experiments are required before the method can be generally applied. It is important to find out in which situations the correction is valid and in which situations it is not.

Conclusion

Sampling errors and instrumental noise are two types of errors that may affect regression and prediction. This paper has presented ways of quantifying, visualizing and handling errors in instrumental measurements when these are used for estimating a reference value by multivariate regression. The methods presented are but a few examples of how the effect of errors can be evaluated and handled.

Acknowledgements

The authors are grateful for financial support through the AQM (Advanced Quality Monitoring) supported by the Danish Ministries of Research and Industry.

References

Andersen, C.M., Bro, R. and Brockhoff, P. Effect of sampling errors on predictions using replicated measurements. (2002) *Submitted.*

Bro, R. and Andersen, C.M. Theory of net analyte signal vectors in inverse regression. (2002) *Submitted*

Carr, H.Y. and Purcell, E.M. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, **94**, 630-638 (1954).

Faber, N.M. Exact presentation of multivariate calibration model as a univariate calibration graph. *Chemometrics and Intelligent Laboratory Systems*, **50**, 107-114 (2000).

Faber, N.M. Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Analytical Chemistry*, **70**, 5108-5110 (1998a).

Faber, N.M. Mean centering and computation of scalar net analyte signal in multivariate calibration. *Journal of Chemometrics*, **12**, 405-409 (1998b).

Faber, N.M., Duewer, D.L., Choquette, S.J., Green, T.L. and Chesler, S.N. Characterizing the uncertainty in near-infrared spectroscopic prediction of mixed-oxygenate concentrations in gasoline: sample-specific prediction intervals. *Analytical Chemistry*, **70**, 2972-2982 (1998).

Faber, K. and Kowalski, B.R. Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares. *Journal of Chemometrics*, **11**, 181-238 (1997).

Faber, K. and Kowalski, B.R. Prediction error in least squares regression: further critique on the deviation used in The Unscrambler. *Chemometrics and Intelligent Laboratory Systems*, **34**, 283-292 (1996).

Ferré, J., Brown, S.D. and Rius, F.X. Improved calculation of the net analyte signal in inverse multivariate calibration. *Journal of Chemometrics*, **15**, 537-553 (2001).

Fuller, W.A. *Measurement error models*. John Wiley. New York, 1987

Goicoechea, H.C. and Olivieri, A.C. A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemometrics and Intelligent Laboratory Systems*, **56**, 73-81 (2001).

Goicoechea, H.C. and Olivieri, A.C. Enhanced Synchronous spectrofluorometric determination of tetracycline in blood serum by chemometric analysis. Comparison of partial least-squares and hybrid linear analysis calibrations. *Analytical Chemistry*, **71**, 4361-4368 (1999).

Lorber, A. Error propagation and figures of merit for quantification by solving matrix equations. *Analytical Chemistry*, **58**, 1167-1172 (1986).

Lorber, A., Faber, K. and Kowalski, B.R. Net analyte signal correction in multivariate calibration. *Analytical Chemistry*, **69**, 1620-1626 (1997).

Martens, H. and Martens, M. Multivariate analysis of quality – An introduction. John Wiley & Sons, Ltd. England, 2001.

Meiboom, S. and Gill, D. Modified spin-echo method for measuring nuclear relaxation times. *The Review of Scientific Instruments*, **29**, 688-691 (1958).

Xu, L. and Schechter, I. A calibration method free of optimum factor number selection for automated multivariate analysis. Experimental and theoretical study. *Analytical Chemistry*, **69**, 3722-3730 (1997).

Appendix 1

Freshness assessment of thawed and chilled cod fillets packed in modified atmosphere using near-infrared spectroscopy

Niels Bøknæs, Kristina N. Jensen, Charlotte M. Andersen and Harald Martens

Lebensm.-Wiss. u.-Technol., 35, 628–634 (2002)

Research Note

Freshness Assessment of Thawed and Chilled Cod Fillets Packed in Modified Atmosphere Using Near-infrared Spectroscopy

Niels Bøknæs*, Kristina N. Jensen, Charlotte M. Andersen and Harald Martens

N. Bøknæs, K. N. Jensen, C. M. Andersen: Technical University of Denmark, Department of Seafood Research (DIFRES), Danish Institute for Fisheries Research, Søltofts Plads, Building 221, 2800 Kgs. Lyngby (Denmark)
H. Martens: Technical University of Denmark, Institute of Biotechnology, Søltofts Plads, Building 221, 2800 Kgs. Lyngby (Denmark)

(Received September 11, 2001; accepted June 20, 2002)

Near-infrared reflectance (NIR) spectra was recorded of 105 samples of cod mince prepared from chill stored thawed cod fillets of varying quality in modified atmosphere packaging (MAP). Traditional chemical, physical, microbiological and sensory quality methods developed for assessing fresh fish products were determined on the same cod fillets. The purpose was to evaluate the potential of NIR spectroscopy for estimating (i) frozen storage temperature, (ii) frozen storage period and (iii) chill storage period of thawed-chilled MAP Barents Sea cod fillets. Furthermore, the potential for measuring of selected quality attributes as drip loss, water holding capacity and content of dimethylamine by NIR was evaluated. The results of the investigation were presented using multivariate modelling methods such as partial least-squares regression (PLSR) and discriminant partial least-squares regression (DPLSR). Systematic differences in the NIR measurements on minced cod fillets were primarily due to the chill storage duration (days at 2 °C) on thawed-chilled MAP fillets. PLSR models based on wavelengths selected by a new Jack-knife method resulted in a correlation coefficient of 0.90 between measured and predicted duration of chill storage period (days at 2 °C). The root-mean-square error of cross-validation (RMSECV) was 3.4 d at 2 °C. NIR measurements provided promising results for evaluation of freshness for thawed-chilled MAP cod fillets completing the traditionally quality methods. However, it is necessary to study the effect of e.g. sample preparation, season, fishing ground and cod size together with more sophisticated pre-treatments of NIR spectra before the NIR method can be integrated as a method for evaluation of thawed-chilled MAP cod fillets.

Keywords: fish; frozen storage; MAP; multivariate analysis; NIR; thawing

Introduction

The use of 'frozen at sea' cod raw material instead of fresh cod fillets for the production of chilled products packed in modified atmosphere packaging (MAP) for retail sale seems promising (Bøknæs *et al.*, 2002). The quality of thawed MAP fish product depends on processing conditions as (i) frozen storage temperature, (ii) frozen storage period and (iii) chill storage period for cod fillets (Bøknæs *et al.*, 2000, 2001, 2002). The limiting factor for the shelf-life of fresh MAP cod products is *Photobacterium phosphoreum* growth and trimethylamine (TMA) production (Dalgaard *et al.*, 1993; Dalgaard, 1995) and the quality can be evaluated using a series of time-consuming physical, chemical and sensory parameters. However, near-infrared reflectance (NIR) spectroscopy for the estimation of processing conditions is an attractive possibility because it is fast, easy to handle, inexpensive to use and has the ability to carry out many samples as routine. NIR spectroscopy is already a well-established method for determining fat and water content in fish (Sollid and Solberg, 1992; Downey, 1996; Wold *et al.*, 1996; Wold and Isaksson, 1997). In addition, recent studies have shown good correlations between NIR spectroscopy and quality attributes of frozen gadoid fish (Jørgensen and Jensen, 1997; Bechmann and Jørgensen, 1998; Pink *et al.*, 1998, 1999). In recent Norwegian studies, NIR has also been used to assess freshness of cod, measured as days on ice (Sigernes *et al.*, 1998). However, the assessment of processing conditions for thawed–chilled MAP cod fillets by NIR spectroscopy has, to our knowledge, never been performed.

The objectives of the present study were to evaluate the potential of NIR spectroscopy to estimate (i) frozen

*To whom correspondence should addressed.
E-mail: nib@dfu.dk

0023-6438/02/$35.00
doi:10.1006/fstl.923

lwt/vol. 35 (2002) No. 7

storage temperature, (ii) frozen storage period, (iii) chill storage period and (iv) measurements of drip loss, water holding capacity (WHC) and dimethylamine (DMA) of thawed cod fillets from the Barents Sea packed in modified atmosphere and stored at 2 °C. Multivariate data analysis was used to determine the correlation between NIR spectroscopy and the selected storage conditions and physical/chemical measurements.

Materials and Methods

Experimental design

A factorial design with two frozen storage temperatures (−20 or −30 °C), four frozen storage periods (3, 6, 9 or 12 mo) as shown in **Table 1** and five chill storage periods (0, 3, 7, 14 or 21 d at 2 °C) was used as previously described by Bøknæs *et al.* (2002). Cod (*Gadus morhua*) were caught, processed and frozen on-board a Russian/Greenlandic freezer trawler in February 1999 in the Norwegian zone (72 °N, 16 °W) of the Barents Sea. Tow duration was 6 h and catch size about 6 metric tonnes, primarily cod for this single catch. Cod of 1.5–3 kg were deheaded and gutted immediately after catch. Cod were bled in seawater for about 30 min at 4 °C and then filleted and skinned by, respectively, a BAADER 190 and BAADER 51 (Nordischer Maschinenbau Rud. BAADER, Lübeck, Deutchland). Cod fillets were trimmed manually for bones, parasites and blood stains. Boneless fillets were manually interleaved packed when separating cod fillets with plastic film in the blocks. The packed fillet blocks, each containing ca 6.8 kg, were frozen in a horizontal plate freezer until a core temperature of −25 °C was reached (ca. 2 h). Frozen cod blocks were packed in cardboard boxes and placed in the cold store on-board the freezer trawler (ca. −30 °C). After 10 wk, the frozen blocks were landed and transported to the Danish Institute for Seafood Research (DIFRES) where they were randomly divided into two portions and kept at −20 and −30 °C, respectively. From each portion, frozen blocks were collected after 3, 6, 9 and 12 mo and sawed into pieces of less than 100 g. Afterwards, frozen cod pieces weighing altogether ca. 250 g were placed in trays with absorbent drip pads and packed in Riloten bags (Danisco Flexible, Lyngby, Denmark) containing a modified atmosphere with 40% CO_2, 40% N_2 and 20% O_2 (AGA, Copenhagen, Denmark). The fish-to-gas ratio was ca. 250 g cod and ca. 500 mL gas. Packed fillet pieces were then thawed for 20 h at 5 °C whereafter they were transferred for chill storage at 2 °C. After different days of chill storage, three packs from each batch were removed and homogenized for NIR analysis. The thawed MAP cod fillets (ca. 100 g) were homogenized for 15 s using a Speedy meat mincer (Krups, No. 7180-70, Ireland). Reference measurements of the selected quality attributes: drip loss, WHC and content of DMA were described in the corresponding study (Bøknæs *et al.*, 2002). For all experiments frozen and chill storage temperatures were recorded by loggers (Tinytag, Gemini Data Loggers Ltd., Chichester, U.K.).

Near-infrared measurements

For NIR analysis, minced cod was placed standardized in a tube (weight of 10 g cod mince, height of 3.5 cm and measured surface area of ca. 3 cm^2). Afterwards, the tubes were placed in a black cylinder and NIR reflectance spectra was recorded from the surface of the cod mince. During measuring, the cod samples were stored on ice. NIR reflectance spectra was measured in duplicate for each pack (three packs per batch) with an InfraProver, II Fourier transform spectrometer (Bran and Luebbe, Germany) using an optical fibre bundle as described by Bechmann and Jørgensen (1998). The spectra were sampled at 12 cm^{-1} intervals from 4,500 to 9,996 cm^{-1} (1,000–2,222 nm), i.e. 459 data points per spectrum. Preliminary data analysis revealed two extreme samples, which were removed as outliers before further data analysis. The six reflectance measurements for each batch were averaged before pre-treatment and modelling of the data. Different types of pre-treatment of the resulting NIR spectra were tested, and the following was chosen due to its simplicity and its

Table 1 Storage characteristics for all batches including frozen storage period and frozen storage temperature and chill storage temperature. All seven batches were stored at −29.0 ± 2.7 °C for the first 3 mo of frozen storage. Results were shown as average ± S.D. CO_2 concentration during chill storage was 33.7 ± 3.5% for all the seven batches

Batch	Storage period (mo)	Frozen storage temperature (°C)	Chill storage Temperature (°C)
A[a] (−30 °C)	3	−29.0 ± 2.7[b]	2.1 ± 0.5[c]
B[a] (−30 °C)	6	−31.5 ± 2.9[b]	2.5 ± 0.6[c]
C[d] (−20 °C)	6	−24.6 ± 4.3[b]	2.5 ± 0.6[c]
D[a] (−30 °C)	9	−32.2 ± 2.6[b]	2.1 ± 0.7[c]
E[d] (−20 °C)	9	−23.5 ± 4.0[b]	2.1 ± 0.7[c]
F[a] (−30 °C)	12	−32.4 ± 3.4[b]	2.5 ± 0.2[c]
G[d] (−20 °C)	12	−23.4 ± 3.6[b]	2.5 ± 0.2[c]

[a]Significant ($P<0.05$) *P. phosphoreum* growth and TMA production during chill storage in MAP.
[b]No significant ($P>0.05$) difference between frozen storage temperatures.
[c]No significant ($P>0.05$) difference in chill storage temperature.
[d]No significant ($P>0.05$) *P. phosphoreum* growth and TMA production during chill storage in MAP.

lwt/vol. 35 (2002) No. 7

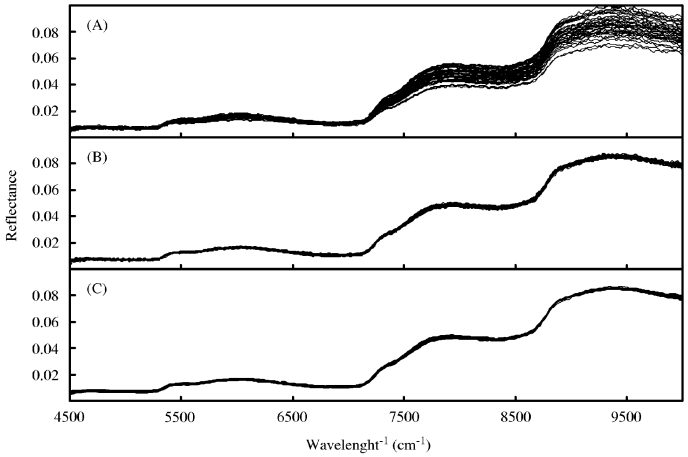


Fig. 1 NIR reflectance spectra measured on minced thawed MAP cod fillets showed: (**A**) all the 35 averaged raw NIR spectra; (**B**) all the 35 averaged and MSC NIR spectra and (**C**) averaging and smoothing of all the 35 averaged and MSC NIR spectra

successful modelling (see below). Each average reflectance spectrum was corrected for additive and multiplicative main effects using multiplicative scatter correction (MSC) (Geladi *et al.*, 1985) for eliminating the effects of physical phenomena like the light-scattering effects of particles with different sizes and shapes (Isaksson and Næs, 1988; Helland *et al.*, 1995). Since the MSC-treated and averaged reflectance spectra was rather noisy, they were averaged and smoothed in the final modelling. Batch mean of NIR spectra on minced cod fillets before and after pre-treatment (MSC, averaging and smoothing) was shown in **Fig. 1**. The number of wavelength channels was reduced down to 230 by averaging of adjacent wavelengths of the MSC-treated NIR reflectance spectra. Afterwards, the spectra was smoothed by a moving average filter based on three neighbouring wavelength channels.

Multivariate data analysis

Partial least-squares regression (PLSR) (Wold *et al.*, 1983) was used for the multivariate calibration in (i) qualitative modelling and (ii) quantitative modelling. In the qualitative modelling, discriminant partial least-squares regression (DPLSR) (Wold *et al.*, 1983) was used in order to gain overview of the data (see also Martens and Martens, 1986; Martens and Jørgensen, 1998; Jacobsen *et al.*, 1999, 2001). In order to investigate the effect of (i) frozen storage temperature, (ii) frozen storage period and (iii) chill storage period for thawed–chilled MAP cod fillets, DPLSR was employed to predict qualitative main effects of the experimental design (Y), from the averaged and smoothed MSC-treated NIR data (X). This means that each frozen storage temperature (−20 or −30 °C), each frozen storage period (3, 6, 9 or 12 m) and each chill storage period (0, 3, 7, 14 or 21 d at 2 °C) were represented by an indicator variable (with values 0 or 1). No standardization (weighting) was used on X and Y variables in the qualitative modelling. The model was validated by full cross-validation.

In the quantitative modelling, PLSR was employed to determine the predictive ability of the NIR spectra (X) and the duration of chill storage period (now represented as one single quantitative variable y, representing the number of days at 2 °C) for thawed MAP cod fillets. Furthermore, PLSR was used to determine the predictive ability of the NIR spectra and, respectively, (i) drip loss, (ii) WHC and (iii) content of DMA. For analysing PLSR results, a newly developed modified Jack-knife method for estimation of parameter uncertainty in PLSR was used (Martens and Martens, 1999). In this case, a manual selection of wavelengths with $P>0.1$ was used for elimination of nonsignificant wavelengths in the NIR spectra. In the quantitative modelling, the NIR data (X variables) was standardized (weighting with 1/standard deviation) since it improved the predictability. The single y-variables were not weighted. The PLSR models were validated by full cross-validation. Multivariate data analysis was performed using The Unscrambler® Ver. 7.6 (CAMO, Norway).

Results

Experimental design

The experimental batch characteristics including temperatures for frozen and chill storage were shown in **Table 1**. The growth of *P. phosphoreum* and TMA production in the thawed–chilled MAP cod fillets was showed in the corresponding study (Bøknæs *et al.*, 2002).

lwt/vol. 35 (2002) No. 7

Multivariate data analysis

Qualitative modelling. The data set consisted of 35 samples (7 batches × 5 periods of chill storage) in an NIR spectra matrix (35 × 459) and a design matrix (35 × 10). The design matrix contained qualitative indicators for the main effects, i.e. 0 or 1 variables for (i) frozen storage temperature (−20 or −30 °C), (ii) frozen storage period (3, 6, 9 or 12 mo) and (iii) chill storage period (0, 3, 7, 14 or 21 d at 2 °C). The DPLSR model contained three factors explaining 64% of the variance in X and 22% of the variance in Y. The scores for the first two factors were shown in **Fig. 2**.

In the scores plot (**Fig. 2**), differences in chill storage between 0 and 21 d expanded the first factor for the DPLSR. Samples with few and many days of chill storage were mostly placed at the left and the right in the plot, respectively (**Fig. 2**). This indicated a weak tendency for the fact that the first factor accounted for most of the variation due to the chill storage period of thawed MAP cod fillets. No systematic differences in MSC-treated NIR spectra for frozen storage temperature and period were observed according to the second and third factor. However, this qualitative modelling indicated that NIR spectra could be used for estimating chill storage period of thawed MAP cod fillets.

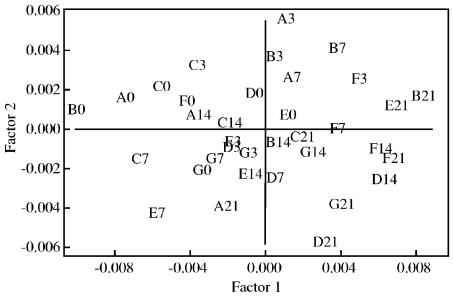


Fig. 2 Scores plot of factor 1 *vs.* factor 2 from DPLSR analysis including NIR spectra and all three design variables. Letters indicate different frozen storage temperatures and frozen storage periods (see **Table 1**). Numbers indicate the period of chill storage in days at 2 °C. Explained variance for X were 44 and 14% and for Y 6 and 7% for factors 1 and 2, respectively

Quantitative modelling. PLSR models using NIR for estimating the period of chill storage (y = days at 2 °C) and selected quality measurements of thawed MAP cod fillets were evaluated. The PLSR models were based on MSC-treated NIR spectra for all 35 samples independent of frozen storage temperature and period. In order to reduce the risk of chance correlations in the Jack-knifed variable selection, the number of wavelength channels was reduced down to 230 by averaging and smoothing wavelengths of the MSC-treated NIR reflectance spectra.

An initial PLSR model was developed, using all 230 wavelength channels of smoothed MSC-treated NIR reflectance spectra as X-variables and chill storage period (0, 3, 7, 14 or 21 d at 2 °C) as y-variable (**Table 2**). The Jack-knife method was used for selecting significant wavelengths in smoothed MSC-treated NIR spectra. All wavelengths that were found to give nonsignificant contributions ($P > 0.1$) estimated by full cross-validation using the Jack-knife method were eliminated in order to remove detrimental X-variables from the model. To further guard against incidental chance correlations, single significant wavelength channels, i.e. with non-significant neighbours on either side, were also eliminated. The regression coefficients for the PLSR models before and after elimination of nonsignificant wavelengths were shown in **Fig. 3**. The significant wavelengths in **Fig. 3B** were not directly connected to water or protein. However, the significant wavelength region between 7600 and 8000 cm^{-1} (connected with C–H groups) could possibly be related to protein changes during chill storage. The PLSR models were evaluated for their number of factors, root-mean-square error of cross-validation (RMSECV), correlation between predicted and measured values, and by the amount of explained validation variance of the Y data. These results were shown in **Table 2**.

When the Jack-knife method was used to select significant NIR reflectance wavelengths, a lower prediction error of the chill storage period by MSC-treated

Table 2 Validation of PLSR-models for the prediction of chill storage period (days at 2 °C), drip loss, DMA and WHC for thawed–chilled MAP cod fillets from NIR data using full spectrum (230 wavelengths) or selected wavelengths of averaged and smoothed MSC–treated NIR data

y-Value	X-values	Resulting range	Number of PLSR factors	RMSECV[a]	Correlation coefficient	% Y var. expl.[b]
Chill storage (d)	Full spectrum	0–21	4	4.9	0.77	61
Chill storage (d)	Selected wavelengths[c]	0–21	3	3.4	0.90	81
Drip loss (%)	Full spectrum	3.2–13.8	3	2.3	0.69	49
Drip loss (%)	Selected wavelengths[c]	3.2–13.8	2	2.0	0.78	63
DMA (mg DMA-N/100 g)[d]	Full spectrum	0–1.5	—	—	<0.1	–
WHC (%)[d]	Full spectrum	62.5–84.8	—	—	<0.1	–

[a]The root-mean-square error of cross-validation (RMSECV) is expressed in the same units as for the y-value.
[b]Amount of variance in the Y-matrix explained by the models (%).
[c]Significant wavelengths selected by the Jack-knifing method with $P < 0.1$.
[d]Since no correlation between NIR and y-value was established, no validation values were reported.

lwt/vol. 35 (2002) No. 7

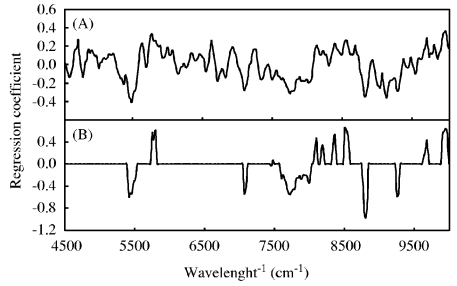


Fig. 3 Regressions coefficients from PLSR models for prediction of chill storage period (days at 2 °C). The regressions coefficients were shown for the optimal number of factors in the model: (**A**) before elimination of nonsignificant wavelengths in a four-factor model and (**B**) after elimination of nonsignificant wavelengths in a three-factor model

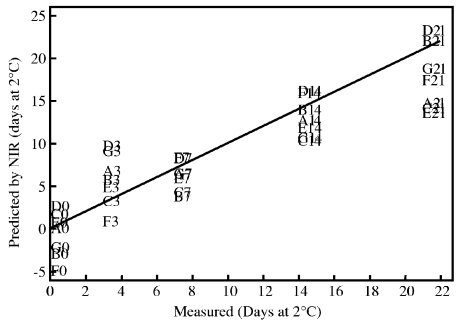


Fig. 4 The relation between duration of chill storage (days at 2 °C) of thawed MAP cod (abscissa) and predicted chill storage period (days at 2 °C) determined from the NIR measurements (ordinate) as obtained by a cross-validated PLSR model based on only significant Jack-knifing wavelengths. The solid line indicates $y = x$. Letters indicate different frozen storage temperatures and frozen storage periods (see Table 1). Numbers indicate the period of chill storage in days at 2 °C

NIR spectra was found (an increase from 61% to 81% of explained variance). The initial use of all NIR wavelengths may thus have included too many noisy and irrelevant X-variables in the model (**Table 2**). Results from PLSR modelling for prediction of WHC, DMA and drip loss were also shown in **Table 2**. These PLSR results showed the same tendency for a quite good prediction of drip loss by NIR as obtained for prediction of chill storage period. However, we obtained very low correlations (<0.1) between NIR and measurements of DMA or WHC (**Table 2**).

Discussion

Systematic differences in the NIR measurements on minced cod fillets, observed in the qualitative calibration models (**Fig. 2**) were primarily due to the chill storage duration (days at 2 °C) on thawed–chilled MAP cod fillets. In contrast, frozen storage temperature and frozen storage period did not influence the NIR measurements very much, at least not in a clear and systematic way. The final quantitative calibration model for prediction of chill storage period, after elimination of nonsignificant wavelength channels, showed quite good predictive ability in the cross validation (**Fig. 3**). Compared to the full spectrum model, the Jack-knifed-based variable selection apparently caused an appreciable improvement (from 61% to 81% of explained variance) in the prediction of the number of days of chill storage at 2 °C (y).

To what extent could this cross-validated improvement be an artefact? The cross-validation by itself is quite safe (Martens and Dardenne, 1998), since the only way the predicted y-variables of the samples have been used, other than for estimating the cross-validated prediction error (RMSECV) itself, is to determine one single parameter—the optimal number of factors. Hence, the initial result (61% prediction ability) is good enough.

However, the combination of Jack-knife-based variable selection and cross-validation needs to be scrutinized. In the Jack-knife-based testing, a high number of more or less independent wavelength channels are assessed with respect to their apparent effect on model stability, and only the best ones are retained in the model. Then, in the cross-validation, the prediction error estimation is based on more or less the same criterion—the X–y agreement between each individual sample and the other samples. Therefore, there is a risk of overfitting that is not detected by the cross-validation. If many incidentally significant X-variables survive the Jack-knife testing (we expect about 10% of nonsense variables to survive the chosen test level), then their combination may create an artificially high cross-validated gain in the predictive ability to be too high.

Preliminary quantitative modelling (results not shown), based on the original, unsmoothed NIR spectra with rather noisy 459 X-variable channels indeed indicated some over fitting even in the cross-validation, because the resulting regression coefficient vector was quite noisy. This was the reason why the number of X-variables was reduced to 230 and smoothed, and finally, that the few retained lonely individual wavelength channels were eliminated manually. With this procedure, we expect the estimated prediction error RMSECV of the final calibration model to be realistic for the models predictive ability in new samples of the same general kind.

This final PLSR model for prediction of chill storage period gave a correlation coefficient of 0.9 and RMSECV on 3.4 d. The prediction of chill storage (days at 2 °C) was obtained more or less independently of the previous frozen storage temperature and frozen storage period. In an earlier study concerning prediction of storage period for fresh cod fillets with no precedent frozen storage treatment, Sigernes *et al.* (1998) found a correlation coefficient of 0.95 and RMSECV on 28 h. Hence, it seems that NIR has a future potential of measuring chill storage duration for the thawed MAP

lwt/vol. 35 (2002) No. 7

cod fillets (**Table 2**). This could be very useful in practice, because the application of traditional methods for the quality assessment of MAP cod fillets including number of *P. phosphoreum* and TMA production were difficult due to differences in microbiological spoilage of thawed–chilled MAP Barents Sea cod (**Table 1**) depending of the precedent frozen storage temperature (Bøknæs *et al.*, 2002).

During chill storage, significant changes in texture for thawed MAP cod fillets have previously been reported (Guldager *et al.*, 1998; Bøknæs *et al.*, 2000, 2001, 2002). However, textural changes obtained for thawed–chilled MAP cod fillets have been related to inevitable processes during freezing, thawing and chill storage such as protein denaturation and enzymatic reactions (Shenouda, 1980; Rehbein, 1988). Such textural changes might be influenced in NIR spectra shown in the present study. This is confirmed in the present study using averaged and smoothed MSC-treated NIR data for predicting drip loss of thawed–chilled MAP cod fillets with a correlation coefficient of 0.78 and RMSEPCV of 2.0%. In addition, a recent study also showed the potential for using visual/near-infrared reflectance spectroscopy for nondestructive estimating of texture in farmed Atlantic salmon (*Salmo salar*) (Isaksson *et al.*, 2001). Recently in Canadian studies, NIR spectroscopy has been used to predict DMA content in frozen minced Atlantic red hake (*Urophycis chuss*) with correlation coefficient higher than 0.95 (Pink *et al.*, 1998, 1999). In contrast, a very low correlation coefficient (<0.1) between NIR spectra and DMA content in thawed–chilled MAP cod fillets was obtained in the present study. This might be related to the very little formation of DMA during frozen and chill storage as DMA content was found to be below 1.5 mg DMA-N/100 g for all batches in the present study (Bøknæs *et al.*, 2002). Furthermore, a very low correlation coefficient (<0.1) on predicting WHC by NIR was also found in the present study. In contrast, recent studies (Jørgensen and Jensen, 1997; Bechmann and Jørgensen, 1998) showed higher correlation using NIR measured on the skin of whole thawed cod or on centrifuged cod mince for determining WHC in thawed cod fillets. However, it was also shown that the use of NIR spectra of mince as in this study was inferior to the use of spectra of the centrifuged mince (Jørgensen and Jensen, 1997). This might explain the very low correlation coefficient between NIR on raw cod mince and WHC obtained in the present study.

In conclusion, NIR is a promising method for estimating duration of chill storage period of thawed–chilled MAP cod fillets. However, before NIR can be introduced for the evaluation of thawed–chilled MAP cod fillets, it is necessary to develop a calibration model spanning the variations related to e.g. chill storage period, sample preparation, season, fishing ground and cod size. Furthermore, data modelling could be optimized using other more sophisticated pre-treatments for NIR measurements on raw thawed MAP cod fillets instead of measuring on minced cod.

Acknowledgements

This project was supported by the Danish Academy of Technical Sciences and Royal Greenland Ltd. The authors wish to thank Carsten Østerberg and Rie Sørensen for their excellent technical assistance during the experiments.

References

Bechmann, I. E. and Jørgensen, B. Rapid assessment of quality parameters for frozen cod using near infrared spectroscopy. *Lebensmittel-Wissenschaft und-Technologie*, **31**, 648–652 (1998)

Bøknæs, N., Østerberg, C., Nielsen, J. and Dalgaard, P. Influence of freshness and frozen storage temperature on quality of thawed cod fillets stored in modified atmosphere packaging. *Lebensmittel-Wissenschaft und-Technologie*, **33**, 244–248 (2000)

Bøknæs, N., Østerberg, C., Sørensen, R., Nielsen, J. and Dalgaard, P. Effects of technological parameters and fishing ground on quality attributes of thawed chilled cod fillets stored in modified atmosphere packaging. *Lebensmittel-Wissenschaft und-Technologie*, **34**, 513–520 (2001)

Bøknæs, N., Jensen, K. N., Guldager, H. S., Østerberg, C., Nielsen, J. and Dalgaard, P. Thawed chilled Barents Sea cod fillets in modified atmosphere packaging—application of multivariate data analysis to select key parameters in good manufacturing practice. *Lebensmittel-Wissenschaft und-Technologie*, **35**, 436–443 (2002)

Dalgaard, P. Qualitative and quantitative characterization of spoilage bacteria from packed fish. *International Journal of Food Microbiology*, **26**, 319–333 (1995)

Dalgaard , P., Gram, L. and Huss, H. H. Spoilage and shelf life of cod fillets packed in vacuum or modified atmospheres. *International Journal of Food Microbiology*, **19**, 283–294 (1993)

Downey, G. Non-invasive and non-destructive percutaneous analysis of farmed salmon flesh by near infra-red spectroscopy. *Food Chemistry*, **55**, 305-311 (1996)

Geladi, P., Macdougall, D. and Martens, H. Linearization and scatter-correction for near-infrared reflectance of meat. *Applied Spectroscopy*, **39**, 491–500 (1985)

Guldager, H. S., Bøknæs, N., Østerberg, C., Nielsen, J. and Dalgaard, P. Thawed cod fillets spoil less rapidly than unfrozen fillets when stored under modified atmosphere at 2 °C. *Journal of Food Protection*, **61**, 1129–1136 (1998)

Helland, I. S., Næs, T. and Isaksson, T. Related versions of the multiplicative scatter correction method for pre-processing spectroscopic data. *Chemometrics and Intelligent Laboratory Systems*, **29**, 233–241 (1995)

Isaksson, T. and Næs, T. The effect of multiplicative scatter correction (MSC) and linear improvement in NIR spectroscopy. *Applied Spectroscopy*, **42**, 1273–1284 (1988)

Isaksson, T. Swensen, L. P., Taylor, R. G., Fjæra, S. O. and Skervold, P. O. Non-destructive texture analysis of farmed Atlantic salmon using visual/near-infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture*, **82**, 53–60 (2001)

Jacobsen, C., Hartvigsen, K., Lund, P., Meyer, A. S., Adler-Nissen, J., Holstborg, J. and Hølmer, G. Oxidation in fish-oil-enriched mayonnaise, 1. Assessment of propyl gallate as an antioxidant by discriminating partial least squares regression analysis. *European Food Research Technology*, **210**, 13–30 (1999)

Jacobsen, C., Hartvigsen, K., Thomsen, M. K., Hansen, L. F., Lund, P., Skibsted, L., Hølmer, G., Adler-Nissen, J. and Meyer, A. S. Lipid oxidation in fish oil enriched mayonnaise: calcium disodium ethylenediaminetetraacetate,

lwt/vol. 35 (2002) No. 7

but not gallic acid, strongly inhibited oxidative deterioration. *Journal of Agriculture and Food Chemistry*, **49**, 1009–1019 (2001)

JØRGENSEN, B. AND JENSEN, H. S. Can near-infrared spectrometry be used to measure quality attributes in frozen cod? In: LUTEN, J. B., BØRRESEN, T. AND OEHLENSCHLÄGER, J. (Eds), *Seafood from Producer to Consumer, Integrated Approach to Quality*. Amsterdam: Elsevier, pp. 491–496 (1997)

MARTENS, H. AND DARDENNE, P. Validation and verification of regression in small data sets. *Chemometrics and Intelligent Laboratory Systems*, **44**, 99–122 (1998)

MARTENS, M. AND JØRGENSEN, B. Multivariate data analysis used for investigation of the sensory quality of fish. In: OLAFSDÓTTIR, G., LUTEN, J. P., DALGAARD, P., CARECHE, M., VERREZ-BAGNIS, V., MARTINSDÓTTIR, E. AND HEIA, K. (Eds), *Methods to Determine the Freshness of Fish in Research and Industry*. Paris: International Institute of Refrigeration, pp. 369–375 (1998)

MARTENS, M. AND MARTENS, H. Partial least squares regression. In: PIGGOTT, J. R. (Ed.), *Statistical Procedures*. London: Elsevier, pp. 293–359 (1986)

MARTENS, H. AND MARTENS, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). *Food Quality and Preference*, **11**, 5–16 (1999)

PINK, J., NACZK, M. AND PINK, D. Evaluation of the quality of frozen minced red hake: use of Fourier transform infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, **46**, 3667–3672 (1998)

PINK, J., NACZK, M. AND PINK, D. Evaluation of the quality of frozen minced red hake: use of Fourier transform near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry*, **47**, 4280–4284 (1999)

REHBEIN, H. Relevance of trimethylamine oxide demethylase activity and haemoglobin content to formaldehyde production and texture deterioration in frozen stored minced fish muscle. *Journal of the Science of Food and Agriculture*, **43**, 261–276 (1988)

SHENOUDA, S. Y. K. Theories of protein denaturation during frozen storage of flesh. In: CHICHESTER, C. O., MRAK, E. M. AND STEWART, G. F. (Eds), *Advances in Food Research*. New York: Academic Press, pp. 275–311 (1980)

SIGERNES, F., ESAIASSEN, K., HEIA, K., WOLD, J. P., EILERTSEN, G. AND SØRENSEN, N. K. Assessment of fish (cod) freshness by VIS/NIR spectroscopy. In: OLAFSDÓTTIR, G., LUTEN, J. P., DALGAARD, P., CARECHE, M., VERREZ-BAGNIS, V., MARTINSDÓTTIR, E. AND HEIA, K. (Eds), *Methods to Determine the Freshness of Fish in Research and Industry*. Paris: International Institute of Refrigeration, pp. 369–375 (1998)

SOLLID, H. AND SOLBERG, C. Salmon fat content estimation by near infrared transmission spectroscopy. *Journal of Food Science*, **57**, 792–793 (1992)

WOLD, S., ALBANO, C., DUNN III, W.J., ESBENSEN, K. HELLBERG, S., JOHANSSON, E. AND SJÖSTRÖM, M. Recognition: finding and using regularities in multivariate data. In: MARTENS, H. AND RUSSWURM, H. (Eds), *Food Research and Data Analysis*. London: Applied Science Publishers, pp. 147–188 (1983)

WOLD, J. P. AND ISAKSSON, T. Non-destructive determination of fat and moisture in whole Atlantic salmon by near-infrared diffuse spectroscopy. *Journal of Food Science*, **62**, 734–736 (1997)

WOLD, J. P., JAKOBSEN, T. AND KRANE, L. Atlantic salmon average fat content estimated by near-infrared transmittance spectroscopy. *Journal of Food Science*, **61**, 74–77 (1996)

Appendix 2

Theory of net analyte signal vectors in inverse regression

Rasmus Bro and Charlotte M. Andersen

Abstract

The net analyte signal and the net analyte signal vector are useful measures in building and optimizing multivariate calibration models. In this paper, a consistent theory for its use in inverse regression is developed. The theory of net analyte signal was originally derived from classical least squares in spectral calibration where the responses of all pure analytes and interferents are assumed to be known. However, in chemometrics, inverse calibration models such as partial least squares regression are more abundant and several tools for calculating net analyte signal in inverse regression models have been proposed. These methods yield different results and most do not provide results that are in accordance with the chosen calibration model. In this paper, a thorough development of a calibration-specific net analyte signal vector is given. This definition turns out to be equivalent to the one recently suggested by Faber [Faber 1998a]. A needed correction of the net analyte signal in situations with negative predicted responses is also offered.

Keywords: multivariate calibration, figures of merit

Introduction

In multivariate inverse calibration, a regression model is sought, predicting the response **y** of size $I\times1$ from the multivariate measurements in **X** ($I\times J$). The regression vector **b** ($J\times1$) is found to minimize the residuals, **e** ($I\times1$), in the equation

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \tag{1}$$

often subject to some additional constraints with respect to **X**. For example, in principal component regression (PCR), **X** is constrained to be in the subspace of the first principal components whereas in partial least squares regression (PLS), a slightly different criterion is used.

The net analyte signal vector is a convenient diagnostic that enables figures of merit (sensitivity, limit-of-detection etc.) to be calculated in a manner similar to univariate regression. The net analyte signal vector was originally defined by Lorber [Lorber 1986] as the part of a measured signal that is orthogonal to the interferents. Although, net analyte signals have mainly been discussed in relation to spectral data and especially in settings where Beers law is assumed to be valid, the principle is general and applicable and useful for any multivariate calibration model. In the following, though, the net analyte signal is discussed under the premise of spectral data.

The underlying idea of methods for calculating the net analyte signal is to separate the contributions in the calibration data matrix, **X**, into one originating solely from the analyte of

interest (called $\mathbf{X}_k$ to reflect that the *k*-th analyte is the analyte of interest), and another from other sources of variability such as interferents ($\mathbf{X}_{-k}$).

$$\mathbf{X} = \mathbf{X}_k + \mathbf{X}_{-k} \tag{2}$$

$\mathbf{X}_k$ denotes the unique part of the analyte signal and $\mathbf{X}_{-k}$ is the matrix describing the signal orthogonal to that. Ideally, the unique analyte part lies in the null-space of the space spanned by the spectra of the interferents. From a matrix spanning the space of the interferents ($\mathbf{X}_{-k}$), the net analyte signal vector of a sample, *i*, is calculated by

$$\mathbf{x}_{k,i}^{*} = (\mathbf{I} - (\mathbf{X}_{-k}^{+})\mathbf{X}_{-k})\mathbf{x}_i \tag{3}$$

where $\mathbf{x}_i$ is the spectrum of the *i*-th sample, $\mathbf{I}$ is an identity matrix and "+" denotes a pseudoinverse. The net analyte signal is usually taken as the norm of $\mathbf{x}_{k,i}^{*}$ and can be used similar to a univariate signal in univariate linear regression [Faber 2000]. For e.g. mean-centered data, a complication arises in calculating the net analyte signal. This will be treated upon deriving a consistent algorithm for the net analyte signal vector.

The matrix $\mathbf{I} - (\mathbf{X}_{-k}^{+})\mathbf{X}_{-k}$ projects the calibration spectra onto the space orthogonal to that spanned by the spectra of all analytes except the sought *k*-th analyte. Thus, in order to find the net analyte signal vector of a certain analyte it is necessary to find the projection matrix, $\mathbf{I} - (\mathbf{X}_{-k}^{+})\mathbf{X}_{-k}$, which involves finding the matrix describing the interferent spectra, $\mathbf{X}_{-k}$. There are several ways to estimate this matrix.

Lorber et al. [Lorber et al. 1997] suggested a method that uses PCR or PLS. First, the calibration matrix $\mathbf{X}$ is rebuilt using *A* significant PCR or PLS components, yielding $\mathbf{X}_{reb}$. Second, a rank annihilation step in the *A*-dimensional space is used for finding the part of the original matrix spanned by the interferents.

$$\mathbf{X}_{-k} = \mathbf{X}_{reb} - \alpha \mathbf{y}_k \mathbf{x}^{T} \tag{4}$$

where $\mathbf{y}_k$ is the projection of the vector of responses $\mathbf{y}$ ($I \times 1$) onto the *A*-dimensional subspace and is given by $\mathbf{y}_k = \mathbf{X}_{reb}\mathbf{X}_{reb}^{+}\mathbf{y}$. The vector $\mathbf{x}$ is a linear combination of the rows of $\mathbf{X}$, which is chosen to include a contribution from the spectrum of the *k*-th analyte. Any reasonable spectrum can be used for this purpose even though it is recommended to use a spectrum that contains maximal information of the analyte. The scalar α can be calculated as

$$\alpha = 1/\mathbf{x}^{T}\mathbf{X}^{+}\mathbf{y}_k \tag{5}$$

Xu and Schechter [Xu & Schechter 1997] proposed another approach where $\mathbf{y}$ is used to define $\mathbf{X}_{-k}$. The calibration matrix $\mathbf{X}$ is scaled by dividing each spectral vector of matrix $\mathbf{X}$ with the

corresponding y-value such that each spectral vector contains the same contribution of the analyte

$$\mathbf{x}_{i,cs} = \frac{\mathbf{x}_i}{y_i} \tag{6}$$

In the next step, the average of the scaled vectors is calculated and subtracted from all the scaled vectors. This gives a mean-centering pre-treatment of the scaled matrix removing the constant contribution of the analyte.

$$\mathbf{x}_{-k,i} = \mathbf{x}_{i,sc} - \bar{X}_{sc} \tag{7}$$

Combining $\mathbf{x}_{-k,i}$ for all samples provides an estimate of $\mathbf{X}_{-k}$. A similar approach is described by Goicoechea and Olivieri [Goicoechea & Olivieri 1999]. The mean calibration spectrum is obtained as in Eq. (8) and the contribution of the analyte is subtracted from the data matrix **X** as shown in Eq. (9).

$$\mathbf{x} = \frac{1}{I}\sum_{i=1}^{I}\mathbf{x}_i \tag{8}$$

$$\mathbf{X}_{-k} = \mathbf{X} - \frac{\mathbf{y}\mathbf{x}^T}{\bar{y}} \tag{9}$$

where $\bar{y}$ denotes the mean calibration concentration of the analyte. Goicoechea and Olivieri [Goicoechea & Olivieri 2001] proposed to define $\mathbf{X}_{-k}$ as the projection of **X** orthogonal to **y** as illustrated in Eq. (10).

$$\mathbf{X}_{-k} = [\mathbf{I} - \mathbf{y}(\mathbf{y}^T\mathbf{y})^{-1}\mathbf{y}^T]\mathbf{X} \tag{10}$$

Booksh and Kowalski [Booksh & Kowalski 1994] also described the net analyte signal vector and suggested a relation to the regression vector but they did not provide an operational method for its calculation. Faber [Faber 1998a] put forward an idea which does not require calculation of $\mathbf{X}_{-k}$. In this method the net analyte signal vector is calculated as

$$\mathbf{x}^*_{k,i} = \mathbf{b}(\mathbf{b}^T\mathbf{b})^{-1}y \tag{11}$$

where y is the concentration of the analyte in the unknown sample (in practice, the prediction of the concentration is used). This method was derived from the fact that the regression vector is the part of the analyte signal orthogonal to the signals of the interferents [Sanchez & Kowalski 1988]. The rationale for this development was to circumvent the computational burden in some of the prior methods and it was argued that the new method was an alternative that gave similar results as the older methods.

Theory

Most of the current approaches for calculating the net analyte signal vector suffer from the same problems. Unless noise-free data are used, the net analyte signal vector will not be proportional to the regression vector. Despite that this proportionality should theoretically be present, the indirect calculation of the net analyte signal vector introduces errors that propagate to the net analyte signal vector. This is unsatisfactory both from a theoretical and a practical point of view.

A further complication with some approaches is that the net analyte signal vector is to some extent independent of the calibration method. Hence, it can be impossible to compare different competing calibration models by their net analyte signal vector as it will be the same regardless of calibration method. This is unsatisfactory because the net analyte signal is a property of a specific model and not of a given data set. Building a calibration model for a specific data set can yield different models depending on the purpose of the model (robustness, accuracy etc.). E.g. when a model is validated to be robust over long periods, less components are likely significant than if the model is tested for highest possible accuracy over shorter periods. Hence, a data set as such can not have an associated net analyte signal vector. Unless idealized, this is only possible for a specific model built with a specific purpose in mind.

In this paper, a very simple calculation of the net analyte signal and its vector is derived. It has the advantage of not relying on any additional assumptions on the structure of the data and to provide a net analyte signal vector which will yield exactly the same predictions as the calibration model for which it is supposed to provide information. Furthermore, it handles mean-centering automatically in a way similar to that suggested by Faber [Faber 1998b].

The matrices $\mathbf{X}_k$ and $\mathbf{X}_{-k}$ can be defined under the premises of a given (rank-reduced) calibration model in the following way. First of all, the null-space of the model can be disregarded. For e.g. PCR, the null-space is given by the projection matrix $\mathbf{I}\text{-}\mathbf{P}\mathbf{P}^+$ where $\mathbf{P}$ ($J\times F$) is the F-dimensional loading matrix This part of the space is not used for predicting and is regarded as noise in the model (used for diagnostic purposes). Instead of dividing the data into two different parts as is usually done in net analyte signal derivations, the model is divided into

$$\mathbf{X} = \mathbf{X}_k + \mathbf{X}_{-k} + \mathbf{X}_e \tag{12}$$

all orthogonal to each other. The matrix $\mathbf{X}_k$ is the part of the signal of the analyte which is orthogonal to the signal of the interferents and the noise and $\mathbf{X}_{-k}$ is the part of the data in the subspace spanned by the interferents (including the part of the analyte signal within that subspace). The residual matrix $\mathbf{X}_e$ is e.g. defined as $\mathbf{X}(\mathbf{I}\text{-}\mathbf{P}\mathbf{P}^+)$ for PCR but follows similarly for other calibration methods.

Within the space of the model given by $\mathbf{PP}^{+}$, the part given by the regression vector **b** is assumed to be in the same direction as the net analyte signal vector. Otherwise, the predictions are influenced by other than analyte variation. Hence the null-space of this direction must be equivalent to the space spanned by $\mathbf{X}_{-k}$. Therefore, the matrix $\mathbf{X}_{-k}$ can be found as

$$\mathbf{X}_{-k} = \mathbf{XPP}^{+}(\mathbf{I} - \mathbf{bb}^{+}) \tag{13}$$

and hence

$$\mathbf{X}_{k} = \mathbf{XPP}^{+}\mathbf{bb}^{+} = \mathbf{Xbb}^{+} \tag{14}$$

This means that the net analyte signal vector is simply found by projecting the data onto the space spanned by the regression vector. The main result in this paper is therefore that the net analyte signal vector of a certain sample, $\mathbf{x}_i$, is found as

$$\mathbf{x}_{k,i}^{*} = \mathbf{bb}^{+}\mathbf{x}_i = \mathbf{b}(\mathbf{b}^{T}\mathbf{b})^{-1}\mathbf{b}^{T}\mathbf{x}_i \tag{15}$$

where **b** is the regression vector from the calibration model. Note that estimated values are used throughout, because the interest is in assessing a specific (estimated) model, not a theoretical construct. The above definition of the net analyte signal circumvents the problems of the older methods and can be seen to be identical to the one proposed by Faber when exchanging $\mathbf{b}^{T}\mathbf{x}_i$ with *y*. Thus, it retains the computational ease of his method. However, as opposed to earlier, it is argued that this is *the* way to calculate the net analyte signal vector and not just an alternative.

The net analyte signal is usually defined as the norm of the net analyte signal vector. However, as pointed out by Faber, this is not valid for models that include centering [Faber 1998b]. In fact, this definition is not even valid for non-centered models in situations where the response is predicted to be below zero (an example is given below). By properly deriving the calculation of the net analyte signal, the problem disappears. The prediction of *y* for a new sample, $\mathbf{x}_i$, is defined as

$$\hat{y} = \mathbf{x}_i^{T}\mathbf{b} = \mathbf{x}_{k,i}^{*\,T}\mathbf{b} \tag{16}$$

because the net analyte signal is the part of $\mathbf{x}_i$ in the direction of **b**. In case of centered data, $\mathbf{x}_i$, corresponds to the centered data and the prediction is of the centered response. This in itself poses no problems, but given that the net analyte signal multiplied by the sensitivity ($1/\|\mathbf{b}\|$) is supposed to equal the prediction of *y*, the definition of the net analyte signal, $x_{k,i}$, follows as

$$\hat{y} = \mathbf{x}_{k,i}^{*T}\mathbf{b} = x_{k,i}^{*}/\|\mathbf{b}\| \Rightarrow$$
$$x_{k,i}^{*} = \mathbf{x}_{k,i}^{*T}\mathbf{b}/\|\mathbf{b}\| \quad (17)$$

With this definition of the net analyte signal, the magnitude remains as when calculated by the norm but the sign is automatically set to be consistent with the obtained prediction of *y*. In practice, the correction above is equivalent to the one suggested by Faber [Faber 1998b] because the magnitude of the net analyte signal in Eq. (17) remains unaltered, whereas the sign changes with the sign of the prediction as also implied in Fabers approach. It is emphasized, though, that it is necessary to use this approach even in situations where there is no centering involved.

Results

In order to verify the adequateness of this definition of the net analyte signal vector, a small example is given to illustrate the differences between some of the methods. The data set consists of five simple laboratory-made samples. Each sample contains different amounts of tyrosine, tryptophan and phenylalanine dissolved in phosphate buffered water.

Table 1 Concentrations of amino acids in calibration samples. The concentrations are given in μM.

Sample	Tryptophan	Tyrosine	Phenylalanine
1	2.67	0.00	0
2	0.00	13.30	0
3	0.00	0.00	900
4	1.58	5.44	355
5	0.88	4.40	297

The samples were measured by fluorescence (excitation 269 nm chosen here, emission 250-450 nm, 1 nm intervals) on a PE LS50B spectrofluorometer with excitation slit-width of 2.5 nm, an emission slit-width of 10 nm and a scan-speed of 1500 nm/s. The data set is hence 5 × 201 (Figure 1).

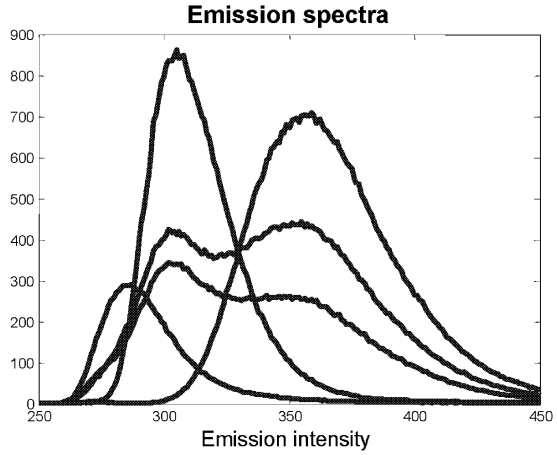


Figure 1 The emission spectra of the five samples.

A calibration model for tryptophan is made using PLS. Three components are expected to be optimal but in order to explore the net analyte signal vectors calculated by different methods, a one- and a two-component model is also evaluated. No centering is performed as zero signal from an analyte is equivalent to zero concentration.

The results of using the new net analyte signal vector and the one proposed by Lorber (as implemented in the PLS_Toolbox ver. 2.1) are shown in Figure 2. For the three-component model, there is virtually no difference between the two methods. For example, for sample two and three, both methods suggest that the net analyte signal is zero which is consistent with the zero concentration of tryptophan (Table 1).

For the two-component model, however, the net analyte signals differ. For example, for sample two and three, the new method suggests that the net analyte signal is still zero while Lorbers method does not. Looking at the predictions of the two-component model (Table 2), it is seen that indeed the predictions for sample two and three are close to zero. This is correctly reflected in the net analyte signal of the new method but not for Lorbers method.

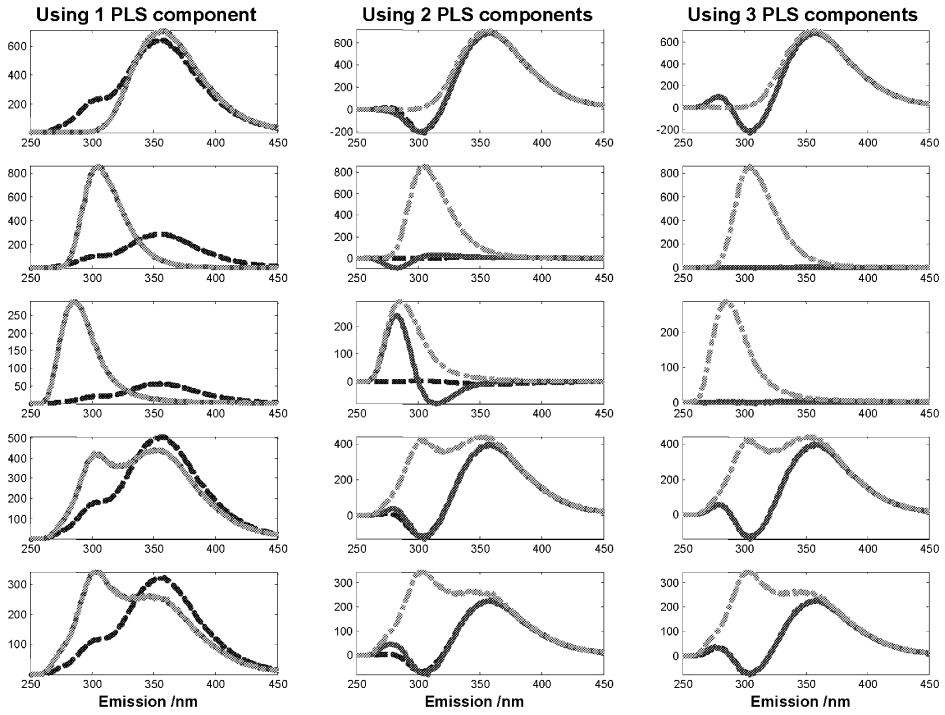


Figure 2 Results from a one-component (left), two-component (middle) and a three-component (right) PLS model for tryptophan is given. Each row represents one sample and the dash-dotted line is the spectrum. The dashed line is the new net analyte signal vector and the solid line is the one calculated by Lorbers method. In the left columns, the solid (Lorber) and dash-dotted (spectrum) lines are overlapping completely, whereas the solid (Lorber) and dashed (new) are overlapping completely in the right-most column.

For the one-component model, another interesting feature is clearly displayed. The net analyte signal from Lorbers method is completely overlapping the original spectrum thus changing dramatically shape from sample to sample. This is counterintuitive as the net analyte signal vector must be *one* specific direction independent of the actual sample. The same problem is not present with the suggested calculation of the net analyte signal vector.

Table 2 Results from a two-component PLS model. Note the negative net analyte signal of sample three.

Sample	True concentration	Predicted	NAS_{new}	NAS_{Lorber}
1	2,67	2,70	4,71	4,71
2	0,00	0,03	0,06	0,41
3	0,00	-0,04	-0,06	1,02
4	1,58	1,54	2,69	2,69
5	0,88	0,87	1,52	1,53

Conclusion

A new consistent and general method for calculating net analyte signal vectors and net analyte signals in inverse calibration has been defined from a theoretical point of view. The theory and method is consistent with earlier literature for ideal cases, but has several advantages in applications on real data.

Reference list

Booksh K S, Kowalski B R, Theory of analytical chemistry. *Anal. Chem.* 1994, **66**, 782A-791A.

Faber N M, Efficient computation of net analyte signal vector in inverse multivariate calibration models. *Anal. Chem.* 1998a, **70**, 5108-5110.

Faber N M, Mean centering and computation of scalar net analyte signal in multivariate calibration. *J. Chemometr.* 1998b, **12**, 405-409.

Faber N M, Exact presentation of multivariate calibration model as univariate calibration graph. *Chemometr. Intell. Lab. Syst.* 2000, **50**, 107-114.

Goicoechea, H C, Olivieri, A C, Enhanced synchronous spectrofluorometric determination of tetracycline in blood serum by chemometric analysis. Comparison of partial least-squares and hybrid linear analysis calibrations. *Anal. Chem.* 1999, **71**, 4361-4368.

Goicoechea, H C, Olivieri, A C. A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemometr. Intell. Lab. Syst.* 2001, **56**, 73-81.

Lorber A, Error propagation and figures of merit for quantification by solving matrix equations. *Anal. Chem.* 1986, **58**, 1167-1172.

Lorber A, Faber N M, Kowalski, B R. Net analyte signal calculation in multivariate calibration. *Anal. Chem.* 1997, **69**, 1620-1626.

Sanchez E, Kowalski B R, Tensorial calibration, I. First-order calibration. *J. Chemometr.* 1988, **2**, 247-263.

Xu L A, Schechter I, A calibration method free of optimism factor number selection for automated multivariate analysis. Experimental and theoretical study. *Anal. Chem.* 1977, **69**, 3722-3730.

Appendix 3

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

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°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°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.

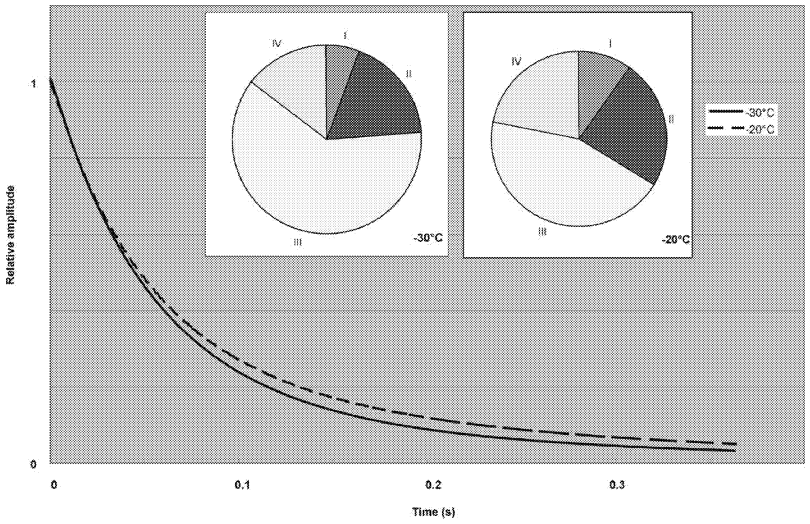


Fig. 1. ^{1}H NMR relaxation signals and water pools in samples from thawed cod. Intact cod muscle was frozen for three months at −30°C and subsequently for three months at the temperature indicated (−30 or −20°C). After thawing, the muscle samples were minced and the spin-spin relaxation signals recorded. 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).

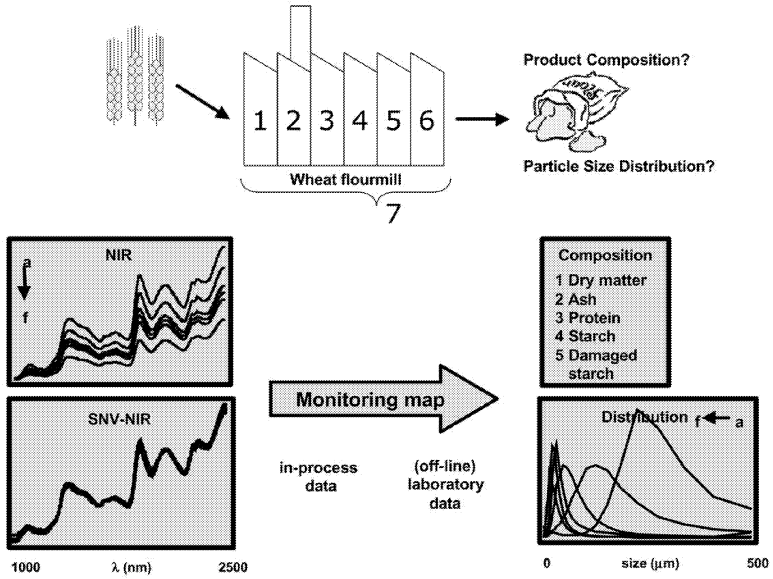


Fig. 2. Building blocks of the wheat flourmill data-set.

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 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

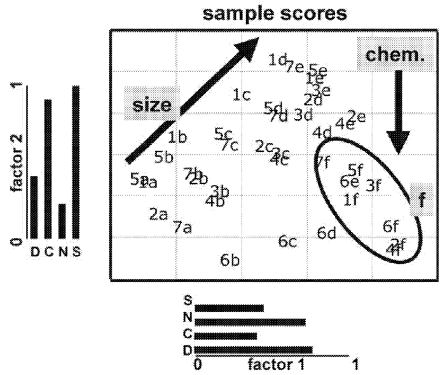


Fig. 3. Multi-block monitoring-map for the wheat flourmill data-set; 'S' is SNV-NIR, 'N' is NIR, 'C' is chemical composition and 'D' is particle size distribution.

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

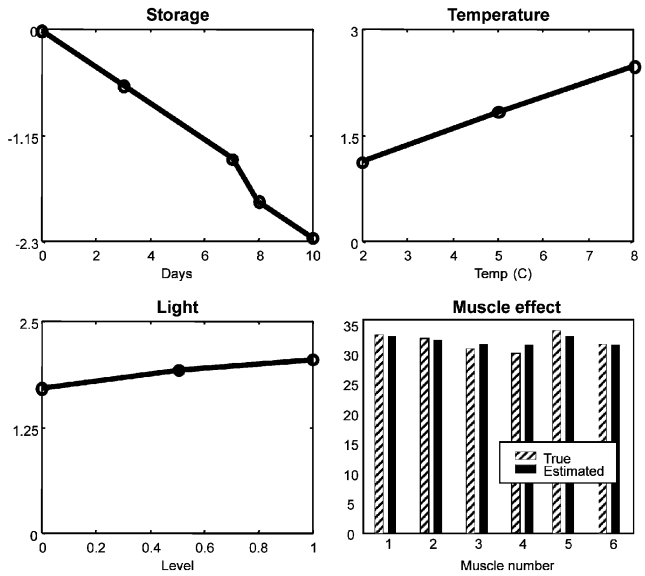


Fig. 4. Results of GEMANOVA model showing the influence of storage, temperature, light and muscle on meat color. Oxygen has no significant effect.

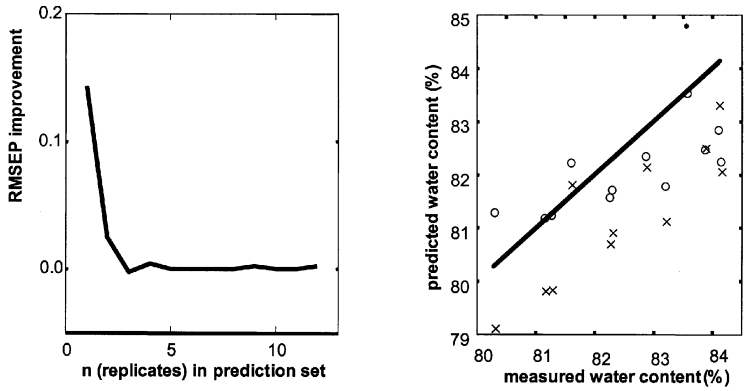


Fig. 5. (left) Difference in the average prediction error, Root Mean Squared Error of Prediction (RMSEP), between the corrected estimation and least squares estimation when six replicates are used for calibration and from one to 12 replicates are used in prediction; (right) an example of the predicted 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.

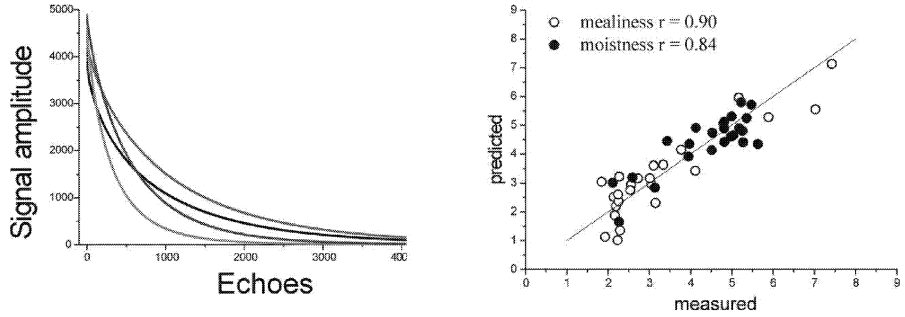


Fig. 6. (left) Low-field ^{1}H NMR relaxation curves of four examples of different potato varieties; (right) prediction curves of mealiness and moistness using NMR.

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

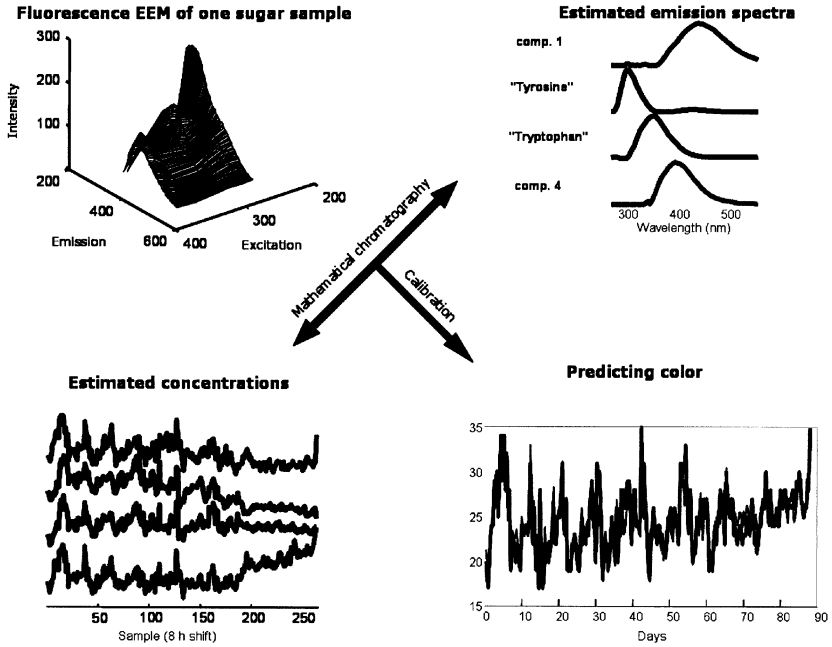


Fig. 7. Sugar samples were taken every eighth hour, dissolved in water and measured spectrofluorometrically. An example is shown upper left. PARAFAC decomposition of the data provides relative concentrations (lower left), relative excitation spectra (not shown) and relative emission spectra (upper right).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