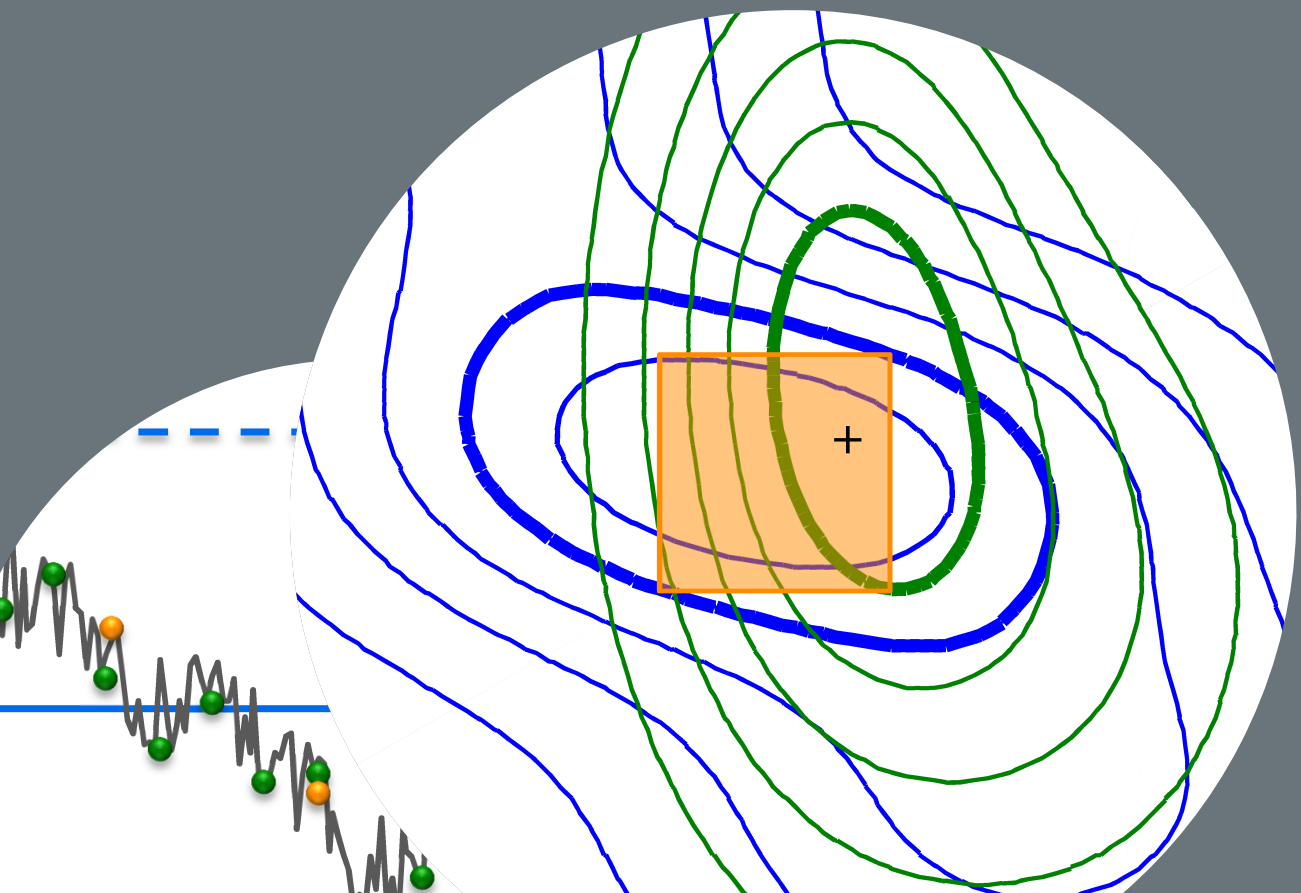
UNIVERSITY OF COPENHAGEN

PHD THESIS 2012 · JONAS HOEG THYGESEN

Dynamic Models and Chemometric Tools for Process Monitoring



Dynamic Models and Chemometric Tools for Process Monitoring

PhD Thesis by

Jonas Hoeg Thygesen

2012

Quality and Technology,
Department of Food Science,
Faculty of Science,
University of Copenhagen

Dynamic Models and Chemometric Tools for Process Monitoring
PhD Thesis 2012 © Jonas Hoeg Thygesen

ISBN 978-87-7611-512-8

Printed by SL grafik, Frederiksberg C, Denmark (www.slgrafik.dk)

Preface

This thesis was written with the aim of fulfilling the requirements for obtaining a PhD-degree from the University of Copenhagen. The presented work is based on work made at The Quality & Technology (Q&T) group, Department of Food Science, University of Copenhagen under the supervision of Associate Professor Frans W.J. van den Berg. I would like to thank Frans for inspiration, good discussions and kind supervision. I am still looking for the field within spectroscopy, chemometrics, statistics and MATLAB programming that Frans does not have experience within. The research was sponsored by the Quality by Design consortium (www.qbd.dk), which was partially funded by the Danish Academy for Science, Technology and Innovation, the consortium is kindly acknowledged for their sponsorship.

As a part of the research for the thesis I had a four month stay at Department of Industrial and Systems Engineering at Rutgers University, New Jersey, USA in the spring of 2010. I would like to thank Professor Susan L. Albin and the rest of the department for hosting my stay, a special thank goes to Associate Professor Melike Baykal-Gürsoy for a thorough introduction to time series modelling, interesting discussions on system identification and her kind and welcoming nature.

I am also thankful to my colleges at Q&T for a perfect working environment, strange ideas during the lunch discussions and perfect Christmas parties - you bring Quality (& Technology) to LIFE :-) My office partner though the last three years Hamid Babamoradi is especially thanked both for fruitful chemometric discussions and interesting observations on the logics of the Danish society.

Finally I would like to thank my family and friends. Particularly my beautiful wife Charlotte Hoeg Thygesen, who is currently carrying our first child, is thanked for her never-ending love and support during the three years. She has the wonderful ability of bringing my thoughts away from eigenvalues, covariance matrices and other high browed subjects, to other more down to earth places, such as how do we cast 130 m^2 of concrete floor in an afternoon?

Jonas Hoeg Thygesen

Frederiksberg, March 2012

Abstract

The food- and pharma-industry is under an ever increasing demand for reduction in energy use, optimal production planning and efficient utilization of raw materials. This has led to the concepts of Quality by Design (QbD) and Process Analytical Technology (PAT). The aim of QbD is to use PAT-tools for obtaining greater process knowledge such that the manufacturer may move away from end-point testing of products, towards building quality into process and thus the products (hence the name Quality by Design). The purpose of this PhD project was to show how spectroscopy based PAT-tools in combination with dynamic predictive models may bring these goals closer to reality. The work presented in this thesis covers the three years research which was also published in four papers:

Paper I investigated how three-way calibrations for Excitation-Emission Matrix (EEM)-fluorescence spectroscopy could be transferred. The study showed that it was possible to develop simple, intuitive transfer methods for three-way EEM fluorescence calibrations. It was additionally shown that though good transfer models could be found for the calibration models with as few as four transfer samples, the results were highly dependent on the selection of the transfer set. The paper thereby illustrated how three-way EEM fluorescence calibration made in an off-line setting (i.e. in the laboratory) with ease could be transferred to an on-line application.

Paper II introduced the state space model and showed how so-called subspace methods allowed state space modelling without *a-priori* assumptions on model shape/form, thereby enabling modelling of the process without the requirement of any prior knowledge on the underlying physics or chemistry. The paper presented how a non-linear milk coagulation process could be approximated by linear state space models. Where conventional control charts reflects the process in a static manner, the control charts proposed in **Paper II** reflected the dynamic behaviour of the process.

Paper III elaborated further on the conclusions from **Paper II.** In this paper a combination of state space models, subspace methods and Kalman filters were shown to have the potential as a versatile tool in batch process modelling and monitoring. A model system of riboflavin breakdown was presented as an example of a batch process. It was shown how the combination of EEM-fluorescence

spectroscopy and PARAFAC modelling allowed direct surveillance of the on-going chemistry in the process. The proposed combination of methods was able to capture and model the dynamics of the batch process. The introduction of the Kalman filter gave the advantage of improved predictions of future process variable trajectories including 95% confidence intervals of the variables. The method was thus shown to be adaptable to new non-NOC conditions and allowed for dynamic control charting of initial condition estimates and current system-states. For end-point prediction a dedicated method based on Partial Least Squares was found to produce slightly better predictions.

Paper IV presented further studies on the model system introduced in **Paper III.** The paper illustrated what is also known as so-called grey box modelling: Modelling in the case where the physics and chemistry governing the process is known or assumed to be known to some extent. In **Paper IV** it was shown how the *a-priori* knowledge on the reaction kinetics governing the process could be implemented during PARAFAC modelling, hereby allowing post-batch charting of the relevant process parameters – the kinetic constants.

The different statistical/chemometric models included in this thesis made it possible to answer different types of questions. The only method able to answer all three questions: *"Where is the process now?"*, *"Where did the process come from?"* and *"Where is the process going?"* was the state space/Kalman method presented in **Paper II** and **III.** The possibility of predicting future process characteristics and variable trajectories opens for the option of model predictive control which in turn may bring the goal of QbD closer to reality.

Resumé

Fødevare- og medicinalindustrien er under et stadigt stigende pres for reduktion i energiforbruget, optimal produktionsplanlægning og effektiv udnyttelse af råmaterialer. Dette har ført til udviklingen af koncepterne *Quality by Design* (QbD) og *Proces Analytisk Teknologi* (PAT). Formålet med QbD er at bruge PAT-værktøjer til at opnå større procesviden, således at producenten kan bevæge sig væk fra slutpunkt test af produkter, til at bygge kvalitet ind i processen og dermed produkterne (deraf navnet *Quality by Design*). Formålet med dette ph.d.-projekt var at vise, hvordan spektroskopi-baserede PAT-værktøjer i kombination med dynamiske prædiktive modeller kan bringe disse mål tættere på virkeligheden. Arbejdet der præsenteres i denne afhandling dækker de tre års forskning som også er offentliggjort i fire artikler:

Artikel I undersøgte hvordan tre-vejs kalibreringer til Excitation-Emission Matrix (EEM)-fluorescens spektroskopi kunne overføres. Undersøgelsen viste, at det var muligt at udvikle enkle og intuitive overførselsmetoder for tre-vejs EEM fluorescens kalibreringer. Det blev endvidere vist, at skønt gode overførselsmodeller kunne findes for kalibreringsmodellerne baseret på så få som fire prøver, var resultaterne meget afhængig af valget af prøvesæt. Artiklen illustrerede dermed, hvordan tre-vejs EEM fluorescens kalibreringer fundet i en off-line situation (dvs. i laboratoriet) med lethed kan overføres til en on-line applikation.

Artikel II introducerede state space modeller og viste, hvordan såkaldte subspace metoder tilladte state space modellering uden *a-priori* antagelser om model type / form, hvorved modellering af processen var mulig, uden krav om forudgående viden om den underliggende fysik eller kemi. Artiklen præsenterede hvordan en ikke-lineær mælkekoagulation proces kan beskrives ved lineære state space modeller. Hvor traditionelle kontrol-kort afspejler processen på en statisk måde, foreslog **Artikel II** kontrol-kort der afspejlede den dynamiske opførsel af processen.

Artikel III uddybede konklusionerne fra **Artikel II**. I denne artikel blev det vist hvordan en kombination af state space modeller, subspace metoder og Kalman filtre har potentiale som et alsidigt redskab i batch-proces modellering og overvågning. Et modelsystem af riboflavin nedbrydning blev præsenteret som et eksempel på en batch-proces. Det blev vist, hvorledes kombinationen af EEM-fluorescens-spektroskopi og PARAFAC modellering tillod direkte overvågning af den

igangværende kemi i processen. Den foreslåede kombination af metoder var i stand til at indfange og modellere dynamikken i batch-processen. Introduktionen af Kalman filteret gav den fordel, at forbedrede prædiktioner kunne opnås for udviklingen i fremtidige procesvariable, dette inkluderede 95% konfidensintervaller for variablerne. Metoden blev således vist at kunne tilpasses nye non-NOC vilkår og tillod dynamiske kontrol-kort for initialbetingelser og nuværende system-tilstande. Til slutværdi prædiktion var en dedikeret metode baseret på Partial Least Squares (PLS) i stand til at frembringe lidt bedre prædiktioner.

Artikel IV præsenterede yderligere undersøgelser af det modelsystem, der blev præsenteret i **Artikel III**. Artiklen illustrerer, hvad der også er kendt som såkaldt *grey-box* modellering: Modellering i det tilfælde, hvor den underlæggende fysik og kemi der styrer processen er delvis kendt. I **Artikel IV** blev det vist, hvordan *a-priori* viden om reaktionskinetik for processen kan implementeres i PARAFAC modellering, hvorved *post-batch* kortlægning kan opnås af de relevante procesparametre - de kinetiske konstanter.

De forskellige statistiske / kemometriske modeller, der indgår i denne afhandling har gjort det muligt at besvare forskellige typer af spørgsmål. Den eneste metode der dog var i stand til at besvare alle tre spørgsmål: *"Hvor er processen nu?"*, *"Hvor kommer processen fra?"* og *"Hvor er processen på vej hen?"* var state space/Kalman metoden præsenteret i **Artikel II** og **III**. Muligheden for at forudsige fremtidige proces karakteristika og variable forløb åbner for muligheden for model prædiktiv regulering, som igen kan bringe målet om QbD tættere på virkeligheden.

List of Publications

Paper I

Thygesen, J. & F. van den Berg (2011): Calibration transfer for excitation-emission fluorescence measurements, *Analytica Chimica Acta*, Vol. 705, no. 1-2, pp. 81 – 87

Paper II

Thygesen, J.H. & F. van den Berg (2012): Subspace methods for dynamic model estimation in PAT applications, *Journal of Chemometrics*, accepted for publication

Paper III

Thygesen, J.H. & F. van den Berg (2012): Dynamic Model Based Monitoring of Batch Processes, *Chemometrics & Intelligent Laboratory Systems*, submitted

Paper IV

Thygesen, J.H., R. Bro & F. van den Berg (2012): Estimation of process characteristics using constrained PARAFAC models, *Chemometrics & Intelligent Laboratory Systems*, In preparation

Additional Publications by the author

Hedegaard, R.V., K. Granby, H. Frandsen, **J. Thygesen** & L.H. Skibsted (2008): Acrylamide in bread. Effect of prooxidants and antioxidants, *European Food Research and Technology*, vol. 227, No. 2, pp. 519-525

Rinnan, Å., L. Nørgaard, F. van den Berg, **J. Thygesen,** R. Bro & S.B. Engelsen (2009): Data Pre-processing. In: D.-W. Sun (ed.): *Infrared Spectroscopy for Food Quality Analysis and Control*. 1. ed. Academic Press, Burlington, MA, USA, pp. 29-50.

Thygesen, J.H. & F. van den Berg (2012): Procesovervågning med Dynamiske Modeller, *Plus Proces*, submitted – popular scientific publication in danish.

List of Abbreviations

ALS	Alternating Least Squares
ARX	Autoregressive model with exogenous input
CI	Confidence Interval
CPAC	Center for Process Analytical Chemistry
CR	Continuum Regression
CVA	Canonical Variates Analysis
CUSUM	Cumulative Sum
DoE	Design of Experiments
DPCA	Dynamic Principal Components Analysis
DS	Direct Standardisation
EEM	Excitation-Emission Matrix
EWMA	Exponentially Weighted Moving Average
FAD	Flavin Adenine Dinucleotide
FIR	Finite Impulse Response
ICA	Independent Component Analysis
ICH	International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
ICH-Q8	ICH eighth quality guideline on Pharmaceutical Development
IR	InfraRed
IUPAC	International Union on Pure and Applied Chemistry
MLR	Multiple Linear Regression
MPC	Model Predictive Control
MSPC	Multivariate Statistical Process Control
N4SID	Numerical algorithm for Subspace State Space System Identification
NADH	Nicotinamide Adenine Dinucleotide
NIPALS	Non-Linear Partial Least Squares
NIR	Near InfraRed
NOC	Normal Operating Conditions
N-PLS	*N*-way Partial Least Squares
PAC	Process Analytical Chemistry
PARAFAC	Parallel Factor Analysis
PAT	Process Analytical Technology
PC	Principal Component
PCA	Principal Components Analysis

PCR	Principal Components Regression
PDS	Piecewise Direct Standardisation
PEM	Predictor Error Method
PI	Proportional-Integral
PID	Proportional-Integral-Derivative
PLS	Partial Least Squares
PRBS	Pseudo Random Binary Sequence
QbD	Quality by Design
SI	System Identification
SNV	Standard Normal Variate
SVD	Singular Value Decomposition
SPC	Statistical Process Control
TLS	Total Least Squares
UF	Ultrafiltration
US-FDA	United States Food and Drug Administration
UV-VIS	UltraViolet-Visual

Table of contents

1 Introduction

The food- and pharma-industry is under an ever increasing demand for reduction in energy use, optimal production planning and efficient utilization of raw materials. The purpose of this PhD project was to show how modern process sensors in combination with statistical predictive models may bring these goals closer to reality.

A common task in industry is continues surveillance of process performance and product quality. Very often this monitoring is done by following classical engineering variables such as pH, temperature or pressure over time. These signals are however seldom the information we in reality are interested in. E.g. in a fermentation plant producing enzymes we are not exactly interested in the pH at which the product was made although this will be a highly relevant process parameter. Instead we may be interested in the enzyme activity of the finished product. Modern process analysers (such as near-infrared/NIR spectrometers) may be used directly in process streams or vessels (so-called "on-line" measurements). When these instruments are combined with multivariate statistics/chemometrics we are able to directly or "real-time" determine the chemistry going on during production rather than rely on inferential information such as pH or temperature.

The chemical information obtained from modern sensors may lead to a higher process understanding by continues observation (e.g. "if pH in the reactor is between 8 and 8.5 then keep temperature between 30 and 32^{o}C to reach the highest enzyme activity"). This is however not the only advantage. A very high sampling frequency is possible with on-line spectroscopy. Classical grab sampling and "off-line" reference measurements in a central laboratory may be a time and money consuming procedure and will thus always be limited to a low frequency (e.g. once every hour or even less). On-line spectrometers on the other hand have a high sampling frequency (e.g. every minute) at a very low cost per measurement. It should however be stressed that on-line implementation of e.g. a NIR-spectrometer is not maintenance-free; the calibrations used for the instruments may every now and then need updating. **Paper I** investigated how such an update or standardization of Excitation-Emission fluorescence spectra could be done for a process fluorescence instrument.

An example of "off-line" ( ) vs. "on-line" ( ) measurements is given in Figure 1-1. The true, unknown value of the process variable is indicated by (|).The figure illustrates how the more frequent on-line measurements allow a closer monitoring of the process.

Figure 1-1 Off-line vs. on-line measurements of a process and possible questions asked during real-time monitoring are indicated

Figure 1-1 indicates the three key questions that may be asked from real-time monitoring during production.

- *"Where is the process now?"* – Are we able to detect that the current product/process stream is within specifications, i.e. quality assurance.
- *"Where did the process come from?"* – When looking back on the past process measurements, how did the process evolve? This information can be used for process optimization, but is always "post-problem"; we can for instance only change the recipe in future production to achieve the right quality.

- *"Where is the process going?"* – If we have a process model that is able to predict future process characteristics we are able to change – by so-called model predictive control (MPC) - the production settings to counteract undesired products not up to quality specification.

Different Multivariate Statistical Process Control (MSPC) methods may be used for answering some of these questions, and the methods have therefore long been, and still are, of research interest within the chemometric society [Skagerberg *et al.*, 1992; Kourti & MacGregor, 1995; Qin, 2003; Laursen *et al.*, 2011]. The main set of methods within MSPC is however primarily suited for feed-back control (post-problem), rather than feed-forward control (pre-problem; see section 4 on control theory). There is need for a new set of methods that enable feed forward control/MPC. Despite the fact that modern sensors allow for direct monitoring of key attributes such as enzyme activity, it is the lack of sufficient understanding of the physics and chemistry of the biotechnological production processes that limits the use of MPC. This insufficient understanding is especially valid for batch-wise production and biological processes such as many food systems. So-called "black box models" are required; models that enable us to identify the process dynamics and predict the process output (e.g. enzyme activity) as a function of the controllable process inputs (i.e. the pH, temperature, etc.). In **Paper II** one class of such algorithms was investigated, state space methods for time series analysis identified using so-called subspace methods It was shown how these methods could capture and model the dynamics of coagulating milk by combining NIR spectroscopy with state space models. It illustrated that modern sensors like NIR give insight into how a normal process behaves, how state space models can be used to follow whether the process is on track ("*Where is the process now?*"), and to predict the development of the coagulation ahead in time ("*Where is process going?*").

A common challenge faced when applying process models is the difference between measurements and dynamic model predictions: Our measurement reports enzyme strength X, but the process model predicts strength Y at this stage of the batch, which do we trust more knowing that both are affected by noise and uncertainty? These were the questions that were investigated in **Paper III** where breakdown of vitamin B2 (riboflavin) was investigated and modelled by state space models. This breakdown process was studied as a generic model system representative for the batch processes in industry. The state space models identified from training or Normal Operating Conditions (NOC) data have the advantage of allowing easy

implementation of a statistical model known as the Kalman filter. The algorithm is an answer to the challenge of measurement vs. model. It enables us to find the best compromise between the two sources of estimation, and also allows us to predict the future process outputs (the concentration of breakdown product one hour from now e.g.) within a Confidence Interval (CI; Figure 1-2). An end-target is often used in batch processes, e.g. reaching a desired percentage conversion of the raw materials in a given amount of time. Once the process is at this target the batch is opened and the product is transferred to a downstream step. Thus, if we see that the predicted interval does not include the target (time 1), we can take a corrective action based on MPC in order to move the future outputs back on track (time 2), thus avoiding products not up to specifications.

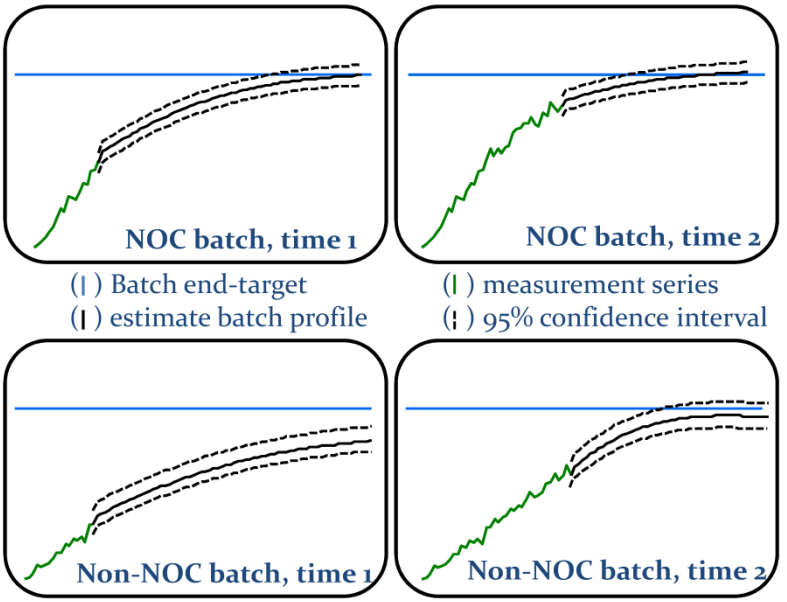


Figure 1-2 The combination of state space models, subspace methods and the Kalman filter allows for Model Predictive Control (MPC)

In some cases supplementary knowledge is available on the process. It could for instance in the case of riboflavin breakdown be speculated that the process follows first order kinetics. Then so-called "grey-box" modelling where some external knowledge is utilized during modelling can be applied, a subject further pursued in **Paper IV**.

1.1 PAT, QbD and other buzzwords

The ideas off applying process analysers on-line in industry is not new and was within the chemometric community in the beginning of the 80'ies approached in a more systematic way with the introduction of so-called Process Analytical Chemistry (PAC). The Center for Process Analytical Chemistry (CPAC) at the University of Washington was a major driver in this [Callis *et al.*, 1987; McLennan, 1995]. Since 1993 a biannual review of PAC with hundreds of references in each has appeared in the peer-reviewed journal Analytical Chemistry [Beebe *et al.*, 1993; Blaser *et al.*, 1995; Workman *et al.*, 1999; Workman *et al.*, 2001; Workman *et al.*, 2003; Workman *et al.*, 2005; Workman *et al.*, 2007; Workman *et al.*, 2009; Workman *et al.*, 2011]. In spite of the huge number of references on process analysers, process chemometrics and on-line chemical analysis, the term PAC was however not widely spread outside chemometrics. This is apparent if a literature search is made on the term "Process Analytical Chemistry" (Figure 1-3) where less than 10 papers appear each year[1].

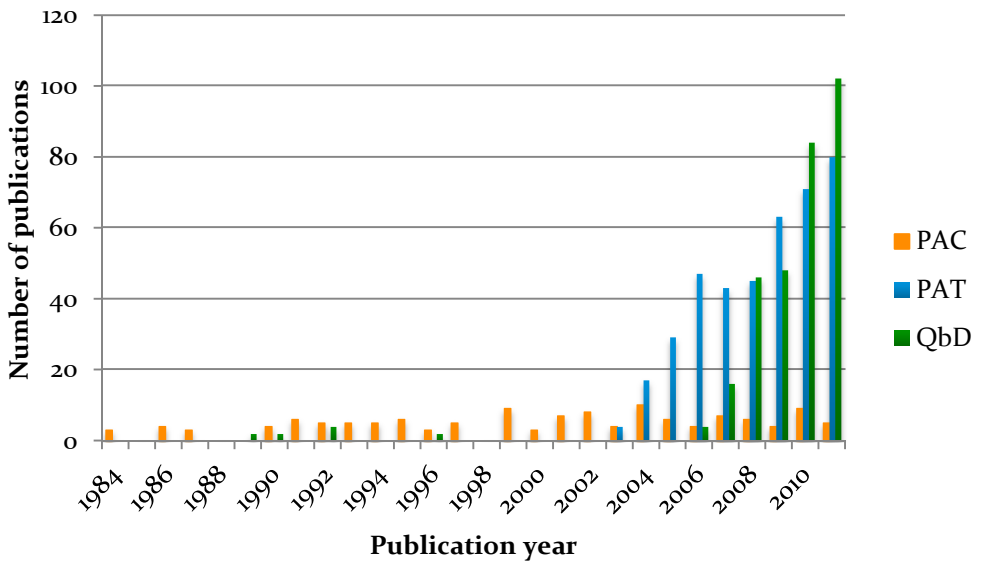


Figure 1-3 Number of publications for the search terms "Process Analytical Chemistry" (PAC), "Process Analytical Technology" (PAT) and "Quality by Design" (QbD)

The advantages of on-line chemical measurements and process chemometrics was however recognized by the United States Food and Drug Administration (US-FDA)

[1] Literature search done February 2012 in Thompson Reuters Web of Science.

as they in September 2004 issued a non-binding recommendation entitled *"Guidance for Industry: PAT – A Framework for Innovative, Pharmaceutical Development, Manufacturing and Quality Assurance"*. The overall principle in the guidance was to use different Process Analytical Technology (PAT) tools to achieve a better process understanding. The FDA saw on-line measurements and multivariate data analysis (i.e. chemometrics) as a part of the PAT solution; this meant that many investigations and accompanying publications were made on on-line measurements and process chemometrics, the papers were however published under the new label PAT rather than PAC (Figure 1-3).

The end-goal of this improved process understanding was to move away from end-point testing of products towards building quality into the products, i.e. *"Quality cannot be tested into products; it should be built-in or should **be** by design"* [U.S.Food and Drug Administration, 2004]. This quality approach was elaborated further in November 2005 as The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) finalized their eighth quality guideline (Q8) on Pharmaceutical Development. The ICH-Q8 termed the quality approach *"Quality by Design"* (QbD). The QbD approach is now widely recognized with more than 100 publications during 2011 (Figure 1-3). The connection between the three terms PAC, PAT and QbD can be illustrated as outlined in Figure 1-4.

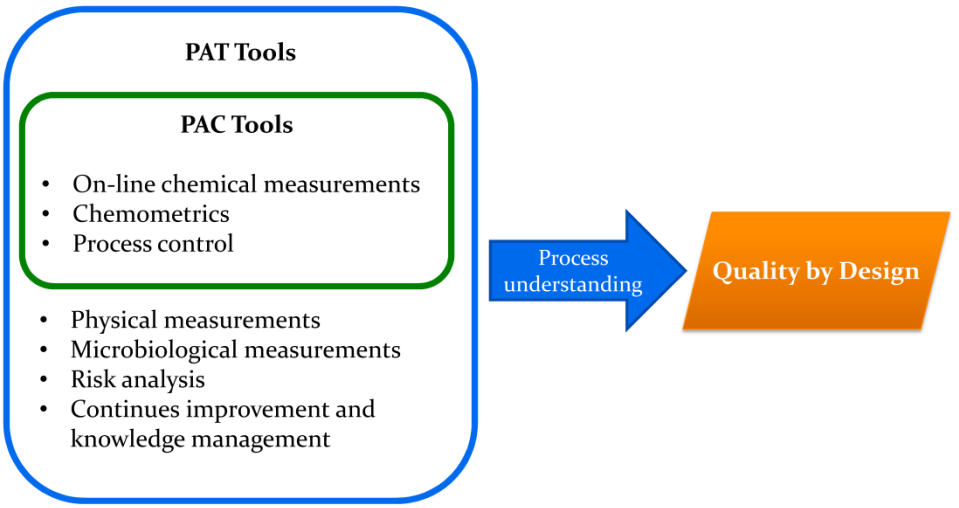


Figure 1-4 Connection between Process Analytical Chemistry (PAC), Process Analytical Technology (PAT) and Quality by Design (QbD)

Two central concepts within QbD are *real time release* and *design space*. Real time release is closely related to the end-goal of building quality into products. It is defined as: *"The ability to evaluate and ensure the acceptable quality of in-process and/or final product based on process data, which typically includes a valid combination of measured material attributes and process controls" [U.S.Food and Drug Administration, 2004; ICH, 2009]*. This means that if a pharmaceutical manufacturer can prove that a given drug has acceptable quality solely based on process data, real time release may be achieved so that no end-point testing is needed. The design space is then defined as: *"The multidimensional combination and interaction of input variables and process parameters that have been demonstrated to provide assurance of quality"* [ICH, 2009]. This means that the concept is tightly connected to response surface modelling known from the field of Design of Experiments (DoE; [Box *et al.*, 1978]). The idea is to use surface response models to find the controllable process variables that allows production of products with the desired quality; the principle is outlined in Figure 1-5 below.

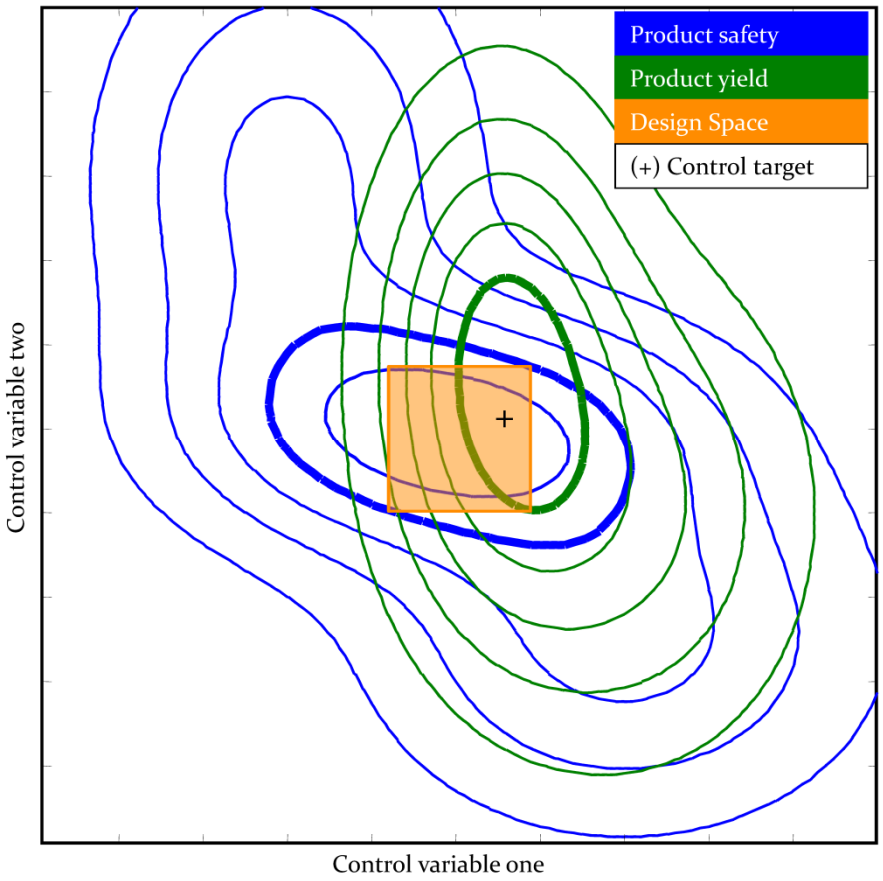


Figure 1-5 Illustration of the design space concept

Based on designed experiments the response surface for product safety is estimated, the acceptable product safety is in Figure 1-5 indicated by the bold blue line. A similar response surface may be made for the product yield, the process target is then chosen as the optimal compromise between the two and the design space is selected so that all combinations of the control variables yields product of an acceptable quality. Moving the process target inside the design space – e.g. if the product yield maximum changes its position over time - is then under the QbD principles not considered a change in the approved process resulting in no need for regulatory actions [ICH, 2009]. A few notes should be made on the appearance of Figure 1-5:

- The two control variables indicated in the figure may be either real measurable entities (e.g. pH or pressure) or linear combinations of several measured variables (e.g. scores in a principal component model – for more details on PCA see section 3.1.1)

- The design space may have linear bounds (as in Figure 1-5) or non-linear bounds. The later could in Figure 1-5 be the case if the complete area inside the bold blue ellipse was chosen as design-space.

- The design space may cover one or several unit operations. The one-unit-operation design space is easier found, but the several-unit-operations design space may allow a higher flexibility. MPC would in such cases be very useful as it could e.g. be the case that an up-stream step yielded product on the edge of the design space. Successful application of MPC in the following down-stream process steps could then ensure that the product was transferred closer to the centre of the design space.

1.2 Definitions and conventions used in this thesis

1.2.1 Inputs, Outputs and Systems

A more general presentation of a production system is given in Figure 1-6.

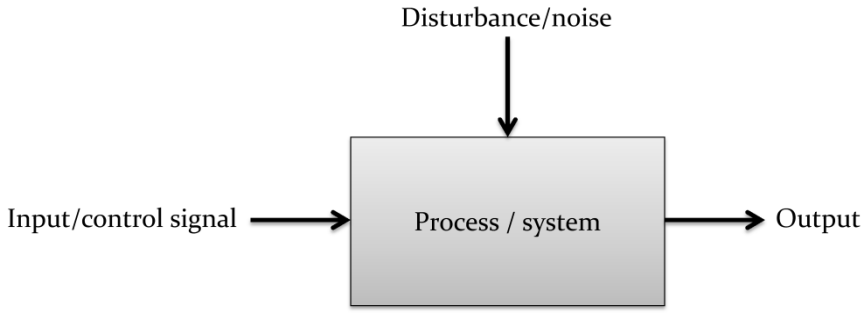


Figure 1-6 Generic representation of process

The terms process or system will be used interchangeable in this thesis, the term covers the process that is being modelled, and the system is therefore often equal to a unit operation. The outputs of the system are the measurable dependent variables that characterize and describe the response of the system [Roffel & Betlem, 2006].

The input or the control signal is also one or more measureable or observable variables; these variables are however under our rule and are used to manipulate the system output. The disturbances or the process noise are external factors influencing the system. They cannot be manipulated but may in some cases be measureable. In such cases the control signals may thus be used to compensate for the disturbances, more on this in section 4.

The disturbances could in the enzyme example from before for instance be equal to fluctuations in raw material composition, the enzyme activity could be seen as the system output, and pH and temperature in the reactor vessel as controllable inputs.

1.2.2 Mathematical convention

The following notation is used:
Scalars are denoted by upper and lower case letters in italics (e.g. *I, J, k* and *t*).
Vectors are denoted by lower case bold letters (e.g. **t, p** and **y**).
Matrices are denoted by upper case bold letters (e.g. **X, A** and **E**).
Tensors are denoted by underlined upper case bold letters (e.g. $\underline{\mathbf{X}}$ and $\underline{\mathbf{E}}$).

2 Process Monitoring

The introduction section indicated some of the advantages of using spectroscopy in process monitoring. This section will elaborate on the principles, advantages and challenges. The section will include a general introduction to spectroscopy and two dedicated sections for the methods included in this thesis: Near Infrared (NIR) and Excitation-Emission Matrix (EEM) fluorescence spectroscopy.

The International Union on Pure and Applied Chemistry (IUPAC) defines spectroscopy as *"the study of systems by the electromagnetic radiation with which they interact or that they produce",* and spectrometry as *"the measurement of such radiation as a means of obtaining information about the systems"* [Sheppard *et al.*, 1985]. The term spectroscopy will however in this thesis be used to cover both the measurement principle as well as the study of interaction between light and matter.

When matter is irradiated, several things can happen [Dahm & Dahm, 2001]:

1) The radiation is reflected or scattered. The incident light may either be reflected as specular reflection (mirror like) where the angle of the incident light is equal to the angle of the reflected light, or as diffused reflection where the light is reflected in many different angles. The scattering of the light is dependent on the particle size with smaller particles giving a higher degree of scatter.

2) The radiation is absorbed. If the electromagnetic radiation corresponds to certain frequencies resonance may occur whereby the radiation is absorbed. Different selection rules apply for the different parts of the electromagnetic spectrum, the specific rules for NIR and fluorescence spectroscopy in the Ultraviolet-Visual (UV-VIS) range will be outlined in the corresponding sections.

3) The light is transmitted. Some light may also pass through the sample without any interaction with the sample.

The relationship between absorbed light, transmitted light and analyte concentration may be explained by Lambert-Beer's law or simply Beer's law (Equation 2-1) [Harris, 2007]:

$$Abs = \log\left(\frac{I_0}{I}\right) = cl\varepsilon \qquad \textbf{Equation 2-1}$$

Where Abs is the absorbance, I_0 the intensity of light entering the sample, I the intensity of the light transmitted, c the concentration of the analyte, l the path length or sample thickness, and ε the molar absorptivity coefficient [Harris, 2007]. Figure 2-1 symbolizes the electromagnetic spectrum.

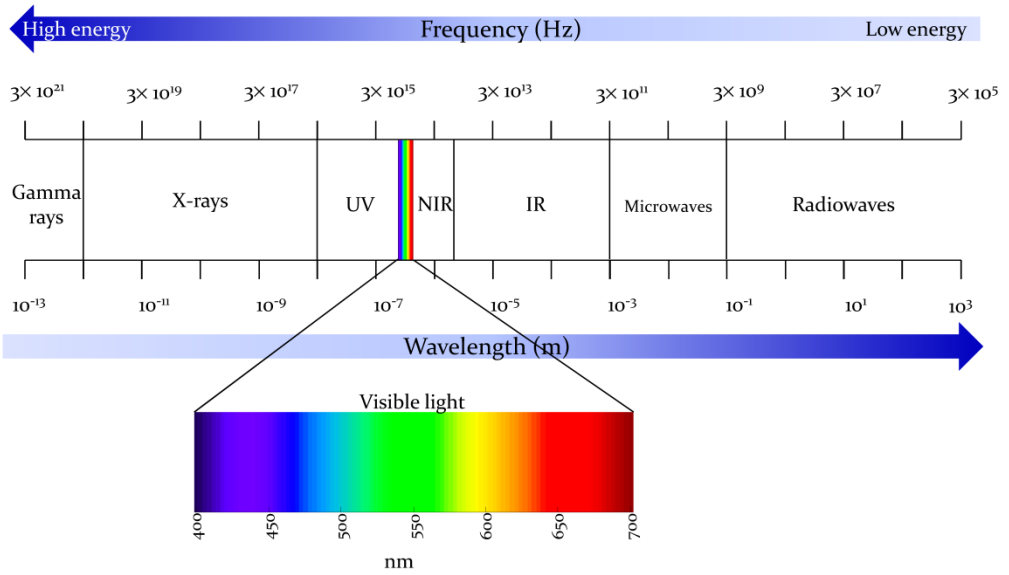


Figure 2-1 The electromagnetic spectrum (data from Harris [2007] and Dolezalek [2012])

The spectroscopic methods used in this thesis applied electromagnetic radiation in the UV-VIS region (**Paper I**, **III** and **IV**), and the NIR region (**Paper II**) of the electromagnetic spectrum.

2.1 Near Infrared Spectroscopy

Figure 2-1 shows that the Near Infrared (NIR) and Infrared (IR) regions are found at the wavelengths[2] between 700 nm and 10^{-3} m; likewise it is illustrated that shorter waves have higher energy. The energy level of NIR and IR is in the range 8 – 40 kJ/mol [Pavia *et al.*, 2000], which corresponds to the vibrational energy of covalent bonds in most molecules. This fact is utilized in IR and NIR spectroscopy - for all molecular vibrations where a dipole moment is displaced, infrared radiation will be absorbed when the frequency of the radiation corresponds to the frequency of the vibrating bond [Pavia *et al.*, 2000]. The two methods are hence - together with Raman spectroscopy - also known under vibrational spectroscopy.

The frequency corresponding to the vibrational energy of a covalent bond is dependent on the mass of the atoms and the strength of the bond, with one or more specific bending and stretching vibrations. The specific vibration-frequency of the bond may partially be explained as a harmonic oscillator: The stronger the bond the higher the vibration-frequency, and the bigger the difference in mass between the vibrating atoms the higher the vibration-frequency [Dufour, 2009]. The molecular vibrations are however not perfectly harmonic. This has the result that overtones are observed in the NIR region approximately at integer multiples of the fundamental vibration frequency of the IR region.

The position of the overtones ($\bar{\nu}_n$) can be found by applying Equation 2-2 if the fundamental wavenumber ($\bar{\nu}_0$) and the dimensionless anharmonicity constant (χ) are known.

$$\bar{\nu}_n = n\bar{\nu}_0(1 - n\chi) \qquad \textbf{Equation 2-2}$$

χ is normally in the range of 0.001 to 0.02, with bonds having larger anharmonicity constants involving hydrogen (e.g. -CH, -NH and -OH) [Miller, 2001; Griffiths, 2002]. Fundamental vibrations with very low χ will only display very weak overtones that might not be detectable [Siesler, 2008].

It is further known that the intensity of each overtone is approximately 10 times smaller than the previous [Miller, 2001], i.e. if the fundamental has the intensity of 1 the first overtone will have an intensity of 0.1, the second 0.01, etc. Since Beer's law

[2] By tradition wavelengths (measured in nm) are normally used in NIR, while wavenumbers (cm^{-1}) are used in IR.

(Equation 2-1) states that the intensity of the signal is directly proportional to the light path length, the lower intensity has the consequence that longer path lengths are allowed/required for NIR than for IR [Siesler, 2008]. This makes NIR spectroscopy easier to implement than IR spectroscopy as a process analyser.

Overtones are not the only peaks found in the NIR range spectrum. Combination bands of fundamentals and overtones or several overtones can often be observed in (and complicate the interpretation of) NIR spectra. These bands will be located approximately at the summation of the wavenumbers for the bands that are combined [Miller, 2001]. The fundamental and overtone signals may originate from different vibrations. It is however required that the signals are originating from the same functional group [Miller, 2001].

Figure 2-2 illustrates some of the normal vibrations that can be observed for CH_2 in IR-spectroscopy.

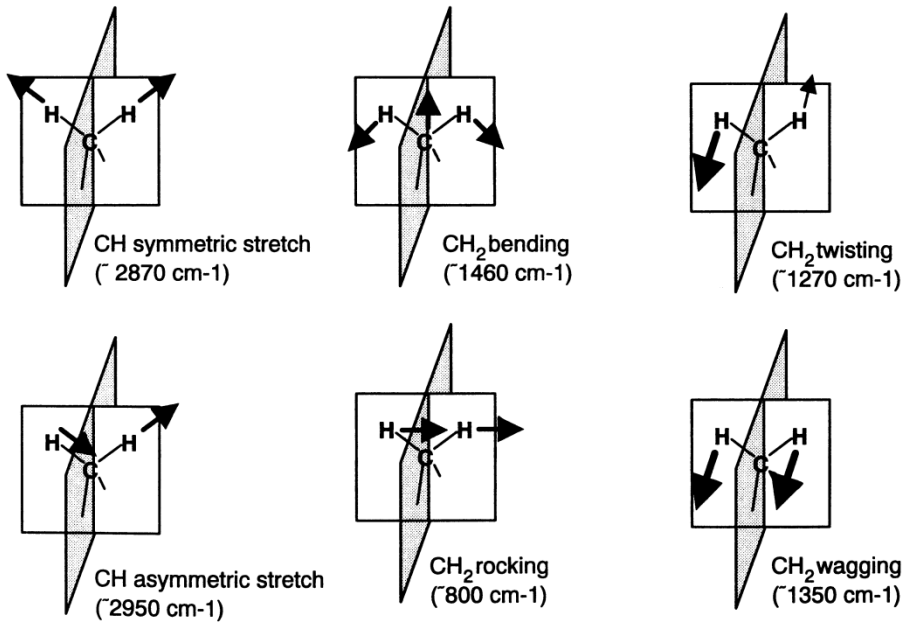


Figure 2-2 Normal vibrations of CH_2 in IR spectroscopy [Miller, 2001].

As an example on combination bands and the selection rules for the combination bands CH_2 may be taken. This functional group has a combination band in the NIR region at ~4310 cm^{-1} (~2320 nm); the band originates from a combination of the

symmetric stretch (2870 cm^{-1}) and the bending vibration (1460 cm^{-1}). It should however be remembered that in spite of the possibility of combination bands, two different modes of vibration of the same group will not always combine [Miller, 2001].

Beer's law (Equation 2-1) stated that the absorption (also in the NIR range) is linearly related to the concentration of any NIR-active analytes. This means that I samples measured at J wavelengths forms a data set of size $I \times J$ that may be approximated by means of bilinear models such as Principal Components Analysis (PCA) or Partial Least Squares (PLS) – methods that will be introduced in section 3.

2.1.1 Sample handing for NIR

There are in general two different methods for sample handling in NIR spectroscopy: Transmission or diffuse reflectance (Figure 2-3)

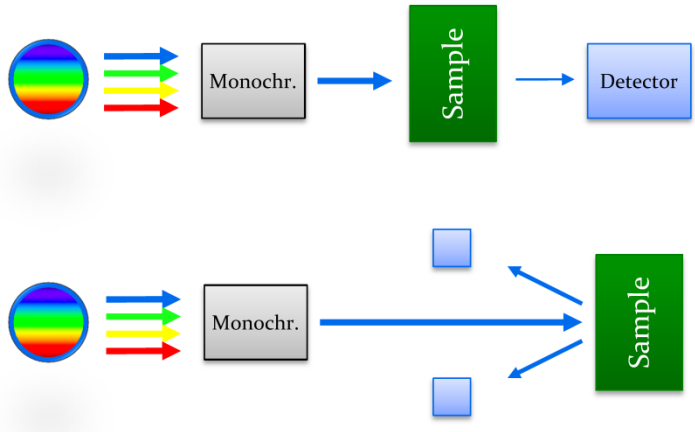


Figure 2-3 Schematic drawings of different sample handling techniques in NIR spectroscopy, transmission (top) and diffuse reflectance (bottom)

In transmission the absorbance/transmittance in the samples is measured, and a blank measurement is made beforehand where water or air is often used as reference [Folkenberg *et al.*, 2008]. In diffuse reflectance the sample is illuminated and the reflected light is measured [Workman & Burns, 2008], a white reflection-standard made of plastic (such as Spectralon®) is often used as standard/blank in this case. Transmission can be used both for IR and NIR. It has the advantage that the total sample thickness is used for the measurement. Errors due to heterogenic

samples (e.g. caused by separation of fat) can hereby be minimized [Workman & Burns, 2008]. There is however also a disadvantage in transmission. The absorbance of liquids (in IR) can be very high, and narrow cuvettes have to be used. This may complicate both correct sampling and cleaning in between different samples [Folkenberg *et al.*, 2008]. Sampling is, as explained above, easier for NIR since absorbencies are lower; transmission is therefore a very common solution for on- and in-line NIR measurements [Wust & Rudzik, 2003; Huang *et al.*, 2008]. Diffuse reflectance is frequently used for NIR of powders and solids. An advantage of this method is that sampling is easy, liquid samples that are too strongly absorbing in transmission may be analysed by this method. A major disadvantage is that the spectra are dependent on particle size (or e.g. the size of fat globules in solution). Uniform particles are therefore required at every measurement in order to obtain reproducible results [Dahm & Dahm, 2001]. Much of the scattering may however be removed by correct pre-processing of the spectra, e.g. by the Standard Normal Variate (SNV) method as presented in **Paper II**. Since only the surface layer of the sample interacts with the light, another disadvantage of reflection measurements are that heterogeneous samples may cause measurement biases [Workman & Burns, 2008].

2.2 Fluorescence Spectroscopy

Fluorescence is the emission of light from a molecule that has been brought to an electronically excited state falling back to the ground state [Lakowicz, 2006]. The molecule is in fluorescence spectroscopy brought from the ground state to the excited state by absorption of light. Fluorescence is commonly measured in the range 250 – 800 nm [Dickens, 2010] meaning that especially the UV-VIS range is used when fluorescence spectra are recorded. Figure 2-4 below illustrates the phenomenon of absorption and fluorescence; the ground state, first and second excited states are designated by S_0, S_1 and S_2 respectively.

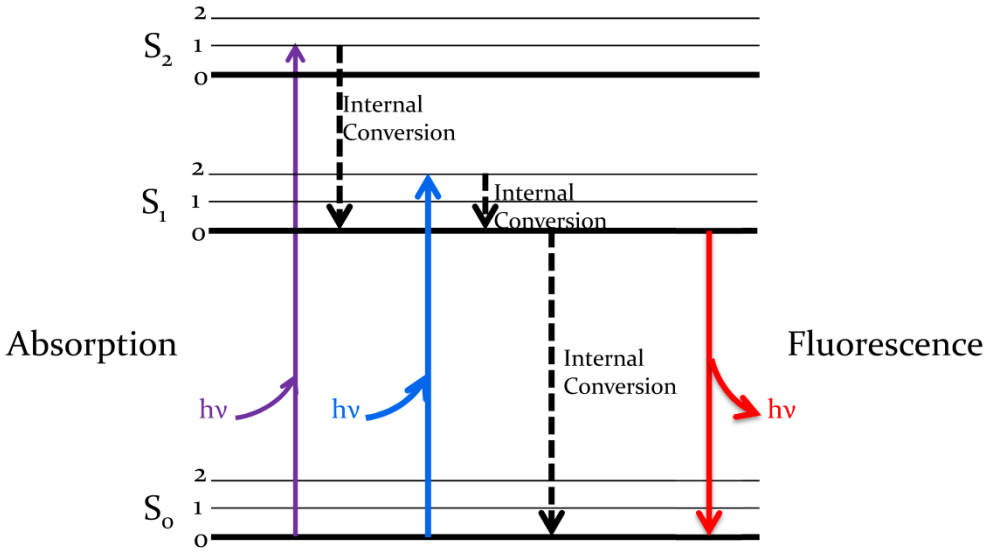


Figure 2-4 Jablonski diagram of the phenomenon of fluorescence, molecules are excited from the ground-state (S_0) to one of the excited states S_1 or S2, fluorescence may occur when the molecule relaxes from S_1 to the ground-state.

As the light excites the molecules, a transition from the ground-state (S_0) to one of the excited states S_1 or S_2 will happen. If the molecule was excited to S_2 it will normally relax to S_1 within 10^{-12} s or less, by transferring the energy to other molecules (e.g. the solvent) though collisions. This process is known as internal conversion [Lakowicz, 2006; Harris, 2007]. The molecule may from S_1 relax to the ground-state either through further internal conversion, or by emission (fluorescence). The first law of thermodynamics states that *"The algebraic sum of all energy changes in an isolated system is zero"* [Smith, 1990], this has the consequence

that only red-shifted fluorescence is seen – the emitted light will always be of longer wavelengths (λ_{em}) than the light used for excitation (λ_{ex}) since this light is less rich in energy and some of the energy is lost through internal conversion. The difference between the two wavelengths is known as the Stokes shift [Lakowicz, 2006].

Fluorescence is typically seen in aromatic or other compounds with cyclic structures where conjugated double bonds are found [Dickens, 2010]. Nicotinamide Adenine Dinucleotide (NADH), Flavin Adenine Dinucleotide (FAD), chlorophyll and many vitamins (including B2 as shown in **Paper I, III** and **IV**) may therefore be measured and quantified with fluorescence spectroscopy [Christensen, 2005]; the method is thereby thus a potential candidate for direct monitoring of the metabolism of microorganisms in bio-reactors.

Different types of fluorescence spectra may be recorded:

- Emission spectra, where one given excitation wavelength is used and the emission is measured at different wavelengths, one vector is recorded per sample.
- Excitation spectra, where the excitation wavelengths are scanned while the spectra are recorded at one single emission wavelength, one vector is recorded per sample.
- Emission-Excitation Matrix (EEM) spectra, where both excitation and emission wavelengths are scanned, a matrix is recorded per sample.

The fluorescence data presented in this thesis (**Paper I**, **III** and **IV**) were recorded as EEM spectra. Beer's law also applies to EEM spectra; they are therefore known to be tri-linear, essentially meaning that a low rank PARAFAC model may be used for a unique decomposition of the data [Smilde *et al.*, 2004], PARAFAC and its uniqueness property are elaborated on further in section 3.1.3.

2.2.1 Sample handling for Fluorescence

The most common sample geometry used in fluorescence spectroscopy is right-angle observation of the sample (Figure 2-5, left) where the fluorescence detector is placed perpendicular to the light source [Lakowicz, 2006].

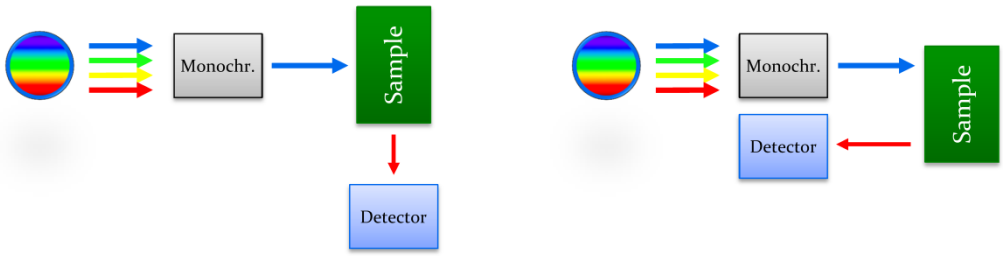


Figure 2-5 Different sample geometries for fluorescence spectroscopy, Left: Right-angle observation of sample, Right: Front-face 180° observation of sample

The EEM spectra presented in **Paper I, III** and **IV** were recorded using the BioView EEM fluorescence process spectrometers (Delta Light and Optics, Hørsholm, Denmark), one of the only fluorescence process spectrometers available on the market [Dickens, 2010]. This instrument records the EEM spectra using by front-face 180° sampling (Figure 2-5, right). The excitation light is via a light-guide sent to a process probe (Figure 2-6) where the sample is illuminated. The emitted light is subsequently collected using the same probe but sent via another light-guide to the detector.

Figure 2-6 BioView process probes mounted on a ultra-filtration (UF) process as monitoring tool. Further details on the UF process and experiments may be found in section 3.3.3.

The instrument is filter based using a combination of 15 excitation filters (equidistantly spaced from 270 to 550 nm) positioned on the light source sequentially and 15 emission bands filters (equidistantly spaced from 310 to 590 nm) positioned on the detector sequentially. The front-face 180° sampling results in a high degree of backscattered light at $\lambda_{em} = n\lambda_{ex}$ where n is an integer larger than zero (so-called Rayleigh scattering). Due to the Stoke shift and to avoid Rayleigh scattering the BioView therefore only measures the emission/excitation combinations where $\lambda_{em} > \lambda_{ex}$. A band of missing values surrounding the Rayleigh scatter band may be used during modelling, since the scatter is inconsistent with the PARAFAC model [Bro & Vidal, 2011]. With the first two excitation wavelength being 270 and 290 nm, 2nd order Rayleigh scattering could be expected at λ_{em} = 540 nm and λ_{em} = 580 nm. The scatter was however not observed in the data presented in this thesis. This could to a certain degree be a result of the much lower intensity of the second order scatter when compared to the first order, but could also be due to the fact that both 2nd order Rayleigh scattering wavelengths are positioned exactly right between the maximum bandpass of the neighbouring filters (as illustrated for the 2nd order Rayleigh at 540nm in Figure 2-7 below).

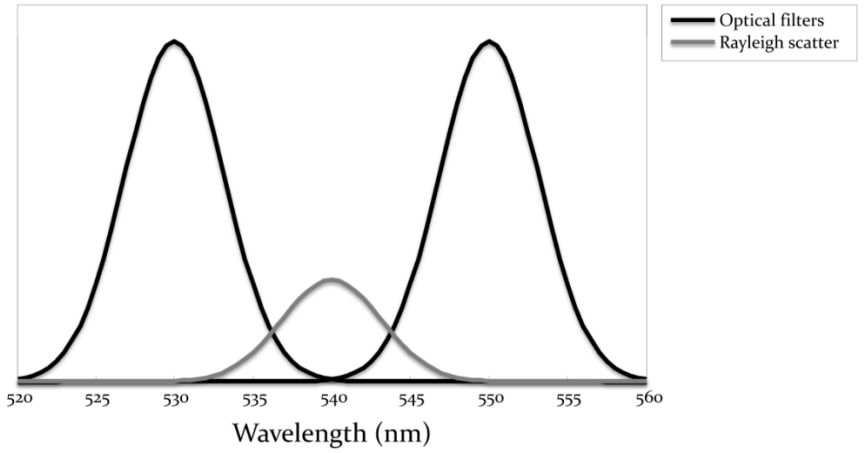


Figure 2-7 Maximum bandpass of optical filters and 2nd order Rayleigh scatter, the influence of the scatter is minimized due to the bandwidth of the filters.

The low intensity 2nd order Rayleigh scatter is hence split on two filters resulting in no or little disturbance of the spectra, and since the emission and excitation spectra in general are much broader than 20 nm, problems with not detecting the analyte due to splitting on two filters are not expected.

3 Process Modelling

Many different reasons for process modelling exist; a key one is to obtain a better process understanding. A good process model will however also offer many other possibilities: Simulation, prediction, optimization, operator training, fault diagnosis, quality and safety monitoring, model-based control and many others [van Overschee & De Moor, 1996]. Various simple rules should however be kept in mind during modelling. Since much information can be obtained simply by plotting the data a thorough inspection of the raw results should be conducted before any modelling. The power of visualization can be easily illustrated by the univariate data-set "Anscombe's Quartet", 4 sets each consisting of 11 observation pairs (x,y) [Anscombe, 1973].

Table 3-1 Anscombe's Quartet

Set	I		II		III		IV	
Variable Obs. no.	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.10	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.10	4	5.39	19	12.50
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89
Mean	*9*	*7.50*	*9*	*7.50*	*9*	*7.50*	*9*	*7.50*
Variance	*11*	*4.127*	*11*	*4.128*	*11*	*4.122*	*11*	*4.123*
R^2	*0.667*		*0.667*		*0.667*		*0.667*	

The four sets have the same mean, variance and correlation between x and y. If linear regression is made on the data, the same equations will be obtained: $y = 0.346x + 4$. Solely based on these statistics no big differences between the sets are

therefore to be expected, if the data are plotted however the differences become apparent (Figure 3-1).

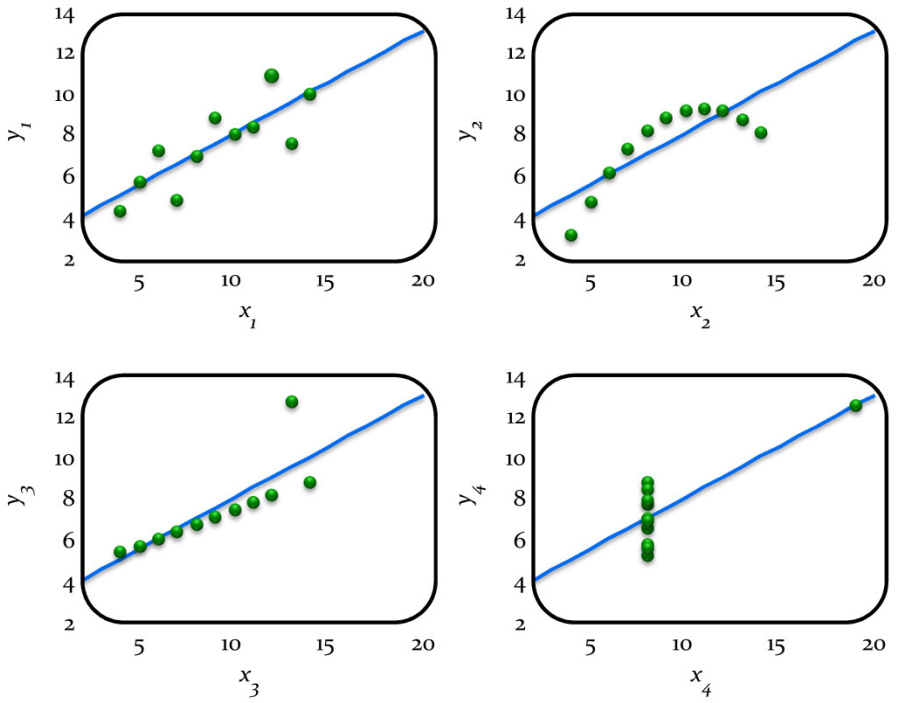


Figure 3-1 Ancombe's Quartet; the four sets has same mean, variance, correlation coefficient and first order least squares regression line fit.

This illustrates two points: 1) Outliers can severely hamper any conclusions made and 2) blindly applying models on data without thorough inspection may lead to false conclusions. A first order polynomial may be appropriate for data-set 1 and 3, but a second order polynomial is probably more suitable for set no.2 – this illustrates the power of data driven modelling. Based on inspection of raw data many patterns can be seen in data, thereby providing a good starting point for modelling. Modern process analysers (such as NIR spectrometers) in combination with more traditional process sensors (e.g. pH-meters, pressure and temperature sensors) can however gather hundreds of variables each minute. Simply just plotting data like these may be informative, but a reduction of the dimensionality is often required in order to facilitate any interpretation. This is where the field of multivariate statistics or chemometrics comes into play.

3.1 Chemometrics

Three different chemometric methods will be outlined in this section: Principal Components Analysis (PCA) [Pearson, 1901], Partial Least Squares (PLS) [Wold, 1966] and Parallel Factor Analysis (PARAFAC) [Carroll & Chang, 1970; Harshman, 1970]. All three methods are based on projection; the large dimensionality of the original data space is reduced by projecting the samples/objects onto underlying or latent components of lower dimensionality. The projection may either be orthogonal (as in PCA) or oblique (as in PLS and PARAFAC). The difference between orthogonal and oblique projection can for a simple 2 dimensional vector be illustrated as in Figure 3-2.

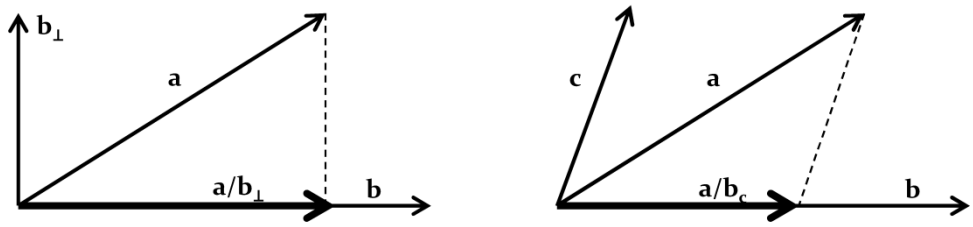


Figure 3-2 Orthogonal (left) and oblique projections(right)

In the case of orthogonal projection the vector **a** is projected in a right angle onto **b** resulting in the vector $\mathbf{a/b_\perp}$ or, stated otherwise, **a** is projected onto **b** along or parallel with $\mathbf{b_\perp}$ the vector orthogonal to **b**. In the case of oblique projection the vector **a** is projected onto **b** along the non-orthogonal vector **c** resulting in the vector $\mathbf{a/b_c}$.

The chemometric methods described in this thesis is much related to the concept of rank of a system. In the case where the measured data **X** is a matrix of size $I \times J$, i.e. J variables measured for I samples or objects, different types of rank can be defined for **X**. The column rank of **X** is defined as the number of linearly independent columns in **X**, the row rank of **X** the number of independent rows. A central theorem in linear algebra states that the row rank is equal to the column rank which overall is just known as the rank of **X,** written as r(**X**) [Strang, 2006]. This means that the (mathematical) rank of **X** is $r(\mathbf{X}) \leq \min(I,J)$, i.e. the rank is smaller or equal to the number of columns or rows whichever is the smaller. If the rank is equal to the number of rows or columns, whichever is smaller, **X** is called full rank. The mathematical rank of a matrix is in many cases not very interesting when modelling

of chemical systems is to be performed. Here the concept of chemical (or practical) rank is more useful [Smilde *et al.*, 2004]. The chemical rank of a system can be defined as the number of observable chemical sources of variation in the system. This can be illustrated e.g. in the case where NIR spectroscopy is used to monitor milk coagulation (**Paper II**). The first batch of coagulating milk was measured at 57 time points at 1400 wavelengths (**X** size 57 × 1400). The mathematical rank of this matrix is 57, a result of the random measurement noise that is present in the NIR data. Moreover, milk does consist of several hundred different chemical compounds [Fox & McSweeney, 1998]; it is however highly doubtable that 57 chemical sources of variation can be observed with NIR spectroscopy during the coagulation. Upon inspection and modelling of the data it could be shown that a chemical rank of 1 was sufficient to describe the system.

3.1.1 Principal Components Analysis (PCA)

PCA is a bilinear method for describing two-way data (**X** size $I \times J$). The original matrix is decomposed into sets of vectors, scores (**t**) and loadings (**p**), plus a residuals matrix (**E**). This principle is illustrated graphically below (Figure 3-3) for a two-component model:

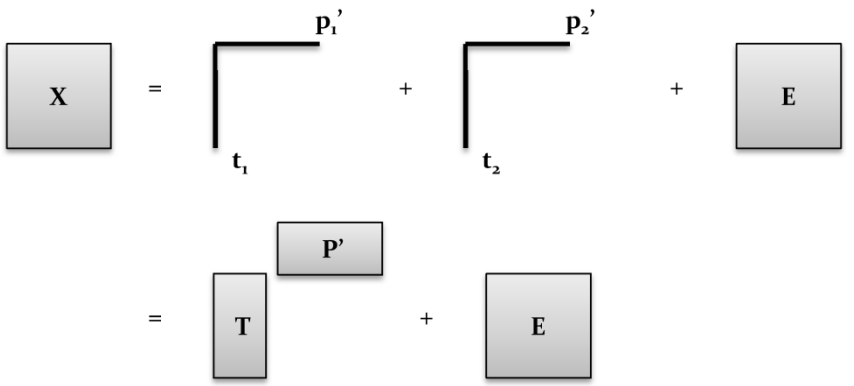


Figure 3-3 Pictorial illustration of the principle behind Principal Components Analysis the original data X is decomposed into scores (T), loadings (P') and residuals (E)

The idea is to have the systematic information in the scores and loadings and the noise in the residuals. The individual element $x_{i,j}$ in the matrix **X** (size $I \times J$) is for an R-component model thus estimated by [Næs *et al.*, 2002; Smilde *et al.*, 2004]:

$$x_{i,j} = \sum_{r=1}^{R} t_{i,r} p_{j,r} + e_{i,j} \;\; i = 1, \ldots, I; \; j = 1, \ldots J$$

Equation 3-1

Or if matrix notation is preferred:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \qquad \textbf{Equation 3-2}$$

where **T** is the scores matrix ($I \times R$), $\mathbf{P}^{\mathrm{T}}$ the transposed loading matrix ($R \times J$) and **E** the residual matrix ($I \times J$). Geometrically PCA can be thought of as orthogonal projections of the original data onto latent variables. The first latent variable or principal component (set of scores and loadings, $\mathbf{t}_1$ and $\mathbf{p}_1$) is therefore selected so that it describes the maximum variation within data. The first loading vector $\mathbf{p}_1$ is found as a linear combination of the original **X**-variables it is selected to have unit length. The first set of scores ($\mathbf{t}_1$) are computed by orthogonal projection of the observed data points onto the first loading. The residual ($\mathbf{E}_1$) for PC#1 is the part of **X** that is not described by the combination of $\mathbf{t}_1$ and $\mathbf{p}_1$, it is found by deflating **X** with the first set of scores and loadings ($\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}^{\mathrm{T}}_1$). A second set of scores ($\mathbf{t}_2$) and loadings ($\mathbf{p}_2$) can then be found from these residuals, the deflation step ensures that the second loading is orthogonal to the first loading (Figure 3-4).

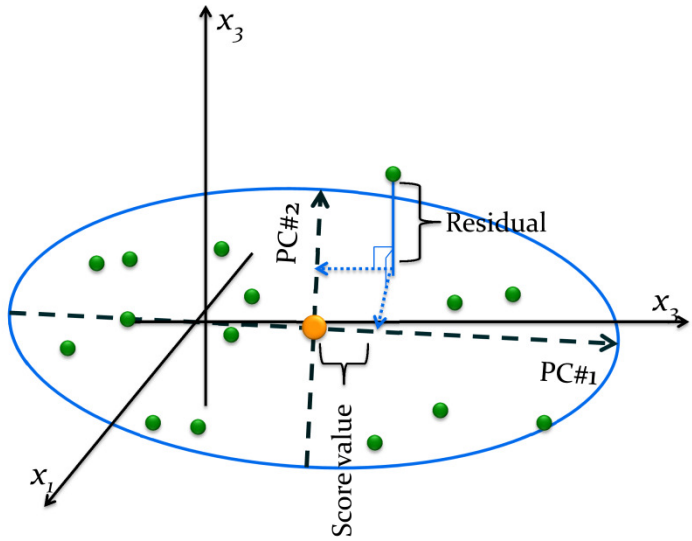


Figure 3-4 Geometrical representation of Principal Components Analysis, (●) original data point, (●) model centre, (--) principal components (PC)

The residuals can in the example illustrated in Figure 3-4 be thought of as the distance from the original data point ($x_{i,j}$) to the plane spanned by PC#1 and PC#2. The distance from the model centre to the point projected onto the plane is can be measured by Hotelling's T^2 value defined as

$$T_i^2 = \sum_{r=1}^{R} \frac{t_{i,r}^2}{s_r^2} \qquad \textbf{Equation 3-3}$$

Where s_r is the standard deviation of the r'th score [Kourti & MacGregor, 1996]. The Hotelling's T^2 is, in combination with plots of the scores and plots of the residuals, a very powerful tool that may be used for outlier detection, one of the many advantages when going from univariate to multivariate statistics [Olivieri, 2008].

Two competing methods are in general used for estimating the scores and loadings, an iterative method (the so-called Non-linear Iterative Partial Least Squares (NIPALS) algorithm [Wold, 1966; Wold *et al.*, 2001]) and the Singular Value Decomposition (SVD) based algorithm [Smilde *et al.*, 2004]. The NIPALS algorithm consists of the steps outlined below:

The NIPALS algorithm for PCA

0) Autoscale or mean center $\mathbf{X}$. For the number of components r=1,2…R compute $\mathbf{T}$ and $\mathbf{P}^T$ by:
1) Choose a column of $\mathbf{X}_{r-1}$ as a start guess of $\mathbf{t}_r$ (random choice or column with largest sum of squares)
2) Solve $\mathbf{X}_{r-1} = \mathbf{t}_r\mathbf{p}_r^T$ with regards to $\mathbf{p}_r^T$:
   $\mathbf{p}_r' = (\mathbf{t}_r^T\mathbf{t}_r)^{-1}\mathbf{t}_r\mathbf{X}_{r-1}$
   and normalize $\mathbf{p}_r$ to length one: $\mathbf{p}_r = \mathbf{p}_r/\|\mathbf{p}_r\|$
3) Solve $\mathbf{X}_{r-1} = \mathbf{t}_r\mathbf{p}_r^T$ with regards to $\mathbf{t}_r$:
   $\mathbf{t}_r = \mathbf{X}_{r-1}\mathbf{p}_r(\mathbf{p}_r^T\mathbf{p}_r)^{-1} = \mathbf{X}_{r-1}\mathbf{p}_r$ (since $\mathbf{p}_r^T\mathbf{p}_r = 1$)
4) Repeat step 2 and 3 until convergence
5) Orthogonalize $\mathbf{X}_{r-1}$ (deflate with $\mathbf{t}_r\mathbf{p}_r^T$):
   $\mathbf{X}_r = \mathbf{X}_{r-1} - \mathbf{t}_r\mathbf{p}_r^T = \mathbf{E}_r$
6) Start again from step 1 with $\mathbf{X}_r$ to extract the next principal component

The NIPALS algorithm is implemented in the commercial software packages Simca-P+ [Umetrics, 2008] and The Uncrambler X [Camo, 2004]. It can however be shown that the PCA model can be estimated based on SVD. The SVD approach is based on the fact that the loading vectors also can be seen as the eigenvectors of the covariance matrix of **X**, cov(**X**) given by

$$\mathbf{Cov(X)}=\frac{\mathbf{X^1X}}{(I\text{-}1)} \qquad \textbf{Equation 3-4}$$

i.e.

$$\mathbf{Cov(X)p}_r=\lambda_r\mathbf{p}_r \qquad \textbf{Equation 3-5}$$

Where λ_r is the eigenvalue corresponding to the eigenvector $\mathbf{p}_r$. The chemical rank of **X** may be estimated by inspecting a plot of these eigenvalues. The chemical rank is estimated as the first R eigenvalues notably larger than the next eigenvalue (R+1). In the case of the NIR data from **Paper II,** a rank of one was chosen since the first eigenvalue of **Cov(X)** was almost a factor 1000 larger than the next two eigenvalues (1.67 vs. $6.3*10^{-3}$ and $1.66*10^{-3}$).

The SVD of **X** is noted as:

$$\mathbf{X=USV^T} \qquad \textbf{Equation 3-6}$$

And since **U** holds the eigenvectors of $\mathbf{XX^T}$, **V** the eigenvectors of $\mathbf{X^TX}$ and **S** the singular values that are equal to the square root of the eigenvalues of both $\mathbf{X^TX}$ and $\mathbf{XX^T}$ [Strang, 2006] it can then be shown that [Smilde *et al.*, 2004]

$$\mathbf{T=US} \qquad \textbf{Equation 3-7}$$

and

$$\mathbf{P=V} \qquad \textbf{Equation 3-8}$$

It is thereby possible to compute the PCA model based on SVD on the covariance matrix. The SVD based approach is the standard method implemented in the commercial software package PLS_Toolbox [Eigenvector Research, 2011].

3.1.2 Partial Least Squares Regression (PLS)

Regression problems are found in almost all fields of science since it is a universal problem: Find the connection between the independent variable **X** (size $I \times J$) and the dependent variable **y** (size $I \times 1$)[3], that is we want to solve

$$\mathbf{y}=\mathbf{Xb} \qquad \text{Equation 3-9}$$

with regards to **b**. It is from linear algebra known that the least square solution may be used if more samples that variables are found in data (i.e. $I >> J$). The solution is then given by [Smilde *et al.*, 2004]:

$$\mathbf{b}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \qquad \text{Equation 3-10}$$

This is known as Multiple Linear Regression (MLR). The requirement of $I >> J$ is due to the calculation of $(\mathbf{X}^T\mathbf{X})^{-1}$. If **X** is not of full rank $\mathbf{X}^T\mathbf{X}$ will be singular or ill-conditioned (i.e. the inverse does not exist or is numerically unstable). This may pose a problem in many real life applications (e.g. spectroscopy) where more variables than samples are measured and/or where the chemical rank is lower than the mathematical rank (i.e. the measured variables are correlated). One solution to this problem is Principal Components Regression (PCR). The idea is to compress the **X** data by PCA followed by an MLR step where the dependent variable **y** is regressed on the PCA-scores **T.** PCR is thereby a two-step solution to the regression problem:

1) PCA on **X** to obtain scores **T** (Equation 3-2)

2) MLR on the scores [Næs *et al.*, 2002]:

$$\mathbf{b} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y} \qquad \text{Equation 3-11}$$

One drawback of PCR is that PCA has the objective of maximizing the variance explained in **X**, the scores and loadings that are found are not necessarily the

[3] A univariate dependent variable is assumed in this section. It should however be noticed that expansion to a multivariate **Y** is straight forward; this is known as PLS2.

optimal for prediction of **y**. Partial Least Squares (PLS) Regression is a solution to this issue. It finds latent variables that are a compromise between explaining the variance in both **X** and **y**. This is done by introducing three underlying models

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad \textbf{Equation 3-12}$$

$$\mathbf{y} = \mathbf{T}\mathbf{q}^T + \mathbf{F} \qquad \textbf{Equation 3-13}$$

$$\mathbf{T} = \mathbf{X}\mathbf{W} \qquad \textbf{Equation 3-14}$$

PLS finds the solution to this system of equations by maximizing the covariance between **X** and **y**. More formally this can be written as:

$$\max_{\mathbf{w}}[\mathrm{cov}(\mathbf{t},\mathbf{y})|\mathbf{X}\mathbf{w}=\mathbf{t} \text{ and } \|\mathbf{w}\| = 1]$$

Equation 3-15

It can be shown that this expression is maximized if [Smilde *et al.*, 2004]

$$\mathbf{w} = \frac{\mathbf{X}^T\mathbf{y}}{\|\mathbf{X}^T\mathbf{y}\|} \qquad \textbf{Equation 3-16}$$

The scores **T** can subsequently be found by Equation 3-14 and the regression vector as [Wold *et al.*, 2001]

$$\mathbf{b} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{q}^T$$

Equation 3-17

Phatak & de Jong [1997] showed that where PCR works by orthogonal projection, the introduction of **W** in PLS rotates the solution so that the projection becomes oblique (Figure 3-2).

The first PLS-algorithm to appear in literature was a modified version of the earlier presented NIPALS for PCA [Wold *et al.*, 2001]. Many different algorithms have since then been published, but it is the author's impression that NIPALS and SIMPLS (an algorithm based on SVD [de Jong, 1993]) are the most commonly used ones. Anderson [2009] gives a review of nine different algorithms, the overall conclusion for NIPALS and SIMPLS are that NIPALS is numerically stable but among the slower algorithms, and SIMPLS fast but numerically unstable for a high number of latent variables. Anderson does however not consider the numerically instability of SIMPLS as a problem with data modelling in practice. SIMPLS is implemented as

standard algorithm in the PLS_Toolbox [Eigenvector Research, 2011], while NIPALS is used in The Unscramber X and Simca-P+ [Camo, 2004; Umetrics, 2008].

3.1.3 Parallel Factor Analysis (PARAFAC)

PARAllel FACtor analysis (PARAFAC) is a method for decomposing *N*-way tensors. The basic idea is the same as in PCA: Find sets of scores and loadings to represent the variation found in the *N*-way tensor **X**. The method applies to tensors with 3 modes or more (hence the name *N*-way). For simplicity is PARAFAC in this thesis explained for *N*=3 (i.e. **X** (size $I \times J \times K$)), it should however be stressed that expansion for 4, 5 or higher orders of tensors are straight forward.

In PARAFAC the original data in the tensor **X**, are decomposed into scores for mode 1 (**A**) and two sets of loadings (**B** and **C** for mode 2 and 3 respectively) [Smilde *et al.*, 2004], this is illustrated below in figure 3-5:

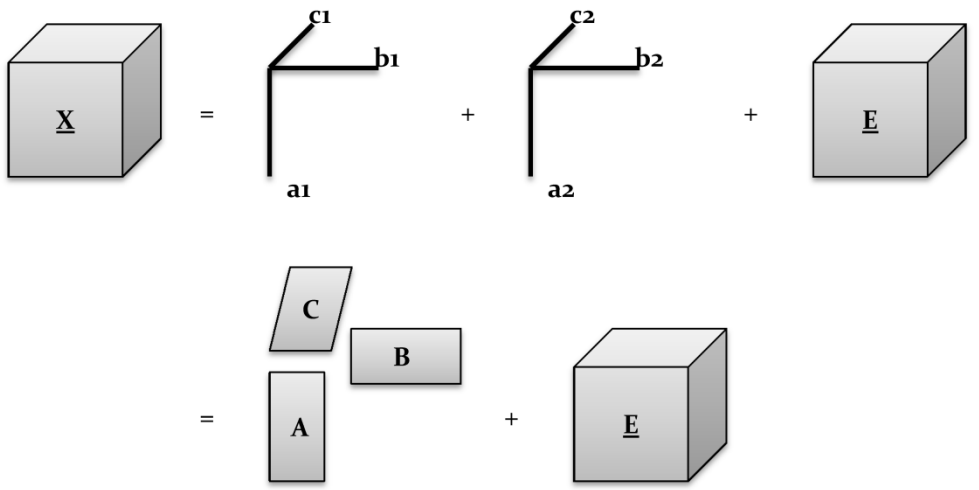


Figure 3-5 Pictorial illustration of the principle behind PARAFAC

The individual element in $\underline{\mathbf{X}}$ (size $I \times J \times K$) is for a R-component PARAFAC model defined by [Smilde *et al.*, 2004]:

$$x_{i,j,k} = \sum_{r=1}^{R} a_{i,r} b_{j,r} c_{k,r} + e_{i,j,k} \quad i = 1, \ldots, I;\ j = 1, \ldots J$$

Equation 3-18

This can in matrix notation be written as:

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}^{\mathrm{T}} + \mathbf{E}_k \qquad \textbf{Equation 3-19}$$

where the individual slab of $\underline{\mathbf{X}}$ ($\mathbf{X_k}$ with the size $I \times J$) is modelled by: $\mathbf{A}$ ($I \times R$) the matrix with the collected first mode scores, $\mathbf{B}^{\mathrm{T}}$ the matrix containing the second mode loadings, and $\mathbf{D_k}$ the third mode loadings. The third mode loadings (or **c**-vectors) are collected in $\mathbf{C}$ ($K \times R$), the k'th column of this matrix is used as the diagonal in the diagonal matrix $\mathbf{D_k}$ [Smilde *et al.*, 2004].

A very appealing feature of PARAFAC is the uniqueness of the model, which results in what is also known under the term "Mathematical Chromatography". This means that for truly tri-linear data, where the systematic structure is dependent on three different phenomena, so the data therefore cannot be collapsed in either dimension without loss of information, a PARAFAC model with the correct number of components will find these underlying phenomena [Bro, 1998]. An example of this can be found in **Paper III** where EEM-fluorescence spectroscopy is used to monitor a model system of riboflavin breakdown. A good correspondence was found between the spectral loadings in the resulting PARAFAC model and the known emission/excitation spectra of riboflavin and its breakdown products lumiflavin and lumichrome (Figure 3-6).

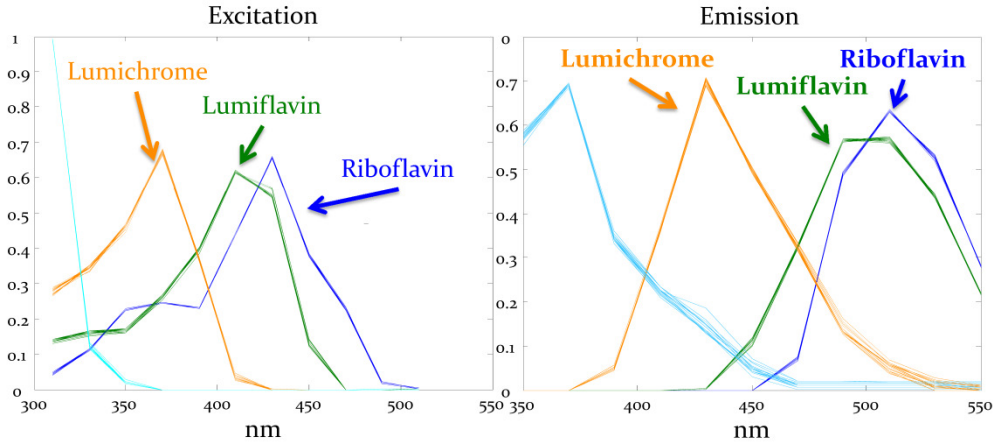


Figure 3-6 PARAFAC emission and excitation loadings of the riboflavin breakdown model system. Literature reports that Riboflavin has emission/excitation maximum at (λ_{ex} 450nm / λ_{em} 520nm), lumiflavin maximum close to riboflaivn but shifted to slightly lower em./ex. wavelengths and lumichrome at (λ_{ex} 360nm / λ_{em} 450nm).

Another interesting feature of PARAFAC was illustrated in **Paper I**, where transfers of three-way calibrations (based on PARAFAC) were studied. The conclusions of the paper was that three-way methods when compared to two-way (PLS) had lower prediction errors, and that as few as four samples could be used for calibration transfer (see sections 3.1.4 *N-way calibration* and 3.1.5 *Calibration Transfer*) . This exemplifies that PARAFAC in the ideal case may use very few samples to estimate the underlying tri-linear phenomena.

The advantages mentioned above are often termed the second order, three-way or multi-way advantage [Olivieri, 2008]. There are though, some challenges in applying PARAFAC, some of which are associated with the algorithms used for finding the models. Most PARAFAC algorithms are based on an Alternating Least Squares (ALS) approach [Bro, 1998]. The ALS algorithms works by splitting the parameters that are to be estimated into sets. ALS then estimates one set of parameters in a least squares sense given initial estimates of the remaining, uses the updated estimates to find the next set and iterates over all the sets until convergence [Bro, 1998]. In case of a PARAFAC model, an ALS algorithm would therefore iterate over the following steps: Estimate **A** given initial estimates of **B** and **C**, then estimate **B** given **C** and the updated estimate of **A,** and finally **C** given the updated **A** and **B** until convergence (hence the name Alternating Least Squares). The NIPALS algorithm for PCA

outlined in section 3.1.1 can thus also be seen as an ALS algorithm (it alternates between estimating **t** and **p**), but where an deflation step is included in NIPALS to ensure orthogonallity between the PC's this is not done in the PARAFAC algorithms since the PARAFAC model does not imply orthogonallity between the PARAFAC components [Smilde *et al.*, 2004]. This has the consequence that where the sequential NIPALS algorithm provides a nested PCA model (PC#1 in a two component PCA model is the equal to PC#1 in a three component PCA model), this is not the case for the ALS based PARAFAC algorithms. It will fit all the components simultaneously, with the result that the first *R* components in a given model are different from the first *R* components in a model with (*R*+1) components for the same data [Bro, 1997]. Since the ALS based algorithms may require long computer time to fit the individual model, it could therefore be tempting to fit a more complex model than expected, and then just inspect the first components, as is often done for PCA. Due to the non-nesting it is necessary to fit and inspect the different models individually [Bro, 1997].

Another effect of the simultaneously fitted components is that if a data set is divided in two, the same loadings are found for two sub sets (if the underlying structure in data is the same for both sub sets and the correct number of components are used). The order of the loadings may however be different for the two sub sets (e.g. loading 1 for sub set1 = loading 3 for sub set 2). The approach with splitting the data set into two is called split half analysis. An example of split half analysis can be found in **Paper III**. Here were the same emission and excitation loadings (Figure 3-6) found for several individual three-way tensors that were independently decomposed by means of PARAFAC, indicating that the estimated loadings reflect the true chemistry in the system.

Finally it should be mentioned that constrains may be put on the ALS solution. In the case presented above, where PARAFAC is applied on EEM-fluorescence data, one would expect the excitation and emission loadings to be positive and non-negativity constrains may therefore be imposed on the mode 2 and 3 loadings to force this situation. More advanced constrains may also be imposed. The PARAFAC scores can be seen as pseudo concentrations due to the uniqueness of the model [Bro, 1997], and in the case where reaction dynamics are followed, constraining the scores to follow e.g. first order kinetics would therefore make sense. **Paper IV** presents how such functional constrains can be applied on the PARAFAC scores.

3.1.4 N-way calibration

In section 3.1.2 it was shown how MLR, PCR and PLS can be applied for regression of a two-way independent variables **X** (size $I \times J$) on a dependent variable **y**. In the case where the independent variables form an N-way tensor ($\underline{\mathbf{X}}$ size $I \times J \times K$), two possible options exists: Either the $\underline{\mathbf{X}}$ data are unfolded to a two-way matrix keeping the sample direction intact (**X** size $I \times JK$) and PLS is applied, or dedicated N-way methods are to be used. In the papers by Ståhle [1989], Bro [1996] and Bro *et al.* [2001] PLS was extended to handle N-way data (N-PLS). It is however also possible to combine PARAFAC with MLR (or another regression method) in a PCR-like manner to obtain an N-way regression model: Decompose $\underline{\mathbf{X}}$ using PARAFAC and regress **y** on the PARAFAC scores **A** (Figure 3-7).

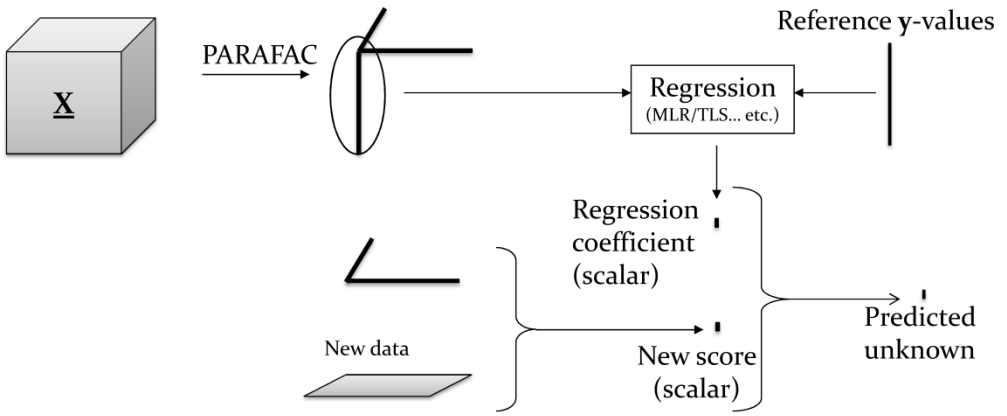


Figure 3-7 Principal behind PARAFAC based *N*-way calibration for a one component PARAFAC model

It is in general accepted that N-PLS produces better predictions (just as PLS outperforms PCR), but also that interpretation of the model is easier for the PARAFAC based regression [Bro *et al.*, 2001; Pedersen *et al.*, 2002]. The PARAFAC based approach was applied in **Paper I.** In this case a non-linear dependence was observed between the PARAFAC scores and **y.** A quadratic function was therefore fitted using total least squares. When the N-way method was compared to 2-way PLS (based on unfolding $\underline{\mathbf{X}}$), a poorer performance was seen for the latter.

3.1.5 Calibration Transfer

Section 2 introduced how spectroscopy may be used to monitor food and pharma production processes. As also mentioned in the introduction, the on-line implementation of e.g. a NIR-spectrometer is not maintenance-free. Apart from the spare part replacements needed due to physical wear and tear on the equipment (it is placed in a process environment!), the calibration also needs to be maintained. This is needed, for instance, due to small continuous changes in the instrument (e.g. drift due to filter bleaching) or more sudden changes (e.g. if the instrument lamp is replaced by a new one). The spectra obtained using the two lamps for the same sample measured under the same conditions will be different and the calibration models will thus not necessarily be valid for the new situation. The most straightforward solution, of course, would be to re-calibrate for the new measurement conditions or to expand the original model for the new situation. Unfortunately, this is also the most expensive solution and sometimes technically impossible. Standardization and calibration transfer methods have been developed aimed at eliminating the need for a full recalibration and to preserve the information collected in an existing model. A commonly used approach for updating the calibration is so-called slope/off-set correction, either directly on the recorded spectra or on the predictions from the calibration model. The method is thus based on simple univariate correction between the spectra recorded on the primary and secondary instrument or the predicted and the actual **y**-value for a given control sample set [van den Berg & Rinnan, 2009]. This approach is however not always sufficient for achieving satisfactory results and more advanced methods have therefore been developed.

The two most popular methods for standardization of two way data are Direct Standardization (DS) and Piecewise Direct Standardization (PDS), introduced in a series of studies conducted by Wang and co-authors [Wang *et al.*, 1991; Wang & Kowalski, 1993a; Wang & Kowalski, 1993b]. The principle behind DS is to find the transfer matrix $\mathbf{F}$ ($J \times J$) that connects the spectra measured on the primary instrument ($\mathbf{X}_p$ - $I \times J$) with the spectra recorded on the secondary instrument ($\mathbf{X}_s$- $I \times J$). This is in essence a regression problem and is in the DS approach solved by using the Moore-Penrose pseudo-inverse:

$$\mathbf{X}_p = \mathbf{X}_s\mathbf{F} \qquad \textbf{Equation 3-20}$$

$$\mathbf{F} = \mathbf{X}_s^{+}\mathbf{X}_p \qquad \textbf{Equation 3-21}$$

Where $\mathbf{X}_s^+$ indicates the Moore-Penrose pseudo-inverse of the transfer sample matrix. Computing the pseudo-inverse may however lead to numerical instabilities especially in the case where more variables than transfer samples are measured. PDS was developed as a solution to this problem. In this approach the transfer function for each variable in the spectrum of the primary instrument is estimated from a (symmetric) window surrounding the same variable on the secondary instrument. This results in a much smaller (and thus more stable) local inversion step [Wang *et al.*, 1991].

While a wide array of different methods for transfer of two-way data can be found in literature, far less work is available for three-way data. **Paper I** therefore had the focus of modifying existing two-way transfer methods and developing new three-way methods for calibration transfer. An uncomplicated local linear method was demonstrated to be the most favourable of the new methods. The method was based on the observation that the on the level of individual emission/excitation channels, the counts on the primary and the secondary instrument had a relationship that could be modelled with a low order polynomial. Similar performance results were obtained for the modified DS/PDS methods and the newly developed three-way methods. It was also shown that the three-way advantages allowed application of very few transfer samples, though with results that was highly dependent on the selection of the transfer set.

3.1.6 Multivariate Statistical Process Control (MSPC)

Statistical Process Control (SPC) has the objective to investigate whether a process is in a state of statistical control [Massart *et al.*, 1997]. It is (in spite of the name) concerned with process *monitoring* rather than process *control*. In order to keep to the terminology of the field the acronym SPC is maintained in the following. The monitoring scheme in SPC is largely based on so-called control charts, charts showing one or more process variable plotted over time (Figure 3-8).

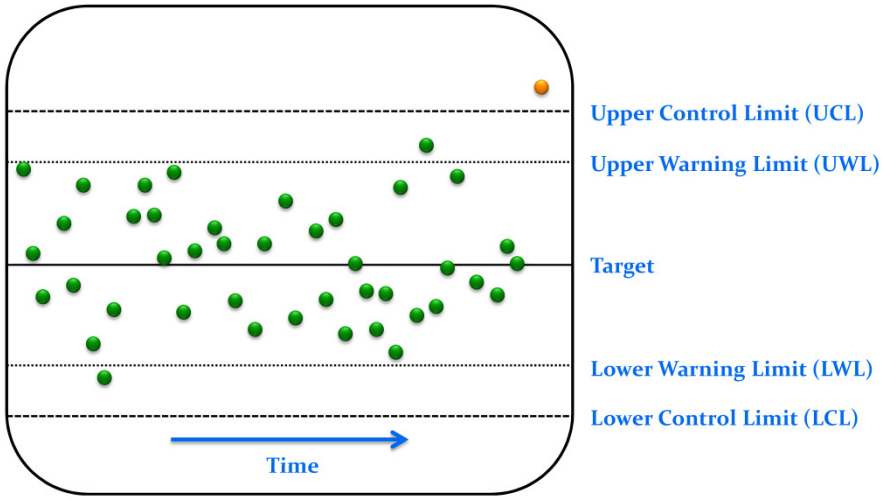


Figure 3-8 Mean or Shewart chart, process variable is plotted over time with process target plus upper and lower warning and control limits.

Figure 3-8 is a so-called mean or Shewart chart; the process variable is charted over time with a target or centre line (CL), upper and lower warning limits (UWL/LWL) and upper and lower action or control limits (UCL/LCL). Typically the target and control limits are based on historical process data where the process was deemed to be in statistical control (so-called Normal Operating Conditions/NOC). The target is then defined as the mean, and the upper and lower warning and control limits as $\pm 2\sigma$ and $\pm 3\sigma$ where σ is the standard deviation around the process mean. Other types of charts, such as the Cumultative Sum (CUSUM) chart (good for detecting drifts), or the Exponentially Weighted Moving Average (EWMA) chart (which puts more emphasis on the last observation than the earlier) may also be used for tracking the process variables. The Shewart chart is however by far the most popular one in industry [Massart *et al.*, 1997], the focus is therefore kept on this chart and the reader is referred to literature for further details on CUSUM or EWMA charts.

The Shewart chart is based on an assumption of normality of the data, meaning that the ± 2 and 3σ limits correspond to 95.5 and 99.7% confidence intervals. This also has the consequence that for a process which is in statistical control, we may expect 9 out of 200 samples to be placed outside the warning limits, and 3 out of 1000 samples to be placed outside the control limits. Different rules for deeming the process in or out of statistical control may therefore be encountered in literature,

the best known are the eight *Western Electric rules* which includes two extra additional lines at $\pm 1\sigma$ in the control chart [Massart *et al.*, 1997; Harris, 2007]:

1. One point outside UCL or LCL.
2. Nine points in a row on one side of the CL.
3. Six decreasing (or six increasing) points in a row.
4. Fourteen points in a row, alternating down and up.
5. Two out of three points outside UWL or LWL.
6. Four out of five points outside the 1σ line on the same side of the CL.
7. Fifteen points in a row within the two 1σ lines.
8. Eight points in a row beyond either of the two 1σ lines

When several (correlated) process measurements are to be monitored, the univariate methods outlined above might not detect if the correlation structure is broken. Methods based on reduction of dimensionality (e.g. PCA [Jackson, 1959] or PLS [Kresta *et al.*, 1991]) might in such cases be useful instead. This is known as *Multivariate* Statistical Process Control (MSPC). In the case where a PCA approach is taken the Hotellings T^2 (Equation 3-3 - distance from the model centre) known as the D-statistic, and the residuals (**E**), the variation not explained by the model, measured in the Q-statistic, are applied. The Q-statistic is for the *i*'th sample based on the sum of squared residuals [Jackson & Mudholkar, 1979]:

$$Q_i = \sum_{j=1}^{J} (x_{i,j} - \hat{x}_{i,j})^2 \qquad \textbf{Equation 3-22}$$

By charting D and Q it is possible to get an answer to the question of whether the process is behaving according to NOC. Since both D and Q always are positive only upper control limits are used for D and Q charts [Massart *et al.*, 1997].

The process illustrated in Figure 3-8 assumes a stationary process where the target is not changing over time. It should be mentioned that charts with a moving mean may be constructed; such charts were presented in **Paper II** and **III**, just as charts with non-symmetrical CI's may be constructed e.g. based on re-sampling of the NOC data [Conlin *et al.*, 2000].

The main set of methods within MSPC is primarily post-problem, rather than pre-problem (i.e. there needs to be an error before we see it). If pre-problem monitoring

is to be achieved time series models are needed. One possible class of these – the state space models – will be introduced in section 3.2 *State Space Models*.

As it was outlined in the introduction an end-target is often used in the case of batch process monitoring. Predicting this end-point (or the quality at the end-point) is therefore of interest. A commonly used method for predicting end-point quality of batches was presented in 1995 by Nomikos & MacGregor [Nomikos & MacGregor, 1995a]. In the case that a number (j = 1, 2... J) of process variables are followed over time (k = 1, 2... K) for several batches (i = 1, 2... I), a three-way tensor of the observed data $\underline{\mathbf{Y}}$ ($I \times J \times K$) can be formed. The method is based on collecting the corresponding quality variables (m = 1, 2... M) in matrix $\mathbf{Z}$ ($I \times M$), unfolding the observed training data in the batch direction to obtain $\mathbf{Y}$ ($I \times JK$) and regressing the autoscaled $\mathbf{Y}$ on $\mathbf{Z}$ using PLS2 (Figure 3-9) [Nomikos & MacGregor, 1995a].

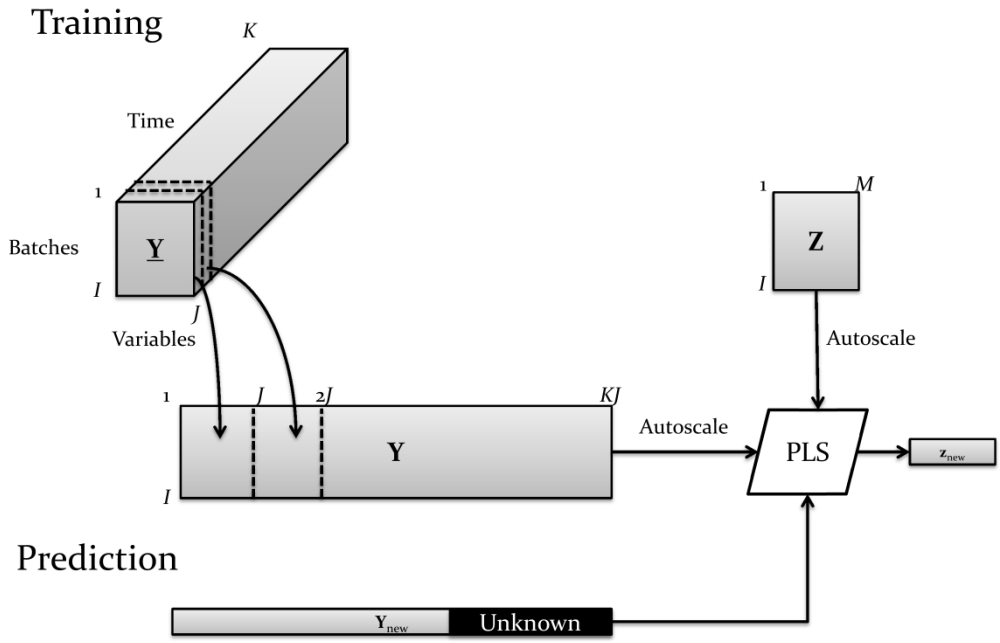


Figure 3-9 Nomikos & MacGregor PLS-approach for end-point prediction

A challenge when applying the method on-line for new batches is the fact that not all data points for the new vector are available. E.g. if the total batch run is 89 steps long (K = 89) and we are currently at k = 20 the remaining 69 time steps are yet unknown. Different methods are available for estimating the process outputs during

the remaining time steps. Nomikos & MacGregor [1995b] showed that simply setting the remaining time steps to missing, and using the ability of PLS to estimate the missing data is the best in a range of methods compared. It does though require the trajectories not to exhibit frequent discontinuities and approximately 10% of the batch history needs to be recorded before reliable results can be obtained [Nomikos & MacGregor, 1995a; Nomikos & MacGregor, 1995b].

3.2 State Space Models

Where chemometric methods has their origin in analytical chemistry [Esbensen & Geladi, 1990], the state space model has its origin in control or systems engineering (especially the fields of aerospace and electrical engineering [Luyben, 1996]). This has the consequence that slightly different notations will be used in this section. In the introduction the concepts of system, input, output and disturbances were introduced. Following the conventional notation in the field of systems engineering the following representation is used: The output vector (at time k) will be labeled $\mathbf{y}_k$, the input $\mathbf{u}_k$, system and measurement noise $\mathbf{w}_k$ respectively $\mathbf{v}_k$ and the state vector, a latent representation of the current state of the system $\mathbf{x}_k$ [van Overschee & De Moor, 1996; Ljung, 1999].

State space models are linear, time-invariant relations between the physical inputs to the system at time k, and the physical outputs (measurements) at time k, connected via the state-vector [van Overschee & De Moor, 1996; Ljung, 1999]. A discrete time state space model can be written via vector/matrix products as shown in Equation 3-23 and Equation 3-24.

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + \mathbf{w}_k \qquad \textbf{Equation 3-23}$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}\mathbf{u}_k + \mathbf{v}_k \qquad \textbf{Equation 3-24}$$

Equation 3-23 is often referred to as the system equation (reflecting that it describes how the system evolves over time via the difference relationship of the state vector $\mathbf{x}_k$) while Equation 3-24 is called the measurement equation (it describes how the measured output is related to the state of the system). The **A**-matrix is called the system matrix which describes how the system (or the states) evolves from one time-step to the next; **A** thereby describes the system dynamics. The input matrix **B** explains how a control input at time-step k would affect the system at k+1. **C** is

referred to as the measurement matrix representing, as stated previously, how the states are reflected in the physically measured outputs ($\mathbf{y}_k$). The fourth matrix (**D**) is called the (direct) feed-though, it explains how a control input (at time step k) can directly be observed in the output at time step k. This term is however seldom included in modelling, and is also not included in the models presented in this thesis. In case of purely stochastic time series (without any inputs such as in **Paper II**) state space equations may also be used to model the data and this is done by leaving out the **B** and **D** terms of the equations. The conceptual links between the input/output formulation of a system and the state space notation can be illustrated as in Figure 3-10.

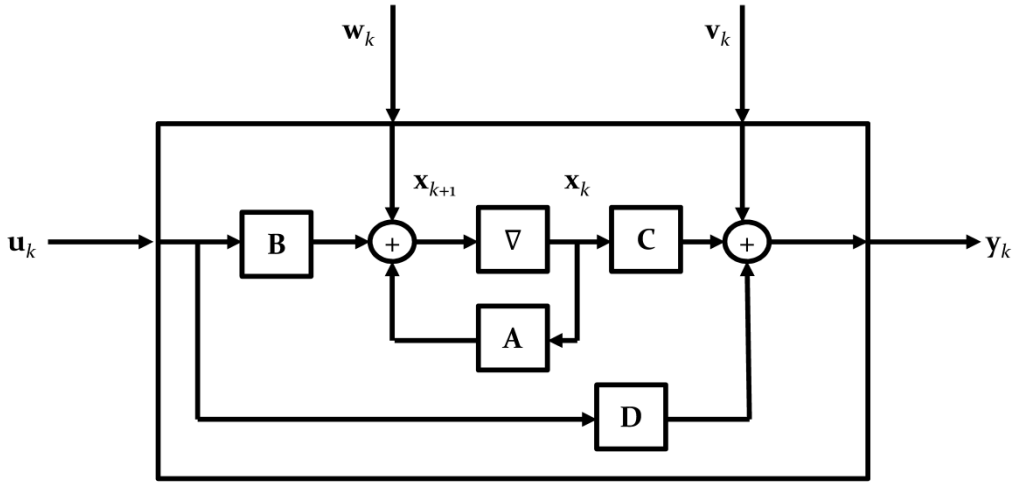


Figure 3-10 Data flow in state space models, only the input ($\mathbf{u}_k$) and the output ($\mathbf{y}_k$) are observed. ∇ is the backward-shift operator ($\nabla(\mathbf{x}_{k+1}) = \mathbf{x}_k$)

Several things are worth noticing in Figure 3-10. First of all it is seen how all system dynamics are collected in **A**, this has the consequence that inspection of the eigenvalues of **A** can be used for assessment of the system or model stability. This is a central diagnostics tool for state space models, further notes on stability and these eigenvalues can be found in section 4.3. Figure 3-10 also illustrates that disturbances affect both the system ($\mathbf{w}_k$) and the measured output ($\mathbf{v}_k$). Finally the figure also illustrates why **D** is referred to as the (direct) feed-though; **D** takes the input ($\mathbf{u}_k$) and directly feeds it to the output ($\mathbf{y}_k$). An important note should be attached to the system states ($\mathbf{x}_k$). Just as loadings and scores in a PCA model not necessarily correspond to e.g. pure compound spectra or concentrations, so do the system states

not necessarily coincide with physical phenomena in the system (e.g. concentrations in a chemical reactor). They should instead be seen as a latent representation of the dynamics spanning the subspace of relevance for the system. It may however sometimes be possible to rotate the states into physically meaningful entries [van Overschee & De Moor, 1996] similarly to that PCA scores and loadings in some cases may be rotated to ease interpretation [Lawaetz *et al.*, 2009].

Just as PCA models may be estimated by the iterative NIPALS or the SVD based method, state space models may either be fitted using iterative predictor-error methods (PEM), or by using so-called subspace methods that are based on projection of data on subspaces using SVD. Section 3.3.1 – on system identification will elaborate on the algorithms and how the models can be estimated.

3.2.1 The Kalman Filter

A common challenge faced when applying process models is the difference between measurement and model prediction: The measurement gives one estimate of the current process output, but the process model suggests another. It is known that both model and measurement are hampered by inaccuracy and uncertainty. One of the strengths of the state space model is easy implementation of what is known as the Kalman filter, a so-called optimal linear observer. It gives an answer to the challenge of measurement vs. model, making it possible on one hand to find the best compromise between the two, but also to estimate confidence intervals of the future process output. The filter was (in the form presented in this thesis) originally published by Rudolph Kalman in 1960 [Kalman, 1960a]. Previous work had been done on filtering of time series (e.g. the Wiener filter [Wiener, 1949]), but only very few practical implementations of the filtering algorithms were found in literature during the 1950'ies. This was however changed with the introduction of the Kalman filter, mainly due to its adaptive nature, but also as a result of the quick adaptation by NASA for the Apollo space program [Simon, 2006]. The filter became hereby well-known and has since spread to other fields of science and engineering.

The Kalman filter works by balancing the system noise ($\mathbf{w}_k$ Equation 3-23) and the measurement noise ($\mathbf{v}_k$ Equation 3-24). Both noise sequences are in the Kalman filter assumed to follow a normal distribution with $\mathbf{w}_k \sim N(0,\mathbf{Q})$ and $\mathbf{v}_k \sim N(0,\mathbf{R})$. This indicates that both the system and the output measurements are affected by

uncertainty at each time step. The Kalman filter combines the noise corrupted measurements of the system output ($\mathbf{y}_k$) with the predicted system output ($\mathbf{C}\mathbf{x}_k$) in a statistical optimal manner [Maybeck, 1979].

Equation 3-23 makes it possible to estimate the state at time-step k ($\hat{\mathbf{x}}_k^-$) from the previous state ($\mathbf{x}_{k-1}$) and the previous input ($\mathbf{u}_{k-1}$), this estimate is known as the *a priori* estimate at time step k (indicated by the "super minus"). The Kalman filter combines the noisy measurements $\mathbf{y}_k$ with the predicted system output $\mathbf{C}\mathbf{x}_k$ by finding the *a posteriori* system state as a linear combination of the *a priori* system state and a weighted difference between measured $\mathbf{y}_k$ and anticipated response $\mathbf{C}\hat{\mathbf{x}}_k^-$.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-) \qquad \textbf{Equation 3-25}$$

The difference ($\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-$) (also known as the innovation) can be computed straight forward, and reflects how large the agreement between the actual system output measurement and the system model is.

The estimation errors for the system states are given by:

$$\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^- \qquad \textbf{Equation 3-26}$$

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k \qquad \textbf{Equation 3-27}$$

With the *a priori* $\mathbf{P}_k^-$ and *a posteriori* $\mathbf{P}_k$ covariances defined as:

$$\mathbf{P}_k^- = E[\mathbf{e}_k^- \mathbf{e}_k^{-T}] \qquad \textbf{Equation 3-28}$$

$$\mathbf{P}_k = E[\mathbf{e}_k \mathbf{e}_k^T] \qquad \textbf{Equation 3-29}$$

$\mathbf{K}_k$ in Equation 3-25 is known as the Kalman gain; it is chosen so that the *a posteriori* system state error covariance ($\mathbf{P}_k$) is minimized. One common definition of the Kalman gain is [Maybeck, 1979]:

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{C}^T(\mathbf{C}\,\mathbf{P}_k^-\mathbf{C}^T + \mathbf{R})^{-1} \qquad \textbf{Equation 3-30}$$

Where $\mathbf{R}$ is the measurement noise covariance (associated with $\mathbf{v}_k$ in Equation 3-24). Equation 3-30 illustrates how the Kalman filter balances the error covariances to give weight to either the actual measurement ($\mathbf{y}_k$) or the predicted measurement ($\mathbf{C}\hat{\mathbf{x}}_k^-$). If the measurement noise covariance ($\mathbf{R}$) is small (the measurements are trusted), the Kalman gain becomes large and Equation 3-25 subsequently weights up the innovation, driving the *a posteriori* system state away from the predicted

measurement ($\hat{\mathbf{x}}_k^-$) towards the actual measurement ($\mathbf{y}_k$). The opposite is of course also the case, the *a priori* system state error covariance ($\mathbf{P}_k^-$) depends on the system noise covariance matrix **Q** (see Equation 3-32 below). If **Q** is small, $\mathbf{P}_k^-$ will also be small, resulting in a likewise smaller Kalman gain, meaning that the predicted measurement will be trusted more. Using the right estimates for **Q** and **R** (or rather the relative size ratio) is therefore of key importance to obtain the right Kalman filter estimates. And while the measurement noise covariance (**R**) is often known or easily estimated, the process noise covariance matrix (**Q**) is less easily available since the states in $\mathbf{x}_k$ themselves are estimates that are not directly observed. Mehra [1970; 1972] showed that a sub-optimal estimate can be obtained as:

$$\mathbf{Q} = \mathbf{P}_o - \mathbf{A}(\mathbf{I}-\mathbf{K}_o\mathbf{C})\mathbf{P}_o\mathbf{A}^T \qquad \textbf{Equation 3-31}$$

With $\mathbf{P}_o$ being the initial error covariance matrix for the system states and $\mathbf{K}_o$ the initial Kalman gain. Procedures for better estimates of **Q** have been published (e.g. Odelson *et. al.* [2006] and Rajamani & Rawlings [2009]), but they do not provide a simple closed form expression. The approach presented in Equation 3-31 was applied in **Paper III;** it was however also shown that more attractive process output estimates could be achieved when the **Q**/**R**-ratio was tuned (discussed below).

The *a priori* error covariance matrix ($\mathbf{P}_k^-$) can at time step *k* be found from the relationship

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q} \qquad \textbf{Equation 3-32}$$

It is updated after a measurement updated to the *a posteriori* error covariance ($\mathbf{P}_k$) by:

$$\mathbf{P}_k = (\mathbf{I}-\mathbf{K}_k\mathbf{C})\,\mathbf{P}_k^- \qquad \textbf{Equation 3-33}$$

Where **I** is the identity matrix, and $\mathbf{K}_k$ the Kalman gain from Equation 3-30. Based on the *a posteriori* error covariance the output covariance matrix $\mathbf{S}_k$ is found by [Mehra, 1972]:

$$\mathbf{S}_k = \mathbf{C}\mathbf{P}_k\mathbf{C}^T + \mathbf{R} \qquad \textbf{Equation 3-34}$$

It can be shown that under the assumption that the process and measurement noise are normal, the state estimates are also normal [Welch & Bishop, 2006]. Furthermore it is well known that a linear transformation of a normally distributed

process results in a process that is also normally distributed. The vector of (1-α) confidence intervals $\mathbf{ci}_k$ for the predicted output at time k can therefore be found as:

$$\mathbf{ci}_k = \hat{\mathbf{y}}_k \pm \Phi_{1-\alpha/2} \sqrt{\text{diag}(\mathbf{S}_k)} \qquad \textbf{Equation 3-35}$$

Where $\Phi_{1-\alpha/2}$ is the (1-α/2) quantile of the standard normal distribution and diag($\mathbf{S}_k$) the diagonal elements of $\mathbf{S}_k$ [Brockwell & Davis, 2002].

Paper III showed how a combination of state space models and Kalman filters could be used for batch process modelling and monitoring. Based on NOC training data, it was shown how the proposed method was able to capture and model the dynamics of a batch process. The implementation of the Kalman filter made the method adaptable to new non-NOC conditions (Figure 3-11).

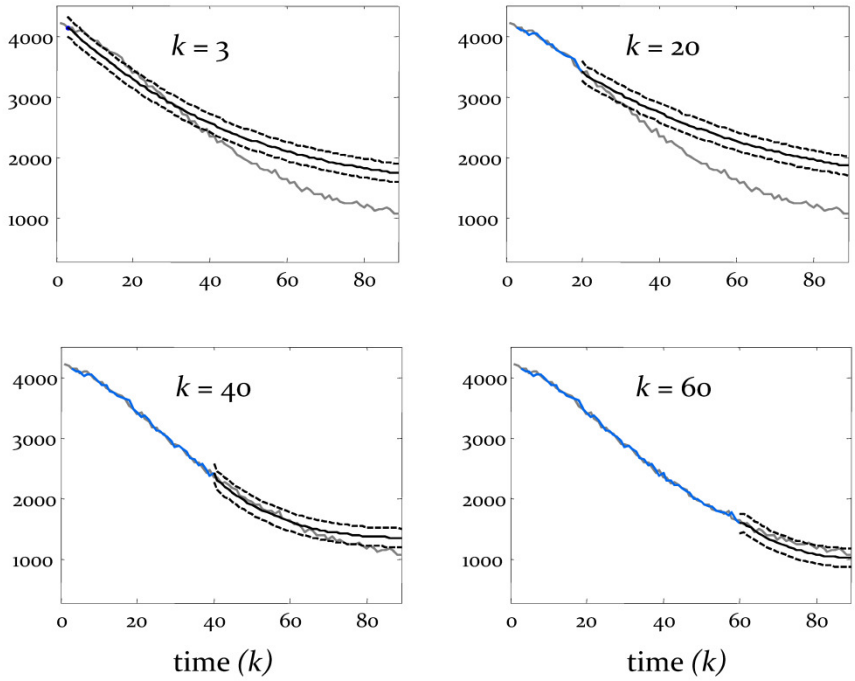


Figure 3-11 Kalman filter estimates and predictions of future process output.

Figure 3-11 shows the observed process outputs (grey), Kalman filter estimates (blue) and predictions of future process output (black) with 95% CI (dotted). The figure illustrates the adaptive nature of the Kalman filter. At k = 3 a bias is seen between the predicted and observed system outputs, likely due to an inaccurate estimate for

the batch boundary conditions $\mathbf{x}_0$. After observing the next data points and correcting the states correspondingly the bias is removed, and there is a good correspondence between the predicted and the observed system outputs. One would of course not have access to the future outputs in a monitoring situation, but the comparison is nevertheless very valuable as a validation tool for the models at hand. Figure 3-11 also indicates why the term Kalman *filtering* is used; the noisy measurements are passed through a filter whereby the noise is reduced, the resulting process output then appears as a smoother trajectory that corresponds well to the trajectory that one would intuitively expect based on chemical insight in the absence of noise. This illustrates that the Kalman filter essentially is a statistically optimal compromise between the observed trajectory and the trajectory predicted by a model.

Paper III also compared methods for end-point prediction in batch modelling. The average of the PARAFAC scores for the last five time points was in this case used as end-point estimate. Figure 3-12 shows control charts for end-point prediction of a NOC batch. The Nomikos-MacGregor PLS method (section 3.1.6) is compared to the Kalman filter with **Q**/**R** estimated with the Mehra approach (Equation 3-31) and a tuned version of **Q**/**R.**

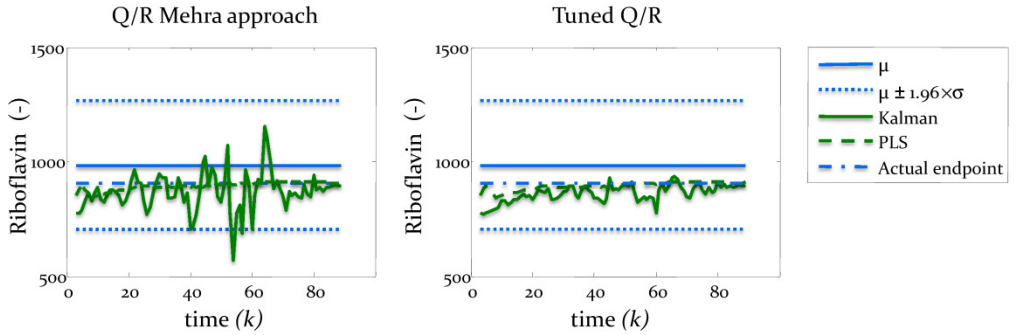


Figure 3-12 Control charts for end-point prediction. Nomikos-MacGregor PLS end-point method is compared to Kalman filter estimates with Q/R estimated using the Mehra approach (Left) and a tuned version (Right).

It is noticed that both the PLS and the Kalman predictions are close to the actual end-point already from the beginning. The figure also illustrates that fairly noisy (or jerky) end-point predictions are obtained by the Kalman method for the Mehra estimate of **Q**/**R**. It is known that the Mehra approach gives a sub-optimal estimate

of **Q**, and it has been shown that the estimates are pessimistic [Odelson *et al.*, 2006]. It was speculated that the filter was able to suppress this noise by tuning the ratio between **Q** and **R**. Warning limits and other parameters are normally tuned in MSPC (section 3.1.6) based on validation performance indicators of the control charts ones in use. The right panel in Figure 3-11 gives an illustration of this where a twenty times reduction of **Q** is used. This comes however with the price of trusting the model more and therefore with the risk of delays in capturing deviating behaviour in a statistical monitoring situation.

It should be kept in mind that though both the state space/Kalman method and the PLS method have the aim of modelling batch data, the objective of the two methods are quite different. Where the first method has the goal of capturing and modelling the dynamics via the system matrices [van Overschee & De Moor, 1996], the unfolding and autoscaling in the latter method has according to Nomikos & MacGregor the objective of *"removing the main non-linear and dynamic components in the data"* [Nomikos & MacGregor, 1995b]. This has the consequence that different questions can be answered with the two methods (Figure 3-13).

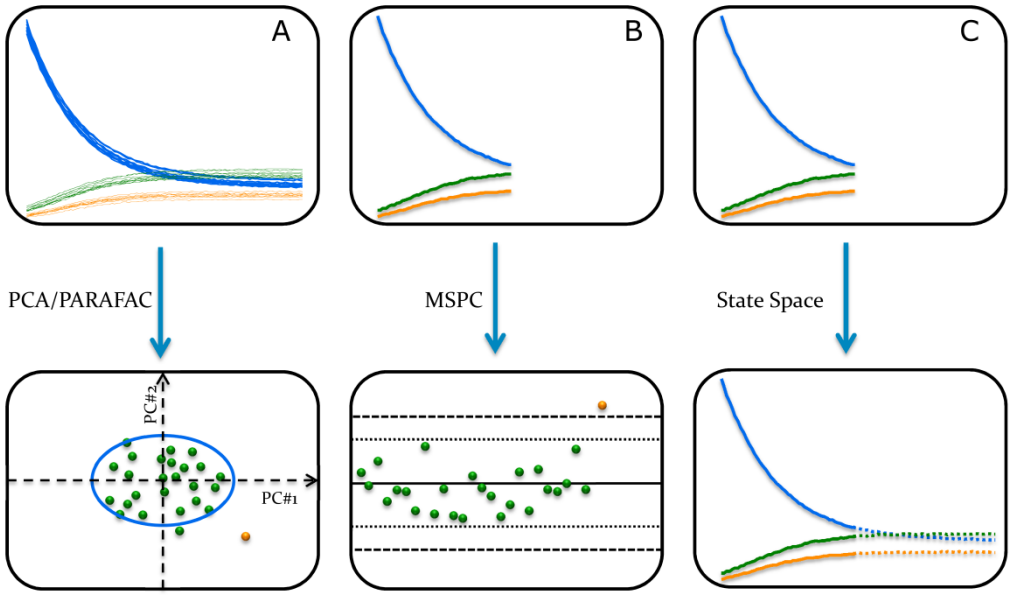


Figure 3-13 Different questions may be answered with the different available modelling tools

As it was outlined in the introduction different questions may be asked during real-time surveillance of processes. One common question to ask is whether the process is on track or not ("*where is the process now?*"). Control charts of the observed outputs (Figure 3-13B) may help in answering this question but do not necessarily reflect if the dynamics are behaving according to NOC, while control charts of the states and dynamics can. **Paper II** presented such control charts. Another question that could be of great interest is what the end-point quality of the batch is going to be ("*where is the process going?*"). The PLS method is especially suited for predicting the end-point quality, but does not include predictions on how the batch will evolve. As Figure 3-13C shows, this is the aim of the state space/Kalman filter model. Via a combination of model predictions and observations the batch trajectory is predicted for the remaining time steps. A final question that may be of interest is whether the initial conditions for the process were within the specifications ("*where did the process start from?*"). PLS – or any other regression algorithm - may be used for predicting the initial conditions, but where the state space model directly gives initial condition estimates, a separate PLS regression model would be required for predictions of the initial conditions because model inversion is not obvious. **Paper II** presented dynamical control charts of the initial conditions estimates based on state space models, directly providing an overview of whether the batch started at NOC. In the case where post-batch analysis is wanted, e.g. to compare long-term performance of a batch-wise production environment, this may be based on either PCA or PARAFAC (Figure 3-13A) of the recorded batch data, dependent on the data structure. **Paper IV** presented how such post-batch charting of the relevant process parameters – the kinetic constants - may be performed. The method thereby made it possible to assess if the batch had followed NOC or not based on PARAFAC, corresponding to the question *"Did my batch do okay?"*.

3.3 System Identification

The state space model was introduced previously. The question of how to identify the matrices in the state space equations (Equation 3-23 and Equation 3-24) remained however open. Overall there can be three levels of process knowledge:

1. Physics and chemistry governing the process is well known (e.g. we know that the process can be described by a first order differential equation with known coefficients)

2. Physics and chemistry governing the process is somewhat known (e.g. we know that the process can be described by a first order differential equation, but we don't know the coefficients)

3. Only limited knowledge is available on the physics and chemistry governing the process (e.g. we can observe the input and the outputs of the system, but don't know how they are connected).

In the case of level 1 knowledge, so-called white box modelling can be applied. White box modelling consists of process models based on mechanistic or first principles. Unfortunately this is often not the case, and only limited knowledge on the physics and chemistry is typically available (level 3). This leads to the need of System Identification (SI) tools; so-called black box methods that allow modelling of processes (or systems) based only on observations of system inputs and outputs. An appealing feature of many of the SI-tools is that they allow inclusion of *a priori* knowledge on the system, i.e. the intermediate knowledge level 2 where some process knowledge is available. This is (for obvious reasons) known as grey box modelling [Roffel & Betlem, 2006].

It was indicated in the introduction to state space models that the estimation algorithms could be specified using the projection based subspace methods, they may however also be specified by means of the iterative predictor-error algorithms. The predictor error method works by iteratively changing the model coefficients in order to minimize the prediction error (hence the name) [Ljung, 1999]. A wide range of different weights, criteria and settings can be used for the PEM. It is out of scope for this thesis to cover them all, the reader is therefore referred to Ljung [1999] for more details on PEM, but some comparing notes on the subspace methods vs. PEM should however be made here (Table 3-2)

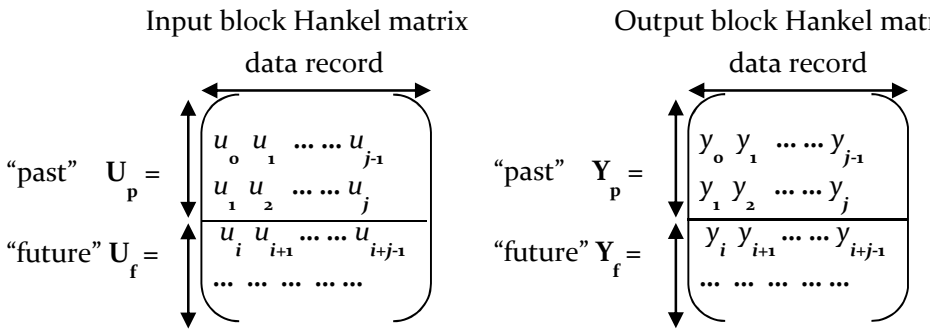
Table 3-2 Comparison of Predictor Error and Subspace Methods

	Predictor Error Methods	*Subspace Methods*
Algorithm is based on	Iterations	Projections (SVD-based)
What needs to be pre-defined?	Full parameterization needed	Size of Hankel matrix and system order
Has the smallest prediction error	Yes	No
Potential convergence problems?	Yes	No
Speed of algorithm	Slow - May require many iterations	Fast – no iterations needed

From Table 3-2 it is evident that there are advantages and drawbacks of both PEM and subspace methods. A full parameterization is needed for the PEM (i.e. we need to know/guess the model structure), this is not the case for the subspace methods, where it is required to select the size of the so-called Hankel matrix (defined below), and the order of the model - both can be determined during modelling. The lack of parameterization can however also be a drawback if grey-box modelling is wanted; in that case PEM is more applicable [Ljung, 2010]. The iterative nature of PEM makes it slower than the subspace methods and also introduces a risk of not achieving convergence. The prediction error is however intrinsically smaller for the PEM than for the subspace methods. Procedures where a subspace method is used for initial modelling followed by PEM optimization has therefore been proposed [van Overschee & De Moor, 1996; Ljung, 2010], experience from industry has however shown that the improvements gained by the PEM step is limited [Wahlberg *et al.*, 2007].

3.3.1 Subspace Methods for State Space Modelling

As Table 3-2 indicates two decisions have to be made during subspace modelling: The size of the Hankel matrices and the model order n. A Hankel matrix is symmetric and has the same elements across the off-diagonals. Written out for the input series (u_0, u_1, u_2 … u_{i+j-1}) and corresponding output series (y_0, y_1, y_2 … y_{i+j-1}) in Equation 3-23 and Equation 3-24, the Hankel matrices would thus be [van Overschee & De Moor, 1996]:



Equation 3-36

A similar Hankel matrix (effectively a row and column time-shifted data representation) can be defined for the states series ($\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$ … $\mathbf{x}_{i+j-1}$) where each entry is a vector of length n (the rank of the system) instead of a scalar. The separation between "past" and "future" data reflects how future inputs, outputs and states can be regressed on past inputs, outputs and states. The selection of the number of block rows (the "past" and "future" horizons) should be made so that i is larger than the expected system order n, while $i+j+1$ is determined by the length of the available training time series.

The input and output Hankel matrices can be combined in a block Hankel matrices **W**. The "past" block Hankel matrix $\mathbf{W_p}$ would thereby e.g. be defined as [van Overschee & De Moor, 1996]:

$$\mathbf{W_p} = \begin{pmatrix} \mathbf{U_p} \\ \mathbf{Y_p} \end{pmatrix} \qquad \textbf{Equation 3-37}$$

The chemical/physical rank of $\mathbf{W_p}$ is an estimate of the true underlying number of dynamic components (which could be called eigenfrequencies) in the system. $\mathbf{W_p}$ can therefore be used to estimate the system order n. Just as the eigenvalues of the covariance matrix can be used for decision on the chemical rank of the system (the number of PCA components), the singular values of the block Hankel matrix may be inspected to assess the number of dynamic components/ the system order. **Paper II** and **III** illustrated how this may be done. The block Hankel matrices for the observed data are closely related to the concepts of observability and controllability

of the system states [van Overschee & De Moor, 1996]. States can in general terms be said to be observable if they can be uniquely determined from the output y_k of the system. A useful system related matrix is the observability matrix $\boldsymbol{\Gamma}$, it was originally introduced by Kalman [1960b] defined as:

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^{2} \\ \dots \\ \mathbf{CA}^{i-1} \end{pmatrix} \qquad \textbf{Equation 3-38}$$

If the rank of $\boldsymbol{\Gamma}$ is equal to n (the system order or number of elements in the state vector $\mathbf{x}_k$) then the system is observable [Kalman, 1960b; van Overschee & De Moor, 1996]. Another useful system related matrix is the controllability matrix $\boldsymbol{\Delta}$. It is, as the name suggests, related to the controllability of the system. The system is controllable if it can be brought to any desired state by the input series u_k. The controllability matrix is defined as [Kalman, 1960b; van Overschee & De Moor, 1996]:

$$\boldsymbol{\Delta} = \begin{pmatrix} \mathbf{A}^{i-1}\mathbf{B} & \mathbf{A}^{i-2}\mathbf{B} & \dots & \mathbf{AB} & \mathbf{B} \end{pmatrix} \qquad \textbf{Equation 3-39}$$

The last system related matrix that needs to be defined is the lower block triangular Toeplitz matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \dots & \mathbf{0} \\ \dots & & & & \\ \mathbf{CA}^{i-2}\mathbf{B} & \mathbf{CA}^{i-3}\mathbf{B} & \mathbf{CA}^{i-4}\mathbf{B} & \dots & \mathbf{D} \end{pmatrix} \qquad \textbf{Equation 3-40}$$

It can be shown [De Moor, 1988] that the original vector/matrix computations in Equation 3-23 and Equation 3-24 can be reformulated in the following format by means of the system related matrices as defined above:

$$\mathbf{Y}_p = \mathbf{\Gamma}\,\mathbf{X}_p + \mathbf{H}\,\mathbf{U}_p \qquad \textbf{Equation 3-41}$$

$$\mathbf{Y}_f = \mathbf{\Gamma}\,\mathbf{X}_f + \mathbf{H}\,\mathbf{U}_f \qquad \textbf{Equation 3-42}$$

$$\mathbf{X}_f = \mathbf{A}\mathbf{X}_p + \mathbf{\Delta}\,\mathbf{U}_p \qquad \textbf{Equation 3-43}$$

The different subspace algorithms available essentially solve this set of equations from which the **A**, **B**, **C** and **D** matrices in Equation 3-23 and Equation 3-24 for a user defined rank *n* of the system are estimated. The term "subspace" refers to the fact that the first step in the algorithms is an oblique (or non-orthogonal) projection **O** of the "future" outputs ($\mathbf{Y}_f$) on the "past" block Hankel matrix $\mathbf{W}_p$ along the future inputs $\mathbf{U}_f$ (Figure 3-14).

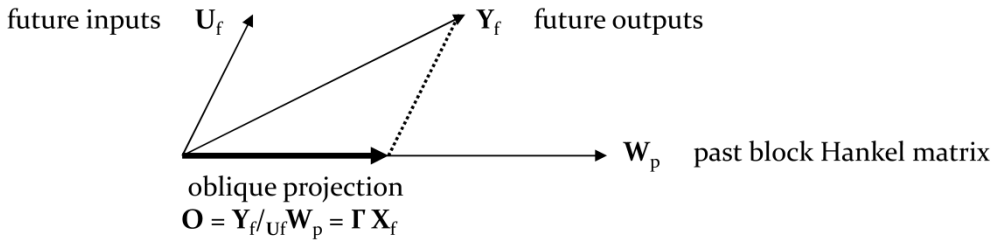


Figure 3-14 Oblique projection of "future" outputs on "past" block Hankel matrix along future inputs

SVD is then calculated for this weighted oblique projection:

$$\mathbf{G}_1\mathbf{O}\mathbf{G}_2 = \boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{V}}^{\mathbf{T}} \qquad \textbf{Equation 3-44}$$

Where $\mathbf{G}_1$ and $\mathbf{G}_2$ are weights determined by the specific algorithm (discussed below). $\boldsymbol{\mathcal{U}}$ and $\boldsymbol{\mathcal{S}}$ are then used to determine the observability matrix ($\mathbf{\Gamma}$) by

$$\mathbf{\Gamma} = \mathbf{G}_1\boldsymbol{\mathcal{U}}\boldsymbol{\mathcal{S}}^{1/2} \qquad \textbf{Equation 3-45}$$

And since the oblique projection is equal to the product of $\mathbf{\Gamma}$ and the states ($\mathbf{X}_k$), is it possible to determine the states by

$$\mathbf{X} = \mathbf{\Gamma}^{+}\mathbf{O} \qquad \textbf{Equation 3-46}$$

where the Moore-Penrose pseudo inverse of the observability matrix is used. The boundary between "past" and "present" can then be shifted one step in order to

determine the states at the next time step ($\mathbf{X}_{k+1}$), making the **A**, **B**, **C** and **D** matrices the only unknowns in the system of linear equations that can thus be solved by least squares.

Paper III showed how a deterministic time series (with input signals) could be modelled. The computations were in this case made using the **N**umerical algorithm for **S**ubspace **S**tate **S**pace **S**ystem **Id**entification (N4SID algorithm) from the Systems Identification toolbox. N4SID uses the data flow outlined above where $\mathbf{G}_1$ and $\mathbf{G}_2$ is identity [van Overschee & De Moor, 1996].

Paper II discussed and presented the case where a purely stochastic time series (without input signals) is modelled. The Canonical Variates Analysis (CVA)-based stochastic Algorithm 3 from the book "Subspace Identification for Linear Systems" by Peter van Overschee and Bart de Moor was used. $\mathbf{G}_1$ contains in this case the inverse square roots of the covariance estimate of the future outputs and $\mathbf{G}_2$ is identity. Since no input signals are available the "past" block Hankel matrix is in this case equal to the "past" outputs ($\mathbf{Y}_p$) and the algorithm then follows the same flow as in the deterministic case: Determine **O** from "future" outputs and the block Hankel matrix (="future" outputs $\mathbf{Y}_f$), determine the observability matrix ($\mathbf{\Gamma}$) from the weighted **O**, determine the states by $\mathbf{X} = \mathbf{\Gamma}^{+}\mathbf{O}$, and solve the system of linear equations by least squares. The algorithm furthermore has the additional feature to produce positive real covariance sequences, making the solutions produced by the algorithm physically/chemically meaningful. The price to pay for this is a bias in the solution [van Overschee & De Moor, 1996].

3.3.2 Designing Experiments for process modelling

Much process knowledge can be gained by inspecting and modelling historical process data. There are however also some drawbacks on only using historical data [Roffel & Betlem, 2006]:

1. The historical data is often collected under closed loop conditions (see section 4.1 for definitions of open and closed loop). Assuming the process is well controlled, this has the consequence that only limited variation is to be seen in data.

2. The process is only operating in a given range, knowledge on possible new operating points outside this range is therefore not available.

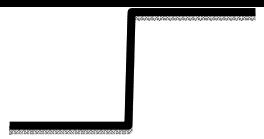
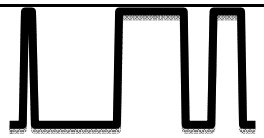
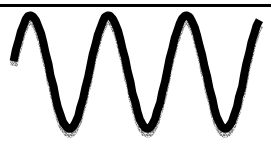
3. Under NOC is there a risk of correlation between the different inputs, this confounding makes it impossible to find the effect of each input.

This illustrates the need for designed experimental data if the full potential of the process models outlined above is to be exploited. How the experiments should be designed is however dependent on the type of model and what the aim of the modelling is.

The first question to ask is if the model should describe dynamics or statics of a process. The statics of a process may on first hand sound uninteresting, but this is far from the case! The statics of a process is what is explored when a design space investigation is made for the process [Roffel & Betlem, 2006]. As also outlined in the introduction response surface modelling [Box *et al.*, 1978] is of key interest in such cases, illustrating that Design of Experiments (DoE) is (or should be) central during modelling. This is also the case when a PLS calibration should be build, since a good model depends on a good calibration data-set, literature on mixture designs (e.g. [Smith, 2005]) could in these cases prove useful.

The dynamics of a process is related to how the system evolves from one time-step to the next. The state space models and subspace methods outlined above are therefore good tools for modelling the dynamics. In the case where such models are wanted, systematic perturbation of the process input is to be done [Roffel & Betlem, 2006]. These perturbations may be done using different patterns (Table 3-3).

Table 3-3 Different perturbation patterns and corresponding outputs

Signal name	Perturbation pattern	Notes
Step		Simple, may be used for estimating rise- and dead-times, often easy to implement in daily production
Pseudo Random Binary Sequence (PRBS)		In general a good pattern for exciting different modes in process, long run-times may however be required
Sinusoid		Good alternative to PRBS, only limited frequencies are however excited

An overall guiding principle is however that the input signal should have the largest possible frequency content [Ljung, 1999], meaning that the only the eigenfrequencies of the system that are excited by the signal can be observed in the corresponding output. The term a "rich" input signal, or a signal of "sufficient excitement" has been used for signals that fulfil this requirement [Ljung, 1999]. One can as an analogy think of fluorescence spectroscopy (section 2.2); with this method we are only able to measure the analytes that we actually excite, and if the excitation frequency does not correspond to a given analyte we will not be able to detect its presence.

The step function is a simple input change that is often used for identification purposes in industry [Wahlberg *et al.*, 2007], it may be used to find the system dead time ("if I change the input now, how long time does it take before I see the output change"), but it can also be used to find the system rise time ("when the output starts to change, how long time does it take before it is at the new equilibrium").

The Pseudo Random Binary Sequence (PRBS) can be seen as a series of step functions at random time points. It is a well-defined signal that can be designed to ensure that the mean and covariance matrix have some mathematically attractive properties (such as easy analytical inversion of the covariance matrix) [Ljung,

1999].The attractive features of the PRBS comes however also with a price to pay: In order to achieve the attractive mathematical properties the PRBS length should be equal to at least five times the major time constant of the process, plus dead time [Roffel & Betlem, 2006]. This may result in long required run-times, plus knowledge on the dead time and the major time constants are needed – which in essence is what we want to find …

Sinusoids may be a good alternative if it is not feasible to use a PRBS. It may for some systems (e.g. due to safety reasons) not be possible to change the input level in steps. Perturbing the input to follow a sinusoid may then be an alternative, a drawback is however that the signal is not necessarily as "rich" as for a PRBS [Ljung, 1999].

It may also be of interest to select the sampling frequency. A differentiation between the sampling frequency done during the experiment, and the actual data points included during modelling phase should be made. Modern process monitors may gather e.g. spectra at a very high frequency almost without any cost. Intuitively one would therefore most likely go for the highest possible sampling frequency. During the modelling phase it may however be required to either up- or down-sample the data (i.e. either resample the data or only use every 10^{th} data-point). Care should be taken if up-sampling is to be done, since this easily may introduce new false dynamic components in the data. **Paper II** showed how so-called zero-order-hold re-sampling could be done to avoid introduction of false dynamic components. Blindly using the highest possible sampling frequency during modelling is also not advisable, since it can be shown [Ljung, 1999] that too high a sampling frequency may lead to a large variance of the estimated coefficients. The optimal sampling frequency is therefore a trade-off between the two: On one hand we want to capture the dynamics of the system, but on the other hand we want a small variance of the estimated model coefficients. This also means that the term "real time" used habitually in literature by itself is meaningless and instead depends on the system dynamics; if the dynamics are slow (e.g. time constants in the range of hours) "real time" may be a sampling frequency of 1 h^{-1}. Seen from a practical viewpoint an estimate of the optimal sampling frequency can be made by making a step change in the input, and then selecting the sampling frequency so that 4-6 samples are recorded during the rise time [Ljung, 1999].

3.3.3 System Identification of an Ultrafiltration Process

During the initial phase of this PhD-study experiments were conducted on an Ultrafiltration (UF) process with the aim of performing System Identification on the process. UF is a membrane-based separation method commonly used as an up-concentration method in the dairy industry [Walstra *et al.*, 2006]. The method is based on feeding the raw material (typically skimmed milk or whey) on a membrane using a high pressure. Depending on the pore size of the membrane some larger molecules (e.g. proteins) will be retained while smaller molecules (e.g. salts or sugars) will be able to pass through the membrane [Walstra *et al.*, 2006]. The feed is hereby separated in two streams, a high concentration stream known as the retentate, and a low concentration stream known as the permeate (Figure 3-15).

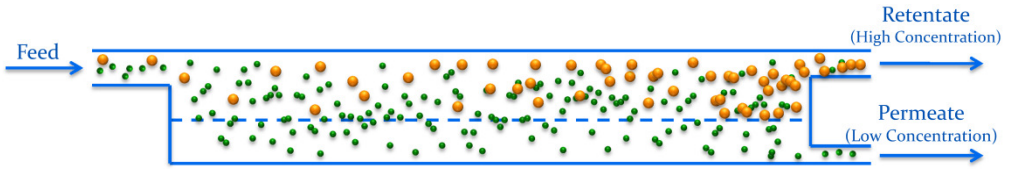


Figure 3-15 Principle in membrane filtration, the feed is separated in a high concentration retentate stream and a low concentration permeate stream (Modified after Walstra *et al.* [2006]).

Experiments were conducted using recombined skimmed milk as raw material. The initial experiments on the process showed that, from a range of candidate methods - EEM-fluorescence, IR and NIR spectroscopy - the first mentioned had the potential to serve as a tool for process monitoring. Two BioView EEM fluorescence process spectrometers were therefore used as in-line spectrometers, one on the permeate stream and one on the retentate stream (Figure 2-6). Based on the initial experiments a PRBS was designed to excite the system, changing the feed pressure (60 or 80% of maximum) and the concentration of the feed (55 or 50% dry matter) Figure 3-16.

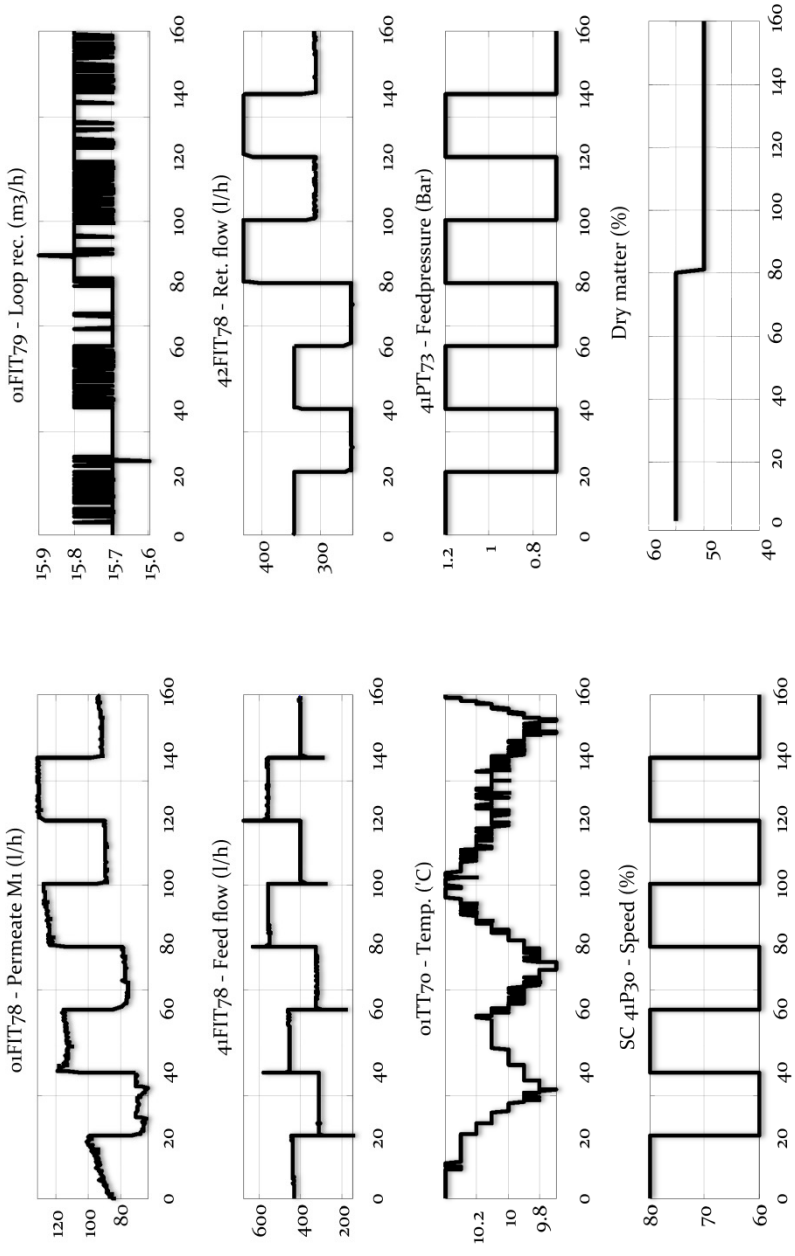


Figure 3-16 Manipulation patterns used in UF-experiments

Fluorescence EEMs were recorded each minute and PARAFAC models were fitted for the spectra. A one component model was made for the permeate stream and a two component model for the retentate stream. Inspection of loadings (not shown) revealed that the permeate component corresponded to riboflavin and the two retentate components to riboflavin and the milk proteins. The corresponding score values are given in Figure 3-17 together with the manipulation profile.

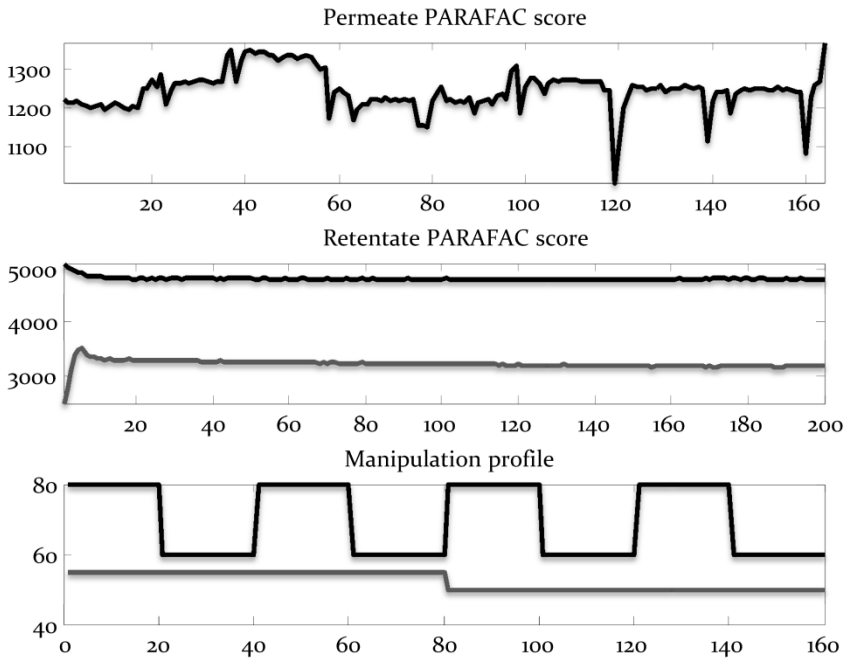


Figure 3-17 PARAFAC scores and PRBS applied in UF experiments

Figure 3-17 shows how no clear effect of the PRBS is seen in neither the retentate nor the permeate signal. The concepts of sufficient excitation, stability and observability were introduced above (section 3.3.1). The experiments conducted on the UF plant showed how the high stability of the system prevented successful identification due to lacking observability of any changing states which in turn corresponds to the fact that the input signal was not of sufficiently rich. Several other experiments were done with other step or PRBS signals as inputs though without obtaining any better results. The process or unit operation in focus for the PhD-study was for this reason (among others) changed towards the other cases covered in this thesis.

3.4 State space and other dynamic models in chemometrics – review of relevant literature

In the introduction to the state space models (section 3.2) it was indicated that the state space model, the Kalman filter and System Identification algorithms originated in the control community. This means that many papers on the three subjects may be found in dedicated control engineering journals such as *Journal of Process Control*, *Automatica, Chemical Enginerring Science, The American Institute of Chemical Engineers Journal* or *Technometrics*. A literature search, March 2012, on the three subjects in Thompson Reuters Web of Science also reveals that an impressive amount of papers are published each year on the three subjects.

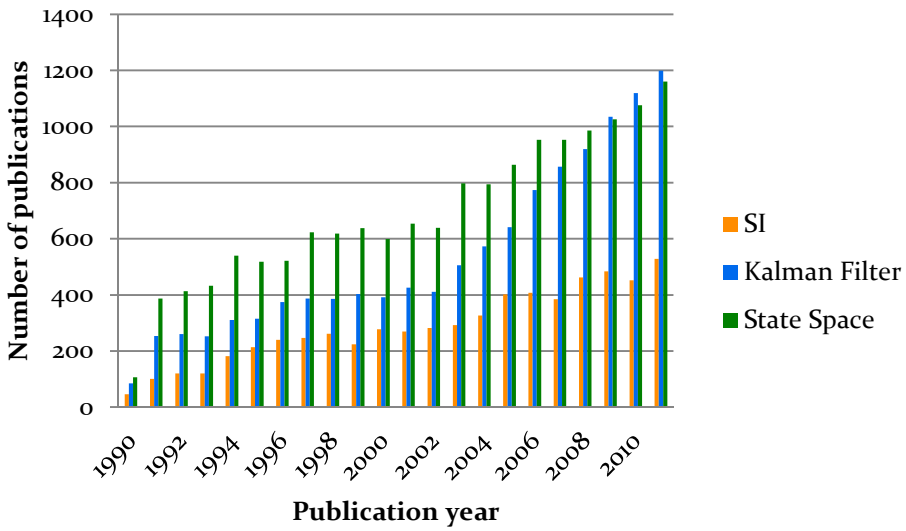


Figure 3-18 Number of publications for the search terms "System Identification" (SI), "Kalman Filter" and "State Space"

The following (non-exhaustive) literature review on state space and other dynamic models is therefore limited to the papers found in the three main chemometric journals: *Journal of Chemometrics, Chemometrics and Intelligent Laboratory Systems* and *Analytica Chimica Acta*.

3.4.1 State Space Models in Chemometrics

The requirement for dynamic models in process monitoring and control was recognized from early on, as Callis *et al.* in 1987 formulated the need for real-time process control based on multivariate statistics [Callis *et al.*, 1987]. The idea of applying state space models in chemometrics is therefore not new, albeit not widely spread. A series of papers on state space modelling were published in the chemometric literature in the late 1990's and early 2000's, but the research area has received less attention during the last 10 years. A chemometric paper on state space models was published in 1997 by Negiz and Çinar [Negiz & Çinar, 1997]. It was shown that PLS can be used to fit state space equations, but it was at the same time shown that modifications of the PLS algorithm were necessary to give useful results. A method based on CVA proved to give the best outcome. Hartnett and co-workers published two different papers in the end of the 1990'ies. In the first of the two papers were genetic algorithms in combination with PCR used to do dynamic inferential estimation of process variables [Hartnett *et al.*, 1998]. The measurement equation of an underlying state space model is used, but focus is not on the state space model itself. This is done in the later paper [Hartnett *et al.*, 1999] where it is shown how a non-linear multivariable production plant can be modelled using a combination of PCA and state space modelling. Wise showed [1991] how observable states could be estimated based on SVD/PCA. This observation is the idea used by Hartnett and co-workers in 1999. PCA is done on the process outputs, the scores obtained are then used as states and the loadings used as **C**-matrix (no input is used in the measurement equation in this paper [Hartnett *et al.*, 1999]). The system equation is subsequently identified by concatenating state- and input-matrices and regressing the concatenated matrix on future states by means of PCR. The PCA based state space model was compared to an analytical state space model. Both had good performance in approximating the non-linear system. The authors note that no prior decision on model order needs to be taken when using this PCA based approach. This is correct, but a decision on the number of principal components in both the PCA and the PCR step needs to be made. Ergon [1998] uses state space equations, PCR and PLS to derive relations that can be used to predict one output variable from another. An example of state space modelling is also given, but this is via PEM. Dynamic system PCR and PLS solutions for output predictions are also presented, again based on a PEM state space model. In a later paper by Ergon and Halstensen [2000] these results are elaborated on for a system with low-sampling-rate reference measurements - a combination of PCA and PEM is utilized to produce

predictions of the reference measurements with a better performance as compared to PLS. Shi and MacGregor give a complete review [Shi & MacGregor, 2000] of different subspace methods and compare them to different latent variable techniques (PCA, PLS and PCR). They come to two overall conclusions: 1) for process monitoring ("Is my process on-track?") latent variable methods are to be preferred, but for process identification ("What are the process dynamics?" or "Where is my process heading?") dedicated subspace identification methods are preferential; 2) CVA and the N4SID algorithm have the best performances of the subspace methods they tested. In a more recent paper Pan *et al.* [2004] showed, contrary to the first conclusion by Shi and MacGregor, that better monitoring performances could be archived if a state space/subspace method was applied compared to PCA-based monitoring. The authors use PCA to reduce the dimensionality of the output, followed by fitting state space models by means of N4SID. A Kalman filter is subsequently used. A large part of the advantage from this model is according to the authors a result of the implementation of the Kalman filter. In most recent paper in the chemometric literature on state space models Odiowei and Cao [2010] presents how a combination of Independent Component Analysis (ICA), CVA and state space models allowed dynamic process monitoring of the non-linear benchmark The Tennessee Eastman Process.

3.4.2 Other dynamic models

State space models are far from the only type of dynamic models encountered in the chemometric literature. A wide array of dedicated dynamic models and modifications of well-known static models may be found. An example of the latter is the dynamic version of PCA, Dynamic Principal Components Analysis (DPCA) [Ku *et al.*, 1995]. The principle in DPCA is simple: Form a Hankel matrix of $\mathbf{X}$ (Equation 3-36) and do standard PCA on this Hankel matrix. Ku *et al.* suggest inspecting the auto- and cross-correlations of the PCA scores to assess the number of PC's and the correct size of the Hankel matrix. Russel *et al.* compared fault detection capability of DPCA to PCA on The Tennessee Eastman Process. These authors applied however an automated selection criteria (based on the Akaike Information Criteria (AIC) [Brockwell & Davis, 2002]) for selection of model complexity and Hankel matrix size. Russel *et al.* found that PCA and DPCA had similar performance in terms of sensitivity, promptness and robustness.

In the case where both process inputs and outputs are included in the Hankel matrix, DPCA may be seen as an autoregressive model with exogenous input (ARX), where the current output is a function of the past inputs and the past outputs. DPCA is however not the only method for estimating ARX models, the iterative PEM method (section 3.2) may be used for estimating both state space, ARX and Finite Impulse Response (FIR) models [Ljung, 1999]. FIR models are relevant when the current output is a function of the past values of the inputs. Wise & Kowalski [1995], showed how a SVD-based method called Continuum Regression (CR) [Wise & Ricker, 1993] in the case of correlated inputs produced better estimates of the true system output.

4 Process Control

The introduction presented how the concept of QbD was related to process control and Model Predictive Control (MPC). This section will elaborate further on these subjects and how knowledge on process control may be used in the pursuit of QbD.

Figure 1-5 illustrated how a design space may be constructed based on response surface modelling. It was shown how the design space and control target should be based on considerations off product safety and yield. **Paper III** and **IV** present a model system of riboflavin breakdown; base was in combination with light used to induce the breakdown reaction. All reagents were added to the reactor simultaneously and pH, temperature and fluorescence measurements were done over time. The process was thus not run as a fed batch - if this was wanted the base could have been fed to the process as it was consumed by the reaction. In this case a scenario as outlined in Figure 4-1 may have been experienced, where the speed of conversion is given as the full line, and the volume of base dosed pr. minute is given as the dotted line.

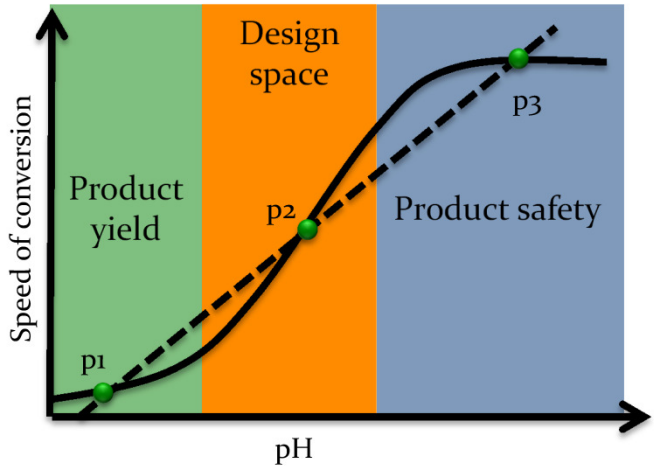


Figure 4-1 Design space based on considerations of product safety and yield (Based on [Roffel & Betlem, 2006] and [ICH, 2009]).

Figure 4-1 illustrates how three equilibrium-points (p1, p2 and p3) exist between the amount of base added pr. min. and the reaction rate. P1 is located at a low flow rate which in turn corresponds to a low pH and thereby a low reaction rate, in a real

world process setting this may not fulfill the existing requirements for a subsequent down-stream step. P3 is located at a correspondingly high pH which may be unwanted seen from an operator/product safety point of view. P2 is thereby the only desirable operation point. In a QbD setting a range surrounding this point would therefore be selected as design space. Unfortunately it is also evident from Figure 4-1 that monitoring and control of the process is needed since p2 is an unstable equilibrium point (a small deviation from p2 could easily lead to the process running to one of the other equilibrium point). A control-loop may therefore with advantage be introduced in the process.

4.1 Control loops

The concepts and notation of input (u_k), output (y_k) and disturbance/noise (w_k/v_k) were introduced in Figure 3-10[4] when state space models were presented. Figure 4-2 illustrates how this notation may be used to introduce a control loop on a generic system as well.

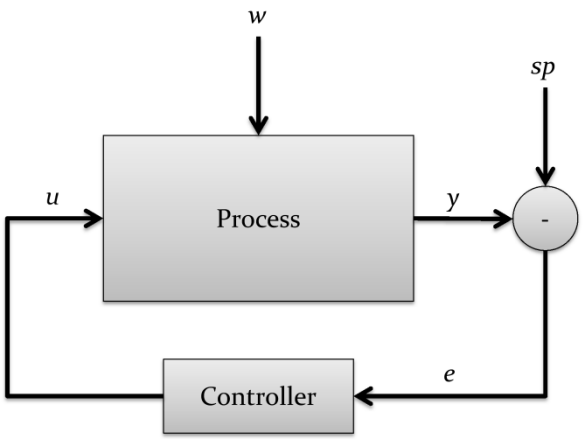


Figure 4-2 Feedback control loop

There are in general two different approaches to control of a process: open or closed loop control. Open loop control is essentially letting the process run without control action, i.e. no feedback is given to the control signal. Open loop control therefore corresponds to selecting an input that is known to result in the desired output and

[4] For ease of notation scalar in- and outputs are used in this section.

letting the process run at these settings [Luyben, 1996]. It would in the fed-batch example from above correspond to selecting a given flow rate of base near p2 and not changing this selected flow rate during processing. Many processes in industry are open loop stabile [Luyben, 1996], meaning that in spite of the apparent over-simplistic appearance of open loop control, it will often do the trick. Closed loop control is however needed in the example above since disturbances working on the system would easily tip the process to p1 or p3. Closed loop control includes a feed-back as illustrated in Figure 4-2. The process output is compared to a set-point (*sp*) and the difference (*e*) is via a controller fed back to the system as a change in the process input [Stephanopoulos, 1984]. Feed-back control is however, as also stated in the introduction, always "post-problem", a deviation from the set-point is needed before any corrective action is taken, resulting in production of sub-standard products. This observation leads to the concept of feed-forward control (Figure 4-3).

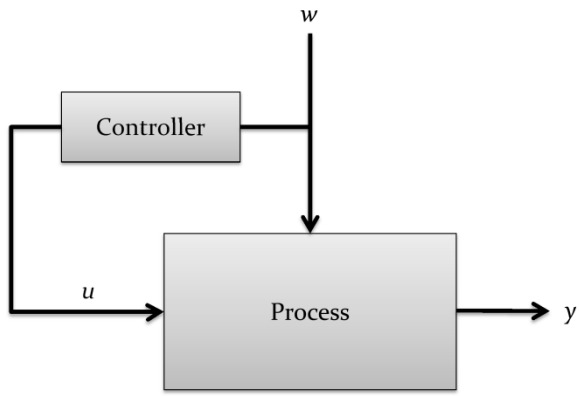


Figure 4-3 Feed-forward control of process

The principle behind feed forward control is based on the fact that some of the disturbances working on the system may be observable. Once a disturbance entering the system is detected the input variables are manipulated so that the system output is kept constant [Luyben, 1996]. Feed forward control is thereby closely related to the core idea of Quality by Design: If I know how to manipulate the input variables (I have a good process knowledge, i.e. a model that explains the connection between input and output), and I am able to detect the disturbances, than I can ensure the quality of my product by designing the control action correctly. Feed-forward control is however most often used for systems with slow dynamics [Roffel & Betlem, 2006]. The milk coagulation system studied in **Paper II** has fairly slow dynamics; feed forward control of this system should therefore be feasible. It is for

instance known that the coagulation-reaction is temperature dependent [Fox & McSweeney, 1998]. In order to control the temperature the coagulation in industry is done in jacket-heated vessels [Walstra *et al.*, 2006]. Feed-forward control could for such a vessel e.g. be implemented by changing the temperature based on raw material measurements to circumvent any changes in raw material composition, the impact of the disturbance acting on the process would hereby be limited.

It is however not always possible to directly measure the output or process variable that we are interesting in controlling. It may be the case that the output can be modelled as a (linear) combination of some easily measurable variables. This is known as inferential control [Stephanopoulos, 1984], and is recognizably closely related to the combination of chemometrics and spectroscopy. **Paper II** may thus be seen as an example of inferential control. The stage of coagulation is not easily measured directly, but NIR spectroscopy may serve as an indirect measure of the coagulation. If a control-loop is wanted for the coagulation process, this may therefore be based on the PCA-scores from the NIR-spectra. It may also be the case that the process variable we want to control is sampled at a low frequency, in these cases the Kalman or state space approach from **Paper II** and **III** could be applied to estimate the process variable between measurements.

4.2 Selecting the control input signal (PID and Model Predictive Control)

Feed-back is the most commonly applied control-loop in industry, mainly in the form of Proportional-Integral (PI) or Proportional-Integral-Derivative (PID) controllers [Roffel & Betlem, 2006]. The controller consists (as the name suggests) of two or three components:

- The proportional term: Ensures that the input change is proportional to the change in error, i.e. a small error change results in a small input change, and a larger error change results in a larger input change. It is thereby a function of the *current* error. Controllers consisting of the P-term alone may however result in an off-set since no change in the error results in no change in the input [Stephanopoulos, 1984].

- The integral term: An integral term in the controller will accumulate the previous errors and ensure that it asymptotically approaches zero. It is thereby a function of the *past* errors. It may however result in overshoot, meaning that the controlled output or process variable may exceed the set-point [Luyben, 1996].

- The derivative term: Include the current change of the error in the controller. The term is thus a very crude estimate of the future errors (i.e. a positive derivative means an increasing error-term) [Luyben, 1996]. It may therefore be used to limit the overshoot produced by an I-term. PI control is nevertheless more commonly applied in industry than PID control, since the D-term also may induce oscillation of the system output, especially in the case when measurement noise is present [Roffel & Betlem, 2006].

The general PID-controller equation may be written as [Roffel & Betlem, 2006]:

$$u(k) = \bar{u} + G\left[e(k) + \frac{1}{\tau_i}\right]\int e(k)dk + \tau_d \frac{de(k)}{dk} \quad \textbf{Equation 4-1}$$

Where $u(k)$ is the controller output at time-point k, $\bar{u}$ the steady-state controller output, $e(k)$ the error to time k, G the controller gain, τ_i the controller integral time constant and τ_d the controller derivative time constant. Different methods exist for tuning the PID controller (finding the gain and time constants) such as the Ziegler-

Nicholos [Ziegler & Nicholos, 1942] or the Cohen-Coon [Cohen & Coon, 1953] methods. It is out of scope to cover these methods within this thesis, the reader is therefore referred to the original papers or other references (e.g. Roffel & Betlem [2006]) for further discussion on the method-principles and the pros and cons of the different tuning methods.

In the case of feed-forward control, no general closed form equations are readily available in the time-domain. Specialized feed-forward control-laws may be obtained based on Laplace transforms of the specific process-model. More details on construction of feed-forward control-laws are available in Stephanopoulos [1984].

In section 3.2.1 it was shown how the Kalman filter in combination with the SI-tools allowed dynamic models of systems. As it was shown in **Paper II** and **III** these models may be used to find deviations from process targets ahead in time. Model Predictive Control (MPC) uses this predicted deviation to find the optimal input [Rawlings, 2000] to minimize the future errors. The input is determined by minimizing a cost function based on process inputs and states. The overall MPC procedure may thus be summarized as [Findeisen *et al.*, 2007]:

1) Obtain estimates of current system-state
2) Estimate optimal input by minimizing cost-function $J(\mathbf{x}(t),\mathbf{u}(t))$
3) Implement input until next sampling step
4) Return to 1)

The cost function is commonly subject to constrains, meaning that fairly complex functions are obtained where numerical solutions are needed. The reader is referred to dedicated literature for details on the different MPC algorithms, constrains and cost functions (e.g. Rawlings [2000] or Findeisen *et al.* [2007]).

Paper II showed how subspace methods allowed on-line system identification of the milk coagulation process. This form of MPC is also known as adaptive control; it may be useful when changing process conditions are expected from one batch to the next [Luyben, 1996]. It could for instance in the milk coagulation example be useful when milk from different farmers with different coagulation properties is used in the process.

4.3 Stability of systems

In the introduction to state space models (Section 3.2) it was indicated that the eigenvalues of the state space system matrix (the **A**-matrix) could be used to access the stability of the system. This section will, among other things, elaborate more on the subject of stability. Figure 4-4 illustrates how the position of an eigenvalue in the complex plane ($\mathbb{C}$) affects the system behavior.

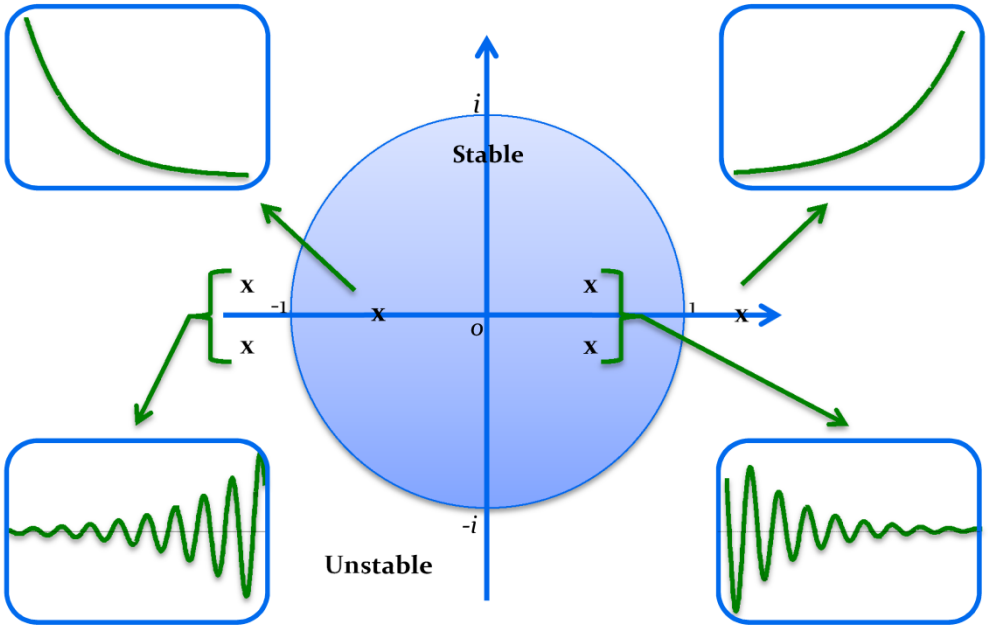


Figure 4-4 Eigenvalues of the state space system matrix (A-matrix) may be used to access the system stability (based on Luyben [1996], Roffel & Betlem [2006] and Simon [2006])

Different types of stability may be found in linear system theory. In this thesis the following definition of stability will be used [Simon, 2006]:

A system is stabile if and only if

$$\lim_{k\to\infty} \mathbf{x}_k = 0 \qquad \textbf{Equation 4-2}$$

for all bounded initial states ($\mathbf{x}_0$).

It can be shown [Simon, 2006] that the system will be stabile if the eigenvalue(s) lie(s) inside the unit-circle. It can also be shown [Roffel & Betlem, 2006] that in the case where a complex pair is found as one or more of the eigenvalues the system will oscillate. The frequency will depend on the imaginary part of the pair and the damping of the real part of the pair [Roffel & Betlem, 2006]. This has the consequence that a complex pair inside the unit-circle will result in an oscillating but stabile system, while a complex pair outside the unit-circle will result in a chaotic oscillation meaning that the system is unstable. The concepts of observability and controllability were introduced in section 3.3.1. A system was defined as controllable if it could be brought to any desired state by the input. Figure 4-1 illustrated a system where operation at p2 was unstable. Correct implementation of a control-loop should however make the system controllable, this illustrates that operation of unstable systems may still be feasible. The concepts of detectability and stabilizability are closely connected to observability and controllability, as a less strict form of the two. An unobservable system may thus be defined as detectable if the unobservable modes are stable, and an uncontrollable system be stabilizable if the uncontrollable modes are stabile (i.e. the eigenvalues are inside the unit circle) [Simon, 2006].

A case study on SI on an ultrafiltration unit operation was presented in section 3.3.3. The concepts of observability, stability and sufficient excitation came to play a major role in this study. The reader is therefore referred to this section for further details on the practical implications of these concepts.

5 Conclusions and future perspectives

The purpose of this PhD project was to show how modern process sensors based on spectroscopy in combination with chemometrics and dynamic models allowed real-time monitoring of processes.

Paper I investigated how three-way calibrations for EEM-fluorescence spectroscopy could be transferred. The study showed that it was possible to develop simple, intuitive transfer methods for three-way EEM fluorescence calibration. An uncomplicated local linear method was demonstrated to be the most favourable of the new methods. When the two- and three-way calibration methods were compared, the three-way method showed slightly lower prediction errors both for calibration and re-calibration, essentially underlining the fact that three-way models are required for three-way data. The new transfer methods were compared to the classical methods found in literature (Direct and Piecewise Direct Standardization on unfolded data). Similar results were obtained for the new and the classical methods. It was additionally shown that though good transfer models could be found for the PARAFAC models with as few as four transfer samples, the results were highly dependent on the selection of the transfer set. When recalibration and calibration transfer was compared for the fluorescence data set used, calibration transfer was better with lower prediction errors and fewer samples needed. The paper thereby illustrated how three-way EEM fluorescence calibration made in an off-line setting (i.e. in the laboratory) with ease could be transferred to an on-line application.

Paper II introduced the state space model and showed how subspace methods allowed state space modelling without *a-priori* assumptions on model shape/form. In this sense the subspace methods enabled the modelling to be data rather than hypothesis driven, essentially allowing modelling of the process without the requirement of any prior knowledge on the underlying physics or chemistry. The paper presented how a non-linear milk coagulation process could be approximated by linear state space models. The models were estimated recursively i.e. the models were re-estimated in real-time as the measurements became available. This adaptable approach to modelling thereby showed that state space models were potential tools for process monitoring. Where conventional MSPC control charts

reflects the process in a static manner, the control charts proposed in **Paper II** reflected the dynamic behaviour of the process.

Paper III elaborated further on the conclusions from **Paper II.** In this paper a combination of state space models, subspace methods and Kalman filters were shown to have the potential of a versatile tool in batch process modelling and monitoring. A model system of riboflavin breakdown was presented as an example of a batch process. It was shown how the combination of EEM-fluorescence spectroscopy and PARAFAC modelling allowed direct surveillance of the on-going chemistry in the process. The proposed combination of state space models, subspace methods and Kalman filters were able to capture and model the dynamics of the batch process. Where recursive modelling allowed adaptive monitoring in **Paper II**, the introduction of the Kalman filter in **Paper III** had the same objective, however with the extra advantage of improved predictions of future process variable trajectories including 95% confidence intervals of the variables. The method was thus shown to be adaptable to new non-NOC conditions and allowed for dynamic control charting of initial condition estimates and current system-states. For end-point prediction a dedicated method based on Partial Least Squares was found to produce slightly better predictions.

Paper IV presented further studies on the model system introduced in **Paper III.** The paper illustrated what is also known as so-called grey box modelling: Modelling in the case where the physics and chemistry governing the process is known to a limited extend. In **Paper IV** it was shown how the *a-priori* knowledge on the reaction kinetics governing the process could be implemented during PARAFAC modelling, hereby allowing post-batch charting of the relevant process parameters – the kinetic constants.

In conclusion this study has presented how the combination of spectroscopy, chemometrics and dynamic models may be used in process monitoring of batch processes. It was shown how different statistical/chemometric models made it possible to answer different types of questions. The post-batch approach presented in **Paper IV** made it possible to assess if the batch had followed NOC or not based on PARAFAC, corresponding to the question *"Did my batch do okay?"*. A classical PLS based MSPC method for end-point prediction was used in **Paper III**. This method provided good end-point predictions and was thus well suited for answering the question of *"Where is the process going?"*. The method was however not able to

provide details on the trajectory towards the end-point. The state space/Kalman method presented in **Paper II** and **III** was able to provide such trajectory estimates, and the method also provided the option of estimating initial conditions and current process states. The method was thus able to answer all three questions: *"Where is the process now?"*, *"Where did the process come from?"* and *"Where is the process going?"*. The possibility of predicting future process characteristics and variable trajectories opens up for the option of model predictive control which in turn may bring the goal of Quality by Design closer to reality.

Future research in line with the present results could therefore include application of the proposed combination of subspace methods, state space models and Kalman filters on a real-world process, such as a fermentor where the combination of EEM-fluorescence spectroscopy and PARAFAC models would allow direct monitoring of the microbial growth conditions. Real world application of the proposed methods would however require some adaptation; a more robust version of the Kalman filter (known as the H_∞ filter [Simon, 2006]) might in such cases be needed. A full blown design space investigation including a MPC-identification would in this a case also be very interesting. From a modelling point of view, a comparative study between the state space models and the other dynamic models (e.g. DPCA, FIR and ARX) would also be of interest. And finally would expanded comparisons on grey-box modelling most likely show interesting results as well.

6 References

Andersson, M. (2009): A comparison of nine PLS1 algorithms. *Journal of Chemometrics*. Vol. 23, no. 9-10, pp. 518-529.

Anscombe, F.J. (1973): Graphs in Statistical Analysis. *American Statistician*. Vol. 27, no. 1, pp. 17-21.

Beebe, K.R., W.W. Blaser, R.A. Bredeweg, J.P. Chauvel, R.S. Harner, M. Lapack, A. Leugers, D.P. Martin, L.G. Wright & E.D. Yalvac (1993): Process Analytical-Chemistry. *Analytical Chemistry*. Vol. 65, no. 12, p. R199-R216.

Blaser, W.W., R.A. Bredeweg, R.S. Harner, M.A. Lapack, A. Leugers, D.P. Martin, R.J. Pell, J. Workman & L.G. Wright (1995): Process Analytical-Chemistry. *Analytical Chemistry*. Vol. 67, no. 12, p. R47-R70.

Box, G.E.P., W.G. Hunter & J.S. Hunter (1978): *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building.* John Wiley & Sons, Hoboken, NJ, USA

Bro, R. (1997): PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*. Vol. 38, no. 2, pp. 149-171.

Bro, R. (1998): *Multi-way Analysis in the Food Industry,* Doctoral Thesis. University of Amsterdam, The Netherlands.

Bro, R. (1996): Multiway calibration. Multilinear PLS. *Journal of Chemometrics*. Vol. 10, no. 1, pp. 47-61.

Bro, R., A.K. Smilde & S. de Jong (2001): On the difference between low-rank and subspace approximation: improved model for multi-linear PLS regression. *Chemometrics and Intelligent Laboratory Systems*. Vol. 58, no. 1, pp. 3-13.

Bro, R. & M. Vidal (2011): EEMizer: Automated modeling of fluorescence EEM data. *Chemometrics and Intelligent Laboratory Systems*. Vol. 106, no. 1, pp. 86-92.

Brockwell, P.J. & R.A. Davis (2002): *Introduction to Time Series and Forecasting.* 2nd. ed. Springer Verlag, New York, New York, USA

Callis, J.B., D.L. Illman & B.R. Kowalski (1987): Process analytical chemistry. *Analytical Chemistry*. Vol. 59, no. 9, pp. 624A-637A.

Camo (2004): *The Unscrambler X Appendices: Method References.*

Carroll, J.D. & J. Chang (1970): Analysis of individual differences in multidimensional scaling via an *N*-way generalization of 'Eckhart-Young' decomposition. *Psychometrika*. Vol. 35, no. 3, pp. 283-319.

Christensen, J. (2005): *Autofluorescence of Intact Food - An Exploratory Multi-way Study,* PhD-thesis. Quality and Technology, Department of Food Science, The Royal Veterinary and Agricultural University, Denmark.

Cohen, G.H. & G.A. Coon (1953): Theoretical considerations of retarded control. *Transactions of the ASME*. Vol. 75, pp. 827-834.

Conlin, A.K., E.B. Martin & A.J. Morris (2000): Confidence limits for contribution plots. *Journal of Chemometrics*. Vol. 14, no. 5-6, pp. 725-736.

Dahm, D.J. & K.D. Dahm (2001): The Physics of Near-Infrared Scattering. In: P.C. Williams & K.H. Norris (eds.): *Near Infrared Technology in the Agricultural and Food Industries*. 2nd. ed. American Association of Cereal Chemists, Inc, St. Paul, Minnesota, USA, pp. 1-17.

de Jong, S. (1993): Simpls - An Alternative Approach to Partial Least-Squares Regression. *Chemometrics and Intelligent Laboratory Systems*. Vol. 18, no. 3, pp. 251-263.

De Moor, B. (1988): *Mathematical concepts and techniques for modeling of static and dynamic systems,* PhD-thesis. Department of Electrical Engineering, Katholieke Universiteit Leuven, Belgium.

Dickens, J.E. (2010): Fluorescent Sensing and Process Analytical Applications. In: K.A. Bakeev (ed.): *Process Analytical Technology*. 2nd. ed. John Wiley & Sons, Chichester, U.K., pp. 337-352.

Dolezalek, H. (2012): Classification of Electromagnetic Radiation. In: W.M. Haynes & D.R. Lide (eds.): *CRC Handbook of Chemistry and Physics*. 92nd - Internet Version. ed. CRC Press, Boca Raton, Florida, USA , pp. 233-234.

Dufour, E. (2009): Principles of Infrared Spectroscopy. In: D.-W. Sun (ed.): *Infrared Spectroscopy for Food Quality Analysis and Control*. Academic Press, Burlington, Massachusette, USA, pp. 1-28.

Eigenvector Research (2011): *Eigenvector Documentation Wiki*. Available at the Internet: http://wiki.eigenvector.com/index.php?title=Pca. [cited January 9, 2012].

Ergon, R. (1998): Dynamic system multivariate calibration. *Chemometrics and Intelligent Laboratory Systems*. Vol. 44, no. 1-2, pp. 135-146.

Ergon, R. & M. Halstensen (2000): Dynamic system multivariate calibration with low-sampling-rate y data. *Journal of Chemometrics*. Vol. 14, no. 5-6, pp. 617-628.

Esbensen, K. & P. Geladi (1990): The start and early history of chemometrics: Selected interviews. Part 2. *Journal of Chemometrics*. Vol. 4, no. 6, pp. 389-412.

Findeisen, R., T. Raff & F. Allgower (2007): Sampled-data nonlinear model predictive control for constrained continuous time systems In: S. Tarbouriech, G. Garcia & A.H. Glattfelder (eds.): *Advanced Strategies in Control Systems with Input and Output Constraints*, Springer Verlag, Berlin, Germany pp. 207-235

Folkenberg, J.R., S.M. Nikolajsen, H.V. Juhl, H. Larsen, D.K. Pedersen, H.V. Andersen & A. Frandsen (2008): Fourier Transform Infrared Spectroscopy as a Tool for Winemaking. *DOPS-NYT* no. 2, pp. 14-18.

Fox, P.F. & P.L.H. McSweeney (1998): *Dairy Chemistry and Biochemistry*. Kluwer Academic / Plenum Publishers, New York, New York, USA

Griffiths, P.R. (2002): Introduction to Vibrational Spectroscopy. In: J.M. Charlmers & P.R. Griffiths (eds.): *Handbook of Vibrational Spectroscopy*. vol. 1. John Wiley & Sons Ltd, Chichester, UK, pp. 33-43.

Harris, C.D. (2007): *Quantitative Chemical Analysis*. 7th. ed. W.H. Freeman and Company, New York, New York, USA

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

Hartnett, M.K., G. Lightbody & G.W. Irwin (1998): Dynamic inferential estimation using principal components regression (PCR). *Chemometrics and Intelligent Laboratory Systems*. Vol. 40, no. 2, pp. 215-224.

Hartnett, M.K., G. Lightbody & G.W. Irwin (1999): Identification of state models using principal components analysis. *Chemometrics and Intelligent Laboratory Systems*. Vol. 46, no. 2, pp. 181-196.

Huang, H.B., H.Y. Yu, H.R. Xu & Y.B. Ying (2008): Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of Food Engineering*. Vol. 87, no. 3, pp. 303-313.

ICH (2009): *ICH Harmonised Tripartite Guideline - Pharmaceutical Development Q8(R2)*.

Jackson, J.E. (1959): Quality Control Methods for Several Related Variables. *Technometrics*. Vol. 1, no. 4, pp. 359-377.

Jackson, J.E. & G.S. Mudholkar (1979): Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*. Vol. 21, no. 3, pp. 341-349.

Kalman, R.E. (1960a): A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME - Journal of Basic Engineering*. Vol. 82, pp. 35-45.

Kalman, R.E. (1960b): Contributions to the Theory of Optimal Control. *Boletin De La Sociedad Matematica Mexicana*. Vol. 5, pp. 102-119.

Kourti, T. & J.F. MacGregor (1995): Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods. *Chemometrics and Intelligent Laboratory Systems*. Vol. 28, no. 1, pp. 3-21.

Kourti, T. & J.F. MacGregor (1996): Multivariate SPC methods for process and product monitoring. *Journal of Quality Technology*. Vol. 28, no. 4, pp. 409-428.

Kresta, J.V., J.F. MacGregor & T.E. Marlin (1991): Multivariate Statistical Monitoring of Process Operating Performance. *Canadian Journal of Chemical Engineering*. Vol. 69, no. 1, pp. 35-47.

Ku, W.F., R.H. Storer & C. Georgakis (1995): Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. Vol. 30, no. 1, pp. 179-196.

Lakowicz, J.R. (2006): *Principles of fluorescence spectroscopy*. 3ed ed., Springer Verlag, Berlin, Germany

Laursen, K., M.A. Rasmussen & R. Bro (2011): Comprehensive control charting applied to chromatography. *Chemometrics and Intelligent Laboratory Systems*. Vol. 107, no. 1, pp. 215-225.

Lawaetz, A.J., B. Schmidt, D. Staerk, J.W. Jaroszewski & R. Bro (2009): Application of Rotated PCA Models to Facilitate Interpretation of Metabolite Profiles: Commercial Preparations of St. John's Wort. *Planta Medica*. Vol. 75, no. 3, pp. 271-279.

Ljung, L. (1999): *System Identfication - Theory for the User*. 2nd. ed., Prentice Hall PTR, Upper Saddle River, New Jersey, USA

Ljung, L. (2010): *System Identification Toolbox™ 7 - Getting Started Guide*. 7.4. ed., The MathWorks Inc., Natick, Massachuessets, USA

Luyben, W.L. (1996): *Process Modeling, Simulation and Control for Chemical Engineers*. International edition. ed. McGraw-Hill Publishing Company, New York, New York, USA

Massart, D., B. Vandeginste, L. Buydens, S. De Jong, P. Lewi & J. Smeyers-Verbeke (1997): *Handbook of chemometrics and qualimetrics - Part A, Data Handling in Science and Technology*. Elsevier, Amsterdam, The Netherlands

Maybeck, P.S. (1979): Introduction. *Stocastic models, estimation, and control*. vol. 1. Academic Press, Inc., New York, New York, pp. 1-16.

McLennan, F. (1995): Process Analytical Chemistry in Perspective. In: F. McLennan & B.R. Kowalski (eds.): *Process Analytical Chemistry*. Blackie Academic and Professional, Bishopbriggs, Glasgow, UK, pp. 1-13.

Mehra, R.K. (1970): On the Identification of Variances and Adaptive Kalman Filtering. *IEEE Transactions on Automatic Control*. Vol. 15, no. 2, pp. 175-184.

Mehra, R.K. (1972): Approaches to Adaptive Filtering. *IEEE Transactions on Automatic Control*. Vol. 17, no. 5, pp. 693-698.

Miller, C.E. (2001): Chemical Principles of Near-Infrared Technology. In: P.C. Williams & K.H. Norris (eds.): *Near Infrared Technology in the Agricultural and Food Industries*. 2nd. ed. American Association of Cereal Chemists, Inc, St. Paul, Minnesota, USA, pp. 19-37.

Næs, T., T. Isaksson, T. Fearn & T. Davies (2002): *A user-friendly guide to multivariate calibration and classification*. NIR Publications, Chichester, U.K.

Negiz, A. & A. Çinar (1997): PLS, balanced, and canonical variate realization techniques for identifying VARMA models in state space. *Chemometrics and Intelligent Laboratory Systems*. Vol. 38, no. 2, pp. 209-221.

Nomikos, P. & J.F. MacGregor (1995a): Multi-way partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems*. Vol. 30, no. 1, pp. 97-108.

Nomikos, P. & J.F. MacGregor (1995b): Multivariate Spc Charts for Monitoring Batch Processes. *Technometrics*. Vol. 37, no. 1, pp. 41-59.

Odelson, B.J., M.R. Rajamani & J.B. Rawlings (2006): A new autocovariance least-squares method for estimating noise covariances. *Automatica*. Vol. 42, pp. 303-308.

Odiowei, P.P. & Y. Cao (2010): State-space independent component analysis for nonlinear dynamic process monitoring. *Chemometrics and Intelligent Laboratory Systems*. Vol. 103, no. 1, pp. 59-65.

Olivieri, A.C. (2008): Analytical advantages of multivariate data processing. One, two, three, infinity? *Analytical Chemistry*. Vol. 80, no. 15, pp. 5713-5720.

Pan, Y.D., C. Yoo, J.H. Lee & I.B. Lee (2004): Process monitoring for continuous process with periodic characteristics. *Journal of Chemometrics*. Vol. 18, no. 2, pp. 69-75.

Pavia, D.L., G.M. Lampman & G.S. Kriz (2000): Infrared Spectroscopy. *Introduction to spectroscopy*. 3ed. ed. Brooks / Cole, Belmont, California, USA, pp. 13-29.

Pearson, K. (1901): On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*. Vol. 2, pp. 559-572.

Pedersen, D.K., L. Munck & S.B. Engelsen (2002): Screening for dioxin contamination in fish oil by PARAFAC and *N*-PLSR analysis of fluorescence landscapes. *Journal of Chemometrics*. Vol. 16, no. 8-10, pp. 451-460.

Phatak, A. & S. Dejong (1997): The geometry of partial least squares. *Journal of Chemometrics*. Vol. 11, no. 4, pp. 311-338.

Qin, S.J. (2003): Statistical process monitoring: basics and beyond. *Journal of Chemometrics*. Vol. 17, no. 8-9, pp. 480-502.

Rajamani, M.R. & J.B. Rawlings (2009): Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. *Automatica*. Vol. 45, pp. 142-148.

Rawlings, J.B. (2000): Tutorial overview of model predictive control. *IEEE Control Systems Magazine*. Vol. 20, no. 3, pp. 38-52.

Roffel, B. & B.H. Betlem (2006): *Process dynamics and control: modeling for control and prediction.* John Wiley & Sons, Ltd, Chichester, UK

Sheppard, N., H.A. Willis & J.C. Rigg (1985): Names, symbols, definitions and units of quantities in optical spectroscopy (Recommendations 1984). *Pure and Applied Chemistry*. Vol. 57, no. 1, pp. 105-120.

Shi, R.J. & J.F. MacGregor (2000): Modeling of dynamic systems using latent variable and subspace methods. *Journal of Chemometrics*. Vol. 14, no. 5-6, pp. 423-439.

Siesler, H.W. (2008): Basic Principles of Near-Infrared Spectroscopy. In: D.A. Burns & E.W. Ciurczak (eds.): *Handbook of Near-Infrared Analysis*. 3rd. ed. CRC Press, Boca Raton, Florida, USA, pp. 7-20.

Simon, D. (2006): *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* John Wiley & Sons, Hoboken, N.J., USA

Skagerberg, B., J.F. MacGregor & C. Kiparissides (1992): Multivariate Data-Analysis Applied to Low-Density Polyethylene Reactors. *Chemometrics and Intelligent Laboratory Systems*. Vol. 14, no. 1-3, pp. 341-356.

Smilde, A.K., R. Bro & P. Geladi (2004): *Multi-way Analysis - Applications in the chemical sciences.* John Wiley & Sons Ltd, Chichester

Smith, E.B. (1990): *Basic Chemical Thermodynamics.* 4th edition. ed. Oxford University Press, New York, New York, USA

Smith, W.F. (2005): *Experimental design for formulation.* Society for Industrial and Applied Mathematics,

Ståhle, L. (1989): Aspects of the analysis of three-way data. *Chemometrics and Intelligent Laboratory Systems*. Vol. 7, no. 1–2, pp. 95-100.

Stephanopoulos, G. (1984): *Chemical Process Control - An Introduction to Theory and Practice.* Prentice Hall, Englewood Cliffs, New Jersey, USA

Strang, G. (2006): *Linear algebra and its applications.* 4th edition. ed. Thomson Brooks/Cole, Belmont, California, USA

U.S.Food and Drug Administration (2004): *Guidance for Industry: PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*.

Umetrics (2008): *User Guide to SIMCA-P+.* Version 12. ed.

van den Berg, F. & Å. Rinnan (2009): Calibration Transfer Methods. In: D.-W. Sun (ed.): *Infrared Spectroscopy for Food Quality Analysis and Control*. 1. ed. Academic Press, Burlington, Massachuessets, USA, pp. 103-118.

van Overschee, P. & B. De Moor (1996): *Subspace Identification for Linear Systems.* Kluwer Academic Publishers, Boston/London/Dordrecht

Wahlberg, B., M. Jansson, T. Matsko & M.A. Molander (2007): Experiences from Subspace System Identification - Comments from Process Industry Users and Researchers. In: A. Chiuso, A. Ferrante, S. Pinzoni & G. Picci (eds.): *Modeling, estimation and control: Festschrift in honor of Giorgio Picci on the occasion of his sixty-fifth birthday*. Springer Verlag, Berlin, Germany, pp. 315-327.

Walstra, P., J.T.M. Wouters & T.J. Geurts (2006): *Dairy Science and Technology.* 2nd ed. CRC Press, Boca Raton, Florida, USA

Wang, Y.D. & B.R. Kowalski (1993a): Standardization of Second-Order Instruments. *Analytical Chemistry*. Vol. 65, no. 9, pp. 1174-1180.

Wang, Y.D. & B.R. Kowalski (1993b): Temperature-Compensating Calibration Transfer for Near-Infrared Filter Instruments. *Analytical Chemistry*. Vol. 65, no. 9, pp. 1301-1303.

Wang, Y.D., D.J. Veltkamp & B.R. Kowalski (1991): Multivariate Instrument Standardization. *Analytical Chemistry*. Vol. 63, no. 23, pp. 2750-2756.

Welch, G. & G. Bishop (2006): *An Introduction to the Kalman Filter*. Available at the Internet: http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf [cited January 6, 2012].

Wiener, N. (1949): *The Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications.* Wiley, New York, New York, USA

Wise, B.M. (1991): *Adapting Multivariate Analysis for Monitoring and Modeling of Dynamic Systems*. University of Washington, Seattle, USA.

Wise, B.M. & B.R. Kowalski (1995): Process Chemometrics. In: F. McLennan & B.R. Kowalski (eds.): *Process Analytical Chemistry*. Blackie Academic and Professional, Bishopbriggs, Glasgow, UK, pp. 259-312.

Wise, B.M. & N.L. Ricker (1993): Identification of Finite Impulse-Response Models with Continuum Regression. *Journal of Chemometrics*. Vol. 7, no. 1, pp. 1-14.

Wold, H. (1966): Estimation of principal components and related models by iterative least squares. In: P.R. Krishnaiah (ed.): *Multivariate Analysis*. Academic Press, New York, New York, USA, pp. 391-420.

Wold, S., M. Sjöström & L. Eriksson (2001): PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*. Vol. 58, pp. 109-130.

Workman, J., K.E. Creasy, S. Doherty, L. Bond, M. Koch, A. Ullman & D.J. Veltkamp (2001): Process analytical chemistry. *Analytical Chemistry*. Vol. 73, no. 12, pp. 2705-2718.

Workman, J., M. Koch, B. Lavine & R. Chrisman (2009): Process Analytical Chemistry. *Analytical Chemistry*. Vol. 81, no. 12, pp. 4623-4643.

Workman, J., M. Koch & D.J. Veltcamp (2003): Process analytical chemistry. *Analytical Chemistry*. Vol. 75, no. 12, pp. 2859-2876.

Workman, J., M. Koch & D. Veltkamp (2005): Process analytical chemistry. *Analytical Chemistry*. Vol. 77, no. 12, pp. 3789-3806.

Workman, J., M. Koch & D. Veltkamp (2007): Process analytical chemistry. *Analytical Chemistry*. Vol. 79, no. 12, pp. 4345-4363.

Workman, J., B. Lavine, R. Chrisman & M. Koch (2011): Process Analytical Chemistry. *Analytical Chemistry*. Vol. 83, no. 12, pp. 4557-4578.

Workman, J., D.J. Veltkamp, S. Doherty, B.B. Anderson, K.E. Creasy, M. Koch, J.F. Tatera, A.L. Robinson, L. Bond, L.W. Burgess, G.N. Bokerman, A.H. Ullman, G.P. Darsey, F. Mozayeni, J.A. Bamberger & M.S. Greenwood (1999): Process analytical chemistry. *Analytical Chemistry*. Vol. 71, no. 12, pp. 121R-180R.

Workman, J.J. & D.A. Burns (2008): Commercial NIR Instrumentation. In: D.A. Burns & E.W. Ciurczak (eds.): *Handbook of Near-Infrared Analysis*. 3rd. ed. CRC Press, Boca Raton, FL, USA, pp. 67-78.

Wust, E. & L. Rudzik (2003): The use of infrared spectroscopy in the dairy industry. *Journal of Molecular Structure*. Vol. 661, pp. 291-298.

Ziegler, J.G. & N.B. Nicholos (1942): Optimum Settings for Automatic Controllers. *Transactions of the ASME*. Vol. 62, pp. 759-768.

Paper I

Thygesen, J. & F. van den Berg (2011)

Calibration transfer for excitation-emission fluorescence measurements

Analytica Chimica Acta, Vol. 705, no. 1-2, pp. 81 – 87

Analytica Chimica Acta 705 (2011) 81–87

Contents lists available at ScienceDirect

Analytica Chimica Acta

journal homepage: www.elsevier.com/locate/aca

Calibration transfer for excitation–emission fluorescence measurements

Jonas Thygesen*, Frans van den Berg

University of Copenhagen, Faculty of Life Sciences, Department of Food Science, Quality & Technology, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

ARTICLE INFO

Article history:
Received 16 November 2010
Received in revised form 5 April 2011
Accepted 12 April 2011
Available online 20 April 2011

Keywords:
Calibration transfer
Fluorescence spectroscopy
Parallel factor analysis
Vitamin B2

ABSTRACT

The main part of the wide array of different calibration transfer methods found in literature is dedicated to two-way data arrangements ($m \times n$ matrices). Less work has been done within the area of calibration transfer for three-way data structures ($m \times n \times l$ tensors) such as calibrations made for excitation–emission-matrix (EEM) fluorescence spectra. There are two possible ways to attack the problem for EEM transfer. Either the tensors are unfolded to two-way data, whereby the existing methods can be applied, or new methods dedicated to three-way calibration transfer have to be developed. This paper presents and compares both.

It was possible to make a local linear *pixel-based* model that could be used for transfer of EEM's. This new method has a similar performance to the *classical* methods found in literature, direct- and piecewise direct standardization. The three-way advantages made it possible to use as few as four samples to build useable transfer models. Care has to be taken though when choosing the samples. When subset recalibration of the systems is compared to calibration transfer, better performance is seen for the transferred calibrations. Overall the three-way calibration transfer methods have a slightly better performance than the two-way methods.

1. Introduction

The subject calibration transfer or standardization in chemometric model building and process analytical chemistry & technology (PACT) really falls under economics. Constructing a high-quality inverse multivariate calibration model such as partial least squares (PLS) in the presence of unknown interfering signals requires tens, hundreds or sometimes up to a thousand samples plus reference analysis, often collected over a long period of time. This is a big investment. Calibration transfer focuses on preserving this investment by keeping the model valid over time for the same instrument (model maintenance) or by sharing the cost where a model developed on one system (the primary or master instrument) is applied to one or more other systems of a similar nature (the secondary or slave instrument). The relevance of the subject is emphasized by the fact that in July 2010 over 20 references with *calibration transfer* or *standardization* in their title appeared in the open literature, and a relatively large number of these indicate connections with industry. Most of these papers are found in the analytical chemistry/chemometric and spectroscopy literature focusing on computational methodologies; three comprehensive reviews are available [1–3]. Moreover, just like chemometric data analysis in general, the potential of calibration transfer is being recognized by the outside world [4–8]. Economics also play a role in the reverse direction: since we are trying to minimize our expenses on the collection of sample and reference analysis, the model will always be suboptimal compared to full recalibration (e.g. on a secondary instrument) and, moreover, an independent evaluation via a test set or uncertainty estimation by re-sampling is typically not feasible due to the small number of samples involved. This makes an understanding of the mathematical operations involved in calibration transfer and the effects on spectroscopic data crucial for proper and safe use.

A number of scenarios could result in a multivariate calibration model being or becoming invalid. This would occur, for instance, if the original instrument is replaced by a new one. The responses from two instruments for the same sample measured under the same conditions will be different and multivariate calibration models will thus not necessarily be valid for this new situation. It must be stated that big improvements have been achieved by instrument manufacturers on hardware harmonization in recent years, especially in Near InfraRed (NIR) spectroscopy. Diode lasers for wavelength alignment, internal (reflection) standards for intensity corrections, and charged-coupled device (CCD) similarity-matching at the manufacturer by characteristics comparison are just a few measures on offer to improve instrument-to-instrument compatibility. Nevertheless, the instability of one and the same unit over time is another problem which can seriously affect the performance of a model. Small continuous changes (e.g. instrumental drift due to filter bleaching in fluorescence instruments) and sudden

* Corresponding author. Tel.: +45 3533 3500.
E-mail address: thygesen@life.ku.dk (J. Thygesen).

0003-2670/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.aca.2011.04.017

changes (response shifts caused by repairs or replacements of light sources for example) in the instrument still cause signals to change, leading to increased prediction errors in the absence of proper model maintenance. The most straightforward solution, of course, would be to re-calibrate for the new measurement conditions or to expand the original model for the new situation. Unfortunately, this is also the most expensive solution and sometimes technically impossible. Standardization and calibration transfer methods have been developed aimed at eliminating the need for a full recalibration and to preserve the information collected in an existing model. Even if instrumental hardware is matched well, sampling for in-process measurements could still render the calibration model invalid. For example, interfacing to a process stream via different routes such as multiplexers and/or fiber optics will not always be in the hands of the instrument manufacturer. Bend angles of the fibers and optical components for different sampling points will differ, all influencing the detector responses in a unique way. This is precisely the direction where new developments in process monitoring and control on NIR (still the *workhorse* in PACT) are heading: various measurement points for similar streams in the factory based on multiplexers or a family of relatively cheap CCD-based systems that depend on one global calibration [9]. Almost all methods for calibration transfer dependent on measuring several samples on both primary and secondary instrument, spanning the variation in responses between the instruments that way. The cost-saving aspect of calibration transfer is thus very much active. From an operational point of view generic standards – is an easy-to-handle and storable form such as polystyrene standard materials, certified reference materials, easily reproducible mixtures, etc. – are preferred. However, it is essential that these generic standards are compatible with regular samples from an information point of view, both in desired (e.g. the concentration to be predicted) and undesired properties (e.g. scatter properties), which is seldom a simple requirement in process monitoring.

It should be noted that almost all of the wide array of different methods found in literature are dedicated to transferring calibrations for two-way data ($m \times n$ matrices). Much less work has been done within the field of transferring calibrations for three-way data ($m \times n \times l$ tensors), such as calibrations made for excitation–emission-matrix (EEM) fluorescence spectra. The objective of this study is therefore to develop and evaluate new, simple methods for N-way calibration transfer. The current work is based on two process EEM fluorescence spectrometers of the same manufacturer. They are produced to be alike, but due to differences in the optical fibers used and inherent, small differences in the filters e.g. due to aging/bleaching, there is a clear distinction between the spectra the two spectrometers record. It is well established that so-called three-way advantages can be obtained when EEM fluorescence spectroscopy is combined with PARAFAC modeling [10]. The three-way advantages allows (among other things) for fewer samples to be used during calibration when compared to two-way factor models such as PLS, and it could also potentially make calibration transfer a less complicated task. It is therefore of interest to see if it is possible to develop new minimal three-way methods for calibration transfer. Alternatively, modification of existing two-way transfer methods is also of interest.

2. Materials and methods

A set of milk samples was spiked with pure riboflavin (vitamin B2 $\geq$ 98%; Sigma Aldrich Inc., Saint Louis, MI, USA). All samples were produced from the same skimmed milk powder base (Arla Foods Amba, Viby, Denmark) and riboflavin was added in levels ranging from 0 to 100% of the nominal value of skimmed milk. 39 different levels of riboflavin were used, the lowest being 0.181 mg100 mL^{-1} and the highest 0.344 mg 100 mL^{-1}, all levels were produced in independent duplicates – a total of 78 milk samples.

All samples were measured on two different BioView EEM fluorescence process spectrometers (DELTA Danish Electronics, Light & Acoustics, Hørsholm, Denmark). Both instruments use a combination of 15 excitation filters (λ_{ex}; equidistantly spaced from 270 to 550 nm) positioned on the light source sequentially and 15 emission bands filters (λ_{em}; equidistantly spaced from 310 to 590 nm) positioned on the detector sequentially. Both instruments work with a weakly non-linear analogue amplifier and an analogue-to-digital-converter with $2^{12} = 4096$ steps/counts resolution. The instrument gains were established before experimentation such that the signal maxima per landscape are between 1000 and 2000 counts. The short-term stability of the instruments was tested on generic materials and no changes were observed over the period of a day, just as no changes in either instrument were observed during experimentation.

The high grade narrow band filters are designed to have their throughput maxima at the same wavelength as accurately as technically possible. However, the primary instrument is older and used more frequent, and bleaching by the strong light source will influence the filter characteristics over time. The primary instrument is equipped with high-throughput compressed liquid filled fibers while the secondary instrument is equipped with conventional normal-throughput quartz–fibers. Samples were measured at milk processing/storage temperatures (5 ± 2 °C).

3. Theory

3.1. Calibration of fluorescence data

There are two general methods for calibration of fluorescence data. The first method would unfold the three-way tensor into a matrix with the format $N_{samples} \times (\lambda_{em} \times \lambda_{ex})$. Two-way methods (such as PLS) can hereafter be used to build the sought calibration. The method has the advantage that many different calibration transfer concepts are available. The disadvantages of unfolding the tensors are that the information in the three-way structure of the data is lost. The second method is therefore to utilize the *N*-way structure of the tensor directly [10]. In this study calibrations were obtained as follows (Fig. 1):

1. The measured EEMs are collected in three-way arrays (samples × emission × excitation).
2. PARAFAC decomposition of the data with the correct number of factors/model complexity is performed.
3. Sample scores from PARAFAC models are combined with reference *y*-values. A non-linear dependence was observed between counts and concentrations of vitamin B2. A quadratic function is therefore fitted using total least squares (B2 as dependent variable).
4. New data can be predicted by projecting the EEM onto the PARAFAC model emission and excitation loadings, and combining the retrieved score with the quadratic function for prediction.

3.2. Calibration transfer methods

An example of the same milk sample measured on the primary and secondary instrument is given in Fig. 2, shown as intensity maps/landscapes.

It can be observed that the two maps are alike but not perfect copies of one another. The most noticeable is the large intensity difference (approximately a factor 2). A simple intensity correction of the entire matrix would therefore remove a large part of the differences. But, at second evaluation, one can also observe that the ratio

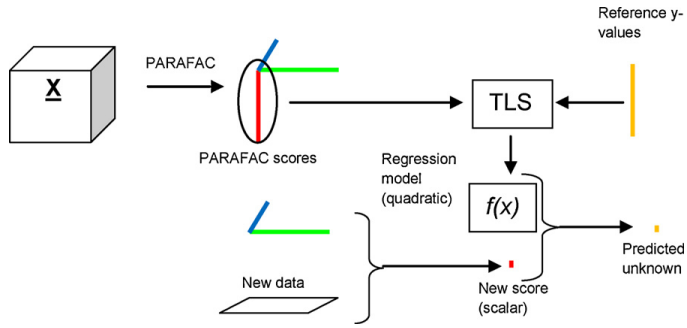


Fig. 1. Principal behind PARAFAC based EEM calibration.

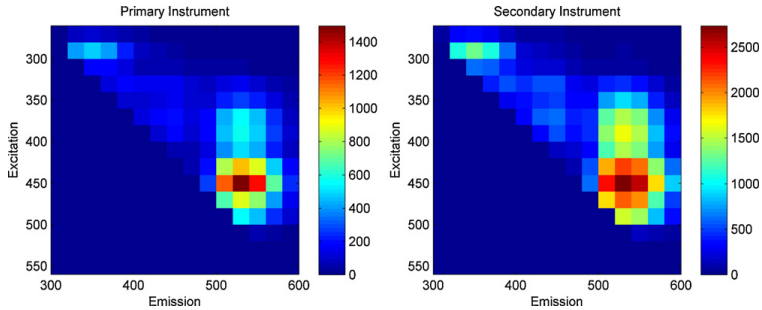


Fig. 2. Intensity maps of the same milk sample measured on the primary and secondary instrument.

between some of the pixels in the matrix is not the same for the two instruments. This is most obvious for the middle part of the map. The pixel at e.g. λ_{em} 450 nm/λ_{ex} 370 nm is relatively more intense for the secondary instrument than for the primary instrument. Fig. 3 shows that the high level of selectivity by fluorescence spectroscopy (in this case towards vitamin B2) makes calibration a relatively easy task. It also makes obvious that not all pixels in the map are vitamin concentration dependent which advocates the combination with uniqueness properties of multi-way modeling. We also observe a modest different tendency in curvature between the analyte signal peak for primary and secondary instrument; this suggests that a more complex correction could be necessary.

One straightforward correction could be on the individual emission/excitation channels level (the individual pixels in the matrix). The counts for different pixels on two different instruments are expected to have a relationship that can be modeled with a low order polynomial. If for instance a sample gives a high signal on the primary instrument, the same is expected for the secondary

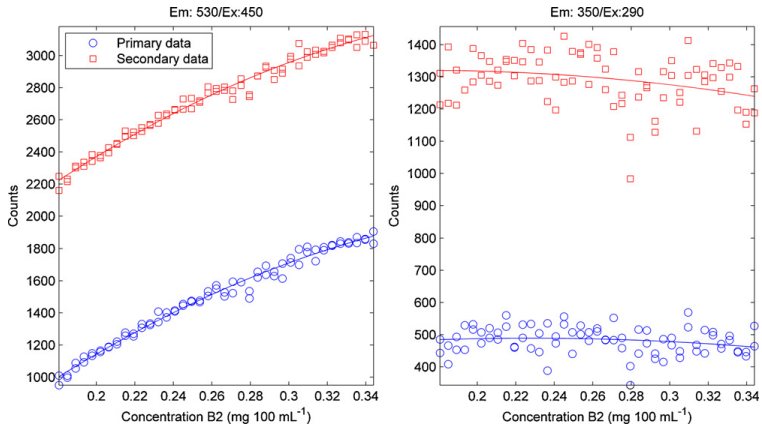


Fig. 3. Single pixel intensities versus known vitamin B2 concentration for analyte signal peak (left) and background signal (right).

instrument. For information rich pixels with a good signal-to-noise ratio (such as the centre pixel at λ_{em} 530 nm/λ_{ex} 450 nm) a linear first order polynomial could be used. But if the pixel has little or no information (such as the pixel at λ_{em} 350 nm/λ_{ex} 290 nm), no real relation between the pixels can be expected, and any linear model fitted to the numbers would therefore carry risk of modeling random noise rather than true information; a weighted regression form could thus potentially improve the models. Based on these observations six different methods for calibration transfers were investigated:

1. Single scalar intensity correction (global).
2. Non-weighted univariate linear (pixel-to-pixel).
3. Weighted univariate linear (pixel-to-pixel).
4. NaN filtering (pixel-to-pixel).
5. Direct standardization (DS; global multivariate).
6. Piecewise direct standardization (PDS; local multivariate).

Each method falls, as indicated, in one of three classes: (1) global methods, working simultaneously on the whole EEM; (2) pixel-to-pixel methods, relating a pixel in the primary instrument EEM to the corresponding secondary, and (3) multivariate, that is not only pixel to pixel, but including the neighbor points. The methods are described in detail below.

3.2.1. Single scalar intensity correction

The single scalar method is the most simple calibration transfer method included in this study. The principle is to find one single scalar (f) that all elements in the secondary tensor are multiplied by to obtain an estimate to use in the model build on the primary instrument. The method can therefore be seen as a global intensity correction. The multiplier is found by unfolding the tensors from the n transfer samples on the two instruments into augmented column vectors ($\mathbf{x}_p$ and $\mathbf{x}_s$), and subsequently finding the least squares solution for f by:

$$\mathbf{x}_p = \mathbf{x}_s \times f \rightarrow f = (\mathbf{x}_s^T \mathbf{x}_s)^{-1} \mathbf{x}_s^T \mathbf{x}_p$$

3.2.2. Non-weighted and weighted univariate linear

For both the weighted- and non-weighted univariate linear regression modeling is done by fitting a slope (b_1) and intercept (b_0) for all the channels seperately (the non-weighted model is treated as weighted model with equal weights for all data points).

$$(\mathbf{w} \times \mathbf{x}_p^c) = b_0 + (\mathbf{w} \times \mathbf{x}_s^c) \times b_1$$

where $\mathbf{x}_p^c$ is the n element vector of counts for n transfer samples for one EMM pixel/channel on the primary instrument, $\mathbf{x}_s^c$ the similar n element vector for the secondary instrument, and $\mathbf{w}$ contains the weights for the different transfer samples – in case of non-weighted regression a vector of ones is used as weights. For weighted regression the norm or the Euclidean length $w(n) = \sqrt{\mathbf{x}_p^c(n)^2 = \mathbf{x}_s^c(n)^2}$ of the combined counts is used. Since counts are positive this norm reflects the amount of information in the channel (the combined distance away from zero); it will be large if much information (=high counts) are present in both the primary and secondary instrument, and small if only little information is present (=noise).

3.2.3. NaN filtering

In Not-a-Number (NaN) filtering PARAFACs ability to handle missing values is utilized [10]. The first step in NaN filtering is to fit a non-weighted univariate pixel-to-pixel linear model. If the correlation between the primary and secondary measurements is below a given cut-off value (R^2 = 90.0%), the pixel is set to missing (NaN). Hereby only information rich channels are used in the calibration transfer step. The method can be seen as an extreme version of the weighted regression with a hard threshold, where pixels with low correlation are given the weight zero.

3.2.4. Direct standardization

Direct standardization (DS), introduced in a series of studies conducted by Wang and co-authors [11–13], is one of the two established multivariate standardization methods applied. It can be seen as a global multivariate version of the non-weighted univariate linear pixel-to-pixel method explained in the section above. The three-way array of EEMs is unfolded in order to apply DS. The data are unfolded to matrices with the format n transfer samples × ($\lambda_{em} \times \lambda_{ex}$) and the empty columns – where the excitation wavelength is higher than the emission wavelength – are removed. The transferred data is subsequently refolded back into a three-way tensor. In DS the dependence between the primary and secondary instruments modeled by a linear model using the Moore–Penrose pseudo-inverse [3]:

$$\mathbf{X}_p = \mathbf{X}_s \mathbf{F} \rightarrow \mathbf{F} = \mathbf{X}_s^+ \mathbf{X}_p$$

3.2.5. Piecewise direct standardization

One major problem with DS is the step where the transfer matrix $\mathbf{F}$ is determined by the Moore–Penrose pseudo-inverse. This step can easily lead to numerical instabilities translating into poor results, especially if the number of transfer samples n is much smaller than the number of variables in the spectrum ($15 \times 15 - 105 = 120$ variables in each EEM in our case). This observation led to the alternative model PDS where the transfer for each variable in the spectrum of the primary instrument is estimated from a (symmetric) window surrounding the same variable on the secondary instrument [11]. A much smaller (and thus more stable) local inversion step is used. A window size of 7 points was used in this study. The window is moved over the total spectrum, and a band diagonal $\mathbf{F}$ transfer matrix with regression vectors equal to the window size is formed [3]. Issues and possible remedies have been reported concerning artifacts introduced in the transferred spectra due to local rank differences [14], but are not employed here.

4. Results and discussion

4.1. Presentation of data and calibration development

A score plot for a two component PARAFAC model of the EEMs for the milk data-set measured on the primary and secondary instrument combined is given in Fig. 4, no samples were removed as outliers.

As expected – two clear groups are found within the data based on both instruments. Within each group a one component PARAFAC model was found to be optimal. Seen from a chemical viewpoint this makes sense since only one chemical component (vitamin B2) is varying in-between the samples. The excitation and emission loadings (not shown) are very similar to pure B2 loadings. The quality of the models is further supported if regression is made on the PARAFAC scores (as outlined in Fig. 1). The model for the primary instrument shows a low leave-one-out cross-validation prediction error (RMSECV = 5.61×10^{-3} mg·100 mL^{-1}) and a good correlation between the measured and the predicted values (R^2 = 99.4%; see Table 1). This model will therefore serve as a basis of comparison. To have a similar basis of comparison for the PLS models, the data tensor is unfolded and a PLS model is made. Also in this case one component is optimal with a low prediction error (RMSECV = 6.78×10^{-3} mg·100 mL^{-1}) and high correlation between measured and predicted (R^2 = 99.1%). The PLS loading vector resem-

Table 1
Performance of calibration and recalibrated models.

	PARAFAC		PLS	
	RMSECV/RMSEP (10^{-3} mg 100mL^{-1})	R^2 (%)	RMSECV/RMSEP (10^{-3} mg 100 mL^{-1})	R^2 (%)
Full calibration (CV 78 samples) primary instrument	6.50	99.1	6.91	99.0
Full calibration (CV 78 samples) secondary instrument	6.66	99.0	8.27	98.5
Calibration primary instrument using 34 calibration samples, test-set validation 36 samples	7.73	97.8	7.94	98.9
Recalibration secondary instrument using 8 samples, test-set validation 36 samples	8.22	98.9	9.54	98.7

bles the unfolded B2 spectra. This model is therefore used as a basis of comparison for the PLS models.

4.2. Recalibration

Calibration transfer would, as mentioned in Section 1, often compete with recalibration of the secondary instrument. It is therefore of interest to compare the methods developed below with simple recalibration of the secondary instrument. Many different algorithms for sample selection can be found in literature, e.g. leverage based methods such as the Kennard-Stone algorithm [15]. In this study a transfer set was selected based on *a priori* knowledge on the vitamin content in the sample. Eight samples were selected to span the vitamin range evenly. Of the remaining 70 samples, 34 were used for calibration development (calibration samples) and 36 were used as test-set. The calibration- and test-set samples were also selected to span the vitamin range evenly. Table 1 summarizes the performance of the calibrated and re-calibrated PLS and PARAFAC based models.

Table 1 show that PARAFAC overall has a lower prediction error than PLS for both instruments. It also shows that the models for the primary instrument are better than the models for the secondary instrument. Recalibration of the secondary instrument is possible if PARAFAC is used, but recalibration of PLS models gives unacceptable large prediction errors. The remainder of this paper will therefore focus on PARAFAC rather than PLS. The PARAFAC calibration made using 34 calibration samples measured on the primary instrument, in the following will be transferred to the secondary instrument via the different transfer methods developed. The actual vs. predicted plot for the calibration is given in Fig. 5.

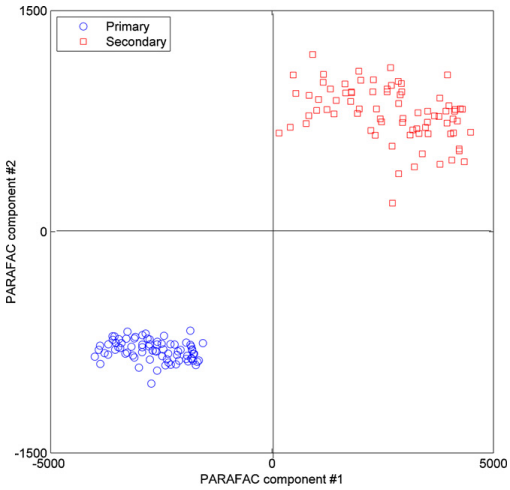


Fig. 4. Score plot for 2 component PARAFAC model of EEMs for all milk samples.

4.3. Initial screening

Several different methods are available when one has to compare the transferred data from the secondary instrument to the data from the primary instrument; in this study the relative residual sum of squares (RSS) is used. When the test-set data from the two instruments (before and after transfer) is augmented into column vectors ($\mathbf{x}_p$, $\mathbf{x}_s$ and $\mathbf{x}_{tr,s}$) RSS is defined as:

$$\mathbf{e} = \mathbf{x}_p - \mathbf{x}_s, \quad \mathbf{e}_{tr} = \mathbf{x}_p - \mathbf{x}_{tr,s}, \quad RSS = \frac{\mathbf{e}_{tr,s}^T \mathbf{e}_{tr,s}}{\mathbf{e}^T \mathbf{e}}$$

A RSS close to one indicates that not much similarity between the two instruments was gained by the transfer; a RSS close to zero indicates that the transferred secondary data is very close to the primary data. The performance of the different transfer methods using the same 8 transfer samples and 36 validation samples as in Table 1 is shown in Table 2.

We noticed that all methods reduce the instrument-to-instrument differences. The simple single scalar intensity correction is not sufficient to correct for all the differences as anticipated from Fig. 3. Furthermore it is observed that weighted regression would be a poor choice. Information is apparently removed by applying the suggested weighing during regression. It is also noticed that the NaN filtered method produces data with a good fit/correction. This is not surprising since the filtering essentially removes all problematic points in the EEM landscapes (those where little information/systematic variation is available). The multivariate methods fit the data nicely, as expected.

Fig. 6 shows three different EEM maps: a recording on the primary instrument, a refolded landscape from PARAFAC modeling and a transferred sample using NaN filtering. Comparison shows that for these data the NaN filtering removes almost all data below emission wavelengths 510 nm. The second sub-plot shows that the PARAFAC model also weights down the data at wavelengths below 510 nm. For these data the NaN-filtering therefore removes data points that anyhow would not be included in the PARAFAC model but in the residuals. No aditional advantage for prediction

Table 2
Transfer method comparison using RSS criterion.

	RSS
No transfer	1.000
Single scalar intensity correction	186×10^{-4}
Non-weighted univariate linear	2.94×10^{-4}
Weighted univariate linear	22.5×10^{-4}
NaN filtered	2.94×10^{-4}
Direct Standardization	3.60×10^{-4}
Piecewise Direct Standardization	2.69×10^{-4}

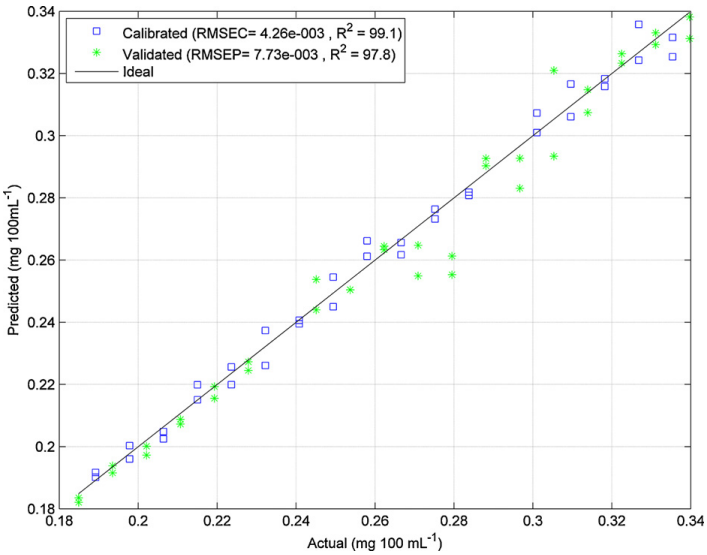


Fig. 5. Actual vs. predicted plot for PARAFAC model using using 34 calibration samples, test-set validation 36 samples.

Table 3
Re-sampling estimate of RMSEP and R^2 for different transfer functions. 200 resampling loops were used.

		4 transfer samples	8 transfer samples	16 transfer samples
		RMSEP/R^2 (10^{-3} mg/100 mL)/(%)	RMSEP/R^2 (10^{-3} mg/100 mL)/(%)	RMSEP/R^2 (10^{-3} mg/100 mL)/(%)
No transfer	Lower	450.20/98.6	444.62/98.7	434.49/98.9
	Median	523.29/97.8	518.18/98.0	516.11/98.0
	Upper	608.75/96.8	604.16/96.8	598.99/96.7
Scalar	Lower	16.45/99.1	17.09/99.2	16.33/99.3
	Median	24.49/98.9	23.47/98.9	23.66/98.9
	Upper	41.02/98.4	35.19/98.4	34.91/98.3
Linear	Lower	6.89/99.2	6.18/99.3	5.81/99.4
	Median	9.31/98.9	7.56/99.0	7.44/99.0
	Upper	$(2.24 \times 10^{15}/10.2)$	9.93/98.4	9.62/98.4
NaN	Lower	6.87/99.3	6.20/99.3	5.78/99.4
	Median	8.93/99.0	7.60/99.0	7.44/99.0
	Upper	38.75/98.5	10.14/98.4	9.94/98.4
DS	Lower	6.48/99.6	5.88/99.6	5.40/99.6
	Median	10.17/98.9	8.35/99.1	7.87/99.1
	Upper	30.24/96.1	13.33/97.6	11.12/98.2
PDS	Lower	6.69/99.3	6.50/99.4	6.33/99.4
	Median	9.02/99.0	7.88/99.0	7.87/99.0
	Upper	28.08/98.4	11.02/98.5	10.63/98.5

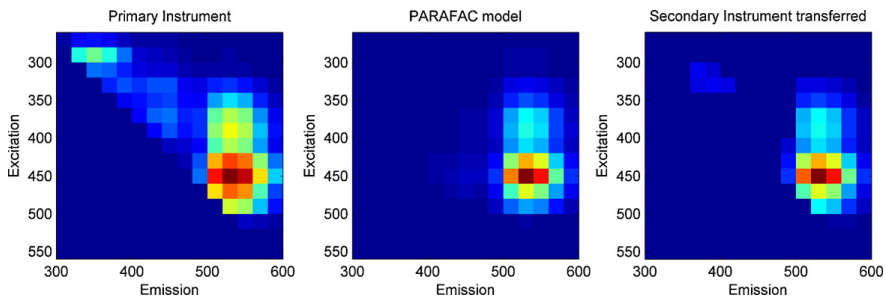


Fig. 6. Fluorescence landscapes of primary instrument, transferred secondary instrument using NaN filtering and outer product of PARAFAC model loadings.

of vitamin B2 via the PARAFAC based method is therefore expected by means of NaN-filtering, when compared to the linear transfer method.

4.4. Robustness of transfer model and number of transfer samples needed

The robustness of the different transfer models are investigated in this section. The sample set was divided into a calibration-, test- and transfer-set for PARAFAC modeling. 34 samples were always used for calibration, 4, 8 or 16 of the remaining samples were randomly selected and used as transfer samples, the remaining samples (not in the test or calibration set)were used as test-set samples. This procedure was repeated 200 times for each transfer method and transfer set size. The resulting RMSEP and R^2 (actual vs. pred.) were sorted according to size. The 95% CI were approximated by removing the ten lowest and highest RMSEP and R^2 values. Table 3 presents the median, the lower and upper bound of the sorted values.

In Table 3 several things are noticed: first that for most models, not surprisingly the median RMSEP decreases as more samples are included for estimating the transfer model. This is though not the case for the scalar transfer model, indicating that this extremely minimal model can be estimated using very few samples. Nevertheless, care has to be taken when making conclusions based on Table 3, since the table is based on resampling and a certain Monte Carlo error can therefore be expected. For the latter four transfer models it is therefore noticed that the median RMSEP is very similar, the difference between the models is especially small for 8 or 16 transfer samples. It is possible to use four samples to build the transfer models. It requires though that care is taken when selecting the transfer set. If the right samples are selected, it is possible to obtain calibration errors very close to the error of the original transfer. If a poor/blind choice is made during transfer set selection large prediction errors can be found. This is especially the case for the linear transfer model, here numerical instabilities can occur if poor choices are made.

When the RMSEP values of Table 3 are compared to the performance of the recalibrated models (Table 1), we see that transfer is to be preferred over recalibration. The median RMSEP values is slightly lower for the transferred data, but for the optimal transfer sets, even lower prediction error scan be obtained with the same transfer set size (size 8). Lower prediction error can also be obtained for just four transfer samples. This means that by making the right choices in transfer-set selection, effort can be saved if calibration transfer is used instead of recalibration. Although fluorescence is a selective measurement principle, unknown matrix effects of future samples (e.g. seasonal variations in protein percentage in the milk) might change the model performance. If this is the case a more elaborate investigation is required.

5. Conclusions

This study showed that it was possible to develop simple, intuitive transfer methods for three-way EEM fluorescence calibration. An uncomplicated local linear method was demonstrated to be the most favorable of the new methods. When the two- and three-way calibration methods were compared, the three-way method showed slightly lower prediction errors both for calibration and re-calibration. The new transfer methods were compared to the *classical* methods found in literature (direct and piecewise direct standardization on unfolded data). Similar results for the new and the *classical* methods were obtained. It was additionally shown that though good transfer models could be found for the PARAFAC models with as few as four transfer samples, the results are highly dependent on the selection of the transfer set. When recalibration and calibration transfer are compared for the fluorescence data set used, calibration transfer is better with lower prediction errors and fewer samples needed.

Acknowledgements

Jonas Thygesen's research is sponsored by The Danish Research and Innovation Council in the QbD research consortium. DELTA Danish Electronics, Light & Acoustics are kindly acknowledged for supplying the secondary instrument.

References

[1] O.E. de Noord, Chemom. Intell. Lab. Syst. 25 (1994) 85–97.
[2] R.N. Feudale, N.A. Woody, H. Tan, A.J. Myles, S.D. Brown, J. Ferré, Chemom. Intell. Lab. Syst. 64 (2002) 181–192.
[3] F.W.J. van den Berg, Å Rinnan, in Da-Wen Sun (Eds.), Infrared Spectroscopy for Food Quality Analysis and Control, Academic Press, Burlington, MA, USA, 2009, pp. 105–118.
[4] L. Duponchel, C. Ruckebusch, J.P. Huvenne, P. Legrand, J. Mol. Struct. 481 (1999) 551–556.
[5] R.S. Park, R.E. Agnew, R.J. Barnes, J. Near Infrared Spectrosc. 7 (1999) 117–131.
[6] J. Fontaine, J. Horr, B. Schirmer, J. Agric. Food Chem. 52 (2004) 701–708.
[7] E.L. Bergman, H. Brage, M. Josefson, O. Svensson, A. Sparen, J. Pharm. Biomed. Anal. 41 (2006) 89–98.
[8] M.C. Alamar, E. Bobelyn, J. Lammertyn, B.M. Nicolai, E. Molto Postharvest, Biol Technol. 45 (2007) 38–45.
[9] E. Bouveresse, C. Casolino, C. de la Pezuela, J. Pharm. Biomed. Anal. 18 (1998) 35–42.
[10] R. Bro, Chemom. Intell. Lab. Syst. 38 (1997) 149–171.
[11] Y.D. Wang, D.J. Veltkamp, B.R. Kowalski, Anal. Chem. 63 (1991) 2750–2756.
[12] Y.D. Wang, B.R. Kowalski, Anal. Chem. 65 (1993) 1174–1180.
[13] Y.D. Wang, B.R. Kowalski, Anal. Chem. 65 (1993) 1301–1303.
[14] P.J. Gemperline, J.H. Cho, P.K. Aldridge, S.S. Sekulic, Anal. Chem. 68 (1996) 2913–2915.
[15] R.W. L.A Kennard, Stone Technometrics 11 (1969) 137–148.

Paper II

Thygesen, J.H & F. van den Berg (2012)

Subspace methods for dynamic model estimation in PAT applications

Journal of Chemometrics, Accepted for publication

Special Issue Article

CHEMOMETRICS

Received: 1 August 2011, Revised: 24 January 2012, Accepted: 30 January 2012, Published online in Wiley Online Library: 2012

(wileyonlinelibrary.com) DOI: 10.1002/cem.2424

Subspace methods for dynamic model estimation in PAT applications

Jonas Hoeg Thygesen and Frans W. J. van den Berg

One primary goal in the application of process analytical technology tools is improved process monitoring and control. A second is to obtain a better understanding of how a normal process behaves (i.e. the normal dynamics). In order to perform feed-forward control, time series models of the process data are required. Such models could be developed on the basis of known physical/chemical knowledge of the system (i.e. first principal or mechanistic modeling). However, very often, this is not possible because of the lack of sufficient information. This leads to the need of system identification (SI). One class of models within SI is the state space models, linear models that relate the input of the system at time *k* to the output at time *k* via estimation of the so-called system states. State space models may be fitted using what is known as the subspace methods. Subspace methods are based on the projection of data on subspaces identified by, for example, the singular value decomposition of time-shifted data during a training phase. This paper introduces state space models, illustrates how subspace methods are closely related to known chemometric tools, and how they can be applied in, for example, model-based feed-forward process monitoring and control. The concepts are illustrated using a data set from an intrinsically nonlinear milk coagulation process that can be approximated well by a linear dynamic model using a small set of virtual (or principal) states. We present an alternative process-monitoring strategy where the dynamic components and boundary conditions of a developing milk coagulation batch are estimated in real-time and compared to normal operating conditions.

Keywords: state space models; subspace methods; process monitoring; dynamic models; milk coagulation

1. INTRODUCTION

A primary goal in the application of process analytical technology (PAT) and quality by design (QbD) tools is the so-called real-time release. Real-time release is, according to the 2004 FDA guidance on PAT, "the ability to evaluate and ensure the acceptable quality of in-process and/or final products based on process data" [1]. This means that (multivariate) process monitoring and control of dynamic (changing) systems is asked for. Different multivariate statistical process control (MSPC) methods have therefore long been of research interest within the chemometric society [2–5]. The main set of methods within the MSPC is however more suited for feed-back control (*post-problem*), rather than feed-forward control (*preproblem*). In order to facilitate feed-forward control, time series models of the process data are desired. This could be achieved on the basis of known physical/chemical knowledge of the system (first principal or mechanistic modeling [6]). Very often, however, this is not possible because of the lack of (sufficient) knowledge on the system, for example in such complex processes as food production, leading to the need of system identification (SI). One class of models within SI is the state space models. They are linear, time-invariant models that relate the input to the system at time *k* to the output at time *k* via estimation of the system states. These states try to capture or model the dynamic behavior or development of a system without having a direct physical meaning, much like the concept principal component or latent variable in, for example, principal component analysis (PCA). State space models may either be fitted using iterative predictor error algorithms or by using the so-called subspace methods that are based on the projection of data on subspaces identified by, for example, singular value decomposition. The aim of this paper is to discuss state space models, illustrate that subspace methods are closely related to known chemometric tools, and show how they can be applied in feed-forward process monitoring and control.

2. THEORY

Discrete time state space models can be written via vector/matrix products as in Equations 1 and 2. They are linear models that link the input to the system at time k (u_k), to the output at time k (y_k), via the system state vector $\mathbf{x}_k$ (size $n \times 1$; for ease of notation, we will assume univarite inputs and outputs here, but expansion is straight forward):

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}u_k + \mathbf{w}_k \quad (1)$$

$$y_k = \mathbf{C}\mathbf{x}_k + \mathbf{D}u_k + v_k \quad (2)$$

The $\mathbf{A}$ matrix (size $n \times n$) is called the system matrix and describes the system dynamics (i.e. how the system states in vector $\mathbf{x}_k$ evolve from one time step to the next $\mathbf{x}_{k+1}$). The order (or rank) of this matrix determines how many distinct

* *Correspondence to: Jonas Hoeg Thygesen, Department of Food Science, Quality and Technology, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C.*
E-mail: thygesen@life.ku.dk

Department of Food Science, Quality and Technology, University of Copenhagen, Faculty of Life Sciences, Rolighedsvej 30, Frederiksberg C, Denmark

components or states are identified in the system and their connectivity (determined by the entries in **A**). As reported, the states do not necessarily coincide with the physical phenomena in the system (e.g. biomass in a bioreactor) but can be seen as a latent representation of the dynamic behavior. **B** is called the input matrix that relates the control input at the current time step to the system states one step ahead. It shows how an input to a system at time k (e.g. feed to a bioreactor) would influence the state of the system at time $k+1$ (e.g. biomass growth conditions in the reactor). **C** is the output matrix that describes how the system states are reflected in the measurable system output (which, typically, is a physically identifiable entity). It is the link between the principal model at time k and the physical world at time k (e.g. between biomass and growth conditions and actual cell count). **D** is called the (direct) feed-through term (e.g. how the feed of the reactor is seen instantaneously in the measured response, hence, not how feed changes the system); this term is often not included in the modeling. The careful reader will notice that for the univariate case **B**, **C**, and **D** in Equations 1 and 2 should officially be lowercase vectors. However, to stay with common notation, we will keep using matrix capitals instead. $\mathbf{w}_k$ and v_k are noise sequences representing model inaccuracy and measurement uncertainty, respectively. Equation 1 is often referred to as the system equation (reflecting that it describes how the system evolves over time), whereas Equation 2 is called the measurement equation (indicating that it describes how the measured output is related to the state of the system). We will only use discrete time state space models in our study where the effective time between two observations (delta-time, $k+1$ minus k) is the clock time, assumed equidistant and decided by the measurement instrumentation. This could, for example, be the measurement frequency of a spectroscopic determination (or, more accurately, the inverse of the sampling frequency, being the time between two measurements, becoming available).

In order to employ Equations 1 and 2 in a time series-based process-monitoring scheme, the system matrices must be know or estimated. As stated previously, this could theoretically be achieved on the basis of known physical/chemical knowledge of the system, but very often, this is not possible because of the lack of (sufficient) knowledge on the system. This, leads to the need of SI, with one class of algorithms within SI being subspace identification. The reader is referred to the Appendix and van Overschee and De Moor [7] for more details on how estimation of the system matrices is performed for the subspace algorithm that is used in this paper. Because several studies have found canonical variates analysis-based (CVA) algorithms outperform others (see subsequent sections for details), the state space model is fitted using this method (see Appendix). The systems studied in our research are pure batch processes with no external inputs, resulting in a so-called stochastic time series (e.g. beer production in a bioreactor is often run as pure batch with no active input). The computations/estimations remain the same, where all input-related parts are canceled by zero-entries. The applied algorithm produces state space models in a forward innovation form, meaning that an optimal least squares gain (**K**) is used as driving term in the prediction (to substitute the deterministic input $\mathbf{B}u_k$, plus purely stochastic part $\mathbf{w}_k$ in Equation 1). The state space model for the (reduced) stochastic case is, thus, to be reformulated as (see Appendix):

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{K}e_k \tag{3}$$

$$y_k = \mathbf{C}\mathbf{x}_k + e_k \tag{4}$$

Here e_k, the innovation at time step k, is equal to the difference between the observed output and the output predicted by the model (the prediction error at time k). Notice that despite their apparent simplicity, stochastic systems can still approximate complex, nonlinear phenomena because of the interaction between the states in $\mathbf{x}_k$ via the entries in system matrix **A**, plus the initial conditions for the system (e.g. in the bioreactor for beer production, for example, the biomass growth and amount of sugar available for conversion into alcohol might lead to a complex development over time on the basis of internal feedbacks, interacting with the raw material properties and quantities at the batch start/charge plus the yeast's biological efficiency).

3. STATE SPACE MODELS IN CHEMOMETRICS

The idea of applying state space models in chemometrics is not new, albeit not widely spread. A series of papers on state space modeling were published in the chemometric literature in the late 1990s and early 2000s, but the research area has received less attention during the last 10 years. Here, we will give an (nonexhaustive) overview. A chemometric paper on state space models was published in 1997 by Negiz and Cinar [8]. It was shown that partial least squares (PLS) can be used to fit state space equations, but it was, at the same time, shown that modifications of the PLS algorithm were necessary to give useful results. A method, based on CVA, proved to give the best outcome. Hartnett and coworkers published two different papers in the end of the 1990s. In the first of the two papers, genetic algorithms are used in combination with principal components regression (PCR), to do dynamic inferential estimation of process variables [9]. The measurement equation of an underlying state space model is used, but the focus is not on the state space model itself. This is performed later in the paper [10], where how a nonlinear multivariable production plant can be modeled using a combination of PCA and state space modeling is shown. The idea is to perform PCA on the process outputs; the scores obtained are then used as states, and the loadings used as the **C** matrix (no input is used in the measurement equation in this paper [10]). The system equation is subsequently identified by concatenating state matrices and input-matrices and by regressing the concatenated matrix on future states by means of PCR. The PCA-based state space model was compared with an analytical state space model. Both had good performance in approximating the nonlinear system. The authors note that no prior decision on model order needs to be taken when using this PCA-based approach. This is correct, but a decision on the number of principal components in both the PCA and the PCR step needs to be made. Ergon [11] uses state space equations, PCR and PLS, to derive relations that can be used to predict one output variable from another. An example of state space modeling is also given, but this is via the predictor error method (PEM). Dynamic system PCR and PLS solutions for output predictions are also presented, again, based on a PEM state space model. In a later paper by Ergon and Halstensen [12], these results are elaborated for a system with low-sampling-rate reference measurements – a combination of PCA and PEM is utilized to produce predictions of the reference measurements with a better performance as compared with the PLS. Shi and MacGregor give a complete review [13] of different subspace methods and compare them to

different latent variable techniques (PCA, PLS, and PCR). They come to two overall conclusions: (i) for process monitoring ("Is my process on track?"), latent variable methods are to be preferred, but for process identification ("What are the process dynamics?" or "Where is my process heading?"), dedicated subspace identification methods are preferential; and (ii) CVA and the N4SID algorithm [7] have the best performances of the subspace methods they tested. In the most recent paper, applying state space models in the chemometric literature, Pan *et al*. [14] showed, contrary to the first conclusion by Shi and MacGregor [13], that better monitoring performances could be archived if a state space/subspace method was applied compared to PCA-based monitoring. The authors use PCA to reduce the dimensionality of the output, followed by fitting state space models by means of N4SID. A Kalman filter is subsequently used. A large part of the advantage from this model is, according to the authors, a result of the implementation of the Kalman filter.

4. MILK COAGULATION MONITORING

In this paper, it is shown how the dynamics of the coagulation of milk can be observed and modeled by combining near infrared (NIR) spectroscopy, PCA, and subspace-based state space estimation. The example, milk coagulation for cheese production, is a purely stochastic time series with no input or control, and is therefore modeled according to Equations 3 and 4. Twelve batches of coagulating milk were monitored by NIR spectroscopy. The data was first published by Lyndgaard *et al*. [15], and for further details on the procedures and measurements, including the batch numbering, we refer to this publication. In that paper, it was shown that scores from a PCA decomposition of the NIR data could be modeled by mechanistic models [15]. In this manuscript, it is shown that it also is possible to model the data via SI without prior assumptions on the process dynamics. Throughout this paper, we will follow the standard notation used in state space literature. $\mathbf{x}_k$ is therefore the process states at time point k, and $y_\mathbf{k}$ the system output at time point k – in our case, it is the first PCA score vector from the decomposition of the NIR spectra recorded during the batch process. Figure 1(A) shows the NIR spectra of a representative normal operating conditions (NOC) batch color-coded by runtime, whereas Figure 1(B) shows the first PCA score of all 12 batches used in this research [15]. It is seen that the main effect over time is a narrowing and an increase of the water band around 1400–1500 nm which can be attributed to gel formation and hardening [16]. This is a general trend for all twelve batches. In order not to mix symbols, this PCA step is written as $\mathbf{Z} = \mathbf{y}\mathbf{p}^\mathbf{T} + \mathbf{E}$ with $\mathbf{Z}$, $\mathbf{y}$, $\mathbf{p}$, and $\mathbf{E}$, respectively, being a set of NIR spectra sorted as a function of time, the first PCA score vector (equaling the output of our state space model), the corresponding PCA loading vector, and the residuals. The data flow for the state space-based process monitoring is shown in Figure A1 in the Appendix.

Eight batch runs will be considered as training NOC batches, four runs as the test set (Batches 1 and 12 as NOC, and Batches 3 and 8 as extremes/non-NOC). The general PCA score trajectory can be described by three different phases: a very short lag-phase plus a decaying sigmoidal curve and an exponential decay, all superimposed on each other with ill-defined boundaries/transition times [15]. Knowing/predicting the development of the last phase (gel hardening) is of great importance for cheese manufacturing because it gives information on the optimal cutting time (the following step in production [15]), and thus, the quality of the end product. It is, in spite of the clear nonlinear tendencies that can be observed, expected that the data can be well approximated by the linear state space models of sufficient rank. It can furthermore be noticed that two batches differ noticeable from the others: Batch 3 has increasing score values during the first 5 min followed by a very short sigmoidal part which results in the highest "end-value" of all the batches. Batch 8 has a low "end-value" and a longer lag-phase. It should be noted that all experimental runs were performed as similar as possible, and outlying behavior is thus caused by unanticipated but natural variation [15].

5. STATE SPACE BASED MONITORING

In our milk coagulation investigation, delta-time is the measurement frequency of the NIR spectrometer (which gives a new outcome every 36 s, hence, k to $k+1$ takes 36 seconds). Each new NIR data collection is scatter corrected by means of standard normal variate scaling right after collection. The new, expanded spectral matrix $\mathbf{Z}$ is centered, and a PCA decomposition is performed. During the calibration phase, it was established that a one-component PCA model using centered data was essentially the same as a two-component model on the noncentered data, with the first principal component being close to the average

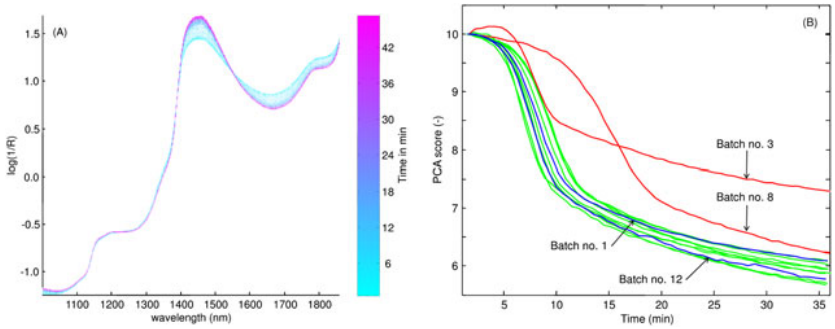


Figure 1. (A) Near infrared (NIR) reflection spectra of coagulating milk in one batch. (B) PCA scores over time from the NIR spectra of 12 different coagulating milk batches.

NIR spectra and the second showing the dynamics of interest in this study. Following the notion of parsimony, a one-component PCA model on the centered data is therefore preferred. In order to avoid numerical problems due to a sign change in the score-values vector during the subsequent state space modeling, all scores are lifted to be positive. This is performed by simply adding the right, same amount to all score values collected thus far to make the first score value equal to 10 ($y_0 = 10$). This operation is performed after each new PCA decomposition. The PCA score time trajectory is zero-order-hold resampled [6] to double the number of data points, after which, a state space model can be fitted. In this procedure, the number of data points is doubled by repeating a measured data value once at the intermediate time for this true value and its proceeding measured neighbor (giving a "staircase" resampled signal that will not introduce false dynamics in the system). This is performed to achieve a more stable estimate of the Hankel matrices (see Appendix) in the beginning of the batch monitoring where only a few measurement observations are available. The models are fitted recursively, meaning that real-time acquisition of data is simulated by stepwise, including more and more observations. The first state space model for each batch is fitted when the first $k = 16$ NIR spectra are collected (corresponding to approximately 9 min into the coagulation process, well pass the initial lag-phase of gel formation, inside the sigmoidal phase for normal batches [15], Figure 1(B)). At the next time step, one more NIR spectrum is included in the data set, and a state space model is determined from the newly computed, offset-corrected and resampled PCA scores for $k = 17$. In this way, it is possible to make an estimate of the **A** matrix (which contains the estimated dynamics of the system) and the initial state vector $\mathbf{x}_0$ (which represents the estimated initial or boundary conditions of the system) at each time step. All these computation steps take less than 1 s on a normal personal computer and can thus be performed in "real-time" for an NIR measurement rate of 36 s (see Figure A1 for the full computational procedure).

The eigenvalues of **A** reflect the dynamics and stability of the system [6]; stable discrete linear time-invariant systems have eigenvalues within the unit circle (where complex eigenvalues indicate an oscillating system [6]). By comparing the eigenvalues of **A** for different runs, the development of different batches can be compared. On the basis of a training set of NOC batches, statistical process control (SPC) charts for the system matrix **A** over time can be constructed and applied to new production runs. The initial state estimate $\mathbf{x}_0$ gives information on the boundary conditions of the difference equation in Equations 1 or 3. This represents the best estimate for the initial conditions in the batch. For monitoring purposes, we propose 95% confidence intervals (95% CI) for the elements of $\mathbf{x}_0$ and the eigenvalues $\boldsymbol{\lambda}$ of **A** on the basis of a Student's *t*-statistic of the values for the NOC training batches. Although it is not guaranteed that normal probabilities are valid for either set of parameters (e.g. because of the mentioned bias in the solution [7], see Appendix), it can serve as a first approximation. The surveillance of the initial conditions in $\mathbf{x}_0$ and the system dynamics in **A**, thus, tracks whether the batch evolves according to NOC or not ("Did the batch start at normal conditions, and is it developing as expected?"). The dynamic representation in Equations 3 and 4 (or Equations 1 and 2 in the nonstochastic case) can further be used to predict progress of the system – by developing, in time, the equations starting from time zero ($\mathbf{x}_0$), an estimate of future states and measurement observations can be made.

6. RESULTS

A critical step in state space modeling is the order or rank selection, the number of states in the system. Different tools can be used for this, and as outlined in the Appendix here, we will use the singular values of the block Hankel matrix (which is built from time-shifted versions of the time series for each batch). As the eigenvalues of the covariance matrix can be used for decision on the number of PCA components, the idea is to inspect the singular values of the block Hankel matrix. Figure 2 presents the average singular values and the approximate 95% confidence limits on the basis of a Student's *t*-statistic for the Hankel matrix of the eight full NOC batch runs with six block rows. Figure 2 shows how the first two singular values are very stable, whereas the remaining four have a higher variance/uncertainty. From this plot, the system order is therefore deemed to be two, which also seems reasonable from the observation of the two main phases in the PCA score trajectory (a sigmoidal and an exponential decay, where the lag-phase is too short and is weakly present for our sampling rate of 36 s to capture). No significant difference in predictive performance of the model was observed for a rank three system (where one real and two complex eigenvalues were found for system matrix **A**), whereas a rank one model severely underperformed with a biased prediction (results not shown).

Equations 3 and 4 enable the prediction of future system outputs. A natural way of validating state space models is therefore to compare the predicted output to the actual system output. Figure 3 shows how the models fitted on each individual batch run handle the one-step-ahead prediction for the four test batches. Data is collected for $K = 1$ to *k;* state space modeling is performed on the collected data, and the one-step-ahead prediction is found using Equations 3 and 4. Figure 3 shows that the one-step-ahead predicted score vector is very close to the observed profile for all the batches, including the non-NOC batches. The order two state space models are thus good at capturing the essential dynamics and producing predictions over a short time horizon. It should, however, be remarked that only testing the one-step-ahead prediction is not very powerful because this may lead to very optimistic prediction errors. In order to test the longer time horizon predictions, the end-score value of each batch, taken here as 36 min into coagulation [15], is therefore

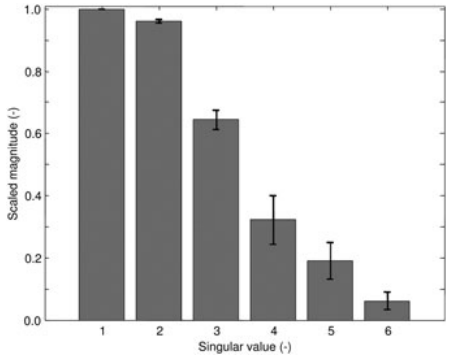


Figure 2. Scaled singular values of the block Hankel matrix for the normal operating conditions data.

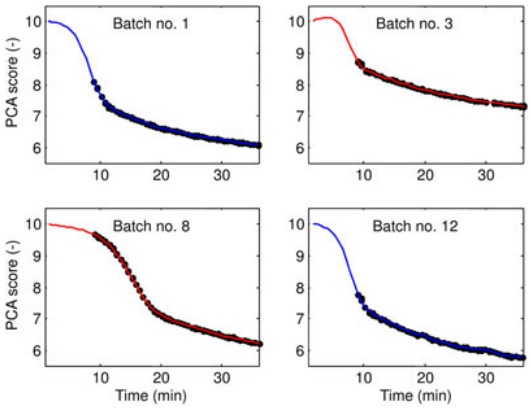


Figure 3. Observed score vector and one-step-ahead prediction (markers) for the four test batches.

predicted after each measurement point. The first prediction is therefore a 27-min horizon prediction (corresponding to a 45 steps-ahead prediction). At the next time step (36 s later), one more data point is obtained, a new system including the initial state identified and the end-value predicted from this information. Figure 4 shows the error for end-value prediction. A challenge when predicting more than one step ahead is that it is not possible to obtain an innovation (e_{k+n}) for future values in Equation 3. The best guess for future time points – used in Figure 4 – is e_k, the last known innovation as a substitute for the remaining time steps.

Several things can be noticed in Figure 4. As anticipated, the end-value estimate gets better as more and more measurements are available for fitting the model and less extrapolation is required, and already after approximately 12 min, an acceptable estimate of the batch end-value can be obtained for Batches 1, 3, and 12. The initial models have clear difficulties in predicting the end-value of the non-NOC Batch 8; the end-value cannot be predicted with a satisfactory small error until 20 min into the batch. This is caused by the longer lag-phase (not present in the NOC set) and delayed response for this batch.

SPC charts for the eigenvalues of **A** and the elements of $\mathbf{x}_0$ are implemented for process monitoring. These charts are shown in Figure 5 for the four test batches. It can be observed that the NOC batches stay inside or close to the proposed 95% confidence limits in all control charts, whereas the two deviating batches clearly break the limits for several of them. For example, it can be observed that both of the non-NOC batches already break the 95% CI in the control chart of the first eigenvalue of **A** after 9–10 min, clearly indicating that these two batches do not follow the NOC dynamics. The two deviating batches are also seen to exceed the control limits for $\mathbf{x}_0$, again indicating that these batches did not obtain the same estimated starting or boundary values as the training batches. It is also worth noticing that the confidence intervals for the first 10–12 min, on the basis of the NOC set, are fairly broad – reflecting the fact that the models based on the first few data points are obviously less well defined than the later models (especially for the exponential decay part representing coagulate hardening) [15].

In Figure 1(B), it was noticed that Batch 3 had a short sigmoidal (or an early onset of the exponential decay) and that Batch 8 was delayed when compared with the NOC batches. This is reflected in the control charts for the eigenvalues. Batch 3 in Figure 5 has an offset that places it outside the confidence limits, but nevertheless, follows the same trajectory as the NOC batches, whereas the slow kinetics of Batch 8 is seen as a delay for the first eigenvalue of **A**.

7. CONCLUDING REMARKS

In this work, state space models and subspace methods for system identification in a PAT application are suggested. It was shown that the subspace methods enabled state space modeling without *a priori* assumptions on model shape/form. In this sense, the subspace methods enabled the modeling to be data-driven rather than hypothesis-driven. The models were able to produce both good short and long time horizon predictions. It was furthermore shown that state space models are potential

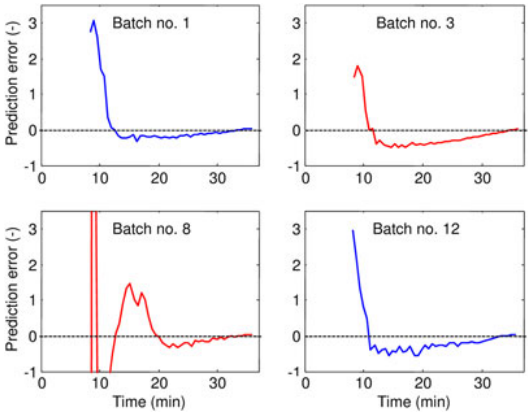


Figure 4. Prediction error for the end-value prediction of the four test batches.

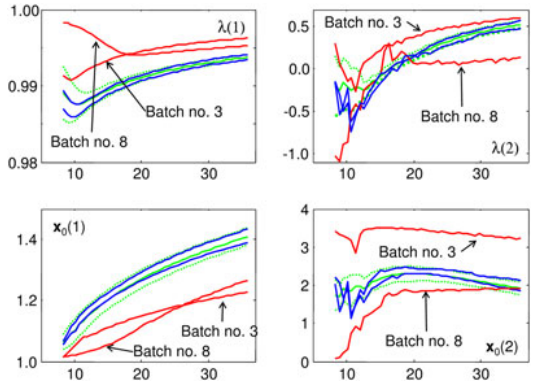


Figure 5. Statistical process control charts for eigenvalues of the **A** matrix and the initial state vector $\mathbf{x}_0$.

tools in process monitoring. Where conventional MSPC control charts reflects the process in a static manner, the control charts proposed in this work reflect the dynamic behavior of the process.

Acknowledgements

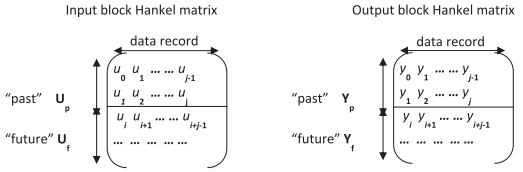
Christian Lyndgaard kindly provided the milk coagulation data. Jonas Thygesen's research is sponsored by the QbD-consortium (www.qbd.dk).

REFERENCES

1. U.S. Department of Health and Human Services Food and Drug Administration. *Guidance for Industry: PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. Food and Drug Administration: Rockville, Maryland, USA, 2004.
2. Laursen K, Rasmussen MA, Bro R. Comprehensive control charting applied to chromatography. *Chemom. Intell. Lab. Syst.* 2011; **107**: 215–225.
3. Qin SJ. Statistical process monitoring: basics and beyond. *J. Chemometrics* 2003; **17**: 480–502.
4. Kourti T, MacGregor JF. Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemom. Intell. Lab. Syst.* 1995; **28**: 3–21.
5. Skagerberg B, MacGregor JF, Kiparissides C. Multivariate data-analysis applied to low-density polyethylene reactors. *Chemom. Intell. Lab. Syst.* 1992; **14**: 341–356.
6. Roffel B, Betlem B. *Process Dynamics and Control Modeling for Control and Prediction*. John Wiley & Sons, Ltd.: West Sussex, UK, 2006
7. van Overschee P, De Moor B. *Subspace Identification for Linear Systems*, Kluwer Academic Publishers: Boston/London/Dordrecht, 1996.
8. Negiz A, Cinar A. PLS, balanced, and canonical variate realization techniques for identifying VARMA models in state space. *Chemom. Intell. Lab. Syst.* 1997; **38**: 209–221.
9. Hartnett MK, Lightbody G, Irwin GW. Dynamic inferential estimation using principal components regression (PCR). *Chemom. Intell. Lab. Syst.* 1998; **40**: 215–224.
10. Hartnett MK, Lightbody G, Irwin GW. Identification of state models using principal components analysis. *Chemom. Intell. Lab. Syst.* 1999; **46**: 181–196.
11. Ergon R. Dynamic system multivariate calibration. *Chemom. Intell. Lab. Syst.* 1998; **44**: 135–146.
12. Ergon R, Halstensen M. Dynamic system multivariate calibration with low-sampling-rate **y** data. *J. Chemometrics* 2000; **14**: 617–628.
13. Shi RJ, MacGregor JF. Modeling of dynamic systems using latent variable and subspace methods. *J. Chemometrics* 2000; **14**: 423–439.
14. Pan YD, Yoo C, Lee JH, Lee IB. Process monitoring for continuous process with periodic characteristics. *J. Chemometrics* 2004; **18**: 69–75.
15. Lyndgaard Ch, van den Berg F, Engelsen SB. Real-time modeling of milk coagulation. *J. Food Eng.* 2012; **108**: 345–352.
16. Dahm D, Lyndgaard Hansen Ch, Hopkins D, Norris K. NIR discussion forum: analysis of coagulating milk. *NIR News* 2010; **21**(5): 16–17.
17. Ljung L. *System Identfication – Theory for the User* (2nd edn). Prentice Hall PTR: Upper Saddle River, New Jersey, USA, 1999.
18. De Moor B. Mathematical concepts and techniques for modeling of static and dynamic systems, PhD thesis, Department of Electical Engineering, Katholieke Universiteit Leuven, Belgium, 1988.

APPENDIX

Different algorithms for state space modeling are available. One method, which is popular within the control engineering society, is the *prediction error method* (PEM). PEM has the advantage that any first principal knowledge on the system can be included during modeling [17]. But this is also the disadvantage of PEM algorithms – they are strongly dependent on the chosen parameterization. The main competitors of PEM are the subspace methods. One class of subspace methods is based on singular value decomposition (SVD). This means that these methods, as opposed to the PEM, are noniterative and require no other parameterization choice than the model order which can be estimated from the singular values of the input/output data [7]. Two decisions should be made during subspace modeling: the size of the Hankel matrices and the model order n. A Hankel matrix is symmetric and has the same elements across the off-diagonals. Written out for the input series (u_0, u_1, u_2 ... u_{i+j-1}) and corresponding output series (y_0, y_1, y_2 ... y_{i+j-1}), in Equations 1 and 2, the Hankel matrices would thus be [7]:



A similar Hankel matrix (effectively, a row and column time-shifted data representation) can be defined for the states series ($\mathbf{x}_0$, $\mathbf{x}_1$, $\mathbf{x}_2$ ... $\mathbf{x}_{i+j-1}$), where each entry is a vector of length n (the rank of the system), instead of a scalar. The separation between "past" and "future" data reflects how future inputs, outputs, and states can be regressed on past inputs, outputs, and states. The selection of the number of block rows (the "past" and "future" horizons) should be made so that i is larger than the expected system order n, whereas $i+j+1$ is determined by the length of the available training time series.

The input and output Hankel matrices can be combined in a block Hankel matrix $\mathbf{W}$. The "past" block Hankel matrix $\mathbf{W_p}$ would thereby, for example, be defined as [7]:

$$\mathbf{W_p} = \begin{pmatrix} \mathbf{U_p} \\ \mathbf{Y_p} \end{pmatrix}$$

The chemical/physical rank of $\mathbf{W_p}$ is an estimate of the true underlying number of dynamic components (which could be called eigenfrequencies) in the system. $\mathbf{W_p}$ can therefore be used to estimate the system order n. The block Hankel matrices for the observed data are closely related to the concepts of observablity and controlablity of the system states [7]. States can, in general terms, be said to be observable if they can be uniquely determined from the output y_k of the system. A useful system-related matrix is the observability matrix $\mathbf{\Gamma}$, defined as:

$$\Gamma = \begin{pmatrix} \mathbf{C} \\ \mathbf{CA} \\ \mathbf{CA}^2 \\ \dots \\ \mathbf{CA}^{j-1} \end{pmatrix}$$

If the rank of $\mathbf{\Gamma}$ is equal to n (number of elements in the state vector $\mathbf{x}_k$), then the system is observable. Another useful system-related matrix is the controllability matrix $\mathbf{\Delta}$. It is, as the name suggests, related to the controllability of the system. The system is controllable if it can be brought to any desired state by the input series u_k. The controllability matrix is defined as:

$$\Delta = \begin{pmatrix} \mathbf{A}^{j-1}\mathbf{B} & \mathbf{A}^{j-2}\mathbf{B} & \dots & \mathbf{AB} & \mathbf{B} \end{pmatrix}$$

The last system-related matrix that needs to be defined is the lower block triangular toeplitz matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{CAB} & \mathbf{CB} & \mathbf{D} & \dots & \mathbf{0} \\ \dots & & & & \\ \mathbf{CA}^{j-2}\mathbf{B} & \mathbf{CA}^{j-3}\mathbf{B} & \mathbf{CA}^{j-4}\mathbf{B} & \dots & \mathbf{D} \end{pmatrix}$$

It can be shown [18] that the original vector/matrix computations in Equations 1 and 2 can be reformulated in the following format by means of the system-related matrices as defined previously:

$$\mathbf{Y}_p = \mathbf{\Gamma X}_p + \mathbf{H}\ \mathbf{U}_p$$
$$\mathbf{Y}_f = \mathbf{\Gamma X}_f + \mathbf{H}\ \mathbf{U}_f$$
$$\mathbf{X}_f = \mathbf{A X}_p + \Delta\ \mathbf{U}_p$$

The different subspace algorithms available essentially solve this set of equations from which the **A**, **B**, **C**, and **D** matrices in Equations 1 and 2 for a user-defined rank *n* of the system are estimated. The term "subspace" refers to the fact that the first step in the algorithms is an oblique (or nonorthogonal) projection **O** of the "future" outputs ($\mathbf{Y}_f$) on the "past" block Hankel matrix $\mathbf{W}_p$ along the future outputs $\mathbf{Y}_f$. Singular SVD is then calculated on this weighted oblique projection: $\mathbf{G}_1\mathbf{O}\mathbf{G}_2 = \mathbf{\ddot{U}}\mathbf{S}\mathbf{V}^T$, where $\mathbf{G}_1$ and $\mathbf{G}_2$ are weights determined by the specific algorithm (where – in the case of a CVA solution – $\mathbf{G}_1$ contains the inverse square roots of the covariance estimate of the future outputs, and $\mathbf{G}_2$ is the identity [7]). $\mathbf{\ddot{U}}$ and **S** are then used to determine the observability matrix (**Γ**) by $\mathbf{\Gamma} = \mathbf{G}_1\mathbf{\ddot{U}}\mathbf{S}^{1/2}$. Because the oblique projection is equal to the product of **Γ** and the states ($\mathbf{X}_k$), is it possible to determine the states by $\mathbf{X} = \mathbf{\Gamma}^+\mathbf{O}$, where the Moore–Penrose pseudo inverse of the observability matrix is used. The boundary between "past" and "present" can then be shifted one step in order to determine the states at the next time step ($\mathbf{X}_{k+1}$), making the **A**, **B**, **C**, and **D** matrices the only unknowns in the system of linear equations that can thus be solved by least squares.

For the batch situation without input signals discussed in this manuscript, CVA–based stochastic Algorithm 3 from the book "Subspace Identification for Linear Systems" by Peter van Overschee and Bart de Moor [7] is used. The "past" block Hankel matrix is, in this case, equal to the "past" outputs ($\mathbf{Y}_p$), and the algorithm then follows the same flow as in the deterministic case: determine **O** from "future" outputs and the block Hankel matrix (="future" outputs ($\mathbf{Y}_f$)), determine the observability matrix (**Γ**) from the weighted **O**, determine the states by $\mathbf{X} = \mathbf{\Gamma}^+\mathbf{O}$, and solve the system of linear equations by least squares. The algorithm, furthermore, has the additional feature to produce positive real covariance sequences, making the solutions produced by the algorithm physically/chemically meaningful. The price to pay for this is a bias in the solution [7]. The equivalent of controllability for the stochastic system, Equations 3 and 4, is sometimes called reachability [11] (those latent states of the system that can be reached by the system dynamics and noise input), whereas the observability is sometimes substituted by detectability (those latent states of the system that can be observed/detected or "are excited by" the system dynamics, plus noise input).

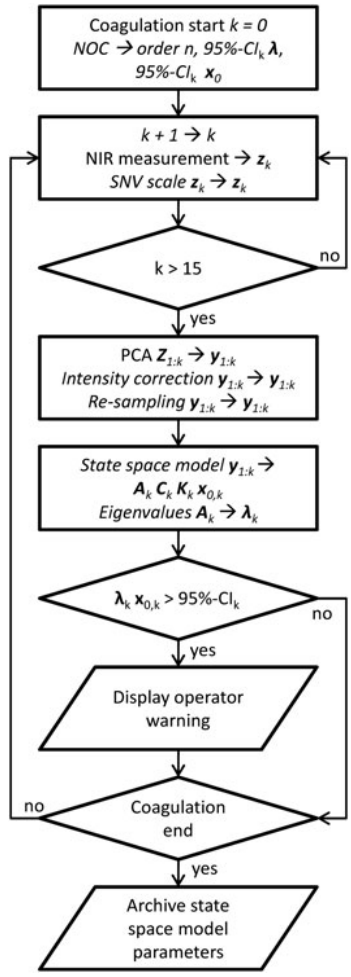


Figure A1. Dataflow for state space-based monitoring.

Paper III

Thygesen, J.H & F. van den Berg (2012)

Dynamic Model Based Monitoring of Batch Processes

Chemometrics & Intelligent Laboratory Systems, Submitted

Dynamic Model Based Monitoring of Batch Processes

Jonas Hoeg Thygesen* and Frans W.J. van den Berg, University of Copenhagen, Faculty of Science, Department of Food Science, Quality and Technology, Rolighedsvej 30, DK-1958 Frederiksberg C, phone: (+45) 3533 3500, fax: (+45) 3533 3245, e-mail: thygesen@life.ku.dk

Abstract

Typically only limited process knowledge is available for batch modeling and monitoring in industry. This paper suggests state space modeling estimated by subspace identification in combination with Kalman filters for application in real-time monitoring and prediction of batch trajectories without the need for such prior process knowledge. A model system of riboflavin (vitamin B2) breakdown is studied. Riboflavin is light sensitive and may, depending on pH, be broken down to the fluorescing compounds lumiflavin and lumichrome at different rates. Excitation–Emission Matrix (EEM) fluorescence spectroscopy is used to monitor the breakdown and it is shown how state space models plus Kalman filters can be used to monitor and predict the different breakdown dynamics. The state space/Kalman method is compared to Partial Least Squares (PLS) regression based monitoring.

Keywords: Subspace identification, state space models, Kalman filtering, dynamic model, process monitoring

1 Introduction

Modeling of batch data for process monitoring and control purposes has long been, and still is, of interest within chemometrics [1-5]. With the increasing demand for Model Predictive Control (MPC) solutions in the process industry [6] time series models are required. Frequently only limited process knowledge is available for modeling, especially in the food and biotechnology areas. Process models based on mechanistic or first principles are therefore seldom an option. This leads to the need of System Identification (SI) tools that allow modeling of batch processes based only on observations of system inputs and outputs. One subclass of such methods within the SI-toolset are state space models that may be found via so-called subspace methods [7-9]. The state space models have the advantage that they allow for easy implementation of the so-called Kalman filter. The Kalman filter is an optimal least squares error solution to the common challenge faced when applying process models: finding the best compromise between the measured process output at time point *k*, and the output predicted by the dynamic process model at this same time point *k*. Since both the measurement and prediction are hampered by noise and error there will be an intrinsic difference between the two. One method of finding the statistically optimal compromise between the two output estimates is via the Kalman filter [10-12]. This paper illustrates how batch trajectories, without any prior process knowledge, can be modeled, monitored and predicted via an approach based on state space models and Kalman filters. The proposed method is compared to a Partial Least Squares (PLS) regression for endpoint prediction, a method commonly applied in literature [2]. It is shown how the two methods in spite of different objectives may complement each other, with the state space methods capturing and describing the dynamics, and the PLS method predicting only the endpoint of the batch trajectories.

2 Theory

2.1 State space notation and the Kalman filter

State space models are linear, time-invariant relations between the physical inputs to the system at time k ($\mathbf{u}_k$), and the physical outputs (measurements) at time k ($\mathbf{y}_k$), connected via the vector $\mathbf{x}_k$ which contains the (most often) virtual or principal states of the system [7-9]. A discrete time state space model can be written via vector/matrix products as shown in Equations 1 and 2.

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\mathbf{u}_{k-1} + \mathbf{w}_k \qquad (1)$$

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k + \mathbf{v}_k \qquad (2)$$

Equation 1 is habitually referred to as the system equation (reflecting that it describes how the system evolves over time via the difference relationship of the state vector $\mathbf{x}_k$) while equation 2 is called the measurement equation (it describes how the measured output is related to the state of the system). The **A**-matrix is called the system matrix which describes how the system (or the states) evolves from one time-step to the next. The input matrix **B** explains how a control input at

time-step k would affect the system at k+1. **C** is referred to as the measurement matrix representing, as stated previously, how the states are reflected in the physically measured outputs ($\mathbf{y}_k$). A fourth matrix (**D**) is sometimes included in equation 2, called the (direct) feed-though, to explain how a control input (at time step k) can directly be observed in the output at time step k. This term is however seldom included in modeling, and is also not included in our work. An important note should be made on the system states ($\mathbf{x}_k$). Just as loadings and scores in a PCA model not necessarily correspond to e.g. pure compound spectra or concentrations, so do the system states not necessarily coincide with physical phenomena in the system (e.g. concentrations in a chemical reactor), they should instead be seen as a latent representation of the dynamics spanning the subspace of relevance for the system. The matrices are in this study identified during a training phase where subspace identification algorithms are applied to a training data set consisting of Normal Operating Condition (NOC) batches. More details on the state space models and subspace identification algorithms can be found in [7].

$\mathbf{w}_k$ and $\mathbf{v}_k$ in equations 1 and 2 are noise sequences assumed to follow a normal distribution with $\mathbf{w}_k \sim N(0,\mathbf{Q})$ and $\mathbf{v}_k \sim N(0,\mathbf{R})$. This indicates that both the system (equation 1) and the output measurements (equation 2) are affected by uncertainty at each time step. The dynamic nature of the state space models are in this study utilized via implementation of a Kalman filter - a so-called optimal linear observer - for the prediction of future system outputs. It combines the noise corrupted measurements of the system output ($\mathbf{y}_k$) with the predicted system output ($\mathbf{C}\mathbf{x}_k$) in a statistical optimal manner [13].

Equation 1 makes it possible to estimate the state at time-step k ($\hat{\mathbf{x}}_k^-$) from the previous state ($\mathbf{x}_{k-1}$) and the previous input ($\mathbf{u}_{k-1}$), this estimate is known as the *a priori* estimate at time step k (indicated by the "super minus"). The Kalman filter combines the noisy measurements $\mathbf{y}_k$ with the predicted system output $\mathbf{C}\mathbf{x}_k$ by finding the *a posteriori* system state as a linear combination of the *a priori* system state and a weighted difference between measured $\mathbf{y}_k$ and anticipated response $\mathbf{C}\hat{\mathbf{x}}_k^-$.

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-) \qquad (3)$$

The difference ($\mathbf{y}_k - \mathbf{C}\hat{\mathbf{x}}_k^-$) (also known as the innovation) can be computed straight forward, and reflects how large the agreement between the actual system output measurement and the system model is.

The estimation errors for the system states are given as:

$$\mathbf{e}_k^- = \mathbf{x}_k - \hat{\mathbf{x}}_k^- \qquad (4)$$

$$\mathbf{e}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k \qquad (5)$$

With the *a priori* $\mathbf{P}_k^-$ and *a posteriori* $\mathbf{P}_k$ covariances defined as:

$$\mathbf{P}_k^- = E[\mathbf{e}_k^- \mathbf{e}_k^{-T}] \quad (6)$$

$$\mathbf{P}_k = \mathrm{E}[\mathbf{e}_k\, \mathbf{e}_k^{\mathrm{T}}] \qquad (7)$$

$\mathbf{K}_k$ in equation (3) is known as the Kalman gain; it is chosen so that the *a posteriori* system state error covariance ($\mathbf{P}_k$) is minimized. One common definition of the Kalman gain is [13]:

$$\mathbf{K}_k = \mathbf{P}_k^{-}\, \mathbf{C}^{\mathrm{T}}(\mathbf{C}\, \mathbf{P}_k^{-}\mathbf{C}^{\mathrm{T}} + \mathbf{R})^{-1} \qquad (8)$$

Where $\mathbf{R}$ is the measurement noise covariance (associated with $\mathbf{v}_k$ in equation 2). Equation (8) illustrates how the Kalman filter balances the error covariances to give weight to either the actual measurement ($\mathbf{y}_k$) or the predicted measurement ($\mathbf{C}\hat{\mathbf{x}}_k^{-}$). If the measurement noise covariance ($\mathbf{R}$) is small (the measurements are trusted), the Kalman gain becomes large and equation 3 subsequently weights up the innovation, driving the *a posteriori* system state away from the predicted measurement ($\mathbf{C}\hat{\mathbf{x}}_k^{-}$) towards the actual measurement ($\mathbf{y}_k$). The opposite is of course also the case, the *a priori* system state error covariance ($\mathbf{P}_k^{-}$) depends on the system noise covariance matrix $\mathbf{Q}$ (see equation 9 below). If $\mathbf{Q}$ is small, $\mathbf{P}_k^{-}$ will also be small, resulting in a likewise smaller Kalman gain, meaning that the predicted measurement will be trusted more. Using the right estimates for $\mathbf{Q}$ and $\mathbf{R}$ (or rather the relative size ratio) is therefore of key importance to obtain the right Kalman filter estimates. And while the measurement noise covariance ($\mathbf{R}$) is often known or easily estimated, the process noise covariance matrix ($\mathbf{Q}$) is less easily available since the states in $\mathbf{x}_k$ themselves are estimates that are not directly observed. Mehra [14,15] showed that a sub-optimal estimate can be obtained as:

$$\mathbf{Q} = \mathbf{P}_o - \mathbf{A}(\mathbf{I}-\mathbf{K}_o\mathbf{C})\mathbf{P}_o\mathbf{A}^{\mathrm{T}} \qquad (9)$$

with $\mathbf{P}_o$ being the initial error covariance matrix for the system states and $\mathbf{K}_o$ the initial Kalman gain. Procedures for better estimates of $\mathbf{Q}$ have been published (e.g. Odelson *et. al.* (2006) [16] and Rajamani & Rawlings (2009) [17]), but they do not provide a simple closed form expression.

The approach presented in equation (9) is applied in this paper. The *a priori* error covariance matrix ($\mathbf{P}_k^{-}$) can at time step k be found from the Lyapunov function:

$$\mathbf{P}_k^{-} = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^{\mathrm{T}} + \mathbf{Q} \qquad (10)$$

With $\mathbf{Q}$ being the covariance matrix for the system noise sequence $\mathbf{w}_k$ in equation (1). It is after a measurement updated to the *a posteriori* error covariance ($\mathbf{P}_k$) by:

$$\mathbf{P}_k = (\mathbf{I}-\mathbf{K}_k\mathbf{C})\, \mathbf{P}_k^{-} \qquad (11)$$

Where $\mathbf{I}$ is the identity matrix, and $\mathbf{K}_k$ the Kalman gain from equation (8). Based on the *a posteriori* error covariance the output covariance matrix $\mathbf{S}_k$ is found by [14]:

$$\mathbf{S}_k = \mathbf{C}\mathbf{P}_k\mathbf{C}^{\mathrm{T}} + \mathbf{R} \qquad (12)$$

It can be shown that under the assumption that the process and measurement noise are normal, the state estimates are also normal [12]. Furthermore it is well known that a linear transformation

of a normally distributed process, results in a process that is also normally distributed. The vector of (1-α) confidence intervals $\mathbf{ci}_k$ for the predicted output at time k can therefore be found as:

$$\mathbf{ci}_k = \hat{\mathbf{y}}_k \pm \Phi_{1-\alpha/2} \sqrt{\text{diag}(\mathbf{S}_k)} \qquad (13)$$

Where $\Phi_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution and diag($\mathbf{S}_k$) the diagonal elements of $\mathbf{S}_k$ [18].

In our case study described in the remainder of the paper discrete time monitoring is prefomed for $k_{start} = 3$ to $k_{end} = 89$. Based on the theory described so far an overall monitoring cycle consisting of the five steps listed below is implemented:

0) Identify system - obtain estimates of **A**, **B**, **C**, $\mathbf{K}_o$, $\mathbf{P}_o$, **Q** and **R** during the training phase from full NOC batches.

1) Initial conditions - estimate $\mathbf{x}_o$, based on $k = 1$ and $k = 2$ [7], $k = k_{start}$

2) Predict state - the *a priori* state estimate $\hat{\mathbf{x}}_k^-$ is found using equation (1), and the error covariance is estimated by equation (10)

3) Perform measurement- obtain $\mathbf{y}_k$

4i) Correct measurement - computing the Kalman gain $\mathbf{K}_k$ by equation (8), use this gain to update the measurement $\hat{\mathbf{x}}_k$ by equation (3), update the *a posteriori* error covariance $\mathbf{P}_k$ by equation (11)

5) If $k < k_{end}$ then $k = k + 1$, return to 2) else end

For the simulation of future system outputs a four-step inner-cycle is included; in order not to mix symbols, the future time-steps are denoted by g:

4ii) $g = k + 1$

4iii) Predict state – the state ahead ($\hat{\mathbf{x}}_g$) is predicted using equation (1), and the error covariance $\mathbf{P}_g$ is predicted by equation (10)

4iv) Translate output - the output $\mathbf{y}_g$ corresponding to the predicted state found using equation (2), and the predicted output confidence intervals $\mathbf{ci}_g$ are found by equation (12) and equation (13)

4v) if $g < k_{end}$ then $g = g + 1$, return to 4iii) else return to 2)

2.2 Partial Least Squares for endpoint prediction

Multivariate Statistical Process Control (MSPC) is a mature research area for which many models and applications are described in literature. A commonly used method for predicting endpoint quality of batches was presented in 1995 by Nomikos & MacGregor [2]. In the case that a number ($j = 1, 2... J$) of process variables are followed over time ($k = 1, 2... K$) for several batches ($i = 1, 2... I$), a

three-way tensor of the observed data $\underline{\mathbf{Y}}$ ($I \times J \times K$) can be formed. The method is based on collecting the corresponding quality variables (m = 1, 2... M) in matrix $\mathbf{Z}$ ($I \times M$), unfolding the observed training data in the batch direction to obtain $\mathbf{Y}$ ($I \times JK$) and regressing the autoscaled $\mathbf{Y}$ on $\mathbf{Z}$ using PLS2 [2].

A challenge when applying the method on-line for new batches is the fact that not all data points for the new vector are available. E.g. if the total batch run is 89 steps long (K = 89) and we are currently at k = 20 the remaining 69 time steps are yet unknown. Different methods are available for estimating the process outputs during the remaining time steps. Nomikos & MacGregor [19] showed that simply setting the remaining time steps to missing, and using the ability of PLS to estimate the missing data is the best in a range of methods compared. It does though require the trajectories not to exhibit frequent discontinuities and approximately 10% of the batch history needs to be recorded before reliable results can be obtained [19,20]. It should be kept in mind that though both the state space/Kalman method and the PLS method have the aim of modeling batch data, the objective of the two methods are quite different. Where the first method has the goal of capturing and modeling the dynamics via the system matrices [8], the unfolding and autoscaling in the latter method has according to Nomikos & MacGregor the objective of "removing the main non-linear and dynamic components in the data" [19]. This has the consequence that different questions can be answered with the two methods. One common question to ask is whether the process is on track or not ("where are we now?"). Control charts of the observed outputs may help in answering this question but do not necessarily reflect if the dynamics are behaving according to NOC, while control charts of the states and dynamics can [7]. Another question that could be of great interest is what the endpoint quality of the batch is going to be ("where are we going?"). The PLS method outlined above is especially suited for predicting the endpoint quality, but does not include predictions on how the batch will evolve. This on the other hand is the aim of the state space/Kalman filter model, via a combination of model predictions and observations the batch trajectory is predicted for the remaining time steps. A final question that may be of interest is whether the initial conditions for the process were within the specifications ("where did we start from?"). PLS – or any other regression algorithm - may be used for predicting the initial conditions, but where the state space model directly gives initial condition estimates, a separate PLS regression model would be required for predictions of the initial conditions because model inversion is not obvious.

3 Material and methods

3.1 Model system design

Riboflavin (vitamin B2) is bright yellow and can in solution be quantified by means of excitation-emission (EEM) fluorescence spectroscopy [21,22]. The vitamin is relatively heat stabile but may, depending on pH, be hydrolysed into lumichrome and lumiflavin when light is present. Lumichrome formation is favored at neutral or acid pH while lumiflavin formation is favored under basic conditions [23]. Multivitamin effervescent tablets (Vitafit Multivitamin) were bought

from a local grocery store (Lidl Stiftung & Co. KG, Necklarsulm, Germany). One tablet weighs 4.5 g (random selected from 6 containers to induce natural variation) with a content of approximately 1.6 mg riboflavin among other vitamins. One tablet was dissolved in 600 mL water and the solution was allowed to settle for 2 minutes before 20 mL 2 M NaOH was added (regulated to pH ≈ 10). Magnetic stirring was applied and three white LED light sources were constantly used as light source during the experiments. Each batch was measured for approx. 60 min and a total of 76 batches were measured over 8 days. 57 of these batches were selected for modeling and test, 44 batches were recorded under NOC, the remaining 13 batches were manipulated by turning off one or more of the light sources for short periods, changing the amount of tablet material or amount of NaOH added. The manipulations were carried out in order to yield modestly trajectories different from the NOC. The remaining 76 - 57 = 19 non-NOC batches have deliberately induced gross errors (e.g. no light sources) and are designed for early detection strategies not pursued in this paper. Instead we focus on the identification of more subtle non-NOC behavior.

3.2 Online monitoring

Temperature (uncontrolled, corresponds to room temperature) and pH in the reactor vessel were continuously logged during the experiments (MadgeTech pHTemp2000 Data Logger, MadgeTech, Contoocook, NH, USA). Fluorescence EEMs were recorded using a BioView EEM fluorescence process spectrometer (DELTA Danish Electronics, Light & Acoustics, Hørsholm, Denmark). The EEMs were recorded using a combination of 11 excitation filters (λ_{ex}; equidistantly spaced from 330 to 530 nm, positioned on the light source sequentially) and 11 emission bands filters (λ_{em}; equidistantly spaced from 370 to 570 nm, positioned on the detector sequentially). A sampling rate of 30 sec/EEM was used. Concentrations of the EMM components are expressed as PARAFAC scores [24].

3.3 Data modeling

State space models were fitted using the n4sid algorithm from the System Identification Toolbox version 7.4.1 in MATLAB R2010b (The Mathworks Inc., Natick, MA, USA). PARAFAC modeling was done using the N-way Toolbox for MATLAB (http://www.models.life.ku.dk/source/nwaytoolbox/ [24]). All other computations were also done in MATLAB using in-house routines.

4 Results and discussion

4.1 Data inspection and training/test set formation

A four component PARAFAC model was fitted for each batch, a representative selection of the resulting scores and spectral loadings are shown in Figure 1 and Figure 2, respectively. Marked in bold are the NOC batches 5 and 29, plus the non-NOC batches 45 (less NaOH added) and 74 (less tablet added). Since the spectral loadings in PARAFAC are normalized to unit length we can safely use the time-scores as (pseudo) concentrations of the four chemical components relying on the uniqueness property [24].

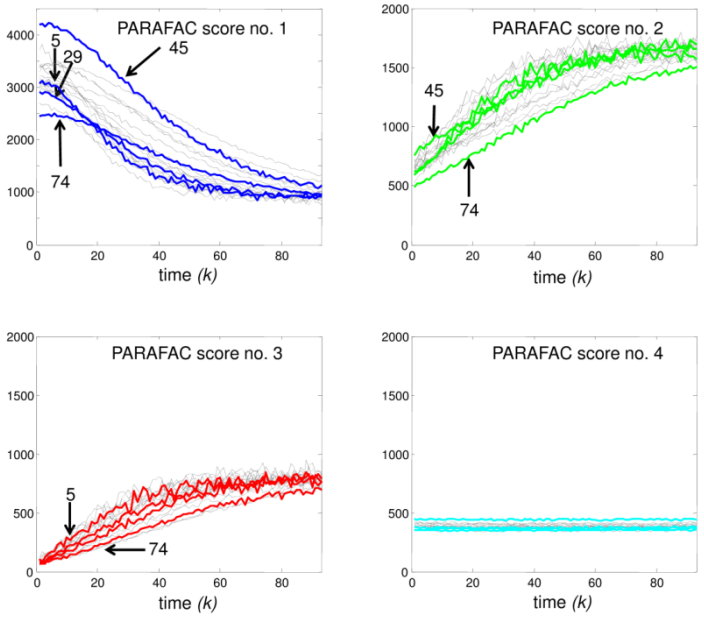


Figure 1 Concentration-scores from a four component PARAFAC model, one model was made for each batch.

It can be seen that one of the PARAFAC components decreases over time, two grow and one stays constant. An inspection of the emission and excitation loadings reveals that the decreasing component corresponds to riboflavin (λ_{ex} 450nm / λ_{em} 520nm [25]). The first of the two increasing components with a slightly shifted peak as compared to the riboflavin peak corresponds to lumiflavin since it is known to be yellow as well but has a peak maximum shifted to slightly lower emission and excitation wavelengths [23]. Lumichrome is from literature known to have a maximum peak at approximately λ_{ex} 360nm / λ_{em} 450nm [23], which corresponds nicely to the last increasing component. The fourth (stable) component is unknown.

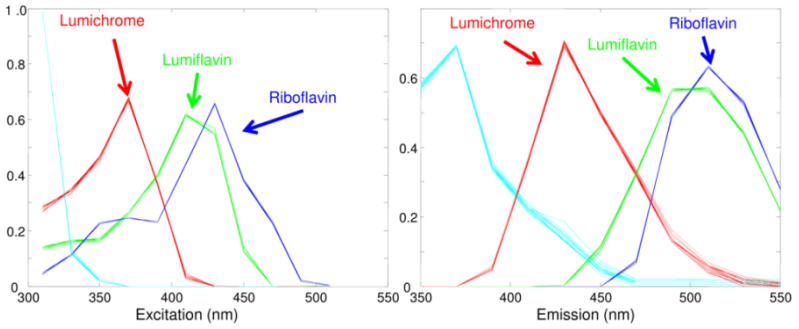


Figure 2 Spectral-loadings from a four component PARAFAC model, one model was made for each batch.

A close inspection of Figure 1 also illustrates that the one of the increasing components (lumiflavin) is formed at a slightly higher rate than the other increasing component (lumichrome). This makes sense seen from a chemical viewpoint since the formation of lumiflavin at pH 10 should be favored over lumichrome. Furthermore it can be observed that the PARAFAC scores for the non-NOC batches especially are deviating for the first and second PARAFAC scores. For batch 74 it is also worth noticing that lower scores in general are seen, corresponding nicely to a lower concentration of both riboflavin and reaction products.

Because the riboflavin hydrolysis reaction is pH-dependent and fluorescence response is known to be temperature dependent [26], the PARAFAC scores will be modeled for the state space case with the three significant scores as outputs ($\mathbf{y}_k$), and the easily measureable reaction solution pH and temperature as inputs ($\mathbf{u}_k$), plotted in Figure 3 for the selected number of batches.

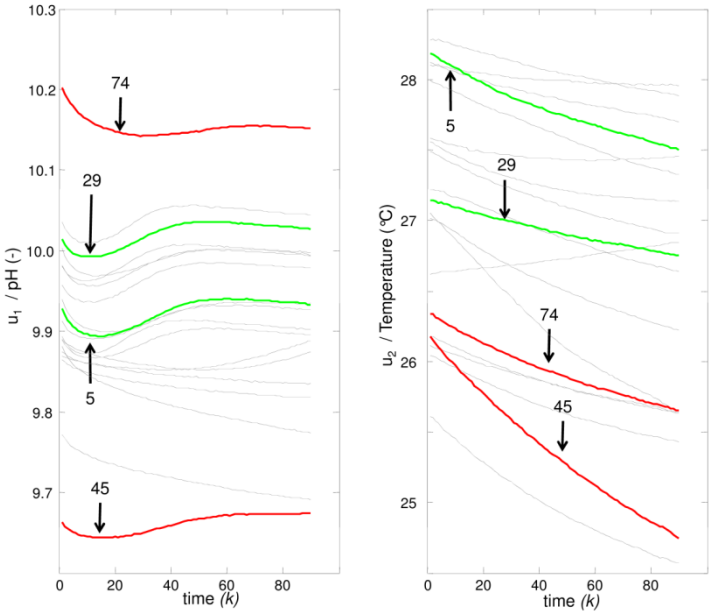


Figure 3 pH and temperature of reaction solutions measured over time.

44 NOC batches were recorded; every fourth was used as test data together with the selected non-NOC batches (24 in total) the remaining 33 NOC batches were used as training data.

4.2 State space modeling and Kalman filtering

In order to determine the number of underlying dynamic components (the order of system matrix **A** in equation 1), time shifted matrices of the input and output data are formed from the training NOC batches, a so-called block Hankel matrices (for more details see Thygesen and van den Berg [7]). An indication of the correct number of system states that can be observed in the data can be found by inspecting the singular values of this Hankel matrix, just as the eigenvalues of the

covariance matrix can be used for deciding the number of PCA components. Inspection of these singular values (Figure 4) shows that a fourth order state space model is suitable.

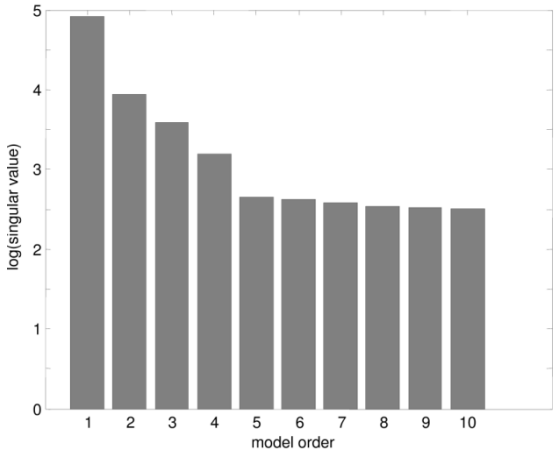


Figure 4 Logarithm of singular values of the block Hankel matrix based on training NOC data.

The state space equation allows prediction of future system outputs (i.e. future riboflavin, lumiflavin and lumichrome score values), this prediction is further enhanced by implementation of a Kalman filter. The filter requires, as outlined in the theory section, estimates of the process noise covariance matrix (**Q**) and the measurement noise covariance matrix (**R**), estimated from the NOC training data. When using the System Identification toolbox in MATLAB for fitting the state space models, an estimate of the measurement noise covariance matrix is directly available. This covariance matrix is used as **R** where the diagonal elements are the assumed value for the variance of the individual measurements (i.e. the different PARAFAC scores), and the square root of the elements are thus the standard deviations on the measured scores. The standard deviations are in this case equal to: 57.1 (riboflavin), 32.6 (lumiflavin) and 29.9 (lumichrome; compare with Figure 1). The process noise covariance matrix (**Q**) is estimated by the method proposed by Mehra, equation (9) [14,15]. The initial Kalman gain $\mathbf{K}_0$ is (just as **R**) available from the estimated model. After fitting the state space model on the training data, the initial state of each batch ($\mathbf{x}_0$) is estimated by finding the estimate that minimizes the prediction error. The variance of this sequence of estimates is subsequently used to form $\mathbf{P}_0$ as a diagonal matrix with the estimated variances as its elements.

Figure 5 and Figure 6 presents the Kalman filter estimates and predictions of riboflavin, lumiflavin and lumichrome PARAFAC scores at time steps k = 3, 20, 40 and 60 for NOC batch 5 and non-NOC test batch 45, respectively.

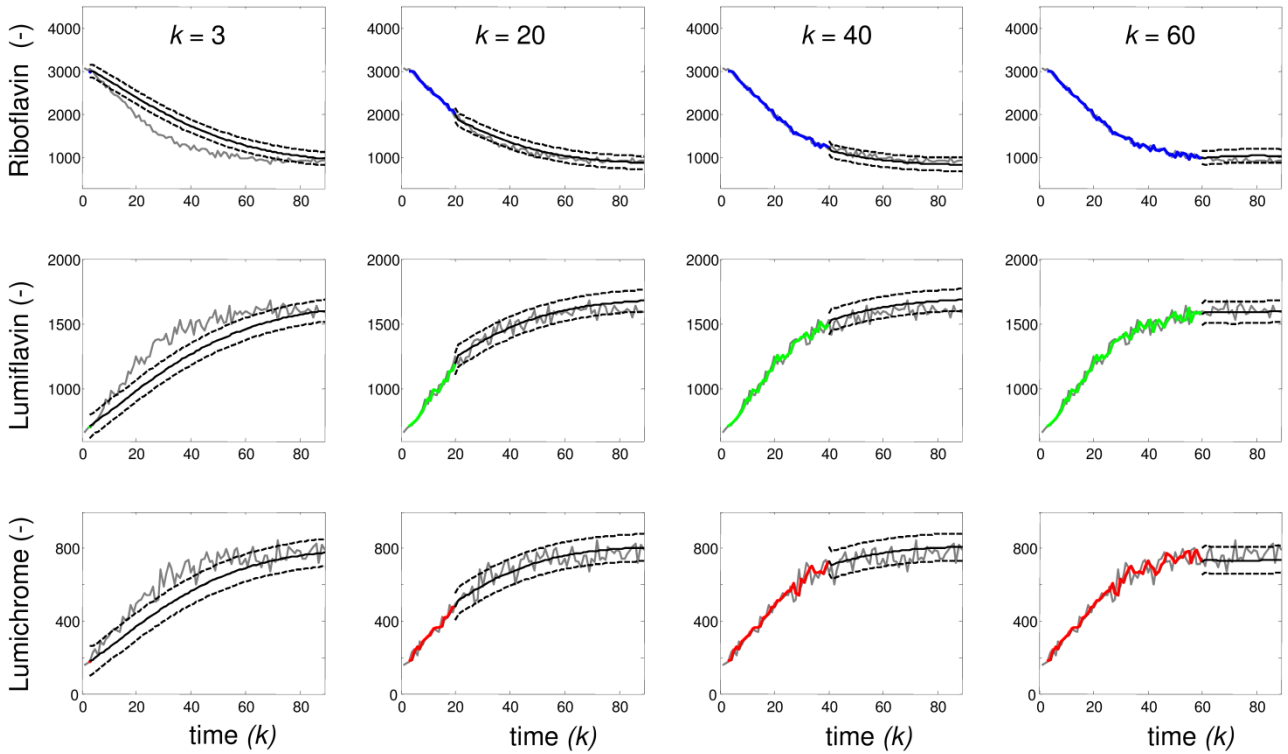


Figure 5 Kalman filter estimates and predictions of future riboflavin, lumiflavin and lumichrome PARAFAC scores at time steps *k* = 3, 20, 40 and 60 for the NOC batch 5.

Several things can be noticed in Figure 5. First of all it shows that the state space model overall captures the system dynamics well. At $k = 3$ a small bias is seen between the predicted and observed system outputs for the middle part of the curve, likely due to a inaccurate estimate for the batch boundary conditions $\mathbf{x}_0$, but the overall trajectory is the same. One would of course not have access to the future outputs in a monitoring situation, but the comparison is nevertheless very valuable as a validation tool for the models at hand. Furthermore, the adaptive nature of Kalman filters can also be noticed. At $k = 3$ a bias is seen, after observing the next data points and correcting the states correspondingly the bias is removed, and there is a good correspondence between the predicted and the observed system outputs. Finally Figure 5 also indicates why the term Kalman *filtering* is used; the noisy measurements are passed through a filter whereby the noise or erratic jumping is reduced, the resulting concentration-profiles thereby appear as a smoother trajectory that corresponds well to the trajectory that one would intuitively expect based on chemical insight in the absence of noise. This illustrates that the Kalman filter essentially is a statistically optimal compromise between the observed trajectory and the trajectory predicted by a model.

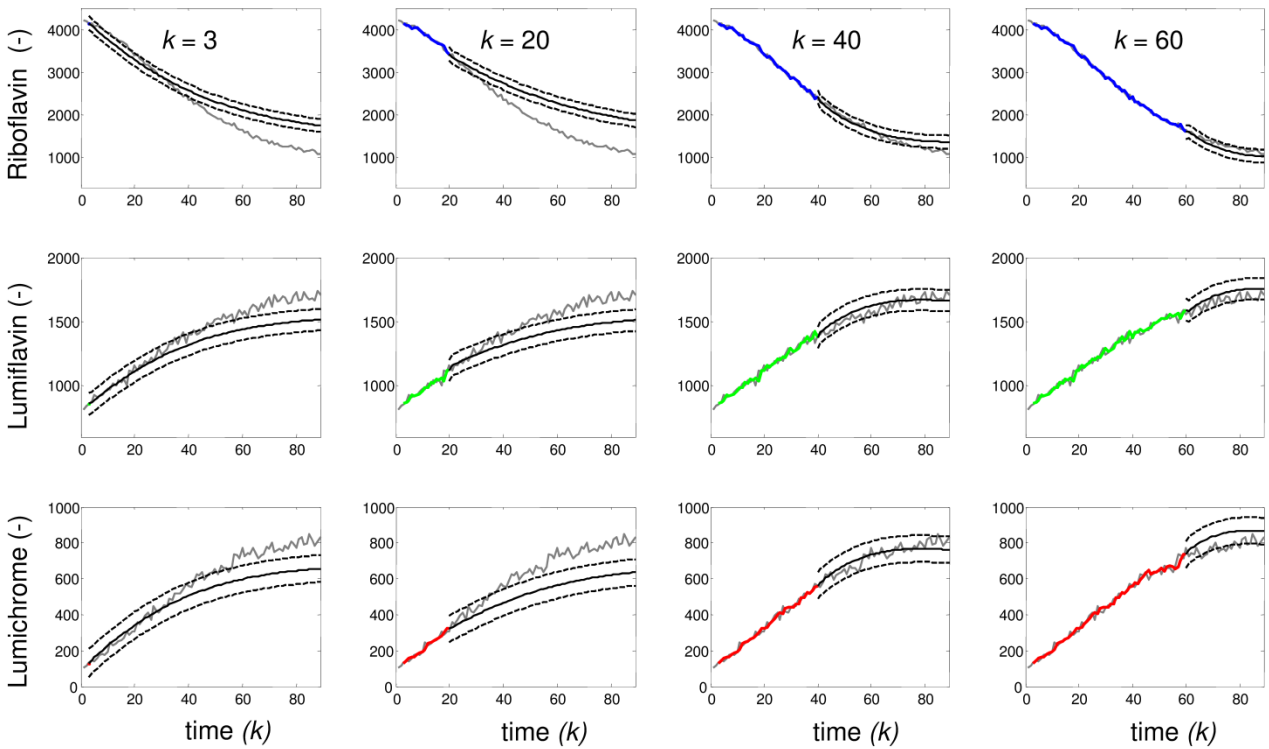


Figure 6
Kalman filter estimates and predictions of future riboflavin, lumiflavin and lumichrome PARAFAC scores at time steps *k* = 3, 20, 40 and 60 for the non-NOC batch 45.

Also in Figure 6 for non-NOC batch 45 a bias is seen, especially for the riboflavin score trajectory. The model clearly over estimates the riboflavin scores as a result of the higher starting values and different dynamics for this non-NOC run. It is however able to correct the predictions as more measurements become available.

4.3 Dynamic control charts

Despite not being directly interpretable as physical quantities the state space model does suggest control charts for the individual states. This chart for the states - the principal behavior of the system - thus reflects the dynamic behavior of the system. If the outputs are evolving according to NOC the states will fall inside the control limits; if the dynamic are different (e.g. faster or slower) the states will fall outside the control limits. Control charts for the four states were made based on the training set (Figure 7). The 95% CI for state i at time k ($ci_{i,k}$) was found as:

$$ci_{i,k} = \bar{x}_{i,k} \pm 1.96\, \sigma_{i,k} \qquad (13)$$

Where $\bar{x}_{i,k}$ is the mean of state i at time k and $\sigma_{i,k}$ the corresponding standard deviation, based on the training NOC data.

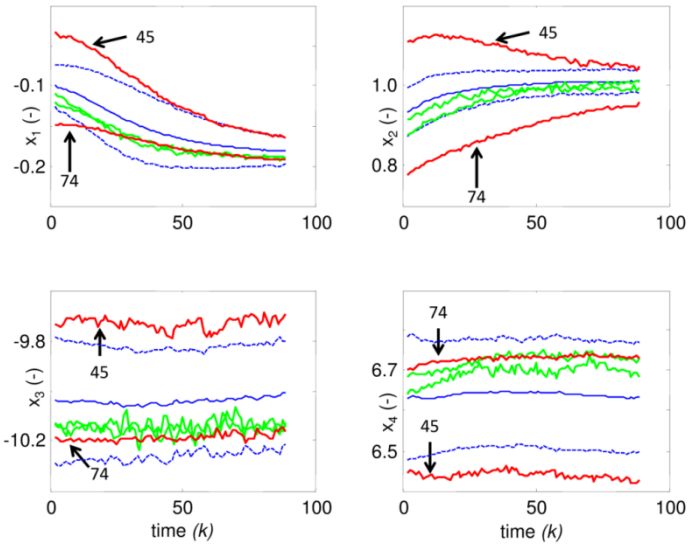


Figure 7 Control charts for states (x_k) of NOC batches 5 and 29 (not identified) and non-NOC batches 45 and 74.

In Figure 7 it is seen that while the NOC test batches fall within the 95% CI for all four states at all times, the two non-NOC batches clearly break one or more of the limits right from the start. In an industrial setting this information would be available to operators already after a few minutes and if desired a corrective action (e.g. dosing extra base to the vessel) could be taken once sufficient confidence is there. It can however also be noticed in Figure 7 that the non-NOC batches slowly

converges towards NOC behavior (this is especially the case for states 1 and 2). This reflects that the same chemistry is present in all batches in spite of the different starting conditions.

4.4 Endpoint prediction

It was shown in the theory section how the state space model in combination with the Kalman filter was able to capture the process dynamics and predict the system output ahead in time. Here endpoint predictions will be compared to a dedicated method such as PLS. The endpoint was in this case taken to be the average PARAFAC scores for k = 85 to k = 89 for the three analytes (riboflavin, lumiflavin and lumichrome). During the training phase a 5 latent variable model was found optimal for the PLS. In order to assess the fit the Root Mean Square of Prediction (RMSEP) found at selected time step was computed, the results are presented in Table 1.

Table 1 RMSEP at selected time steps for endpoint prediction by the state space/Kalman method and the PLS method.

RMSEP Kalman/PLS	k = 3	k = 20	k = 40	k = 60
NOC batches (N = 11)				
Riboflavin	142 / 144	138 / 88	112 / 66	125 / 38
Lumiflavin	66 / 76	62 / 44	48 / 37	50 / 27
Lumichrome	42 / 29	41 / 21	38 / 19	45 / 18
Non-NOC batches (N = 13)				
Riboflavin	339 / 135	287 / 111	164 / 70	124 / 28
Lumiflavin	127 / 77	101 / 44	55 / 45	44 / 31
Lumichrome	94 / 55	86 / 57	49 / 45	42 / 32

Table 1 shows how the dedicated PLS method is better at estimating the endpoint. It is worth noticing that the PLS predictions improves dramatically going from k = 3 to k = 20 steps. This is of course a result of the shorter prediction horizon but also corresponds to the knowledge from literature that approximately 10% of the batch history (in this case $k \approx 9$) should be known before reliable results are to be obtained.

Figure 8 gives a closer assessment of the riboflavin predictions for two selected batches in the form of a control chart of the endpoint predictions: NOC batch 29 (Figure 8A) and the non-NOC batch 45 (Figure 8B). As process target the mean of the NOC training data is given together with 95% CI (found by equation 13), plus the actual endpoint of the batch.

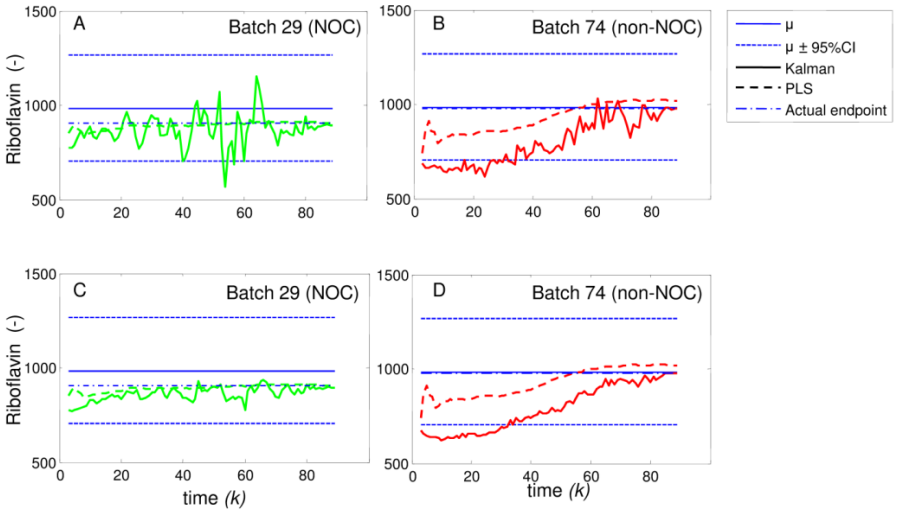


Figure 8 Control charts for endpoint prediction. A: NOC batch 29, B: non-NOC batch 74, C: NOC batch 29 with 20 times reduced process noise covariance matrix Q for Kalman filter, D: non-NOC batch 74 with reduced Q.

While both PLS and the Kalman predictions are close to the actual endpoint already from the beginning for the NOC batch, more data points are needed for the non-NOC before accurate endpoint predictions can be made. Figure 8A and 8B also illustrates that fairly noisy (or jerky) endpoint predictions are obtained by the Kalman method for this particular NOC batch. This is the result of the choice of process- and measurement noise covariance matrix (**Q** and **R**). A sub-optimal estimate of **Q** was applied in this paper based on theory not explicitly on training experience, and it is known that the estimates of **Q** are pessimistic [16]. It would therefore be possible to adjust the filter to suppress this noise by tuning the ration between **Q** and **R**. This adjustment in SPC is normally based on validation performance indicators of the control charts ones in use. One illustration is given in Figure 8C and 8D where a twenty times reduction of **Q** is used. This comes however with the price of trusting the model more and therefore with the risk of delays in capturing deviating behavior in a statistical monitoring situation.

5 Conclusions

In this paper it was shown how a combination of state space models and Kalman filters can serve as a versatile tool in batch process modeling and monitoring. The proposed method was able to capture and model the dynamics of a batch process. The method was also shown to be adaptable to new non-NOC conditions. The method allowed for dynamic control charting of initial condition estimates, current system-states as well as predictions of process variable future trajectories. For endpoint prediction a dedicated method based on Partial Least Squares was found to produce slightly better predictions.

6 Acknowledgements

Jonas Thygesen's research is sponsored by the QbD-consortium (www.qbd.dk) which is partially funded by the Danish Academy for Science Technology and Innovation.

7 References

[1] T. Kourti and J. F. MacGregor, Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods, Chemom. Intell. Lab. Syst., 28 (1995) 3-21.

[2] P. Nomikos and J. F. MacGregor, Multi-way partial least squares in monitoring batch processes, Chemom. Intell. Lab. Syst., 30 (1995) 97-108.

[3] S. Wold, N. Kettaneh, H. Friden, and A. Holmberg, Modelling and diagnostics of batch processes and analogous kinetic experiments, Chemom. Intell. Lab. Syst., 44 (1998) 331-340.

[4] J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, Generalized contribution plots in multivariate statistical process monitoring, Chemom. Intell. Lab. Syst., 51 (2000) 95-114.

[5] S. Rannar, J. F. MacGregor, and S. Wold, Adaptive batch monitoring using hierarchical PCA, Chemom. Intell. Lab. Syst., 41 (1998) 73-81.

[6] J. H. Lee, Model Predictive Control: Review of the Three Decades of Development, Int. J. Control. Autom. Syst., 9 (2011) 415-424.

[7] J. H. Thygesen and F. W. J. van den Berg, Subspace methods for dynamic model estimation in PAT applications, J. Chemom., Submitted (2011).

[8] P. van Overschee and B. De Moor, Subspace Identification for Linear Systems, KLUWER ACADEMIC PUBLISHERS, Boston/London/Dordrecht 1996.

[9] L. Ljung, System Identfication - Theory for the User, 2nd, Prentice Hall PTR, Upper Saddle River, New Jersey 1999.

[10] S. D. Brown, The Kalman Filter in Analytical Chemistry, Anal. Chim. Acta, 181 (1986) 1-26.

[11] R. E. Kalman, A New Approach to Linear Filtering and Prediction Problems, Trans. ASME - J. Basic. Eng., 82 (1960) 35-45.

[12] Welch, G. and Bishop, G., An Introduction to the Kalman Filter, http://www.cs.unc.edu/~welch/media/pdf/kalman_intro.pdf, Last updated: 24-7-2006, Accessed: 6-1-2012

[13] P. S. Maybeck, Introduction, Stocastic models, estimation, and control, Vol. 1. Academic Press, Inc., New York, New York, 1979, pp. 1-16.

[14] R. K. Mehra, Approaches to Adaptive Filtering, IEEE Trans. Autom. Control., 17 (1972) 693-698.

[15] R. K. Mehra, On the Identification of Variances and Adaptive Kalman Filtering, IEEE Trans. Autom. Control., 15 (1970) 175-184.

[16] B. J. Odelson, M. R. Rajamani, and J. B. Rawlings, A new autocovariance least-squares method for estimating noise covariances, Automatica, 42 (2006) 303-308.

[17] M. R. Rajamani and J. B. Rawlings, Estimation of the disturbance structure from data using semidefinite programming and optimal weighting, Automatica, 45 (2009) 142-148.

[18] P. J. Brockwell and R. A. Davis, Introduction to Time Series and Forecasting, 2nd, Springer Verlag, New York, New York, USA 2002.

[19] P. Nomikos and J. F. MacGregor, Multivariate Spc Charts for Monitoring Batch Processes, Technometrics, 37 (1995) 41-59.

[20] E. N. M. van Sprang, H. J. Ramaker, J. A. Westerhuis, S. P. Gurden, and A. K. Smilde, Critical evaluation of approaches for on-line batch process monitoring, Chem. Eng. Sci., 57 (2002) 3979-3991.

[21] J. Thygesen and F. W. J. van den Berg, Calibration Transfer for Excitation-Emission Fluorescence Measurements, Anal. Chim. Acta, 705 (2011) 81-87.

[22] J. Christensen, E. M. Becker, and C. S. Frederiksen, Fluorescence spectroscopy and PARAFAC in the analysis of yogurt, Chemom. Intell. Lab. Syst., 75 (2005) 201-208.

[23] J. B. Fox and D. W. Thayer, Radical oxidation of riboflavin, Int. J. Vitam. Nutr. Res., 68 (1998) 174-180.

[24] C. A. Andersson and R. Bro, The *N*-way Toolbox for MATLAB, Chemom. Intell. Lab. Syst., 52 (2000) 1-4.

[25] Christensen, J. Autofluorescence of Intact Food - An Exploratory Multi-way Study. (2005), PhD Thesis, Quality and Technology, Department of Food Science, The Royal Veterinary and Agricultural University, Denmark.

[26] J. R. Lakowicz, Principles of fluorescence spectroscopy, 3ed edition, Springer, 2006.

Paper IV

Thygesen, J.H, R. Bro & F. van den Berg (2012)

Estimation of process characteristics using constrained PARAFAC models

Chemometrics & Intelligent Laboratory Systems, In preparation

Estimation of process characteristics using constrained PARAFAC models

Jonas Hoeg Thygesen*, Rasmus Bro and Frans W.J. van den Berg, University of Copenhagen, Faculty of Science, Department of Food Science, Quality and Technology, Rolighedsvej 30, DK-1958 Frederiksberg C, phone: (+45) 3533 3500, fax: (+45) 3533 3245, e-mail: thygesen@life.ku.dk

[Manuscript under preparation]

Abstract

This paper presents how previous knowledge on reaction kinetics may be incorporated during PARAFAC modelling; this is done by constraining the scores of the PARAFAC solution to follow exponential decay/growth. The method is illustrated by fitting 3-way and 4-way constrained PARAFAC models on a dataset of EEM-fluorescence. The 3-way analysis showed that chemically meaningful scores and loadings could be obtained together with kinetic parameters of the individual batch. The 3-way analysis was not able to clearly differentiate between batches recorded under normal operating condition and batches recorded under deviating conditions, the method was however able to detect drift in the data. The 4-way analysis showed how comparable loadings and score trajectories were obtained when compared to the 3-way analysis. The drift in data was for this analysis not as evident as for the 3-way analysis, the method was however better at capturing the batch to batch variation in the form of differences between NOC and non-NOC batches. A larger computational effort was required for the 3-way analysis than for the 4-way analysis, the two methods should however in the authors' opinion not be seen as competing but rather as complementing methods that may provide answers to different types of questions.

Keywords: PARAFAC, constraints, process monitoring, grey-box models

1 Introduction

The food- and pharmaceutical-industry is under an ever increasing demand for reduction in energy use, optimal production planning, product consistency and efficient utilization of raw materials. Several papers [1,2] have shown how successful implementation of process surveillance (e.g. via spectroscopy) and process models may bring these goals closer to reality. Different categories of models are available each with advantages and drawbacks, and depending on the objective of modelling and the available process information and knowledge one or the other should be chosen. Three levels or stages of process understanding can be distinguished:

1. The physics and chemistry governing the process is well known (e.g. we know that the process can be described by a first order differential equation with known coefficients)

2. The physics and chemistry governing the process is partially known (e.g. we understand how the process can be described by a first order differential equation, but we don't know the coefficients)

3. Only limited knowledge is available on the physics and chemistry governing the process (e.g. we can observe the input and the outputs of the system, but don't know how they are connected from a causal point of view).

In the case of level 1 understanding so-called white-box modelling can be performed. It consists of process models based on mechanistic or first principles (e.g Newtonian laws of physics). Unfortunately this is often not the case, especially in food processing, and only limited knowledge on the physics and chemistry is available (leaning towards level 3). This leads to the need of System Identification (SI) tools; so-called black-box methods that allow modelling of processes (or systems) based only on observations of system inputs and outputs. An appealing feature of many of the SI-tools is that they allow implementation of *a priori* knowledge on the system, i.e. the intermediate knowledge level 2, where some process understanding is available. This is - for obvious reasons - known as grey-box modelling [3-7]. In this paper we will describe how prior knowledge on reaction kinetics may be incorporated during modelling of spectroscopic data. It is an example of how grey-box modelling may be applied in Multivariate Statistical Process Control (MSPC) were the estimated parameters of the kinetic profiles serve as indicator variables for batch performance. PARAllel FACtor analysis (PARAFAC) [8,9] is the method used for decomposing the multi-linear *N*-way tensors resulting from measurements recorded over batch-time. The model is known to be well suited for excitation-emission (EEM) fluorescence spectroscopy measurement data [10]. This paper illustrates how grey-box based MSPC modelling may be achieve by imposing functional constraints during PARAFAC modelling on a set of EEM fluorescence spectroscopy data.

2 Material and Methods

2.1 Data-set

Riboflavin (vitamin B2) is bright yellow and can in solution be quantified by means of EEM fluorescence spectroscopy [2,11,12]. The vitamin is relatively heat stabile but may, depending on pH, be hydrolysed into lumichrome and lumiflavin when light is present. Lumichrome formation is favored at neutral or acid pH while lumiflavin formation is favored under basic conditions [13]. The data-set studied in this paper originates from a model system of riboflavin where EEM fluorescence spectroscopy is applied to monitor the breakdown. The data-set consists of 62 batches that were measured over 8 days. Each batch is measured for approximately 60 minutes, corresponding to 63 equally spaced measurement time points. The spectra were recorded using a BioView EEM fluorescence process spectrometer (DELTA Danish Electronics, Light & Acoustics, Hørsholm, Denmark). The spectrometers uses a combination of 15 excitation filters (λ_{ex}; equidistantly spaced from 270 to 550 nm) and 15 emission filters (λ_{em}; equidistantly spaced from 310 to 590 nm). The data can thus be seen as one 4-way tensor (size: $63 \times 15 \times 15 \times 62$, time $\times$ excitation $\times$ emission $\times$ batch) or as 62 individual 3-way tensors (size: $63 \times 15 \times 15$, time $\times$ excitation $\times$ emission) 38 of the batches were recorded under Normal Operating Conditions, the remaining 18 were manipulated to yield process conditions different from the NOC (e.g changing pH, adding extra tablet material, dimming the light etc.). Part of this data-set was previously presented by Thygesen & van den Berg [2], we referrer to this paper for further information on experimental setup and design.

2.2 Algorithms

PARAFAC modelling was done using the PLS_toolbox (Eigenvector Research Inc., Wenatchee, WA, USA) in MATLAB R2010b (The Mathworks Inc., Natick, MA, USA); example Matlab code to apply functional constraints in the PLS_toolbox surrounding is presented in the appendix. All other computations were performed in MATLAB using in-house routines.

3 Theory

3.1 Grey-box models

Thygesen & van den Berg [2] have previously shown how 3-way four component PARAFAC models on the Riboflavin data-set provided chemically meaningful results. The paper showed emission and excitation PARAFAC loadings that corresponded to the pure spectra of riboflavin, lumichrome, lumiflavin and a fourth unknown component that was stable/unchanged during batch reaction. The (unconstrained) scores obtained seemed to follow an exponential decay/growth for the three chemicals of interest. The present paper will present how this observation may be incorporated during PARAFAC modelling by constraining the PARAFAC scores to follow this exponential decay/growth. The scores are thus constrained to follow Equation 3-1.

$$a(t) = \frac{\beta_1}{1 + \exp(-k(t - \Delta t))} + \beta_2 \qquad \textbf{Equation 3-1}$$

Where $a(t)$ is the score value at time point t, k the reaction rate constant, Δt a parameter responsible for shifting the point of maximum inflection, and β_1 plus β_2 parameters that determines the starting intensity and end-point off-set of the curve.

3.2 PARAFAC

PARAFAC [8,9] is a method for decomposing N-way tensors $\underline{\mathbf{X}}$. The method applies to tensors with three or more modes, hence the general term N-way. For simplicity is PARAFAC in this theory section is explained for N=3 (i.e. $\underline{\mathbf{X}}$ (size $I \times J \times K$)). The individual element in $\underline{\mathbf{X}}$ (size $I \times J \times K$) is for an R-component PARAFAC model defined by [14]:

$$x_{i,j,k} = \sum_{r=1}^{R} a_{i,r} b_{j,r} c_{k,r} + e_{i,j,k} \;\; i = 1, \ldots, I; \; j = 1, \ldots J \qquad \textbf{Equation 3-2}$$

This can in matrix notation be written as:

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}^T + \mathbf{E}_k \qquad \textbf{Equation 3-3}$$

where the individual slab of $\underline{\mathbf{X}}$ ($\mathbf{X}_k$ with the size $I \times J$) is approximated by $\mathbf{A}$ ($I \times R$), the matrix with the collected first mode scores, $\mathbf{B}^T$ the matrix containing the second mode loadings, and $\mathbf{D}_k$ the third mode loadings. The third mode loadings (or $\mathbf{c}$-vectors) are collected in $\mathbf{C}$ ($K \times R$), the k'th column of this matrix is used as the diagonal in the diagonal matrix $\mathbf{D}_k$ [14].

Most PARAFAC algorithms are based on an Alternating Least Squares (ALS) approach [15]. The ALS algorithms works by splitting the parameters that are to be estimated into sets, one set for each mode in the original tensor $\underline{\mathbf{X}}$. ALS then estimates one set of parameters (one mode in the tensor) in a least squares sense given (initial) estimates of the remaining sets by isolating this mode by rearranging Equation 3-3. The

updated estimates are then used to find the next isolated set, and iteration over all the sets (or modes) is continued until convergence [15]. In case of the PARAFAC model in Equation 3-2 or 3-3, an ALS algorithm would therefore iterate over the following steps: estimate **A** given initial estimates of **B** and **C**, then estimate **B** given **C** and the updated estimate of **A,** and finally **C** given the updated **A** and **B**, then go back to **A**, etc. until convergence. The ALS concept may readily be extended with functional constraints, not essentially different from other constraints such as non-negativity or unimodality [5,6,10,16]. In our implementation, if a functional constraint is applied in the **A**-mode for each of the *R* different score vectors, Equation 3-1 is fitted by non-linear least squares on these scores [17]. This will result in a set of four model parameters - k, Δt β_1 and β_2 – for each of the *R* kinetic score profile. The fitted score-values, estimated from these model parameters - *updated* to more closely follow the expected kinetic profiles - now substitute the estimates from the PARAFAC-ALS estimation. The *updated* **A**-matrix is then used to estimate **B** and **C,** and this extended ALS process continues until convergence.

4 Results and Discussion

4.1 3-way PARAFAC

The data was arranged as single batch run 3-way tensors of size 63 × 15 × 15, (time × excitation × emission), and one PARAFAC model was made for each batch. Functional constrains were implied column-wise in the first mode, meaning that the first three PARAFAC components were forced to follow Equation 3-1 while the fourth component – corresponding to a unknown chemical not changing during reaction – was unconstraint. Based on the spectral profiles presented in previous work [2] the following combination of constrains were used for the 3-way model:

- Mode 1 (Scores): Column-wise functional constrains for component 1-3 following Equation 3-1
- Mode 2 (Excitation wavelengths): all components non-negative
- Mode 3 (Emission wavelengths): all components unimodality and non-negativity constraints

To increase computational speed an initial model was fitted on a single batch without functional constraints (first batch in the series). The scores and loadings obtained from this initial model were used as initial values for the subsequent PARAFAC models. Figure 4-1 presents the obtained scores and loadings.

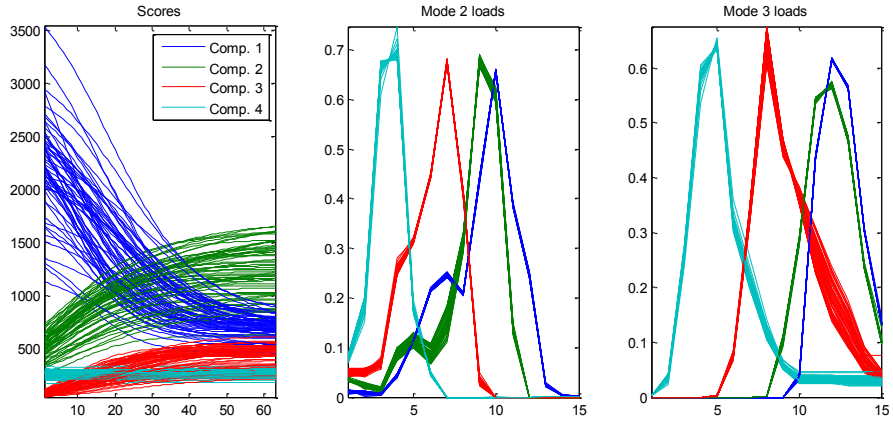


Figure 4-1 Scores, excitation and emission loadings (left to right) from 62 individual 3-way PARAFAC models

Several observations may be done in Figure 4-1. The different batches have different score trajectories, the overall pattern of the scores seem however to follow the same kinetic trend: the first component decreases, while the second and third component grows, and the fourth component is (as expected) constant over batch time. The batch to batch variation in both wavelength loadings (centre and right frame in Figure 4-1) is however very small, indicating that the constrained four component PARAFAC model is suitable for the data set.

As was mentioned in the introduction the kinetic parameters of Equation 3-1 are estimated inside our PARAFAC ALS algorithm. The reaction rate constant *k* is together with the time-shift constant Δ*t* speculated to be the parameters of main interest for this process, shown Figure 4-2 for the three reaction paths.

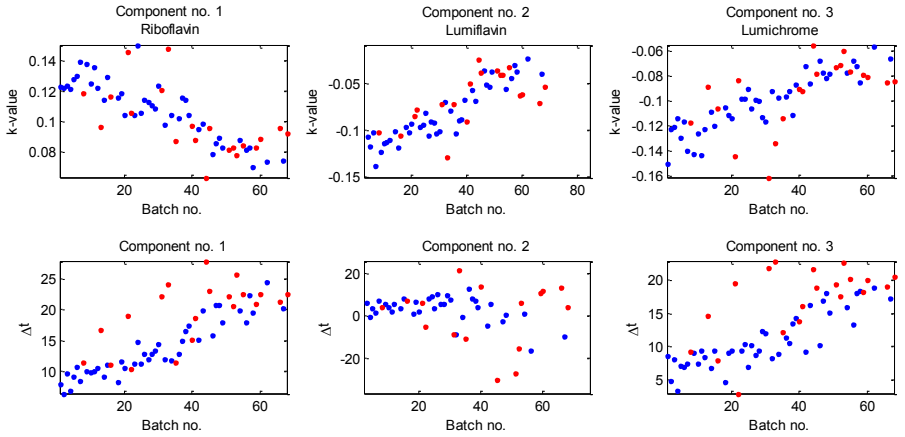


Figure 4-2 reaction rate constant (k-value) and time shift constant (Δt)

Figure 4-2 illustrates that the batch to batch variation (the difference between NOC and non-NOC) is not clearly captured in the presented kinetic constants. Some of the non-NOC batches are deviating from the bulk (e.g. Δt for component 1 and 3), but overall is no clear separation between the groups is seen. It is however from previous work on the data known that some of the manipulations done on the non-NOC batches yielded process data close to normal, it is therefore not surprising that some overlapping of groups are seen. A clear trend in the data is may nevertheless be observed in Figure 4-2, the k-value is systematically decreasing for the first component over the batches just as it is increasing over the batches for the second and third component, suggesting the Riboflavin breakdown reaction speed and formation of the breakdown products lumiflavin and lumichrome decreases over time for the 63 batch runs since a smaller numerical value of the k-value corresponds to a slower reaction rate. Likewise the time delay Δt seems to increase systematically with batch no. for component 1 and 3, also indicating wear on the process equipment since a larger value of Δt results in a later inflection point which in turn again results in a delayed reaction.

4.2 4 way

For 4-way PARAFAC analysis, the data were arranged as one 4-way tensor (size: 68 × 15 × 15 × 68, time × excitation × emission × batch), the wavelength modes were constrained to be non-negative and unimodel (as in the 3-way analysis) while the fourth mode (the batch mode) was left unconstrained. The PARAFAC algorithm was again initialized with unconstrained models to increase the computational speed. Figure 4-3 presents the scores and loadings of the corresponding model together with the two central kinetic parameters of Equation 3-1 *k* and Δt.

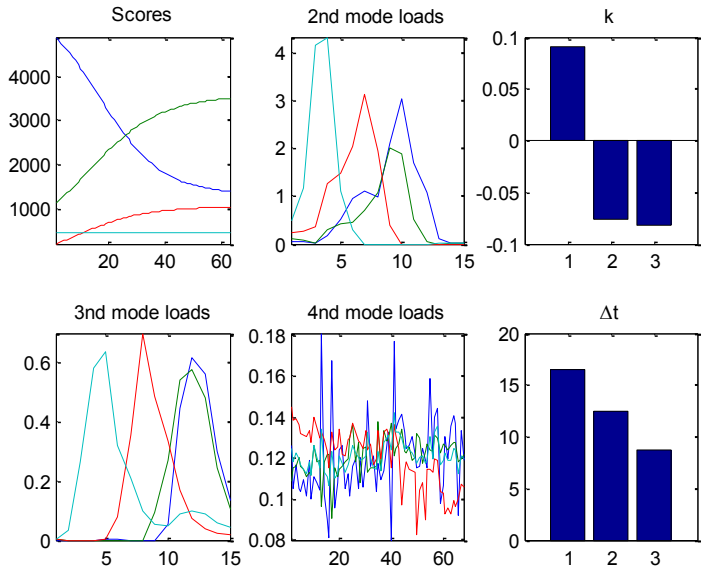


Figure 4-3 Scores, loadings and kinetic parameters of 4-way PARAFAC

The 4-way PARAFAC are one set of common scores/reaction profiles, common excitation and emission loadings and set of parameters obtained for all the 68 batches which can be considered contributions of each batch on the set of common components. For the kinetic parameters the reaction rate constant *k* is presented together with the time-shift constant Δt. The *k*-value shows (not surprisingly) that the first component is decreasing while the other two components are increasing. It may further be observed that the two increasing components have very similar *k*-values which at a first glance seems surprisingly since the second component is growing at an apparently faster rate than the third. This is however an effect of the initial off-set and the end-value of the score trajectory, the larger off-set of the second component when compared to the third, means that the effect of the k-value is less obvious. The effect of Δt may be seen if the point where score trajectories flattens out is assessed. The third component, which has the smallest Δt-value, flattens out already around t=40 while the first component, which has the largest Δt, only barely reaches the point where score trajectories flattens out.

When Figure 4-1 is compared to Figure 4-3 many of the same patterns are recognizable. The overall score trajectory is similar for both figures just and the wavelength loadings are highly comparable. One important note should be made when comparing the 3-way and 4-way analysis: the 3-way analysis requires computation of a PARAFAC model for each batch, while the 4-way analysis requires computation of a single model. In turn means that the 3-way analysis is much more computational demanding, adding further to this is the fact that the 3-way model on average required approx. 400 iterations, while the 4-way only required 170 in spite of the larger dataset, it could be speculated that this difference in computations is an indication that the batch data are indeed 4-way.

The batch to batch variation is captured in the fourth mode loadings. These loadings are however less easily interpreted; the values of the batch mode loading may be seen as batch scores, indicating that the fourth mode loading should be inspected if investigation of batch to batch variation is present. The squared Mahalanobis distance of the fourth mode loading could be used to assess the batch to batch variation within the model. Figure 4-4 presents a plot of the squared Mahalonobis distance.

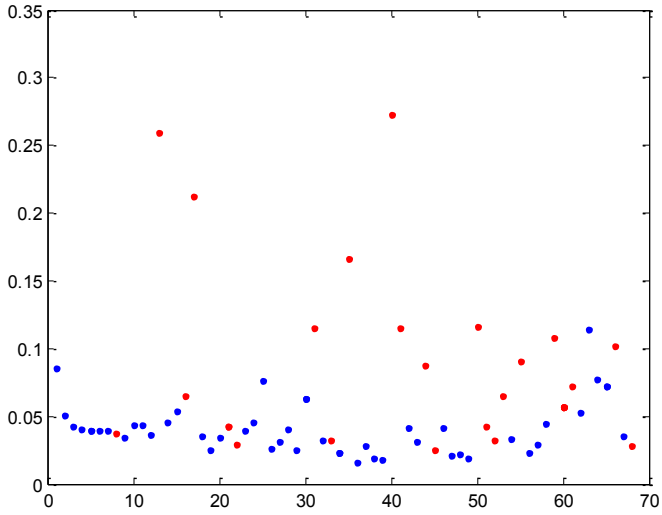


Figure 4-4 Malahanobis distance of batch mode loading, NOC batches in blue, non-NOC in red

In Figure 4-4 it may be observed that the NOC batches are close to the model centre. A few non-NOC batches are situated together with the NOC close to the model centre reflecting that some of the manipulations done on the non-NOC batches yielded batches close to normal. Other NOC-batches are clearly deviating with a large distance from the model centre.

5 Conclusions

This paper presented how previous knowledge on reaction kinetics may be incorporated during PARAFAC modelling; the presented work is thus an example of how so-called grey-box modelling may be done. A dataset of EEM-fluorescence data was presented, the dataset could both be modelled by 3-way and 4-way PARAFAC. The 3-way analysis showed that chemically meaningful scores and loadings could be obtained together with kinetic parameters of the individual batch. The 3-way analysis was not able to differentiate clearly between batches recorded under normal operating condition and batches recorded under deviating conditions, the method was however able to detect drift in the data. The 4-way analysis showed how comparable loadings and score trajectories were obtained when compared to the 3-way analysis. The drift in data was for this analysis not as evident as for the 3-way analysis, the method was however better at capturing the batch to batch variation in the form of differences between NOC and non-NOC batches. A larger computational effort was required for the 3-way analysis than for the 4-way analysis, the two methods should however in the authors' opinion not be seen as competing but rather as complementing methods that may provide answers to different types of questions.

6 Appendix

Two different m.-files are need for imposing functional constraints during PARAFAC modelling: One that defines / evaluates the function and one for imposing the PARAFAC constraints. The function *expfunc* returns the values of an exponential function with the parameters *param* at a series of datapoints (*beta*), outputs are squared residuals (*error*) and the evaluated values (*newbeta*):

```
function [error,newbeta]=expfunc(param,beta)
% function [error,newbeta]=expfunc(beta,param)
% 111208 JT
% exponential function to be evaluated of form c=c0/(1-exp(-k*t))+c1
%

id = [1:length(beta)]'; % Vector where exponential should be evaluated
idnew=-param(2)*(id-param(3)); % Time shift and multiply with reaction const.
newbeta = (param(1)./(1+exp(idnew)))+param(4); %Evaluate id-vector
error = sum( (beta(:) - newbeta(:)).^2); % Find sum of squared residuals

end
```

The following m.-code is an example of how the functional constraints may be imposed

```
load data; % Load data
op = parafac('options'); % Define options-structure to be used during modelling

% Needs to be defined
NumberFactors=4;    % Number of PARAFAC components

ModeToFix = 1;      % Constraint should be on scores (i.e. first mode)

ToFix = [1 2 3];    % This constraint is for the first, second and third...
                    % column, first three components should be constrained,...
                    % fourth is unconstrained

options=op.constraints{ModeToFix}; % Second options-structure for defining
                                   % constraints
options.type='columnwise'; % Constraints are columnwise
options.functional=cell(NumberFactors,1); % Cell for defining functional const.

% Form cell with column constraints, set cell to 0 if unconstrained, 20 if
% functional constraint

options.columnconstraints=cell(0,1,NumberFactors);
for i=1:NumberFactors
    if any(ToFix==i)
        options.columnconstraints{i}=20;
    else
        options.columnconstraints{i}=0;
    end
end

% Form matrix with initial guesses of parameters
parMatrix(1,:)=[-3000 0.1 15 4000]; %Initial guesses for comp no. 1
parMatrix(2,:)=[-1000 -0.1 15 1500]; %Initial guesses for comp no. 2
parMatrix(3,:)=[-1000 -0.1 20 800]; %Initial guesses for comp no. 3
```

```
for i=1:length(ToFix)
    % Provide function handle
    options.functional{ToFix(i)}.functionhandle = @expfunc;
    % Define starting parameters
    options.functional{ToFix(i)}.parameters = parMatrix(i,:);
    options.functional{ToFix(i)}.additional=[]; % no additional input
end
op.constraints{ModeToFix}=options; % Use the defined constrains in options array

% Ready go!
m=parafac(data,NumberFactors,op);
```

7 Reference List

[1] J. H. Thygesen and F. W. J. van den Berg, Subspace methods for dynamic model estimation in PAT applications, J. Chemom., Accepted (2012).

[2] J. H. Thygesen and F. W. J. van den Berg, Dynamic Model Based Monitoring of Batch Processes, Chemom. Intell. Lab. Syst., Submitted (2012).

[3] S. Bijlsma and A. K. Smilde, Estimating reaction rate constants from a two-step reaction: a comparison between two-way and three-way methods, J. Chemom., 14 (2000) 541-560.

[4] B. Roffel and B. H. Betlem, Process dynamics and control: modeling for control and prediction, John Wiley & Sons, 2006.

[5] E. Bezemer and S. C. Rutan, Three-way alternating least squares using three-dimensional tensors in MATLAB, Chemom. Intell. Lab. Syst., 60 (2002) 239-251.

[6] S. P. Gurden, J. A. Westerhuis, S. Bijlsma, and A. K. Smilde, Modelling of spectroscopic batch process data using grey models to incorporate external information, J. Chemom., 15 (2001) 101-121.

[7] E. N. M. van Sprang, H. J. Ramaker, J. A. Westerhuis, A. K. Smilde, and D. Wienke, Statistical batch process monitoring using gray models, AIChE J., 51 (2005) 931-945.

[8] J. D. Carroll and J. Chang, Analysis of individual differences in multidimensional scaling via an N-way generalization of 'Eckhart-Young' decomposition, Psychometrika, 35 (1970) 283-319.

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

[10] R. Bro, PARAFAC. Tutorial and applications, Chemom. Intell. Lab. Syst., 38 (1997) 149-171.

[11] J. Thygesen and F. W. J. van den Berg, Calibration Transfer for Excitation-Emission Fluorescence Measurements, Anal. Chim. Acta, 705 (2011) 81-87.

[12] J. Christensen, E. M. Becker, and C. S. Frederiksen, Fluorescence spectroscopy and PARAFAC in the analysis of yogurt, Chemom. Intell. Lab. Syst., 75 (2005) 201-208.

[13] J. B. Fox and D. W. Thayer, Radical oxidation of riboflavin, Int. J. Vitam. Nutr. Res., 68 (1998) 174-180.

[14] A. K. Smilde, R. Bro, and P. Geladi, Multi-way Analysis - Applications in the chemical sciences, John Wiley & Sons Ltd, Chichester 2004.

[15] Bro, R. Multi-way Analysis in the Food Industry. (1998), PhD Thesis, University of Amsterdam, The Netherlands.

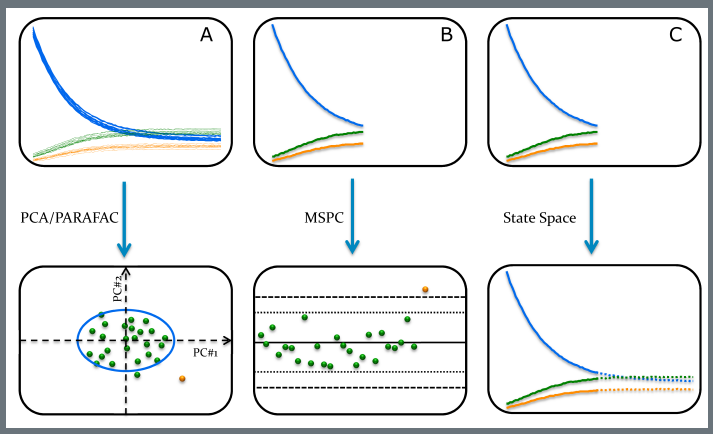
[16] R. Bro and S. de Jong, A fast non-negativity-constrained least squares algorithm, J. Chemom., 11 (1997) 393-401.

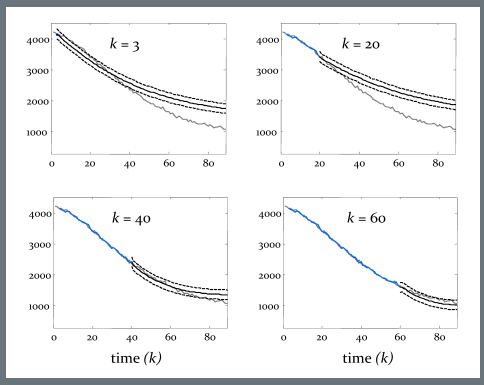
[17] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, Numerical Recipes in C, 2nd, Cambridge University Press, Cambridge, UK 1992.

QUALITY AND TECHNOLOGY
DEPARTMENT OF FOOD SCIENCE
FACULTY OF SCIENCE
PHD THESIS 2012 · ISBN 978-87-7611-512-8

JONAS HOEG THYGESEN

Dynamic Models and Chemometric Tools for Process Monitoring





SL grafik