Process Analytical Technologies in Applied Biotechnology - biomass conversion, 2nd generation bioethanol, & specialty product fermentation

Lomborg, Carina Juel

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Process Analytical Technologies in Applied Biotechnology

- biomass conversion, 2nd generation bioethanol, & specialty product fermentation

Ph.D. Thesis

By

Carina Juel Lomborg

ACABS Research Group

(Applied Chemometrics, Analytical Chemistry, Applied Biotechnology, Bioenergy & Sampling)

Esbjerg Institute of Technology, Niels Bohrs Vej 8, DK-6700 Esbjerg Denmark

Thesis submitted to the International Doctoral School of Science and Technology, Aalborg University, Denmark, for the degree of Doctor of Philosophy

April 2009

Carina Juel Lomborg

ACABS Research Group

Ph.D. Thesis Aalborg University, 2009

ISBN: 978-87-90033-56-9

Published by:
Aalborg University, campus Esbjerg
Niels Bohrsvej 8
DK-6700 Esbjerg
Denmark
www.aaue.dk

Printed by:

Cover design by Martin Helsing

Abstract

This Ph.D. surveys application of Process Analytical Technologies (PAT): Near Infrared Spectroscopy (NIR), Electronic Tounge (ET), and acoustic chemometrics (a.c.), for improved monitoring (and control) of complex biotechnological systems, both for anaerobic and aerobic systems. Sampling within existing and new systems is validated following the Theory Of Sampling (TOS) and where necessary improvements are suggested. The Ph.D. program, therefore, falls naturally into three parallel pathways.

1) Two anaerobic digestion processes applying glycerol and maize silage based substrates were monitored applying either at-line or on-line NIR and a.c. For testing the feasibility of NIR and a.c. two anaerobic digestion trials were carried out in laboratory scale reactors with an active volume of 5 L. Trial 1 was carried out as a co-digestion of glycerol-spiked manure and food waste, while trial 2 was a co-digestion of maize silage-spiked manure and food waste. The overall conclusion was that the a.c. model for total solids had fair prediction abilities (Slope: 0.85, r^2: 0.85). The NIR models obtained from the two trials had good prediction abilities for total and volatile solids (slope: 1.04-1.06, r^2 0.96-0.98) and glycerol (slope: 1.0, r^2: 0.92). For volatile fatty acids generally fair to good prediction abilities were established (slope: 0.83-1.1, r^2 0.89-0.97), though some problems with a few of the models from trial 1. This work was reported in two published papers (1 and 2).

2) The second pathway was directed at production of methyl ketones by *Penicillium roqueforti*. This process is an industrial multi-phase fermentation; the task was to redefine the process into a single-phase fermentation, adding the product pre-curser in a fed-batch configuration, specifically with PAT as a possible new means of improved monitoring and control. A suitable substrate suggested itself from an initial series of experiment and process parameters were outlined. A second series of experiments, planned to optimise this solution were carried out, but severe fermentation problems persisted despite extensive experimentation. In the end no satisfactory solution could be found. Before terminating this pathway a feasibility study on quantitative determination of octanoic acid on fermentation broth applying ET was successfully carried out however, resulting in paper 5.

3) The final pathway work took place in close collaboration with RISØ (now DTU-RISØ): Pilot surveys of the potential of near infrared monitoring raw wheat straw composition with respect to cellulose, hemicelluloses, and lignin specifically for 2nd generation bioethanol production. For this field samples were collected from several sites in Denmark, from which 44 were selected here. Alternative pre-processing's were evaluated; 1st and 2nd order derivative spectra (Savitzky-Golay) and Multiplicative Scatter Correction (MSC) was found optimal for the carbohydrates and lignin respectively. Full spectrum PLS-1 regression models resulted in less good prediction abilities than the models based on wave number variable selection applying a jack-knife

approach. The present first foray models were validated using 4-segment cross validation leading to fair to good accuracies (slope 0.76-0.90) and fair precisions (r^2 0.77-0.83) for the carbohydrates. The lignin model also showed a fair accuracy (0.84) but a distinctly less good precision (0.72). The carbohydrate models were also test-set validated, resulting in an accuracy of 0.85-0.94 and precision estimates 0.86-0.87. This corresponded to a relative root mean square error of prediction in the interval 8 -10 % which is satisfactory for a feasibility study, although the models should be optimized further before application in a routine monitoring context. This work is reported in papers 3 and 4.

This thesis serves as a general introduction to the trinity of process monitoring: sampling, sensor technologies, and chemometrics, with a natural focus on the parts most relevant for the work carried out. Furthermore, it includes chapters introducing each pathway/project.

After the introductory chapters five papers follow:

I. Lomborg CJ (60%), Holm-Nielsen, JB (20%), Oleskowicz-Popiel P (10%), Esbensen KH(10%): Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage. Bioresource Technology 100. p. 1711-1719. 2009

II. Holm-Nielsen, JB (60%), Lomborg CJ (20%), Oleskowicz-Popiel P (10%), Esbensen KH (10%): On-line Near Infrared monitoring of glycerol-boosted anaerobic digestion processes – evaluation of Process Analytical Technologies. Biotechnology and Bioengineering 99. no.2. p. 302 – 313. 2008

III. Lomborg CJ (80%), Jensen ES (10%), Esbensen KH (10%): NIR- characterization of cut wheat straw for bioethanol production – feasibility study. Journal of NIRS (*submitted March 2009*)

IV. Lomborg CJ (80%), Thomsen MH (10%), Esbensen KH (10%): Power plant intake quantification of wheat straw composition for 2nd generation bioethanol optimization – a Near infrared spectroscopy (NIRS) feasibility study. Bioresource technology (*submitted April 2009*)

V. Lomborg CJ (70%), Wiebe L (10%), Esbensen KH (20%). At-line determination of octanoic acid in cultivation broth - An electronic tongue (ET) feasibility study. Journal of Biotechnology 133. p. 162-169. 2008

Synopsis

Dette Ph.d.studie undersøger anvendelsen af Proces Analytiske teknologier (PAT), Nær InfraRød spektroskopi (NIR), Elektronisk tunge (ET) og akustisk kemometri (a.c.), til forbedret overvågning (og kontrol) af komplekse bioteknologiske systemer; både anaerobe og aerobe processer. Prøvetagning i eksisterende og nye systemer er undersøgt med baggrund i teorien om prøvetagning (TOS), og hvor nødvendigt er forbedringer foreslået. Ph.d.forløbet var naturligt inddelt i tre parallelle delprojekter:

1) En anaerob udrådningsproces, der anvendte glycerol og majsensilage baseret substrater blev monitoreret med at-line eller on-line NIR og a.c. For at teste anvendelsesmulighederne af NIR og a.c. to anaerobe udrådningsforsøg blev udført i en 5 L fermentor. Forsøg et blev udført som en sam-udrådning mellem gylle tilsat glycerol og madaffald, mens forsøg to anvendte gylle tilsat majsensilage og madaffald. Den overordnede konklusion fra de to forsøg var, at a.c. modellerne for total tørstof gav en rimelig prædiktion (hældning 0,85; r^2 0,85). NIR modellerne fra de to forsøg gav gode prædiktioner for total og flygtig tørstof (hældning 1,04-1,06, r^2 0,96-0,98) og glycerol (hældning 1,00; r^2 0,92). For de flygtige fede syre blev der opnået rimelige til gode prædiktioner (hældning 0,83-1,1, r^2 0,89-0,97). Der var dog problemer med enkelte modeller i forsøg et. Dette arbejde er publiceret i artikel 1 og 2.

2) Det andet delprojekt var produktionen af methyl ketoner ud fra *Penicillium roqueforti*. Denne proces forløber som en industriel multifase fermentering. Opgaven her var at redefinere processen til en enkelt-fase fermentering, hvor tilsætning af produkt præcursoren skulle ske ved fed-batch proces. Specielt skulle PAT undersøges om en mulighed for overvågning og kontrol. Ud fra indledende forsøg blev et substrat sammensat og procesparametre fundet. Yderligere forsøg for at optimere processen blev udført, men på trods omfattende forsøg var der problemer med disse fermenteringer. Ingen tilfredsstillende løsning blev fundet på disse. Før dette delprojekt blev afsluttet, blev der udført et studie hvor muligheden for bestemmelsen af octansyre i fermenteringsmedie vha. ET blev undersøgt. Konklusion var, at dette var muligt og resulterede i artikel 5.

3) Det tredje delprojekt blev defineret i samarbejde med Risø (nuværende DTU-Risø) og var pilotundersøgelser af NIRs potentiale for overvågning af sammensætningen af rå hvedehalm mht. cellulose, hemicellulose og lignin til anden generations bioethanol produktion. Til denne undersøgelse prøver blev samlet over hele Danmark, og 44 blev valgt. Diverse præ-processerings metoder blev undersøgt. Anvendelse af første og anden ordens spektre derivater (Savitzky-Golay) og multiplikativ sprednings korrigeret spektre (MSC) var optimale for henholdsvis kulhydrater og lignin. Fuld spektral PLS-1 regressions modeller resulteres i mindre gode prædiktioner, i forhold til modeller baseret på enkelte bølgetal udvalgt ved anvendelse af jack-knife. De præliminære modeller blev valideret ved anvendelse af 4-segment kryds-

validering, hvilket resulteres i kulhydrat modeller med rimelige til gode nøjagtige (hældning 0,76 – 0,90) og rimelig præcision (r^2 0,77-0,83). Ligning modellen havde også en rimelig nøjagtighed (hældning 0,84), men en mindre god præcision (r^2 0.72). Kulhydratmodellerne blev desuden valideret med et test-sæt, hvilket resulterede i en nøjagtighed på 0,85-0,94 og en præcision på 0,86-0,87. Dette svarede til en relativ prædiktions fejl (RMSEP%) på 8-10%, hvilket var tilfredsstillende for et potentialestudie. Modellerne skal dog optimeres før anvendelse i en rutinemæssig monitorerings sammenhæng. Dette arbejde resulterede i artikel 3 og 4.

Denne afhandling er opbygget som en generel introduktion til de tre enheder der udgør proces-overvågning; prøvetagning, sensorteknologier og kemometri. I hvert afsnit er der et naturligt fokus på de dele, der er relevante for de enkelte del-processer. Efter denne introduktion findes der tre kapitler, som introducerer de tre delprocesser, der er arbejdet på under Ph.d.studiet. Til sidst er placeret de fem artikler.

I. Lomborg CJ (60%), Holm-Nielsen, JB (20%), Oleskowicz-Popiel P (10%), Esbensen KH(10%): Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage. Bioresource Technology 100. p. 1711-1719. 2009

II. Holm-Nielsen, JB (60%), Lomborg CJ (20%), Oleskowicz-Popiel P (10%), Esbensen KH (10%): On-line Near Infrared monitoring of glycerol-boosted anaerobic digestion processes – evaluation of Process Analytical Technologies. Biotechnology and Bioengineering 99. no.2. p. 302 – 313. 2008

III. Lomborg CJ (80%), Jensen ES (10%), Esbensen KH (10%): NIR- characterization of cut wheat straw for bioethanol production – feasibility study. Journal of NIRS (submitted March 2009)

IV. Lomborg CJ (80%), Thomsen MH (10%), Esbensen KH (10%): Power plant intake quantification of wheat straw composition for 2nd generation bioethanol optimization – a Near infrared spectroscopy (NIRS) feasibility study. Bioresource technology (submitted April 2009)

V. Lomborg CJ (70%), Wiebe L (10%), Esbensen KH (20%). At-line determination of octanoic acid in cultivation broth - An electronic tongue (ET) feasibility study. Journal of Biotechnology 133. p. 162-169. 2008

Acknowledgements

I have spent five years working on the projects described in this thesis. This has been five challenging years, through which I have learned a lot not only about the different specific aspect of process monitoring but also the more general methodology of research. I've learned a lot about collaborating with other people as well as - about myself. Many people have contributed during my doctoral studies.

Through all projects in this study professor Kim H. Esbensen has been my supervisor. Throughout the project numerous experimental problems have been encountered, resulting in a prolonging of the studies. At times everything really seemed hopeless but Kim with your constant support and talent "to see the glass as half full instead of half empty", I got through these rather difficult times. Thank you so much - I really appreciate it. You and I are of nature very different persons, at times also resulting in a somewhat bumpy ride, but the Ph.D. study of course prevailed. I must say I through the last five years have learned a lot from you in and outside science, about chemometrics and sampling but even more importantly you have introduced me to the world of science at large. For that I'm truly great full, thank you so much!

When entering ACABS five years ago, being the only woman in the crowd I entered a somewhat male-filled, dry humor environment. At first it resulted in a lot of red cheeks but everything got back to normal, I believe. I really appreciate it guys, for you letting me in to the group - without all our discussions, social events and every day chat the studies would have been dull and less colorful. Especially I would like to thank my dear friends Hans Henrik Friis-Petersen and Lars P. Julius. Friis for introducing me to life at campus as a Ph.D. student and for the many hours you spent on helping me with program debugging. Lars P. for many fruitful discussions and enjoyable moments at conferences. From ACABS last but not least I would like to thank my good friend and mentor Lars P. Houmøller for always having an open door and mail box. I have really learned a lot from you from chemometrics and analytical chemistry to beer and tea, none of which I would live without.

Campus leader Torben Rosenørn is warmly thanked for ensuring a prolongation of the Ph.D. studies, when experimental problems made sure that it could not be terminated within the given timeframe. Furthermore he is thanked for always having an open door and taking time to discuss small and large issues. I really appreciate it.

I would like to thank very much the staff at AAUE for making the days a bit more joy-full. A few persons deserve special acknowledgment. Thanks to the laboratory technicians; Dorte Spangsmark and Linda Madsen without your skilled guidance in the laboratory, everything would have been much more difficult. Furthermore, thanks to Henry Enevoldsen for numerous debuggings of the electrical system and to Carsten Andreasen for helping me build several experimental setups. Without all of you the studies would most likely not have been possible.

Many thanks to Anne Jodan Cole for her very valuable linguistic support and help with several papers and this thesis.

The project behind paper one and two was carried out in collaboration with a fellow Ph.D. student, Jens Bo Holm-Nielsen and research assistant Piotr Oleskowicz-Popiel. Both of you are thanked. It has been a somewhat stinky but also enjoyable ride. Thanks for never forgetting to smile – even when the biogas reactors exploded.

The project behind paper three and four was carried out in collaboration the Biosystems department at Risø National Laboratory. Thanks to my co-supervisors Erik Steen Jensen for his help and support through the project; I really appreciate it. A special thanks to my co-supervisor and good friend Mette Hedegaard Thomsen for the many hours we spent on discussing the project, on conferences, and everything else. You have truly made my doctoral studies rewarding and memorable. Thanks to all the people at the Biosystems department for making my stays at Risø rewarding and enjoyable. Furthermore, the laboratory technicians; Tomas Fernqvist, Ingelis Larsen, and Annette Eva Jensen are thanked for all their laboratory guidance and help with carrying out analysis.

The project behind paper five was carried out in collaboration with Lars Wiebe, Danisco and he is warmly thanked for all his help and fruitful comments along the way.

Last but not least I would like to thank the people closest to my heart, my friends and all my family. Thanks to my dear friend Lea Jensen for putting up with my constant work chat and giving things another perspective. Special thanks to my loving husband Johannes for sharing my joys and tears with me. For always helping and supporting me, even when it meant you going with me to the laboratory in the middle of the night for sample extraction. I would never be where I am today if it weren't for you.

List of papers and contributions

I. Lomborg CJ (60%), Holm-Nielsen, JB (20%), Oleskowicz-Popiel P (10%), Esbensen KH(10%): Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage. Bioresource Technology 100. p. 1711-1719. **2009**

For this paper I participated in the experimental planning and in carrying out the experiment. I contributed to the introduction, did most of the method section and all the data handling (incl. all modelling).

II. Holm-Nielsen, JB (60%), Lomborg CJ (20%), Oleskowicz-Popiel P (10%), Esbensen KH (10%): On-line Near Infrared monitoring of glycerol-boosted anaerobic digestion processes – evaluation of Process Analytical Technologies. Biotechnology and Bioengineering 99. no.2. p. 302 – 313. **2008**

For this paper I participated in the experimental planning and in carrying out the experiment. I contributed to the introduction and method section. The modeling was carried out by the first author.

III. Lomborg CJ (80%), Jensen ES (10%), Esbensen KH (10%): NIR- characterization of cut wheat straw for bioethanol production – feasibility study. Journal of NIRS **(*submitted, March 2009)***

For this paper I planned and carried out all of the experiments. I did the data handling and wrote the majority of the paper.

IV. Lomborg CJ (80%), Thomsen MH (10%), Esbensen KH (10%): Power plant intake quantification of wheat straw composition for 2nd generation bioethanol optimization – a Near InfraRed-Spectroscopy (NIRS) feasibility study. Bioresource technology **(*submitted, April 2009*)**

For this paper I planned and carried the experiments. I did the data handling and wrote the majority of the paper.

V. Lomborg CJ (70%), Wiebe L (10%), Esbensen KH (20%). At-line determination of octanoic acid in cultivation broth - An electronic tongue (ET) feasibility study. Journal of Biotechnology 133. p. 162-169. **2008**

For this paper I planned and carried out all of the experiments. I did the data handling and wrote the majority of the paper.

Contents

Symbols and Abbreviations

Bold phase lower case letters are vectors, Bold phase upper case letters are matrix.

α	Baseline offset
A	Absorbance
a_i	Concentration in the ith sample
a_L	True concentration in the lot
a_s	Concentration in sample
a.c.	Acoustic Chemometrics
β	Linear slope of baseline
C	Sampling constant
CH_L	Constitutional heterogeneity
COD	Chemical Oxygen Demand
CSE	Correct Sampling Error
C.V.	Coefficient of Variation
d	Top particle size
DH_L	Distributional Heterogeneity
$\mathbf{e_i}$	Residual for i^{th} object
E	Error matrix
ET	Electronic Tongue
FDA	U.S. Department of Health and Human Services Food and Drug Administration
FFT	Fast Fourier Transform
FIA	Flow Injection Analysis
FIR	Far Infrared
FSE	Fundamental Sampling Error
GEE	Global Estimation Error
GC	Gas Chromatography
GSE	Grouping and Segregation Error
h_i	Heterogeneity contribution (index refer to fragments or increments)
h	Leverage
HPLC	High Performance Liquid Chromatography
I_0	Intensity of the reference beam
I	Intensity of the beam after passing through the sample
I	Number of samples/observations
IDE	Incorrect Delimitation Error
IEE	Incorrect Extraction Error
IH_L	Invariant Heterogeneity
IPE	Incorrect Preparation Error

ISE	Incorrect Sampling Error
K	Number of variables
$M_{ave,Fi}$	Average mass of all fragments
m_e	Average of replicated samples
$M(h_i)$	Average of the heterogeneity contributions = 0
M_i	Mass of the fragment F_i
MIR	Mid Infrared
MSC	Multiplicative scatter correction/multiplicative signal correction
N_F	Number of fragments in lot
N_G	Number of groups in lot
NIR	Near Infrared, Near Infrared Spectroscopy
P, p	Loading matrix, vector (X)
PAC	Process Analytical Chemistry
PAT	Process Analytical Technology
PC	Principal Component
PCA	Principal Component Analysis
PLSR	Partial Least Square regression
PSE	Primary Sampling Step
Q, q	Loading matrix, vector (Y)
r^2	Correlation
R	Reflectance
RMSECV	Root Mean Square Error of Cross Validation
RMSEP	Root Mean Square Error of Prediction
RPD	Ratio of standard error of Performance to standard Deviation
r_{sig}	Signal of interest
σ_e^2	Variance of the sampling error
s	Standard deviation
SEP	Standard Error of Performance
SUO	Sampling Unit Operations
SSE	Secondary Sampling Step
T, t	Score matrix, vector (X)
T	Transmittance
TAE	Total Analytical Error
TENIRS	Transflexive Embedded Near Infrared Sensor
TeSe	Tertiary Sampling Step
TOS	Theory Of Sampling
TS	Total Solids
TSE	Total Sampling Error
U, u	Score matrix, vector (Y)

Var	Variance
VFA	Volatile Fatty Acids
VS	Volatile Solids
W, w	Loading-weight matrix, vector
X	Data matrix with dimensions (I x K)
Y	Grouping factor
Z	Segregation factor

1 Introduction

Currently, a shift is taken place where an increasing awareness of the climate changes have brought to the forefront salient questions, such as: how much of this is caused by human activity?, is the process reversible?, and even more importantly: what can be done to reduce the future impact on the global climate and its effect on nature? As a logical consequence more attention is now placed on improving already existing production processes and designing new processes so that waste and byproduct formation and energy consumption are reduced to a minimum – irrespective of the sum-total of their specific contributions to the world carbon-budget issues. At the same time a reduction in the overall cost, a higher quality, and a more consistent product is always desirable.

To obtain this, more thorough knowledge and understanding of the process is needed, which calls for a different process strategy, than what is often applied.

The strategy applied is industry dependent. In the pharmaceutical and other high-tech biotechnological industries, production is often running at predefined settings, from which samples are extracted and delivered to the quality laboratory for analysis of different relevant components. These samples are analyzed and the result is delivered back to the production line within hours or days. Often the delay time from the sample extraction to the lab result allows that the results are seldom used for active process control, but merely as a quality parameter telling if the given production is within specifications and can be sold – or not. In cases where the results are critical for the next process step, the process is often designed to accommodate a delay time[72,74].

In more low-tech biotechnological industries like anaerobic digestion plants (biogas production), the strategy is more of an ad hoc nature. Often the operator has only little knowledge of the material streams coming in (manure, food waste ect.) and what is going on in the tank, and is, thus, running the biogas production more-or-less from experience alone. This of course at times leads to a decrease in the yield, when new feed stocks are coming in. Common for both the low and high tech industry strategy is that it is *suboptimal*, thus leading to lower yield and productivity. For both the question arises: what technique/strategy can be applied for solving this delay problem?

A strategy that comes to mind is the Process Analytical Chemistry (PAC), or the closely related Process Analytical Technologies (PAT). Both PAC and PAT are problem driven, not technology driven and addresses the question; what techniques can be applied for solving my specific measurement problem[89]? The basic concept in PAC is that process analysers should be placed as close to the production as necessary, whether this is in a centralized laboratory (off-line), right next to the process (at-line) or in the process (on-line)[72].

The term *process analytical chemistry* emerged in ca. 1911 and was originally based on taking samples from the process stream, which were transported to the laboratory for analysis[57,124]. In the 1950s the petroleum and petrochemical industry started applying it. But it was not widely used until the 1980s, where *Callis et al.*[17] made the following statement:

> "*The goal of process analytical chemistry is to supply quantitative and qualitative information about a chemical process. Such information can be used not only to monitor and control a process but also to optimize its efficient use of energy, time, and raw materials. In addition it is possible to simultaneously plant effluent release and improve quality and consistency"*

In the 1980s the incorporation was due to a maturation of the manufacturing process i.e., the competitive advantage shifted from product innovation to process innovation. This resulted in a higher product quality and a more efficient process[57,89]. This was further enhanced in 2004 where the American Food and Drug Administration (FDA) described a new comprehensive process understanding in the Process Analytical Technology (PAT) initiative[36]. PAT is more than a mere re-compilation of the PAC approach. PAT also includes guidelines and tools for quality assurance and risk management a.o. However, in 2007 *Workman et al.*[124] stated that PAT was a more appropriate term as measurements were expanding to include physical characterization tools i.e., PAC converged towards PAT.

The goal set by first *Callis et al.*[17] and later on by the FDA initiative is generally obtained by looking at different process signatures, from which the key parameters influencing the process are identified, and used for process monitoring and control[6]. The methods being applied can be simple univariate methods such as pH, temperature, and density measurements, but there is a very clear shift towards more comprehensive chemical analysis, chemometrics, and multi-channel sensor technology.

From this it transpires that PAT is based on a threefold platform:

1. Obtaining representative samples (sampling)
2. Measurement of relevant process parameters (relevant sensor technologies)
3. Conversion of these measurements into process info (data analysis, chemometrics)

The latter two have been an accepted part of PAT from the start, whereas sampling was put in with the redefinition by FDA. A basic understanding of the general principles of the Theory of Sampling (TOS) has, however, not yet been incorporated but is never the less essential for obtaining models with reliable prediction performance. *Esbensen & Mortensen 2009*[32] state:

"Chemometric data models must closely adhere to reliable performance validation, e.g. regarding prediction, classification, time forecasting. Without representative process sampling the chain of evidence vs. lot characteristics is flawed because of inherent materials' heterogeneity at all scales. This applies both to samples as well as sensor signals."

And furthermore

"If the quality of both X- and Y-data involved is suspect, how can a multivariate calibration be expected to be trustworthy? This also includes the issue regarding proper validation of the chemometric multivariate calibration(s) involved, which can only be resolved based on proper understanding of the phenomenon of heterogeneity. The TOS delivers answers to all these issues. The TOS constitutes the missing link in PAT."

Furthermore, *Esbensen and Mortensen 2009*[32] emphasise that it is but a mere misconception that the sampling errors can be eliminated through the subsequent chemometric data analysis: once the sample is obtained nothing can be done to correct for the errors associated with the extraction. The comprehensive platform is here therefore considered as the trinity of PAT, see Figure 1.1, where all three elements are equally important for a successful PAT strategy.

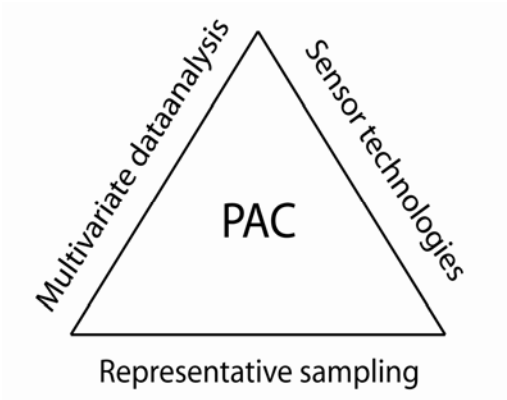

Figure 1.1: Trinity of sampling (adapted from 74)

From the above definitions it is evident that many current and future productions would benefit from an implementation of the full PAT strategy.

For a successful implementation of PAT in an industrial process, contributions from many branches, such as analytical chemistry, sensor technology, chemical engineering, control (automation), are needed thus making PAT a multidisciplinary field.

Representative sampling, the measurements techniques relevant for the current work, and the relevant data analytical approaches will be introduced in chapter 2, 3, and 4 respectively. The

PAT strategy will be applied in three different contexts; anaerobic digestion, 2nd generation bioethanol production, and aroma fermentation.

Anaerobic digestion and 2nd generation bioethanol production are both processes receiving increased attention as methods for reducing the carbon dioxide emission and their contribution in reducing global warming. Both processes are still being developed and optimized, and will most likely gain from the implementation of the PAT strategy. In the current dissertation, the process in question will be looked upon and suggestions made to which key parameters might be of the most relevant interest for monitoring and control. Furthermore, laboratory scale experiments will be carried out to test the feasibility of the parameters and the technology arrived at. This is done in chapter 5 and 6 and in papers 1 through 4.

The last paper, number five, concerns itself with a desired redefinition of an aroma fermentation based on the filamentous fungi *Penicilllium roqueforti.* Currently, the process is carried out applying a multiphase substrate of which only a small part is directly turned into product. A redefinition of the process into a fed-batch -, or a continuous fermentation, applying a specific substrate with precursor addition, is more desirable from an optimization point of view. However, the product precursor is toxic to the organism used; thus, a strict control of the concentration in the substrate is needed. Before carrying out any fermentation optimizations the process is looked upon, and a method for measuring the precursor is suggested and tested in laboratory scale fermentations. This is done in chapter 7 and paper 5 respectively.

2 Sampling

In 1950 Pierre Gy started developing the Theory Of Sampling (TOS) from a practical need in the mining industry on how to obtain a measure of the average concentration in an ore body or in a tailings dump. Since then TOS has found a wide array of applications in various other fields, and is now starting to be considered in chemometrics as the first unit in the trinity of process analytical technologies – *Every multivariate model is critically dependent upon representative samples*.

Multivariate calibrations are seldom based on analyzing the entire system/material at hand, as this would be too expensive and some analyses are destructive. For process monitoring purposes, calibrations are typically based on extracted samples that are analysed at the laboratory (reference values, Y) and/or measurements made directly in the system (X): examples of the latter would be near infrared spectroscopy, acoustic chemometrics, fluorescence, etc. A realization is that only a minute fraction of the system/material at hand is used for establishing the calibration relation, making it essential that the samples considered indeed represent the entire system. Current procedures for sample extraction and sensor placement do not necessarily consider this simple fact, with the consequence that uncontrollable and unknown sampling biases are produced, which at best may yield unnecessary inferior calibration relations and at worst will lead to calibration failure. To avoid this TOS should be contemplated a priori to any sample extraction and handling or placement of a sensor[32].

To understand TOS a minimum of basic definitions and terms are given in Table 2.1.

Table 2.1: Basic definitions and terms in TOS.

Terms	Definitions
Lot	The original material/system, e.g. the entire content in a fermenter, a hay bale etc.
Sample	Material that is extracted from the lot without sampling bias. This process is termed "correct sampling"
Specimen	Material that is incorrectly extracted from the lot (see above).
Increment	A part of a sample unit that combined with other sub-samples yields a composite sample.
Fragment	The smallest unit obtainable in the lot i.e., grain, straw, cell**.**
Dimensions	The sampling process/lot is categorized as 0,1,2,3 dimensional respectively depending of the nature and geometry of the lot. Often a zero dimensional or one dimensional situation is encountered.

The aim of TOS is to reduce the mass of the lot material without changing its characteristic (intrinsic) properties (chemical, physical), thus extracting a sample with the same properties as the lot i.e, a representative sample.

For a sample to be representative it has to be extracted probabilistically *correct*, which means that each virtual fraction of the material must have an identical (non-zero) probability of ending up in the sample and that fractions that do not belong to the material have exactly zero probability of ending up in the sample. Furthermore, all increments/samples have to be extracted by a sampling process that is both accurate and precise (reproducible (r_e)).

$$r_e^2 = m_e^2 + s_e^2$$

Where *m_e is the average of replicated samplings*
s_e^2 is the variance of the sampling error

The sampling process is termed *accurate* when the average error (m_e) is statistically equal to zero, or below or equal a predefined low acceptable value (m_0). The sampling process is characterized as precise (TOS speaks of reproducibility) when the variance of the sampling error (s_e^2) is below a small predefined value (s_0^2). A non-representative specimen is an increment delineated and extracted in any way that contradicts the above characteristics, e.g., parts of the material lot have zero probability of ending up in the sample.

The sampling procedure consists of two parts, 1) sample delineation & extraction, followed by 2) mass reduction & handling (which must also be representative[79,81]). The first step may be

considered as the primary sampling step (PSE), whereas the latter may be a single or a series of steps termed the secondary (SSE), tertiary (TeSE), … sampling step.

Each of the sampling steps (PSE; SSE, TeSE etc) are associated with a relative sampling error:

$$SE = \frac{a_S - a_L}{a_L}$$

Where a_s *is the concentration in the sample*
a_L *is the true concentration in the lot*

The relative sampling error is a sum of correct and incorrect errors associated with the material and the sampling procedure applied.

The overall error is termed the "Global Estimation Error" (GEE), which consists of a contribution from the "Total Sampling Error" (TSE) and the "Total Analytical Error" (TAE). The latter generally has a lot of focus in the analytical laboratory and is constantly being reduced by improvements made in the equipment, analytical procedure, and others. The total sampling error is just as, if not more important but is not receiving similar attention, maybe because it is impossible to ascertain from the sample itself whether it is representative or not. Only an honest, complete delineation of the full procedure by which the sample was obtained and handled gives full disclosure. In the present thesis TSE will be in focus.

2.1 Total Sampling Error

The total sampling error (TSE) is caused by the material properties (heterogeneity) interacting with the specific sampling procedure applied (which may contain "incorrect" sampling errors, or not). TSE consist of a total of five errors as listed in Figure 2.1. This indicates that TSE is reduced mainly by evaluating and changing the sampling process (to get rid of any "incorrect" errors), though changing the material properties by comminution is possible. This is but only rarely possible during process sampling.

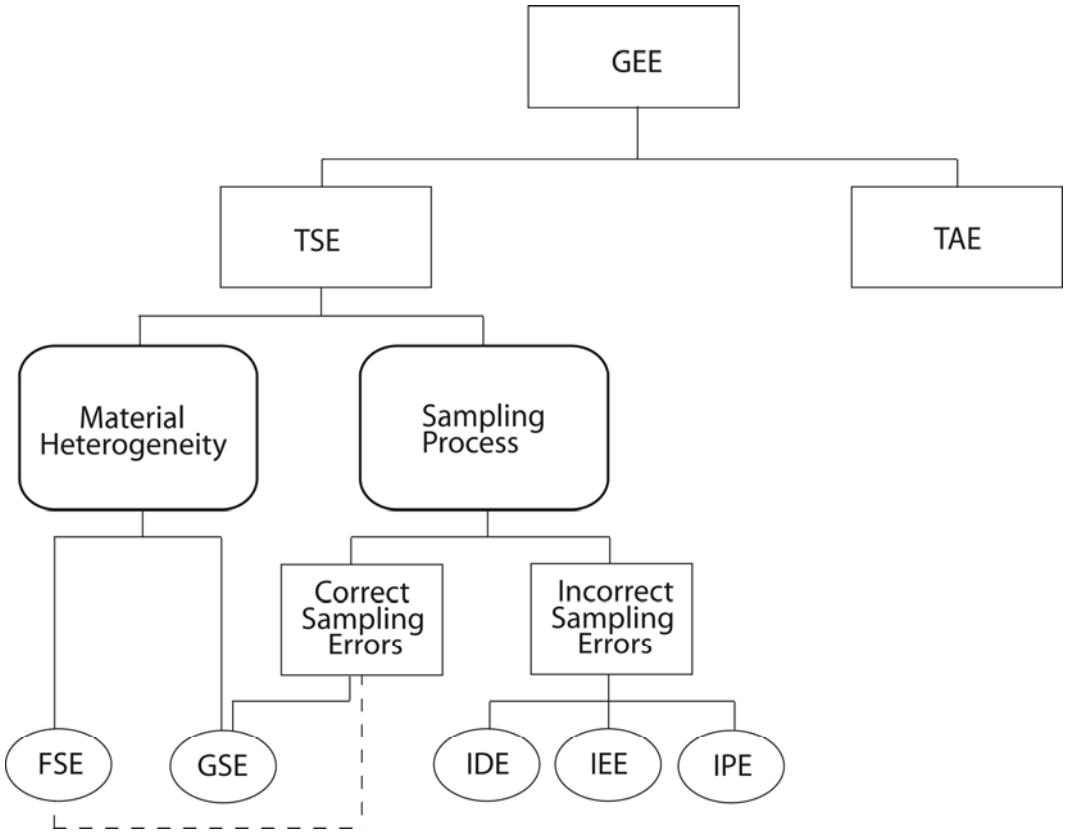

Figure 2.1: Sampling errors involved in 0-D and 1-D sampling.TSE: total sampling error, TAE: total analytical error, FSE: fundamental sampling error, GSE: grouping and segregation error, IDE: Increment delimitation error, IEE: Increment extraction error, and IPE: Increment preparation error. The latter three are collectively known as the "incorrect" sampling errors – opposed to the "correct" sampling errors (FSE + GSE). (Adapted from reference 82)

TSE comprises both the correct (CSE) and incorrect sampling errors (ISE). The correct sampling errors are structural errors caused by the heterogeneity of the material being sampled, resulting in smaller or larger difference between the primary samples extracted if/when sampled in a replication fashion. CSE is the sum of the Fundamental Sampling Error (FSE) and the Grouping and Segregation Error (GSE), which are the two basic sampling errors encountered in all types of lots;

$$m_e(CSE) = m_e(FSE) + m_e(GSE)$$

Or

$$s_e^2(CSE) = s_e^2(FSE) + s_e^2(GSE)$$

Where m_e *is the average of replicated samplings*
s_e^2 *is the variance of the sampling error*

The incorrect (ISE) sampling errors are introduced when applying a non-correct delimitation and/or extraction of the sample and/or when the sample is not handled with care.

It is the primary objective of any sampling process to exclude the bias-generating ISE, preferentially completely – or if not, to document the quantitative effects of any residual ISE.

2.1.1 Correct Sampling Error

The fundamental sampling error (FSE) is inherent to the material properties; size, shape, density of the unit fragments i.e., the compositional heterogeneity of the sample. For a given state of the lot, the FSE is constant and cannot be altered, but by changing the lot system physically (e.g., by comminution) it may be reduced as FSE is always lower for smaller fragments. However, generally FSE can only be reduced to a certain minimum level, as nullification would demand for the whole lot to constitute the sample or the material to be strictly homogenous (see below for definition). This is never the case for any naturally occurring material; thus, FSE is considered the absolute minimum sampling error obtainable if/when all other errors are nullified. In practice this is never carried to completion; therefore, FSE will always have a contribution from the sampling process. See the fragmented line in Figure 2.1.

The Grouping and Segregation Error (GSE) is influenced by both the material heterogeneity i.e., by the compositional heterogeneity (especially by the spatial distribution of the lot) and the sampling process itself (amongst others the sampling mass).

To enable a reduction of TSE it is important to understand all the factors that are influencing the heterogeneity of the material. Counteracting GSE is often the most practical important issue in representative sampling - following successful elimination of all ISE.

2.1.1.1 Material Heterogeneity

Heterogeneity is a universal characteristic of all materials and systems, and can be split in two parts; the constitution heterogeneity of the lot (CH_L) and the distributional heterogeneity of the lot (DH_L).

CH_L depends on the composition, size, and shape of the fragments in the lot alone. This means that materials having a large fragment to fragment difference have a large CH_L, whereas materials with more similar-sized fragments have a lower CH_L. The CH_L can never be reduced to zero, as this would demand that all fragments were indeed strictly identical. It can however be reduced to a practical minimum by changing the physical properties of the sample in example by comminuting. Mixing and blending has no effect on the CH_L. It is due to the constitutional heterogeneity that the FSE exist.

Mathematically the CH_L is defined as the variance of the heterogeneity contributions (h_i) from the lot:

$$CH_L = s^2(h_i)$$

$$CH_L = \frac{1}{N_F}\sum_i (h_i - m(h_i))^2 = \frac{1}{N_F}\sum_i h_i^2$$

$$CH_L = \frac{1}{N_F}\sum_i \left(\frac{a_i - a_L}{a_L} \cdot \frac{M_i}{M_{ave,Fi}}\right)$$

Where *N_F is the number of fragments in the lot*
h_i is the heterogeneity contributions for each fragment
$m(h_i)$ is the average of the heterogeneity contributions = 0
a_i is the concentration in the ith increment
a_L is the concentration in the lot
M_i is the mass of the fragment F_i
$M_{ave,Fi}$ is the average mass of all fragments

DH_L depends on the spatial composition of the lot (CH_L) i.e., the tendency of fragments to group together (characterised by a grouping factor, Y) as well as the tendency of particles to segregate (characterised by a segregation factor, Z):

$$DH_L = \frac{1 + YZ}{1 + Y} CH_L$$

Note that DH_L is defined at the level of the group of fragments, i.e., at the scale level of the practical sampling increment – indeed, the only scale relevant for any sampling process. The larger the difference is in composition between increments in the spatial makeup of the lot, the larger DH_L.

The grouping factor (Y) is dimensionless parameter ($Y \geq 0$), characterising the size of the increments. It is defined as the number of fragments in the lot (N_F) to the number of groups (N_G):

$$Y = \frac{N_F - N_G}{N_G - 1}$$

Where *N_F is the number of fragments in the lot*
N_G is the number of groups in the lot

The numbers of fragments are often much larger than the number of groups, thus yielding the following approximation:

$$Y \approx \frac{N_F}{N_G}$$

Decreasing the number of fragments in the increments/groups will reduce the segregation factor, in principle down to zero when each group contains one fragment only ("ideal sampling", not realisable in practice). This approximation also indicates that carrying out compositional sampling in the manner of keeping the sample mass constant and applying many, small increments will reduce the grouping factor.

The segregation factor (Z) is also a dimensionless parameter (1 ≥ Z ≥ 0) characterising the stratification/segregation status of the lot. Z will be one, if the material is totally segregated and zero when it is totally homogenised, see Figure 2.2.

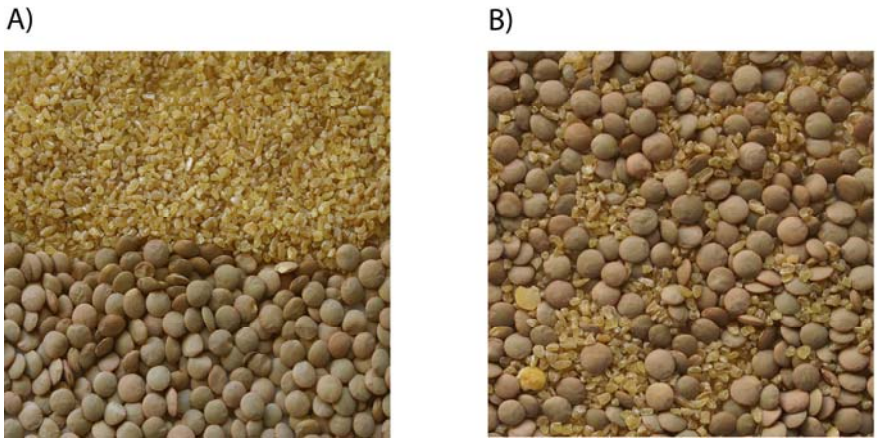

Figure 2.2: Segregation within a material. a) Totally segregated (Z=0), b) Homogenised

When Z = 0 the material is characterised by there being no correlation between the placement of a fragment and its physical and chemical nature. This indicates that the segregation factor may be reduced by mixing and blending.

It is due to the distributional heterogeneity GSE exist, indicating that GSE may be reduced by changing either, the grouping factor, the segregation factor or the constitutional heterogeneity. This indicates that GSE may be reduced by comminution (CH_L), composite sampling (Y), increment reduction (Y), and/or homogenisation (Z). The factors in parenthesis are the ones being modified.

2.1.2 Incorrect Errors

The incorrect sampling errors (ISE) arise when the practical, operational rules of representative sampling are not followed. ISE constitutes the Increment Delimitation Error (IDE), Increment Extraction Error (IEE), and Increment Preparation Error (IPE) respectively. These errors are the

sources of sampling bias, and should always be eliminated or at least reduced to a minimum commensurate with m_0 above.

$$m_e(ISE) = m_e(IDE) + m_e(IEE) + m_e(IPE)$$

IDE is introduced when the material is delimited so that the fragments or increments in the material have an unequal probability of being sampled. Before sampling, the material at hand is delimited into different increments, of which one or more are to be extracted. To avoid a delimitation error, each should be delimited so its probability of ending up in the sample is never zero, and such that no increment has a higher probability of ending up in the sample than any other.

IEE is the result of a non-correct extraction i.e., when a set of practical rules are not followed during sample extraction, thus in practice yielding an increment that deviates from the correctly delimited one. An example of this is seen in Figure 2.3.

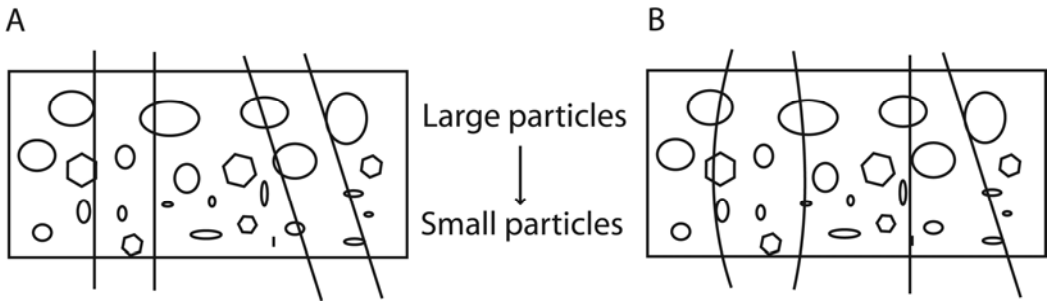

Figure 2.3: Schematic representation of the delimitation of the sample. A) Correct delimitation B) Incorrect delimitation. In applying the incorrect delimitation samples having a larger concentration of the medium sized and small size particles will be obtained in this example of a significantly segregated material. (Adapted from refrence 82)

For example, it is important that the rule of center-of-gravity is followed when handling particulate material i.e., the particles with their center of gravity within the delimited area are always found in the increment after extraction.

IPE covers situations where the integrity of the sample is not upheld after extraction, and thus has several contributors (error sources) like contamination, loss of material, human error, and fraud. According to TOS this is the only error that is non-statistical.[43,84]

Both the correct and the incorrect sampling errors are encountered in all sampling steps, so that the TSE at each step is:

$$m_e(TSE) = m_e(FSE) + m_e(GSE) + m_e(IDE) + m_e(IEE) + m_e(IPE)$$

2.2 Avoiding Sampling Errors

From the general knowledge obtained via the theory and the experience obtained through applying TOS in more than 50 years, seven sampling unit operations (SUO) have been formulated, which should be applied to avoid incorrect sampling, see Table 2.2.

Table 2.2: Sampling Unit Operations (SUO). 1-D, 2-D, 3-D: 1, 2, and 3 dimensional lots. [82]

	Sampling Unit Operation
1	Perform a heterogeneity characterization of new material
2	Mix (homogenise) well before all further sampling steps
3	Use composite sampling whenever possible
4	Use only representative mass reduction
5	Reduce particle size whenever necessary
6	Characterization 1-D heterogeneity using variographic analysis
7	Turn 2-D and 3-D- lots into 1 –D equivalents whenever possible

The SUO may be divided in the more practical SUO (2-5) and the guiding principles SUO (1, 6-7). Not all of them necessarily have to be invoked in a particular sampling situation, but they constitute the arsenal available for solving sampling problems.

2.2.1 Guiding SUO

The first sampling unit operation deals with the overall characterisation of the lot. Before handling any lot, the empirical heterogeneity should be known, thus enabling the delineation of a correct sampling procedure, with respect to primary sample size and the need for; composite sampling, mass reduction, homogenisation and particle size reduction. The heterogeneity of the lot could be obtained – at a very fundamental level – from the estimation of the constitutional heterogeneity (CH_L). However, to do so it is necessary to know the number of fragments (N_F) constituting the lot. In most practical cases the numbers of fragments are impossibly way too high to be counted; therefore, the heterogeneity is often estimated at a practical level commensurate with the actual sampling process in existence, or being designed and implemented.

The famous "Gy's formula" states an important empirical relationship:

$$s^2(FSE) \approx \frac{C \cdot d^3}{M_s}$$

Where *C is a sampling constant*
d is the top particle size (d_{95} in material science terms)
M_s is the sample size

C is a compound "materials' sampling constant" that dependents on four factors; the average density of the particles, the concentration of the analyte in the lot, whether the analyte is incorporated in other components, and the shape of the analyte. For more details see *Gy 1998*[43]. The above equation requires detailed knowledge about the physical properties of the material to be sampled and has mainly found its base in the mining industry. For some materials the knowledge of the physical properties is limited and not easily estimated making the determination of the variance of FSE very difficult, if not impossible. A more practical approach for estimating the heterogeneity may be applied for such materials.

Through replicated extraction of ten samples from the lot (using the current sampling procedure), including the estimation of the analyte concentration, it is possible to estimate the empirical sampling variance and thus the coefficient of variation (C.V.). The result obtained is a sum of all the sampling errors from the primary step down to the analytical technique. Often it is possible by replicating at each sampling step to estimate the contribution (variance) at each step.

Usually the summary TSE is the main interest. If TSE is too high, TOS's principles must be invoked in order for relevant improvements to be made in the sampling procedure. Interestingly, in the most recent TOS contribution it has been possible to derive a highest acceptable threshold for the practical sampling variability as expressed by the C.V. thus arrived at: Any sampling process with a C.V. higher than 16% (20% in practice) is shown to reflect uncontrolled Poisson process characteristics and will have to be worked on further[85]. The threshold level is quoted here for obvious practical reasons; all theoretical argumentation is the responsibility of the other doctoral thesis cited[85].

The coefficient of variation obtained at the primary sampling step is as such not a direct measure of the heterogeneity, but of the practical sum-of-errors caused by the heterogeneity interacting with the specific sampling procedure. In cases where the primary sampling has been carried out in accordance with TOS (eliminating all ISE as well as having reduced GSE as much as needed) it can be seen as the minimal practical error obtainable (MPE) with the current procedure. As such it has a very practical meaning as it is only theoretically possible to reduce

the sampling error all the way down to the magnitude originating from the compositional heterogeneity contribution CH_L.

The sixth and seventh SUO is applied in process sampling when dealing with one-dimensional process streams or time series. By applying these, the effect of long and short term periodic fluctuations are reduced. These will not be dealt further with in the current thesis. For more detail see reference 32, 34, 43, 79, & 84.

In order to obtain a correct primary sample, SUO 2-5 comes into play, as they specify the possibilities for how the sample should be handled practically to eliminate all ISE's and reduce the contributions from the GSE and/or FSE. Contemplating SUO 2-5 it is seen that they deal exactly with reduction of the three contributing factors to FSE and GSE.

Particle size reduction is done by reducing the size of the fragments by comminution. From the Gy's formula it is seen that the particle size is raised to the power of three, making it evident that comminution has a significant effect on the reduction of the heterogeneity of the material and thus on both the FSE and GSE. The contribution from FSE to the TSE may only be reduced by reduction of particle size.

"Homogenisation" (which in all practical cases is mixing only) is carried out to increase the probability of obtaining a sample concentration near the average concentration of the lot. It may be carried out applying different techniques. At the laboratory a mechanical approach applying a mixer or shaking the material is often applied. Another method applicable to particulate material is the long pile method. In this method the particulate material is laid up in a pile by a snake-wise movement of the sample container from one end of the pile to the other, as many times as practical. The composite sample is then extracted as a full, planar-parallel cross section cut of the pile laid up. This method is not standard laboratory practice (yet), but it is highly effective for breaking down the spatial correlation between the different fragments.

Composite sampling is carried out by extracting many, small increments and combining them in one sample. By applying this procedure the effect from the GSE is reduced. By applying this approach the total mass of the sample may increase above the mass needed for analysis. However, this problem is easily overcome by carrying out appropriate representative mass reduction.

Mass reduction may be carried out in a number of different ways; grab sampling, fractional shoveling, riffle splitters, rotational dividers, and others of which not all yield representative subsamples. Description and a comprehensive testing of the different methods may be found in *Petersen et al. 2004*[81]. The overall conclusion from the tests carried out was that grab sampling (the simple extraction of the mass needed for analysis from one point of the lot) yielded the

worst results, and should never be applied; it is instructive to contemplate that the use of a spatula in the analytical laboratory actually is a grab sampling technique.

A method not tested by *Petersen et al. 2004*[81] paper is the long pile method. After homogenising the material by laying it up as described above, a number of increments may be extracted as a cross section of the pile applying a scoop. The number of increments extracted is subsequently combined in a single sample. This is repeated until the mass of the sample is reduced to the size needed for analysis.

2.3 Different Sampling Situations

During this Ph.D. study two completely different type of sampling situations were encountered; sampling of liquid/solid samples from a bioreactor and sampling of straw from agricultural fields, and all necessary mass reduction procedures needed in both cases. Samples were extracted from two different microbial systems; aerobic fermentation representing a well defined system and anaerobic digestion representing a rather under-defined system.

To clarify the task at hand each study was preceded by studying in detail the complete process from primary sampling to analysis, thereby revealing all potential pitfalls. From this analysis, a relevant, optimised sampling procedure was designed applying the SUO assessed as necessary. The procedure outline for each system is summarised in each chapter covering the three processes below.

3 Analytical Measurement Technologies

The analytical measurement technologies are considered the second leg in the trinity of process analytical technologies.

In the biotech industry different analytical technologies (chemical, physical, and biological) are applied for measuring different variables. In Table 3.1 different sensor type examples are given. Chemical and physical sensors are the most widespread, as the biological sensors are much more complex and difficult to implement[58].

Table 3.1: Examples of analytical technologies. HPLC: High performance liquid chromatography, FIA: Flow injection analysis, ELISA: Enzyme linked immunosorbent assay, PCR: Polymerase chain reaction

	Sensor type	
Chemical	Physical	Biological
HPLC	Temperature	ELISA
FIA	Pressure	PCR
Ion selective electrodes	Photometric	DNA/RNA

The pre-request for good bioprocess monitoring and control is the use of a relevant methodology, making the method and the implementation strategy (at-line, in-line or on-line) very process dependent. The bioprocess industry counts several products and application fields: pharmaceutical, food, and energy production. As a consequence the demands span from a very high demanding mammalian cell culture where the asepticity of the system is of outmost importance, and a small fluctuation in temperature or pH may be fatal for the quality of the product, to anaerobic digestion plants where no such critical demands are seen[22, 58, 121].

In this thesis, the frame is set by material characterisation and two different fermentations: anaerobic digestion and aerobic aroma production. The anaerobic digestion is currently being monitored by pH, temperature, flowmeters, and calorimeters, which is inadequate for establishing process instabilities and predicting possible process failures. The aerobic aroma production is being redefined, for which a method is needed for the monitoring of the product precursor. The characterisation of wheat straw is needed for optimising the 2nd generation bioethanol production prior to the establishment of a production facility. For handling these three different scenarios, three analytical techniques were evaluated and their feasibility tested: near infrared spectroscopy, acoustic chemometrics, and electronic tongue.

3.1 Near Infrared Spectroscopy

Near infrared spectroscopy (NIRS) has a very diverse capability of analyzing different organic constituents such as sugars, fat, specific organic compounds, moisture etc. and has been applied to a number of different products such as grain, polymers, dairy products, coffee to mention some. In Table 3.2 a short list of different applications reported in the bioprocces industry is given; more examples are given in reference 9, 16, 52, & 95.

Table 3.2: Application fields reported in the bioprocess industry.

Application field
Mycelial biomass from *P. Chrysogenum and S. Fradiae*[118,119]
Biomass glycerol and methanol in high density complex fed-batch[25]
Oil quality for biodiesel production[7]
Ground barley for use as a feedstock in biofuel production[105]
Chemical composition of rice straw[53]
Moisture, ash, and calorific content in biofuels (wood chips)[68]
Monitoring the dynamics of biogas production[76]

The reasons for the popularity and broad application range are many. In Table 3.3 some of the advantages and disadvantages of NIRS reported in the literature are given. These are strongly process dependent and some are not present in all application areas.

Table 3.3: Advantages and disadvantages of applying NIR[9,95,100].

Advantages	Disadvantages
- May be used in laboratory as well as in industry - Fiber probes available, allowing the instrument and operator to be situated far from aggressive environments - No chemicals - Quick while maintaining accuracy - Non-invasive, inexpensive - Multi component analysis applying a single sensor - Minimized or no sample preparation - Online approach possible - Real time control possible	- Probe fouling (in-situ application) - Calibrated models are not necessarily robust with respect to drifts, upsets and interferences - Gas phase effects (bubbles)

3.1.1 Electromagnetic Spectrum

NIRS is fundamentally classified as molecular spectroscopy and is based on the interaction between electromagnetic radiation and matter. The electromagnetic spectrum spans from the high energy γ-rays (10^{-13} m (10^{11} cm^{-1})), where inner electron transitions in high atom number elements is possible, to the radio waves (10^{5} (10^{-7} cm-1)). The infrared region is situated in the interval 12,500 – 10 cm^{-1}. The region is further divided into near (NIR), mid (MIR) and far (FIR) infrared, NIR constituting the range 2,500 – 800 nm (4,000 to 12,500 cm^{-1}), see Figure 3.1.

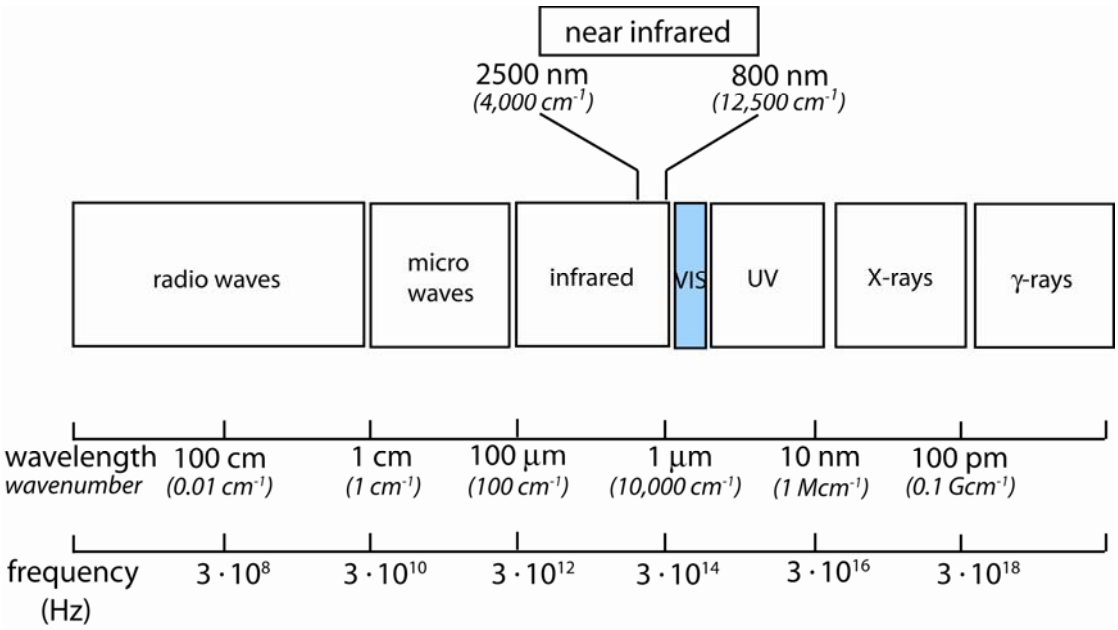

Figure 3.1: Electromagnetic spectrum (adapted from reference 95).

A molecule can occupy different quantum energy levels, the major level corresponding to the different electronic orbitals. Furthermore, for each major energy level a series of quantum vibrational and rotational energy levels exist. The activity seen in the infrared region is due to the fact that infrared radiation has a frequency corresponding to that of the vibrational levels of many chemical compounds. The vibrations taking place may be divided into stretching and bending, see Figure 3.2.

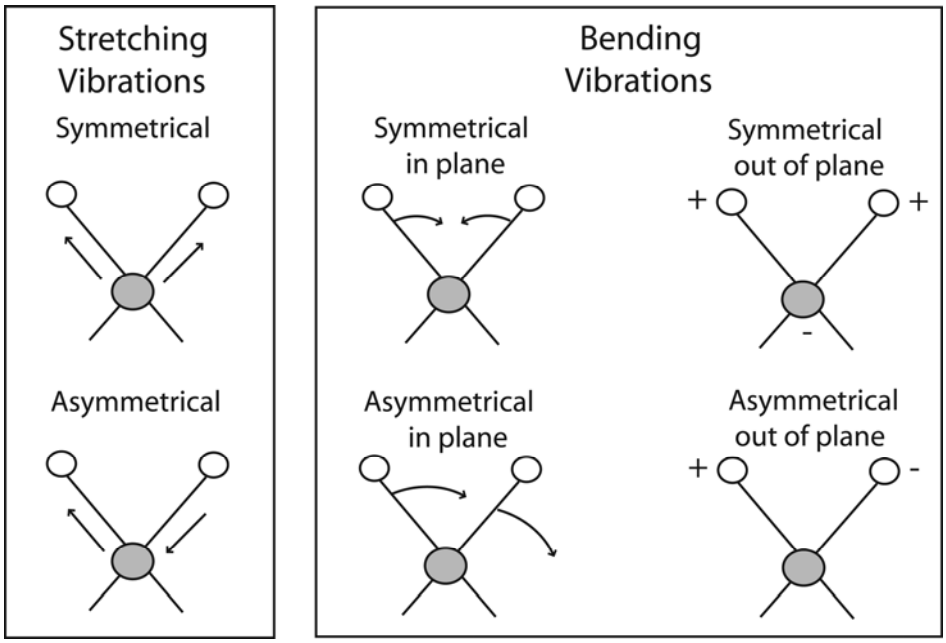

Figure 3.2: Symmetrical and asymmetrical vibrational modes for a molecule containing three atoms.

Six types of fundamental vibrations exist, and the absorption bands in the mid-infrared region are due to these. Besides these fundamental vibrations combinations of vibrations between atoms connected to the same central atom are possible, thus yielding the so-called combination bands.

The absorption bands in the near infrared region are caused by these combination bands and additionally by the presence of overtones. Generally, the absorption in a molecule can be explained by the combination of a simple harmonic oscillator and a quantum mechanical treatment by the Schrödinger equation[99]. The overtones are caused by a deviation from this harmonic state encountered when the vibrating atoms are either close together or far apart. Being close together, a repulsive force between the two atoms exist, working in the direction of the restoring force, resulting in a faster increase in the energy than predicted by the harmonic model. Being far apart the opposite situation is observed, i.e., the restoring force is weak, and as a consequence the decrease in energy is slower. These anharmonic motions yield energy transitions with frequencies of approx. two, three, four etc. times that of the fundamental vibration. Like combinations of fundamental vibrations, combinations of overtones are possible. The probability of the overtones decreases significantly with their order. As a consequence the intensities of these overtones are much lower than that of their fundamental analogs[99,103].

3.1.2 Absorption Bands

For absorption of infrared light to occur, and thus for the molecule to be IR active, two conditions need to be met. The frequency of the radiation must correspond to the transitions between vibrational energy levels in the molecule and a change in the polarity/dipole of the

molecule is needed, i.e., only asymmetric vibration leads to an absorption of infrared light[106]. As a consequence, essentially only molecules containing covalent bonds are IR active and can be measured by NIR. Common bonds contributing to the NIR spectrum are C-H, N-H, O-H, and S-H, making the method versatile and applicable for measurements of many organic compounds. A list of typical absorption bands is given in Table 3.4. The exact location of the band is molecule dependent.

Table 3.4: Typical absorption bands for functional groups in NIR[102].

Group	Combinations	Overtones 1st	2nd	3rd
		[cm-1]		
CH	4098-4405	5618-5917	8130-8547	10695-11050
CH_2	4149-4444	5714-5988	8264-8696	10811-11111
CH_3	4167-4545	5848-6135	8368-8929	10989-11494
H_2O	5208-5405	6920-7177	10101-10471	13072-13514
$R\text{-}NH_2$	4535-4637	6623-6711	9542-9852	12195-12821
ArCH		6061-6211	9091-9259	11494-11765
R-OH	4785-4854	6757-7092	10638-10811	13369-13793

By looking at these intervals it is evident that the absorption bands in the NIR region are broad and overlapping. As a consequence the specificity of NIR is low, making direct identification of specific bands difficult if not impossible.

3.1.3 Spectral Acquisition

Two general modes of spectral acquisition exist; transmission and reflection. Furthermore, a combination of the two is possible, thus yielding a third option termed trans-reflectance or simply transflectance, see Figure 3.3.

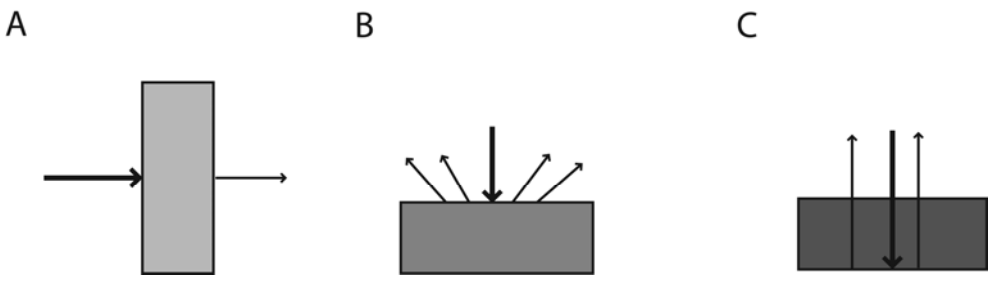

Figure 3.3: Three modes for spectral acquisition. A) transmission, B) reflectance and C) transflectance.

In transmission the near infrared light is passed through the sample, and the absorbance calculated from the ratio of the intensity of the reference and the sample (the transmittance);

$$A = \log\left(\frac{1}{T}\right) = \log\left(\frac{1}{\frac{I}{I_0}}\right) = \log\left(\frac{I_0}{I}\right)$$

Where A is the absorbance
T is the transmittance
I_0 is the intensity of the reference beam
I is the intensity of the beam after passing through the sample

Transmission is normally used for tablets and liquid samples.

In the reflection mode the light is reflected by the sample instead of going through, and a parameter (log 1/R) corresponding to the absorbance is calculated from the reflectance.

$$\log\left(\frac{1}{R}\right) = \log\left(\frac{1}{\frac{I}{I_0}}\right) = \log\left(\frac{I_0}{I}\right)$$

Where A is the absorbance
R is the reflectance
I_0 is the intensity of the reference beam
I is the intensity of the beam being reflected by the sample

Reflection is normally used for solids, e.g., powders.

The trans-reflectance mode is as mentioned a combination of the two where light passes through the sample, into a mirror and is then reflected back through the sample. Trans-reflectance has proven useful when measuring liquid samples and liquid samples containing solids.

3.1.4 Instrumentation

A number of different NIR instruments exist, and a list of various companies, type of instrument, and wavelength coverage's was put together by *Workman & Burns, 2008*[123]. In the current thesis a Quant (ABB) and Corona 45 NIR (Carl Zeiss) were applied, the former being a Fourier transform instrument and the latter a dispersive instrument.

A Fourier transform instrument principally consists of a light source, a Michelson interferometer, a detector, and a laser. The near infrared light is emitted from the source to the Michelson interferometer. Here it is divided into two by a beam splitting mirror and reflected

by two mirrors prior to being recombined. One of the reflecting mirrors is moving, thus changing the intensity of the near infrared light. The light is passed through the sample and into the detector. The laser is applied for controlling the data acquisition in the detector. In the Quant, a quartz halogen lamp and an InAs detector are applied. The signal is obtained as an interferogram, which is the output as a function of the difference between the distances traveled by the two beams in the interferometer. This is transformed into the frequency domain by a fast Fourier transformation algorithm (FFT).

A dispersive instrument consists of a light source, a slit, a dispersion device, and a detector. The near infrared light is emitted from the source and sent through the sample. From the sample it is passed through a slit, dispersed in a monochromater (usually a grating) and sent to the detector. In the Corona 45 NIR a halogen lamp (10 W) and a diode array detector are applied.

3.2 Acoustic chemometrics

Acoustic chemometrics (a.c.) is a multivariate analytical technique which amongst others has shown potential for determination of particle size and quantification of oil in water. Table 3.5 is a short list of applications reported in the literature.

Table 3.5: Application fields for acoustic chemometrics.

Application field
A pipeline transporting a dense slurry containing silica particles[50]
In-line prediction of powder particle size distribution[44]
On-line control tool for a urea fertilizer production[46]

None of the cases reported is from the bioprocess industry, but from these and the advantages and disadvantages reported for the method (Table 3.6) it is evident that a.c. should also be of value in biotechnological applications, e.g., for monitoring of oil in water, particle size distribution, and total solids. The fact that this sensor modality is non-intrusive may be of significant value to the fermentation industry where ascepticity is critical (not an issue for anaerobic fermentation).

Table 3.6: Advantages and disadvantages of applying acoustic chemometrics.

Advantages	Disadvantages
- Non-intrusive sensors - One sensor may be applied for monitoring of several process parameters - Relative inexpensive equipment - On-line/real time approach	- Not necessarily as accurate and precise as laboratory methods - Calibrated models are not necessarily robust with respect to process drifts, upsets and interferences

3.2.1 Vibrations

Vibrations are dynamic phenomena associated with transportation and/or processing, for example, caused by the transfer of kinetic energy to a pipeline wall. During transportation of particles in a pipeline, kinetic energy will be transferred to the wall each time a particle hits. This energy is absorbed and released as vibrational energy. Vibrations created within a process are process and materials dependent, and in acoustic chemometrics these are measured and applied for quantifying different process and product parameters.

For measuring the vibrations a vibration transducer is applied, for example a piezoelectric accelerometer, see Figure 3.4.

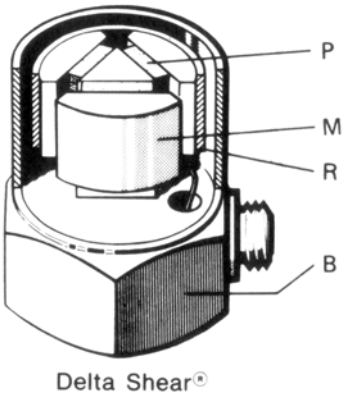

Figure 3.4: Schematic drawing of piezoelectric accelerometer. P) Piezoelectric crystals, M) seismic mass, R) mounting ring, and B) accelerometer base[98].

The accelerometer consists of a base which is connected to a seismic mass by piezoelectric elements. During process monitoring the base is e.g., glued (or otherwise fixed) to, for example, a pipeline and thus subjected to vibrations. Due to the vibrations a force corresponding to the product of the acceleration of the seismic mass and its mass acts on each piezoelectric crystal, and as a result each piezoelectric crystal generates a charge proportional to the applied force.

As the mass of the seismic mass is constant the charge generated is proportional to the acceleration of the seismic mass.

The acceleration of the seismic mass has a magnitude and phase corresponding to that of the accelerometer base, and as a consequence the acceleration measured is that of the surface onto which the accelerometer has been mounted.

The charge generated by the piezoelectric crystal is converted to voltage or a current and registered by a computer. The vibrations are measured in the time domain (change of amplitude of the vibration as a function of time) which is transformed to the frequency domain by a fast Fourier transformation.

3.2.2 Instrumentation

The acoustic chemometric experimental setup is given in Figure 3.5.

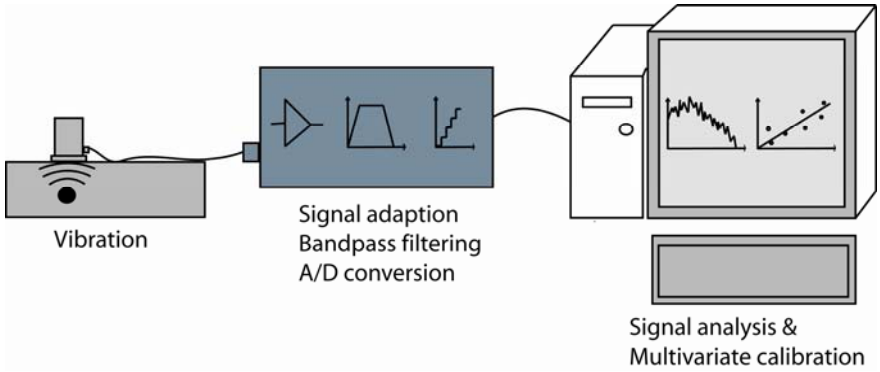

Figure 3.5: Experimental setup and signal conditioning during acoustic chemometric measurements.

In the present thesis, the vibrations are measured by the accelerometer Deltatron®4396, which generates a current that is preprocessed prior to entering the computer.

The first step in the preprocessing is a signal adaption step, where the acoustic signal is amplified. The second step is a filtering. This is done by an analog pass-band filter, which passes frequencies within a certain range (bandwidth), and retains those outside the bandwidth. By applying the right filter, signals with only the frequencies of interest can be obtained. The third step is an analog to digital conversion of the signal, thus allowing it to be transferred to the computer.

In the computer the signal is transformed from the time domain into the frequency domain by applying a fast Fourier transformation (FFT) algorithm. This frequency spectrum is applied as the X data in the multivariate analysis[45].

3.3 Electronic Tongue

An array applying different chemical sensors was first introduced in 1985, and has since been termed taste chip, taste sensor/system, and electronic tongue. The electronic tongue has since been applied successfully for food, environmental, and industrial analysis, and a short list of applications reported in the bioprocess industry is given in Table 3.7. More examples are given in reference 21, 91, and 120.

Table 3.7: Application fields for electronic tongue reported in the bioprocess industry.

Application field
Prediction of Port wine age (2 to 70 years old)[92]
Identification of green tea grade level[19]
Off-line measurements on rapid *Eschericia coli* batch fermentation[117]
Monitoring the production of fermented food[56]
Quantification of organic acids (citric, lactic, and orotic) in a starting culture for cheese production[33]

Most of the reported applications reported are off-line analyses, a few explanations to this may be found in Table 3.8. From the disadvantages it is evident that the contamination of sensor surfaces in complex media, the risk of drifts in characteristics, and problems with the calibration drift, need to be seriously addressed before implementing the technology for routine analysis. The advantages are of course sensor type specific and not all may be relevant to all chemical sensor types.

Table 3.8: Advantages and disadvantages applying electronic tongue[67,91].

Advantages	Disadvantages
- Minimized or no sample preparation - Relatively simple and inexpensive equipment - Easy automation of measurements - Low concentrations can be quantified - Recognition of complex liquids possible - Better detection limit and selectivity than for specific electrodes - Real time control possible	- Contamination of sensors - Temporal calibration drifts may be seen

3.3.1 Chemical Sensors

The electronic tongue consists of an array of sensors and as such a wide variety may be applied in the design: electrochemical, optical or enzymatic. The most applied systems and also the one applied in the current thesis are electrochemically based potentiometric sensors. Different types of potentiometric sensors exist: ion-selective or cross sensitive and non-specific, where experiments have shown that an array of cross sensitive or non-specific electrodes is optimal compared to an array containing only specific ion selective electrodes.

The cross selective or non-specific electrodes applied in the electronic tongues consists of a piece of PVC tubing to which a membrane has been glued. The sensor either has liquid inner filling or solid inner contact. See Figure 3.6.

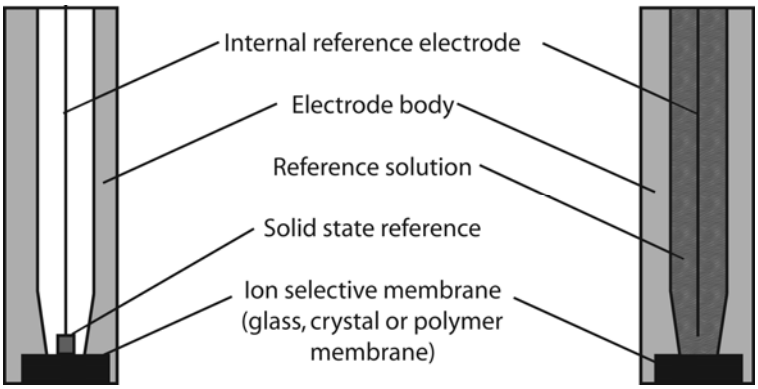

Figure 3.6: Schematic drawing of electrodes with liquid inner filling (right) or solid inner contact (left).

The membrane material applied in the sensors determines the selectivity of the sensors, thus making the choice of membrane task specific. A wide variety of different materials from inorganic compositions to organic polymers have been reported, for example chalcogenide glasses doped with different metals, PVC polymers containing various active components as ionophors, metalloporphirines, and crystalline compositions[66]. In the sensor array design phase 10-30 sensors with different membrane compositions and thus working principle are normally tested for their cross sensitivity and their responses. The cross sensitivity, i.e., sensitivity towards as many components in the solution as possible, should be high and the response reproducible.

The key element in the ET approach is the powerful multivariate calibration of the complement of selected sensor signals. Sometimes a non-linear PLS approach has been found useful, matching non-linear dose-response relationships for some sensors. Mild non-linearities may of course be well compensated by one, or a few, more PLS-components.

A comprehensive and very useful overview of the state-of-the-art ET scene can be found in the proceedings from the latest ISOEN conference in 2007[65]. In the present thesis 20 sensors were tested for their responses to octanoic acid and six chosen for the final array, which all had

sensitivities to octanoic acid from 10^{-5} mol/L. The sensors chosen were all potentiometric non-selective electrodes with solid inner contact. No information regarding the actual composition of the membrane could be obtained, making the overall selectivity of the membranes in the array unknown.

3.3.2 Instrumentation

The array applied consisted of six non-specific potentiometric sensors with solid inner contact and a pH electrode. The array was connected to a custom-made multi-channel digital voltmeter with high input impedance, which was connected to a computer for data acquisition. The signals obtained were the potential difference between each electrode and a reference electrode. The reference applied was a double junction electrode, with an inner reservoir of saturated potassium chloride and an outer of ammonium nitrate. See Figure 3.7.

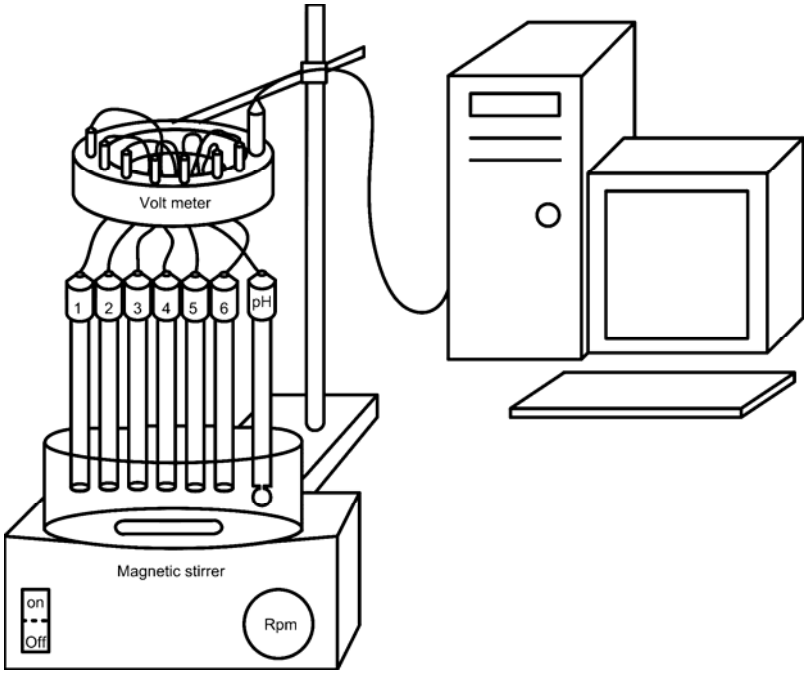

Figure 3.7: Experimental set-up for the electronic tongue.

4 Chemometrics

Chemometrics is here considered the third leg in the trinity of process analytical technologies.

Several definitions on chemometrics exist, the present one given by the journal of *Chemometrics & Intelligent Laboratory Systems*[18]:

"Chemometrics is the chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments and to provide maximum chemical information by analysing chemical data"

The general approach in chemometrics is to consider the world as multivariate instead of the univariate approach often applied. When several measurements are made on the same system, some of these may correlate in some way. By finding these correlations, new knowledge about the system will almost always be obtained.

Three major application fields in chemometrics are: explorative data analysis, discrimination & classification, and regression & prediction. In this thesis a short introduction to explorative data analysis and regression & prediction is given.

Before carrying out the data analysis, the data is often assessed to see if pre-treatment is needed. Several pre-treatment types exist of which a few are described in section 4.5.

4.1 Principal Component Analysis

For explorative data analysis, the backbone of multivariate data analysis is applied: Principal Component Analysis (PCA). A short definition of PCA can be given as follows:

> *"Principal component analysis - a mathematical procedure for resolving sets of data into orthogonal components whose linear combinations approximate the original data to any desired degree of accuracy. As successive components are calculated, each component accounts for the maximum possible amount of residual variance in the data set. In spectroscopy, the data are usually spectra, and the number of components is smaller than or equal to the number of variables or number of spectra, whichever is less."*[5]

The starting point for resolving spectral data is a data matrix (**X**) consisting of *I* rows and *K* columns, where the rows are the observations; samples, chemical reactions, etc. and the columns the variables: electrode responses, wavelengths in a spectrum, etc. See Figure 4.1

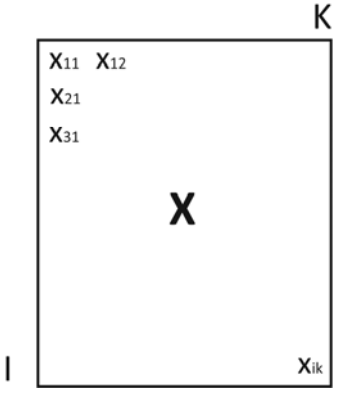

Figure 4.1: Raw data matrix.

The purpose of PCA is to reorganize these K variables in such a way that a new system describing the important variation in as few factors as possible is obtained. This is done by constructing a coordinate system that is orientated along the directions of maximum variance. For example a 25 x 3 matrix may be plotted as seen in Figure 4.2.

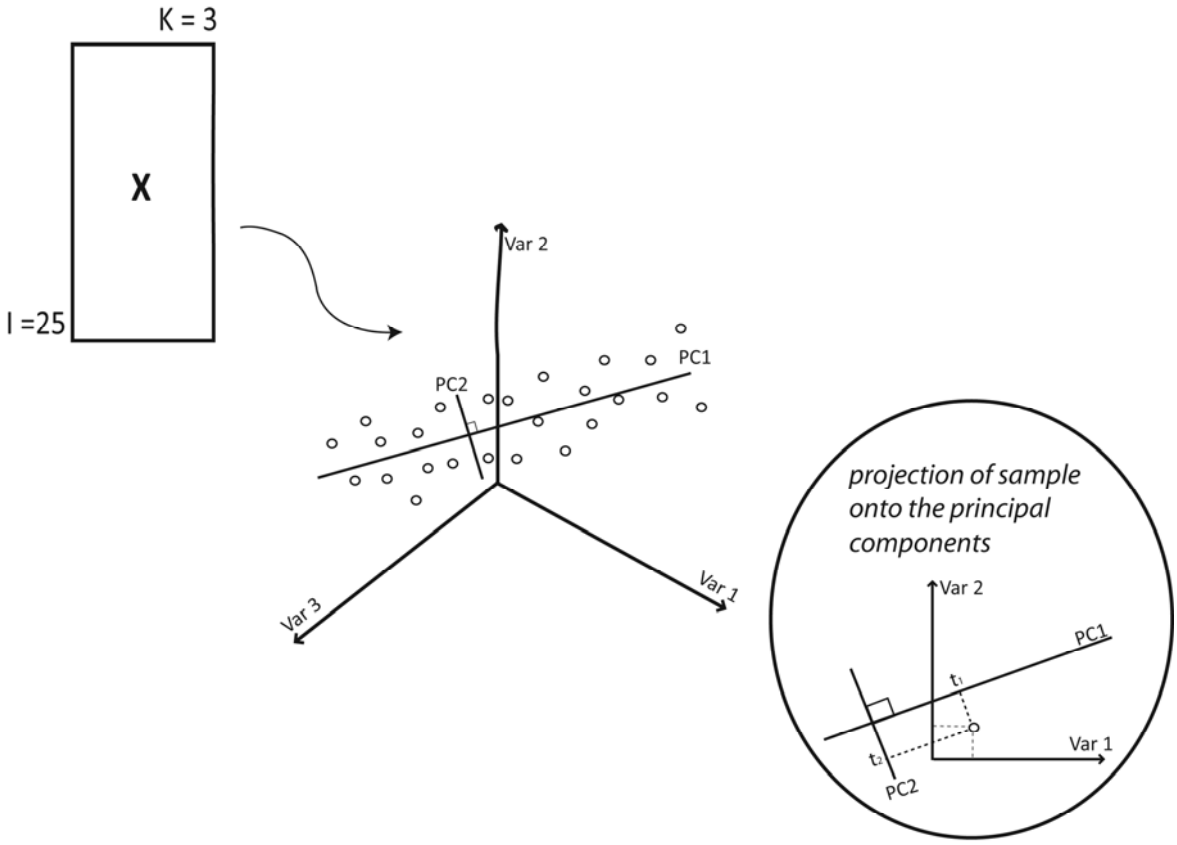

Figure 4.2: A data matrix plotted in a three dimensional variable space (K=3). The intersection between all PC's is found as the average observation.

By placing a line along the direction of the largest variance, more information regarding the relationship between samples is obtained than by viewing along any of the three original axes. This direction is termed the 1st principal component (PC_1). Successive directions are placed perpendicular to all previous directions, which all go though the PC_0 (calculated as the average observation). Thus, the second highest variance will be modelled as the 2nd principal component, etc. By always calculating the components perpendicular to each other, unique components are obtained in the sense that they alone describe a certain amount of total variation in the data set.

The PCs can be found by many equivalent numerical procedures, e.g., fitted by reducing the sum of all squared projection distances e_i (residuals) to a minimum.

For any data matrix, *centering* is often performed by moving the center of the model from the origo of the original coordinate system to the average observation. In general the average observation does not correspond to any real-world observation (object).

By perpendicular projecting the observations onto the principal components their coordinates in the new system, termed *scores* (t), are obtained, which are collected in a score matrix (**T**). A row in the score matrix makes up the scores for a single observation, while the columns are the projection points of each observation unto a single PC. Scores may be either positive or negative depending on their location in relation to the model center (average observation). The scores are related to the original coordinate system by weights, called *loadings* (p), which are collected in the loading matrix (**P**). A row in the loading matrix is the contribution of a single variable to the PCs. See Figure 4.3.

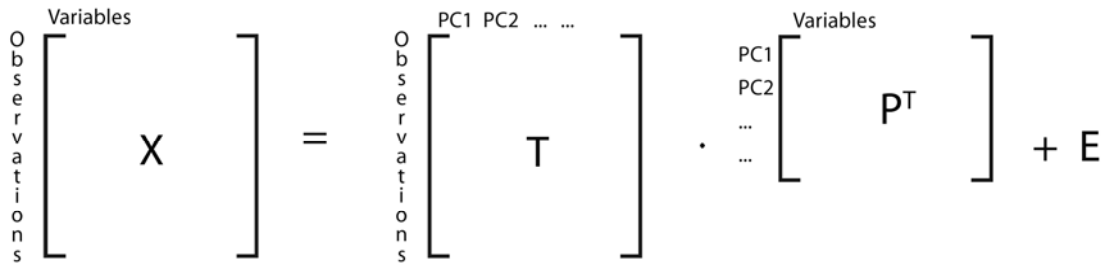

Figure 4.3: Matrix representation of a centred PCA model.

Mathematically, the contribution of a variable to a PC is the cosine of the angle between the variable axis and the PC. If a PC points in exactly the same direction as the original axis, the angle between them is zero, thus yielding a loading of 1, i.e., the PC is a total explanation of the variance given by this one variable alone. If the PC is placed perpendicular to original axis, a loading of zero is obtained, i.e., the PC explains nothing of the variation seen in that variable.

E is the error matrix representing the residual part of **X**, which is not explained (modelled) by the model. This part is generally regarded as the *noise part* of **X**, and is used as a measure of the

model fit, i.e., how closely the model corresponds to **X**. If **E** is large, the PCA fits the data poorly and vice versa.

In interpreting a model the scores and loadings are of special interest. In plotting the different scores against each other the different observations are related to one another, while plotting the loadings against each other gives a graphical view of the inter-variable relationship. [8, 29, 30, 104]

4.2 Partial Least Square - Regression

The output from a given instrument (**x**) may be related to the properties of a sample (**y**) by establishing a regression model (calibration), thus enabling the possibility for prediction based on future samples. As an example, a NIR spectrum may be related to the concentration of a specific sugar component in a sample, thereby enabling future measurements of this particular sugar concentration by NIR. Several methods for doing this exist; the most popular however is the Partial Least Squares (PLS) regression, originally suggested by Herman Wold in the mid seventies[54].

One of the advantages by using PLS is that it can handle colinearity in the data, which is very often encountered, e.g., in NIR spectra. This is done by applying new component variables, *latent variables,* for expressing the relevant information found in **X**. The information residing in the **X** matrix is termed the independent data, whereas the data in **Y** is dependent.

For carrying out calibration it is crucial that the samples applied span the range expected in future samples, i.e., samples similar to the ones that are going to be obtained in the future have to be applied. Generally the PLS regression consists of three steps, see Figure 4.4.

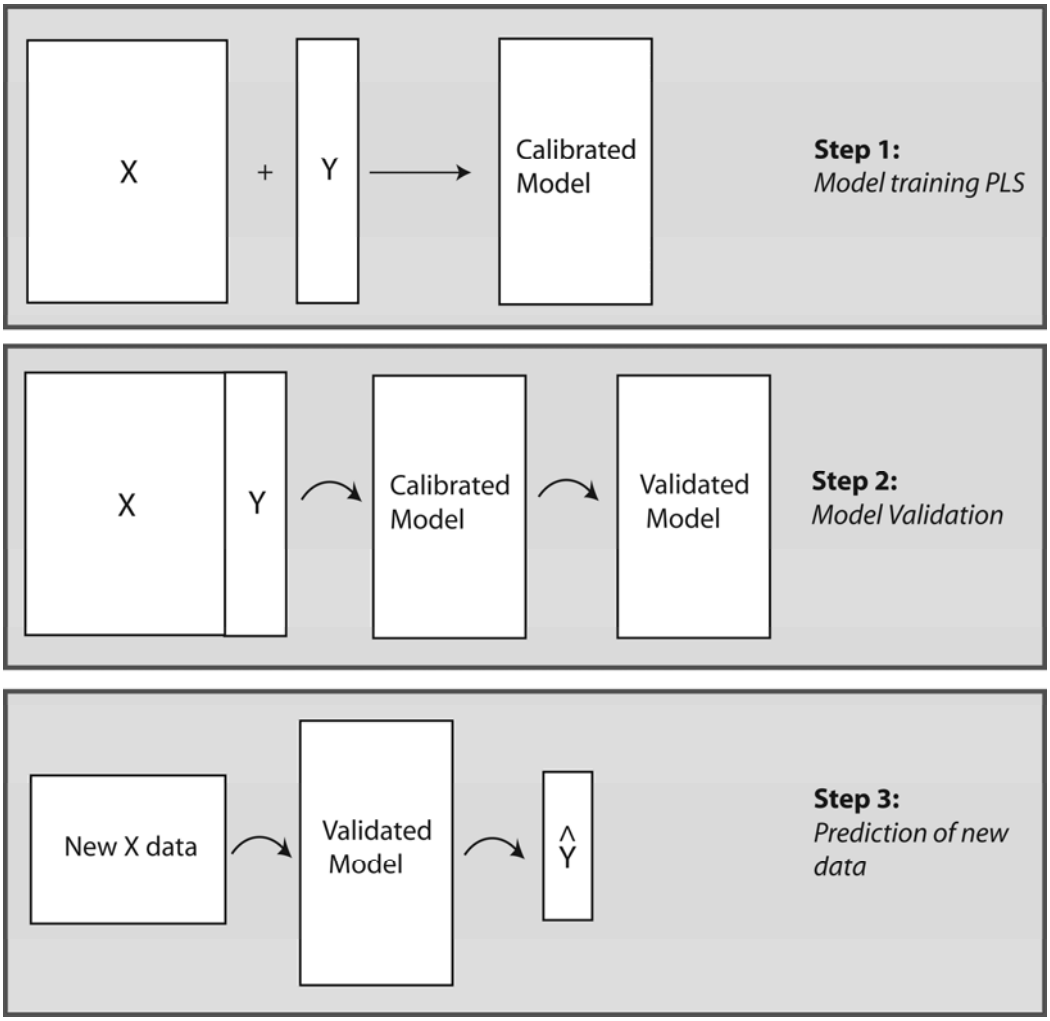

Figure 4.4: Three steps in Partial Least Square regression.

The calibration step carried out in a PLSR is analogous to PCA from a projection point of view. However, in the PLS the y-data is applied directly in decomposition of **X**, thereby obtaining a regression model that contains exactly the information from **X** that is most relevant for modelling the property sought in the samples. This is obtained by applying y-data information (y-score vectors) instead of the scores (t) in calculating the X-loadings. Since these are principally different from their PCA counterparts they are termed loading-weights (**w**). By doing so the least-squares solution ($\mathbf{w} = \mathbf{X}^T\mathbf{y}/|\mathbf{X}^T\mathbf{y}|$) to the following equation is obtained:

$$\boldsymbol{X} = \boldsymbol{y} \cdot \boldsymbol{w}^T + \boldsymbol{E}$$

The X-data is now projected onto this loading weight, establishing the corresponding score values (**t**). The first PLS component is now established in both the X- and the y-space, from which the residual part (non-explained) of **X** and **Y** is calculated and used in establishing the next component. The procedure is repeated until the optimal number of components describing the sought for regression relation is obtained (see validation below).

One algorithm applied for establishing the PLS components is termed the NIPALS-algorithm (Nonlinear Iterative Partial Least Square) and is depicted in Figure 4.5.

$\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are centered*

$$f = 1, \boldsymbol{X}_f = \boldsymbol{X}, \boldsymbol{y}_f = \boldsymbol{y}$$

$$1.\ \ \boldsymbol{w}_f = \frac{\boldsymbol{X}_f^T \boldsymbol{y}_f}{\left|\boldsymbol{X}_f^T \boldsymbol{y}_f\right|}$$

$$2.\ \ \boldsymbol{t}_f = \boldsymbol{X}_f \boldsymbol{w}_f$$

$$3.\ \ \widehat{q_f} = \frac{\boldsymbol{t}_f^T \boldsymbol{y}_f}{\boldsymbol{t}_f^T \boldsymbol{t}_f}$$

$$4.\ \ \boldsymbol{p}_f = \frac{\boldsymbol{X}_f^T \boldsymbol{t}_f}{\boldsymbol{t}_f^T \boldsymbol{t}_f}$$

$$5.\ \ \boldsymbol{X}_{f+1} = \boldsymbol{X}_f - \boldsymbol{t}_f \boldsymbol{p}_f^T \ \ \&\ \ \boldsymbol{y}_{f+1} = \boldsymbol{y}_f - \widehat{q_f} \boldsymbol{t}_f$$

$$6.\ f\ equal\ to\ f + 1\ and\ step\ 1\ to\ f\ is\ equal\ to\ F$$

$$7.\ \widehat{\boldsymbol{X}} = \boldsymbol{T}\boldsymbol{P}^T \ \ \&\ \ \widehat{\boldsymbol{y}} = \widehat{\boldsymbol{q}}\boldsymbol{t}$$

Figure 4.5: NIPALS algorithm for PLS[14].

For the final model the following is obtained: scores (**T**), loadings (**P**), and loading weights (**W**). The loading weights are obtained as a result of y-guided decomposition of the X-data, and describe the *effective loadings* that are needed in constructing the X-Y regression relation. For this reason, it is only the loading weights which are of interest for interpretation of the X-Y model relationships.

After calibration, an evaluation of the fitted model is carried out, i.e., the model is validated. This is done by evaluating the size of the residuals in calculating a prediction error estimate. Different methods for establishing more or less realistic validation exist: leverage correction, cross validation, or test-set validation. These methods will be described further in section 4.3. Validation should furnish *reliable* information as to the realistic magnitude of the average prediction error for future samples. From the validation it is decided if the model quality is good enough for the purpose for which it is intended. This is done by applying different measures, see section 4.4. [14, 30, 70, 104]

4.3 Validation

As a part of modelling, validation must always be carried out. The objective is to establish the optimal number of component needed and to establish the validity of the final model obtained, which will facilitate a measure of the prediction error as well. Validation is carried out by estimating the predicted residuals, and may be performed applying one of the three following methods; test set validation, cross validation, or leverage corrections. The only realistic measure of the residuals is obtained by applying the first method.

4.3.1 Test Set

Test set validation is the most stringent and best, as entirely new observations/samples only are used for evaluation. In practice this is done by collecting new samples, on which both **X** and **Y** are measured. The X-data obtained is fed into the calibrated model, and the residuals estimated as the difference between the predicted y-value and the y-value measured for the observation/sample. From these individual residuals the total average prediction error for the model is estimated. The optimal number of components determined as where the prediction error is minimal (see 4.4.2).

In applying an entirely new data set for validation, the future situation, for which the model is intended, is tested in as realistic scenario possible, i.e., the performance of the model on future samples is tested under the most relevant conditions.

The reason why it is so important to apply new data for the validation procedure may be illustrated by the Theory of Sampling. According to TOS no such thing as a constant sampling bias exists, due to the heterogeneity of the lot (possibly augmented by non-vanishing ISE). In practice this means that a variation from sample to sample exist, which is manifested in the total sampling error (TSE). The ensemble TSE for two data sets will never be identical, and definitely not in cases were a material heterogeneity exists. By drawing two independent data sets, two estimations of this TSE manifestation are obtained, thus enabling estimation of the overall prediction error including the contribution from the sampling procedure realizing varying ensemble TSE. The logical consequence of this argument is that since the TSE is never identical for two drawings, even the estimation obtained by test-set validation will not be identical to the one obtained for future samples. In order to fully cope with this variability an infinite number of data sets should be used for test set validation, but in practice one or two is usually considered enough (for obvious “practical reasons”).

To get a reliable prediction error estimate it is always important that the test set(s) spans the same range as that of the model and that all other influential factors and conditions are as identical to the future situation in which the prediction model is to be used as possible.

4.3.2 Cross Validation

In cross validation there is no test set. In cross validation the calibration data set is simply divided into n segments. One segment at the time is kept out during calibration and subsequently applied for validating the model; this is repeated n times corresponding to n segments having been temporarily excluded. For each sub-model, in which a segment is excluded, the predicted residuals are calculated and later summarized giving an alternative measure of the average prediction error. The number of samples in a segment can be set between 1 and I/2 of the total number of samples in the calibration set, I. The first option, termed full cross validation, gives a very clear overoptimistic prediction error estimate. The latter, termed 2 segment cross validation, gives a prediction error estimate closest to the one obtained using a test set. However, it is critical to acknowledge even in the case of a 2 segment cross validation that no test set has ever been included in the validation. As a consequence there is no way of knowing how large the full ensemble TSE contribution would have been, i.e., how close the prediction error obtained by cross validation is to the realistic, full test set level and thus how good the future prediction performance of the model really is. All cross validation results are actually just a measure of the internal subset stability of the training set modelling[31,32].

Cross validation is often applied for establishing the optimal number of components in a model and subsequently for validation as well with the dangers outlined above always present.

A situation where cross validation can be said to be acceptable does exist. In cases where a small data set is at hand, applying a cross validation procedure may in fact be the only solution possible. If this is done, the number of samples per segment should be kept as high as possible.

4.3.3 Leverage Correction

A *leverage* (h) is the effect of an observation on the model i.e., observations having a high leverage is situated far from the model centre. Using the leverage for model validation is possible, but it has been termed a “quick and dirty” method as only one data set is used for both calibration and validation, and the prediction error obtained is over optimistic. In leverage correction the error estimates are corrected by the leverage of each observation/sample before calculating the prediction error. This is done by dividing the residual for an observation by (1 – h), thus giving observations far from the model a higher contribution to the overall error estimate. This seems fair; however, these far-away observations will often be situated close to the PC direction, thereby giving too low residual estimations and thus an overoptimistic prediction error estimate. Therefore, leverage correction should only be applied – if at all – in the initial validation steps like outlier removal, and not in any final validation step.[30, 80, 104]

4.4 Model Evaluation

As a means of evaluating the prediction models obtained, different measures may be applied. In the present work the correlation and the slope obtained from the prediction statistics are applied. Furthermore, different measures of the average prediction error, estimated from the residuals, are applied: the "Root Mean Square Error of Prediction" (RMSEP), % RMSEP, and finally the ''ratio of standard error of performance to standard deviation" (RPD), which is a special favourite in spectroscopy.

4.4.1 Slope and Correlation

The correlation and slope obtained from the predicted vs. measured plot are measures of the precision i.e., "degree of reproducibility" and accuracy i.e., "closeness to the actual result" respectively.

The squared sample correlation coefficient (r^2) is modelled by the specific relation, i.e., the proportion of total variation in the values of the variable Y that may be accounted for by a linear relationship with values of a variable X:

$$r^2 = \frac{{S_{xy}}^2}{\sqrt{S_{xx}S_{yy}}}$$

Where the entries for S_{xx}, S_{yy}, and S_{xy} are defined as:

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$S_{xy} = \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

The squared sample correlation may be applied for comparing different models, as it is conveniently independent of the measuring units involved.

4.4.2 Root Mean Square Error of Prediction

The average prediction error termed the *Root Mean Square Error of Prediction* (RMSEP) for a test set validated model is defined as:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_{i\,\mathrm{predicted}} - y_{i\,\mathrm{reference}})^2}{n}}$$

Where $y_{ipredicted}$ *is the predicted y value of the* i^{th} *sample*
$y_{ireference}$ *is the measured y value of the* i^{th} *sample*
n is the number of samples in the model

The numerator is an expression of the error obtained when fitting the y values, and will generally decrease as components are added to the model, thereby decreasing the RMSEP. Adding an extra component to the model also increases the error associated with estimating the regression parameters. By plotting the RMSEP as a function of the number of PLS components, the two errors are summarised, thus yielding a more or less defined minimum, see Figure 4.6.

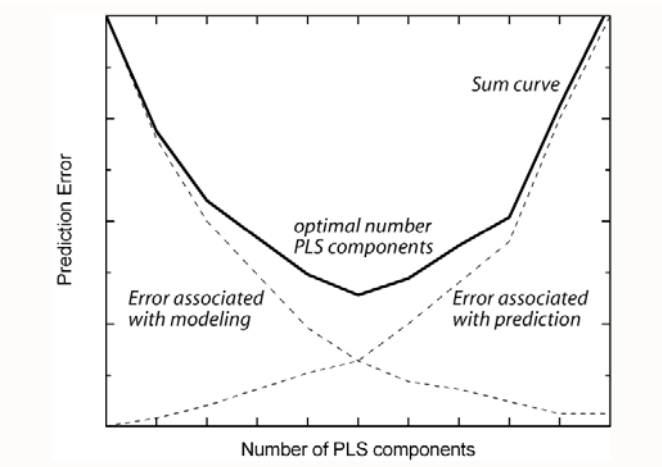

Figure 4.6: Prediction error as a function of the number of PLS components in the model.

The error of prediction is often termed the RMSEP or RMSECV; the latter is an abbreviation of *Root Mean Square Error of Cross Validation*, the suffix being an indication of the validation method from which it originates[8,30,70]. In the published articles in this thesis it was decided to use the term RMSEP, independent of the validation method applied however.

The RMSEP is dependent on the measuring units applied and has the same unit as the y-variable. To enable model comparison the % RMSEP may be calculated.

$$\%\,\mathrm{RMSEP} = \frac{\mathrm{RMSEP}}{y_{max} - y_{min}} \cdot 100\%$$

As an alternative, sometimes the average y-level can be used in the denominator. The important issue is always to be cognizant on the definition of the specific % RMSEP in use, and of course only compare when identical definitions are documented.

4.4.3 RPD

RPD is an abbreviation of ‘‘*ratio of standard error of performance to standard deviation*”, which can also be used for model evaluation and comparison, as it is a dimensionless parameter. The RPD is defined in two different ways:

$$\mathrm{RPD} = \frac{s_Y}{\mathrm{RMSEP}} \quad or \quad \mathrm{RPD} = \frac{s_Y}{\mathrm{SEP}}$$

Where s_Y *is the standard deviation on the reference data used for validation*
SEP is the standard error of performance

SEP is the RMSEP corrected for the bias, and it is a measure of how good the model would be if the bias problem could be solved, and not how good it actually is. Therefore, it is here chosen to apply the first of the two equations.

The values for the RPD may be interpreted as the ratio of natural variation in the samples to the size of the likely prediction errors. Ideally the RMSEP should be much lower than the standard deviation, thus yielding RPD values above five[37,122], see Table 4.1.

Table 4.1: RPD statistics[122].

RPD	Classification	Application
0.0 – 2.3	Very Poor	Not recommended
2.4 – 3.0	Poor	Very rough screening
3.1 – 4.9	Fair	Screening
5.0 – 6.4	Good	Quality control
6.5 – 8.0	Very good	Process control
8.1 +	Excellent	Any application

One should not put emphasis on the scalar RPD values alone; RPD expresses the descriptive “aspect ratio” of width vs. length of any data point spread, within a particular prediction validation context. Together with graphical illustration of prediction vs. reference relationships, RPD conveys a convenient short hand message besides other standard statics employed in spectroscopy and chemometrics.

4.5 Preprocessing

During data gathering, different factors (light scattering effects, noise, absorptions) may influence the samples or variables. Often this contribution gives no chemical information, and thus merely yields models that are more complex (more PLS components are needed), if a decent model can be obtained at all.

Many different methods for reducing or eliminating these effects have been reported. In the present work scaling, multiplicative scatter correction, and Savitzky–Golay derivatives have been applied.

4.5.1 Scaling

A question to ask is to scale or not to scale? Scaling is performed by correcting each value with the standard deviation of the given variable, by doing so each scaled variable gets the same variance as all others:

$$x_{scaled} = x \cdot \frac{1}{s_X}$$

Where x_{scaled} *is the scaled x value*
x is the original x value
s_X *is the standard deviation*

When the x-variables are measured in different units or at different levels, e.g., mg and kg, this effect may be dominating giving a model that is an expression of this instead of the underlying chemistry. In such situations a clear advantage of scaling **X** is encountered. When the X-data consists of spectra, the choice would appear no longer to be obvious, as the measuring units are here identical. The absorption at the different wavelengths may vary between the different samples, thus justifying a scaling. In the present work the effect of scaling has therefore been investigated individually for each model and applied whenever it actually contributed to decreasing the complexity of the model.

4.5.2 MSC and Derivatives

In spectra flat baselines, as well as baselines with slopes ≠ 0, may be encountered, see Figure 4.7.

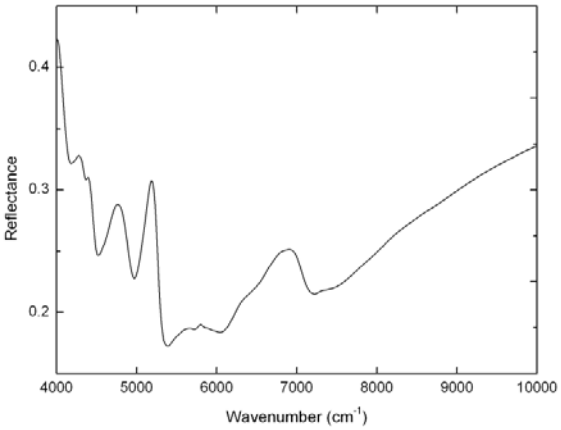

Figure 4.7: Raw NIR spectrum with a baseline slope≠ 0.

These effects are often not a result of the underlying chemistry. An example would be light scattering caused by interaction of the near infrared light with the particle size. Often such a light scattering effect has no correlation with the response variable in question, and thus removing it prior to calibration yields simpler multivariate models. A sample vector may be split into a sum of the different effects:

$$r = r_{sig} + \alpha + \beta x$$

Where r_{sig} *is the signal of interest*
α *is the baseline offset*
β *is the linear slope of the baseline*

By applying *Multiplicative Scatter Correction* (MSC) both effects may be removed individually or together according to:

Common offsett $r_{new\_ik} = r_{ik} - \alpha_i$

Common amplification $r_{new\_ik} = \frac{r_{ik}}{\beta_i}$

Full MSC $r_{new\_ik} = \frac{r_{ik} - \alpha_i}{\beta_i}$

The offset and slope is termed the common offset and common amplification respectively, and are obtained by fitting a regression line in a plot of the individual spectral values against the average spectral values.

MSC was originally suggested as a technique to solve the light-scattering effect seen in reflectance spectroscopy[42] but it has also found more general use and is currently rather termed multiplicative signal correction.

Offset and/or slope may also be removed by applying 1st or 2nd derivatives (Savitzky-Golay). By applying these each point in the spectrum is replaced by the 1st or 2nd derivative of a polynomial approximation in that point, yielding:

$$r'_{ik} = r'_{sig\_ik} + \beta \qquad \text{and} \qquad r''_{ik} = r''_{sig\_ik}$$

From this it is evident that applying a 1st derivative will only remove the offset, whereas the 2nd derivative removes both the offset and slope.

Besides choosing the derivative order, the polynomial order and window size also have to be specified. In the present work generally a 2nd order polynomial and a window size of 3 – 7 have been applied; where appropriate the results of derivation have always been investigated by visual inspection.

The result of applying MSC and 1st and 2nd order derivative (Savitzky-Golay) on NIR spectra from Figure 4.7 is seen in Figure 4.8. As seen the derivatives changes the overall spectrum appearance dramatically. The first derivative may be interpreted as the slope of the spectrum at each wavelength, whereas no easy interpretation of the second derivative exists. The effects of alternative pre-processing's can in some situations be gauged from general experiences, while in (many) others it is entirely related to the specifics of each data set and its alternative modelling. Thus, it is often necessary to carry out a more-or-less extensive experimental pre-processing pilot study in order to find the optimal choices.

With the "mechanics" of multivariate calibration mastered, the remaining prime scientific issue often relates directly to this pre-processing theme. There are nearly always many alternatives available – the number usually greatly augmented by pre-processing being sensitive to the presence/absence of outliers. Often alternative pre-processing results are not very different from one-another[8,30,38].

In the present work the different preprocessing techniques have been investigated and the one giving the maximum reduction in model complexity and RMSEP chosen.

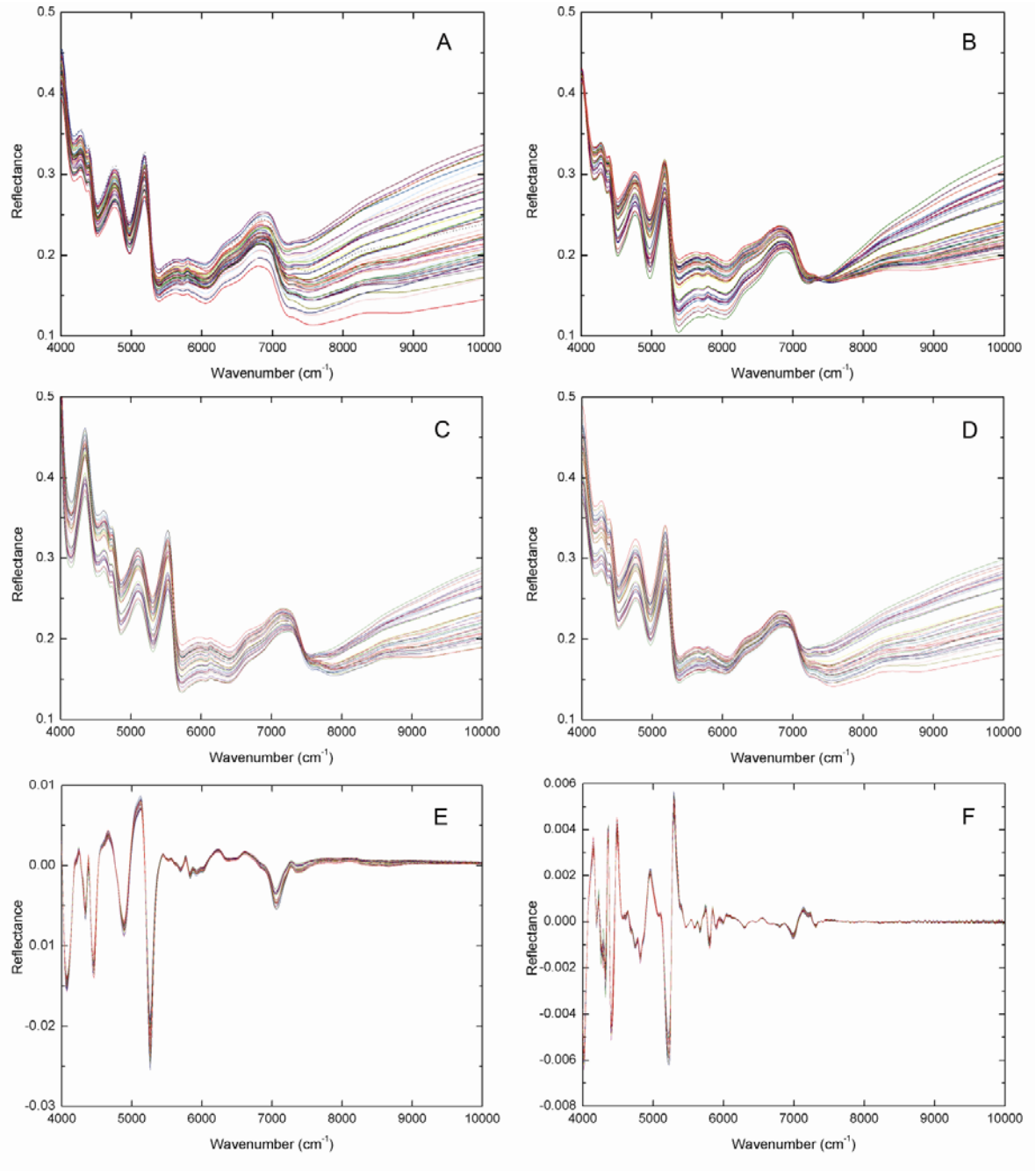

Figure 4.8: NIR spectra: a)Raw, b) MSC, c) MSC alfa, d) MSC beta, e) 1st derivative (S. Golay), and f) 2nd derivative (S. Golay).

5 Anaerobic Digestion

Anaerobic digestion is a complex microbiological process where organic carbon is converted to mainly carbon dioxide and methane under anaerobic conditions in a complex multi-step process. Besides carbon dioxide and methane small traces of hydrogen sulphide and water are also always present in the gas produced. Anaerobic digestion is the formative basis for all biogas production.

The steps involved in the digestion are: hydrolysis, acidogenesis (fermentation), acetogenesis, and methanogenesis. See Figure 5.1.

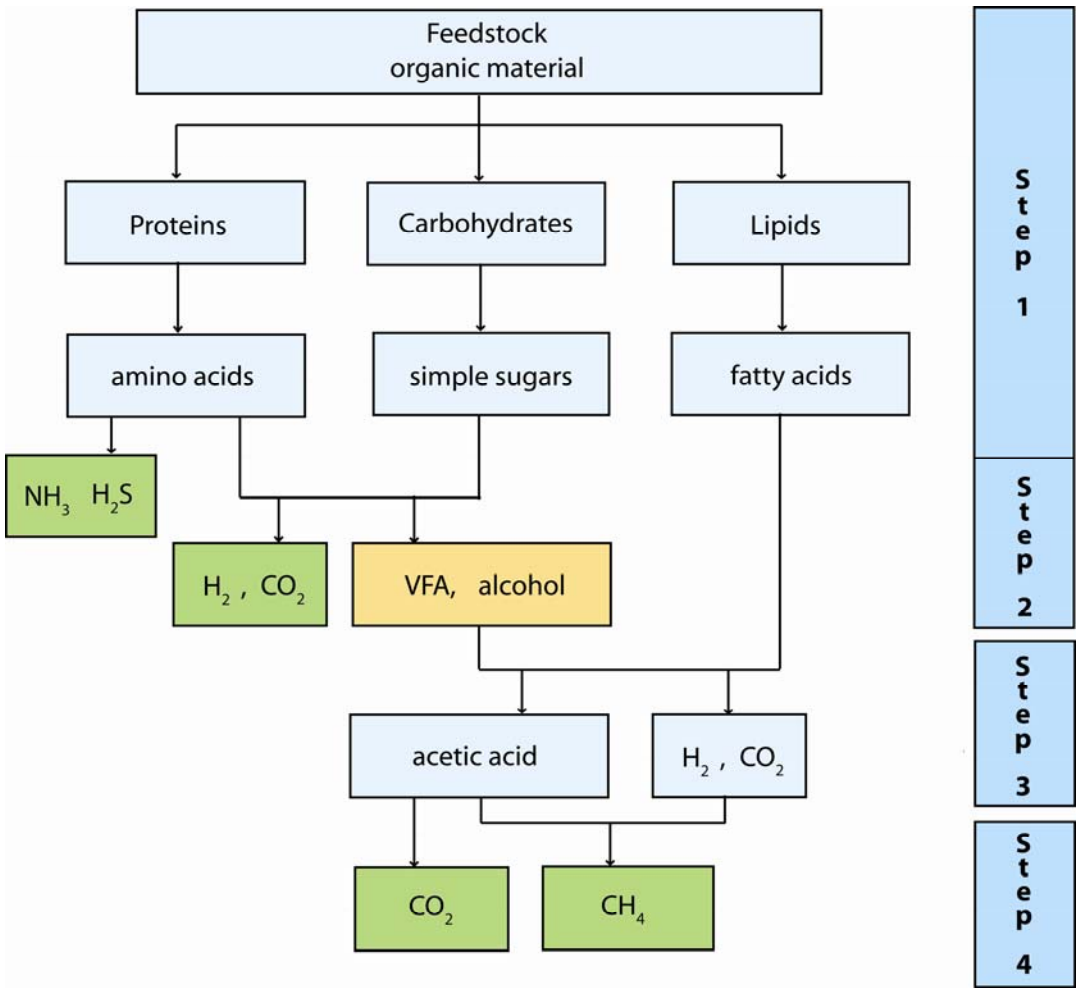

Figure 5.1: Anaerobic digestion of organic material to biomass. Step 1: hydrolysis, Step 2: acidogenesis, Step 3: acetogenesis, and Step 4: methanogenesis (adapted from reference 2008). VFA: volatile fatty acids.

The different steps utilize different bacterial cultures; consortia, between which a balance should be obtained to ensure a satisfactory environment in the biogas reactor and an acceptable biogas yield. Many process instabilities are caused by a failure to maintain the

balance between the consortia carrying out the acetogenesis and the methanogenesis[20], as these two consortia differs in their nutritional needs, growth kinetics and sensitivity to environmental conditions.

In the hydrolysis step, high-molecular compounds, e.g., proteins, fat, and polysaccharides are digested to lower molecular compounds; amino acids, fatty acids, and mono-, di-, tri-, and oligosaccharides. The degradation is carried out extracellularly typically by excretion of enzymes; lipases, cellulases, and amylases from hydrolytic bacteria like *Bacteroides*, *Clostridium*, *Acetivibrio*, and *Fibrobacter*. The hydrolysis involves several steps; enzyme production, diffusion, adsorption, reaction, and enzyme deactivation. This step has been identified as the rate-limiting step in processes applying high particulate substrates. For small chain components however, the acidogenesis is carried out directly[28,71,96].

In acidogenesis the sugar monomers from the hydrolytic step or already present in the substrate are converted to puryvate ($C_3H_4O_3$), ATP, and NADH (electron carrier molecule) via the glycolysis or pentose phosphate pathway. The pyruvate and amino acids are subsequently converted to variety of short chain fatty acids (acetic, propionic, and butyric acids), alcohols, hydrogen, and carbon dioxide through various fermentation pathways, see Figure 5.2. The pathway chosen is different for the different microorganisms and some microorganisms even have branched pathways. In the latter case the pathway chosen is the one most thermodynamically favourable under the given conditions; substrate concentration, pH and dissolved hydrogen in the bioreactor. Acidogenesis is performed by many of the micro organisms also responsible for the hydrolysis; *Bacteroides* and *Clostridium*. Besides organisms like *Lactobacillus* and Anaerolineae carry out acidogenesis[28,96].

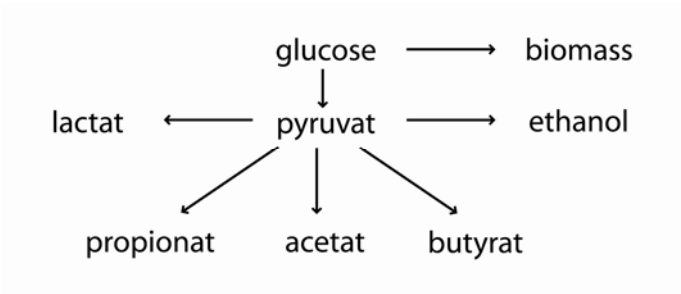

Figure 5.2: Fermentation products from glucose during anaerobic digestion (adapted from refrence 90)

The CO_2, H_2 and other one carbon compounds obtained via acidogenesis are turned into methane directly by methanogenic bacteria. Longer chain fatty acids (C>2), alcohols (C>1) and branched chain and aromatic fatty acids however are oxidized to acetate and hydrogen in acetogenesis. See Table 5.1 for volatile fatty acid oxidation reactions. Acetogenesis is carried out by obligate proton reducing bacteria such as homoacetogenes living in syntrophy with hydrogen utilising methanogenic bacteria. The syntrophic relation is necessary as all oxidations

are thermodynamically unfavorable at standard conditions. Having a low hydrogen pressure (<10^{-4} atm), shifts the equilibrium of the reaction towards product formation making the reaction feasible. Different organisms are active during acetogenesis, e.g., different species from the genus *Syntrophomonas and Pelatomaculum*[28].

Table 5.1: Volatile fatty acid degradations.(adapted from reference 96)

Substrate	Reactions
Propanoic acid	$CH_3CH_2COOH + 2H_2O \rightarrow CH_3COOH + 3H_2 + CO_2$
Butanoic acid	$CH_3CH_2CH_2COOH + 2H_2O \rightarrow 2CH_3COOH + 2H_2$
Valeric acid	$CH_3CH_2CH_2CH_2COOH + 2H_2O \rightarrow 2CH_3COOH + 2H_2$

The pathway utilised for oxidation is different for the different fatty acids. In example propinate is mainly oxidised via the methyl-malonyl-CoA pathway, butyrate via the β-oxidation pathway, and iso-butyrate is isomerised to n-butyrate and oxidized in the β-oxidation pathway[3,26]. Of the volatile fatty acids oxidation of propionate has the lowest rate[83].

In the final step of the anaerobic digestion (methanogenesis) acetat, formate and hydrogen is converted to methane and carbon dioxide by methanogenic archaea, which are specialised in degrading these substrates. Different methanogenes exist, some of which are able to utilise several substrates whereas others are able to utilise a single substrate only[108]. Two dominating routes for methane production exist. Through the first route carbon dioxide is reduced to methane by applying hydrogen as an electron donor. This reaction is carried out by lithotrophic hydrogen oxidizing methanogens like *Methanobrevibacter, Methanobacterium, Methanogenium, and Methanospirillium*. The second route is a fermentation of acetic acid to methane and carbon dioxide. This fermentation is carried out by organotrophic acetoclastic methanogens like *Methanosaeta* and *Methanosarcina*. The two species are favoured under low and high acetate concentrations respectively.

$$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$$

$$CH_3COOH \rightarrow CH_4 + CO_2$$

Approximately two-thirds of the methane produced originates from the fermentation of acetic acid[28,39].

5.1 Feedstock

The anaerobic digestion is carried out applying a wide variety of different feedstocks. Theoretically the process can be run applying only farmyard manure. The dry matter content in cattle and pig manure, however, is only around 2-6 %, which is mainly plant fibres that are not

easy degradable, and as a consequence a low biogas yield is obtained. In practice the anaerobic digestions are, therefore, run as co-digestion of manure and other organic materials like industrial waste, energy crops, food waste, etc.[69,126,127]. In Denmark the legislative demand is that manure should constitute 75 % of the feedstock applied.

The overall microbial demand for the feedstock is different macro and micro nutrients. The macro nutrients constitute carbon, hydrogen, nitrogen, oxygen, sulphur, phosphorus, calcium, potassium, magnesium, and iron. The concentration of these present in the cell should be around 10^{-4} M. The micronutrients are nickel, cobalt, and copper that are required in smaller amounts (< 10^{-4} M). It is important for the bioreactor to be in balance with respect to all nutrients, as all of them can become inhibitory if present in too high concentrations. In practice it is often the ratio between carbon and nitrogen (C/N) that is considered. A too high C/N ratio would increase the danger of nitrogen depletion, whereas a too low ratio would ultimately lead to ammonium inhibition. A C/N ratio around 25 is considered optimal[4,71]. In practice the favourable C/N ratio is obtained by mixing feedstock with different C/N ratios. Examples of feedstocks having a low C/N ratio are pig and poultry manure, whereas energy crops like silage and straw have a high C/N ratio.

5.2 Monitoring

A stable biogas production is based on maintaining the fragile balance between the different consortia making up the anaerobic digestion. To enable this process, changes and instabilities should be observed and reported as quickly as possible. In general a fermentation process may be monitored by measuring substrate conversion (COD, VS removal), intermediates accumulation (VFA, pH, alkalinity) product formation (gas composition, production rate), and microbial communities (activities, populations) applying a variety of different methods. The anaerobic digestion being a complex process has made it difficult to find simple and reliable control parameters.

Some of the most commonly used process parameters for the anaerobic digestion are gas production, gas composition, pH, and volatile solids reduction. In general these are adequate for detecting gradual changes but only the gas production is applicable for detecting sudden changes[111]. Sudden changes caused by for example a different feedstock, pH or temperature may induce critical changes, which should be spotted as quickly as possible to prevent process instabilities or failure. In general intermediates like volatile fatty acids are being acknowledged as important indicators of process instabilities, and has therefore been suggested as a monitoring parameter[1,10,12].

For monitoring of volatile fatty acids (VFA) on-line continuous determination by GC[4,27] and IR [109] have been suggested. Another method; near infrared spectroscopy has been found applicable to monitor complex mixtures - one of which is fermentation broth from anaerobic digestion

processes[107] and monitoring of the anaerobic digestion of a mixture of cellulose, albumin, and minerals[76]. Furthermore, earlier off-line NIR experiments have shown good results for monitoring of several parameters amongst others VFA[48].

In the current thesis it was therefore decided to investigate the feasibility of applying NIR for at-line and on-line determination of the VFA concentration when applying different substrates for the co-digestion.

Additionally, it was decided to test the feasibility of applying acoustic chemometrics (a.c.) for monitoring the total solids in the complex matrix applied for biogas production.

5.3 Experiment

For testing the feasibility of NIR and a.c. two anaerobic digestion trials were carried out applying reactors with an active volume of 5 L. Trial 1 was carried out as a co-digestion of glycerol spiked manure and food waste, and trial 2 as a co-digestion of maize silage and spiked manure and food waste. Before carrying out the experiments the sampling procedure and handling was outlined see Figure 5.3.

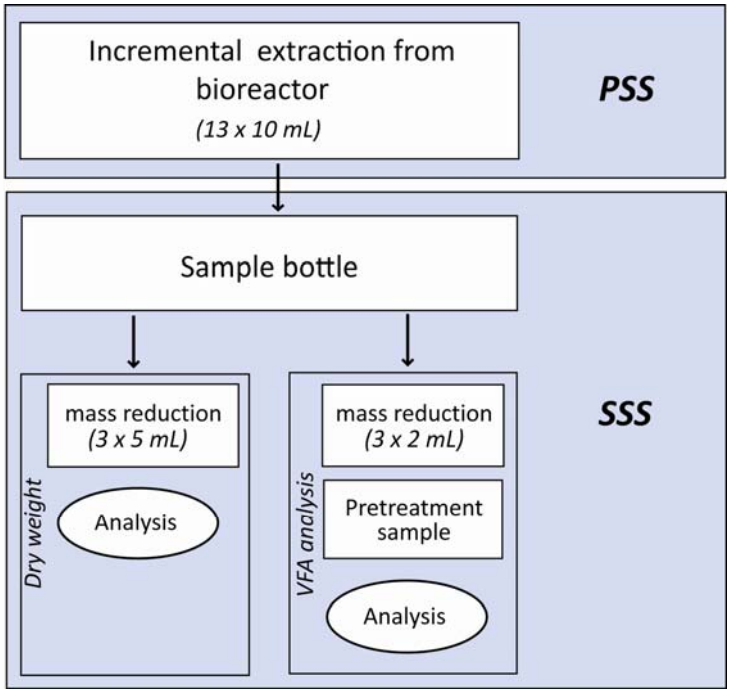

Figure 5.3: Sampling procedure and handling during anaerobic digestion trials.

During the fermentation monitoring trial a 100 mL/130 mL sample (primary sample) was collected from the bioreactor for reference analysis and simultaneously at/on-line measurements were carried out applying NIR and a.c. This resulted in two potential sampling error contributions; sample extraction and spectral recording. In the system, the heterogeneity

was being caused by a density difference in the substrate, thus demanding extraction of a total top to bottom column, for the sample to be representative. This was not easily realisable, so it was chosen to transform the sampling situation to 1-dimension, SUO no. 7.

The 1-dimensional sampling situation was obtained for both extraction and spectral recordings by implementation of the TENIRS system. In earlier studies the TENIRS system had been applied successfully, for off-line NIR analysis of samples extracted from a bioreactor[48]. In short the system works by pumping bioslurry from a 1L bottle, mounted onto the system into a horizontal loop, through a macerator, through a flow-through measuring cell for spectral recordings and back into the bottle. A more through description of the system is given in reference 2.

For the current purpose the TENIRS system was rebuilt and connected directly to the bioreactor (at an intermediate height), pumping the bioslurry from the reactor into the loop, thereby enabling on-line spectral recordings. Samples were obtained by a sampling aggregate implemented right after the flow through cell, thereby ensuring that samples measured by the TENIRS and the volume withdrawn for reference analysis were pair-wise comparable. As a consequence multivariate calibration with an absolute minimum of sample-spectrum mismatch was obtained, although not eliminating IDE mismatch completely. The system had a length 230 cm in total and a diameter of 2 cm. The system is seen in Figure 5.4.

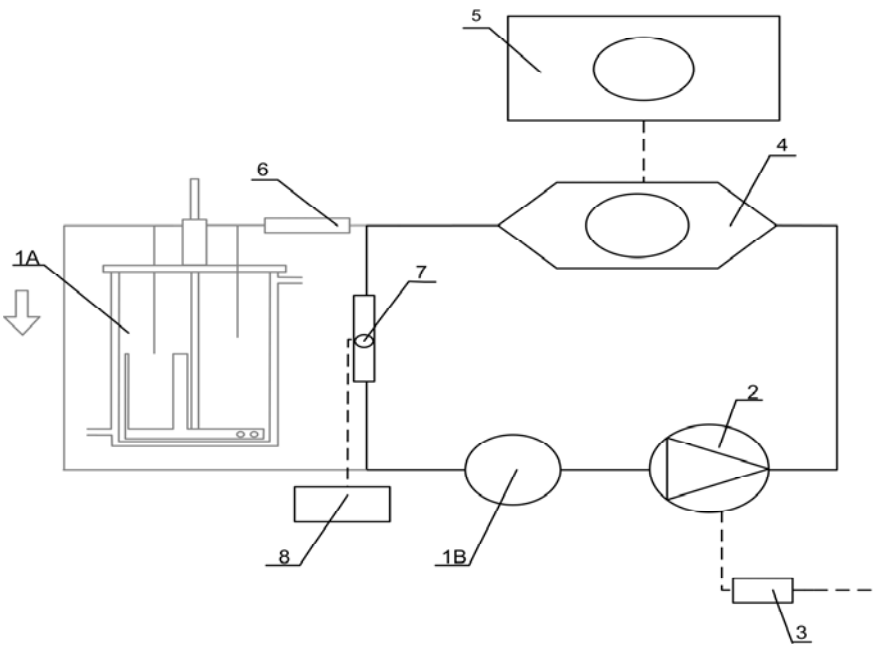

Figure 5.4: Experimental set-up during anaerobic digestion trials. 1A and 1B represents the two different operation modes. 1A represents the on-line fermentation, the grey line representing the connection of the loop to the fermentor, and 1B is applied in the at-line trial; (2) Impeller pump with motor; (3) Frequency inverter; (4) Cell for NIR measurement; (5) Zeiss Corona NIR instrument; (6) Sampling device; (7) constriction for a.c. attachment; (8) Signal amplifying module for a.c.

To avoid flow segregation, and thus obtaining samples in total compliance with TOS the sample extraction and spectral recordings should have been implemented in an upwardly flowing stream[32,49]. In these first studies in the system this was not obtainable, and a horizontally placed loop had to be accepted. The consequence was sample extraction from the bottom of a horizontally flowing pipe, introducing a certain risk of an overrepresentation of the heavy particle in the sample i.e., an IEE. As an attempt to compensate composite sampling, applying the 10 mL bottle repeatedly as several increments, was effectuated.

Recently this particular sampling device has been further evaluated by *Boland (2008)*[13], who showed that highly significant sampling bias will develop if circulation speed and solid loads are unbalanced. The characterisation was carried out applying different frequencies (20, 30 Hz) and solid loadings (2.5, 5.0, and 7.5) and overall conclusion was that the reproducibility of the sampling device was very good, close to zero, but the accuracy was very much frequency and solid load dependent, thus leading to non representativeness. With respect to representativeness Boland concludes:

> *"Once again the TENIRS valve is precisely wrong and the composite samples only lead to slightly better results than the grab samples."*

In the current anaerobic digestion trials, the pump frequency applied was 40 Hz, meaning that the accuracy was probably better than the ones reported by *Boland 2008*[13]. This of course does not change the fact that the samples were not representative and that the loop should be changed before carrying out any more experiments.

The primary samples were thoroughly mixed before being subdivided into secondary samples having volumes suitable for reference analysis in the laboratory. Subsampling was achieved by vigorous shaking followed by fast pouring the amount needed for reference analysis, into a beaker. This technique was employed based on slurry sampling development work by *Holm-Nielsen et al. 2006*[49], who demonstrated that this particular procedure produced minimum contributions to the overall sampling error.

From these samples the VFA, total and volatile solids (trial 2 only) and glycerol (trial 1 only) concentrations were determined and used as y parameters in a PLSR model. The models obtained for the VFA in the two experiments are seen in Table 5.2, whereas the NIR models for glycerol and the NIR and a.c. models for total and volatile solids may be seen in paper 1 and 2.

Table 5.2: VFA models obtained in trial 1(upper) and 2(lower).

	Components	Slope	r^2	RMSEP	RPD
Acetic acid	3	1.1	0.95	1645	4.3
Propionic	1	0.37	0.54	1364	2.2
i-butanoic	6	0.97	0.97	26.46	6.2
Butanoic	5	0.51	0.55	271.9	1.5
i-valeric	6	0.88	0.94	55.82	3.9
Valeric	3	0.66	0.86	24.31	2.4
Total VFA	3	1.0	0.95	2333	3.7
Acetic acid	3	0.89	0.89	913	3.1
Propionic	3	0.89	0.91	206	3.4
i-butanoic	3	0.83	0.89	43.8	2.8
Butanoic	3	0.87	0.92	400	3.6
i-valeric	3	0.86	0.89	76.4	3.1
Valeric	3	0.92	0.92	43.2	3.5
Total VFA	3	0.87	0.90	1594	3.3

Contemplating the results from the two trials it is seen that trial 2 yield models with comparable prediction statistics for the different compounds. From this it was concluded that the NIR could be applied for the monitoring of all contributors to the total VFA.

The trend seen from trial 1 is not exactly the same. The models obtained for acetic, i-butanoic, i-valeric, and total volatile fatty acid also had good prediction statistics and RPD values. In contrast the models obtained for propanoic, butanoic, and valeric acid had inferior prediction statistics. From the knowledge that these could in fact be modelled in trial 2 and that their corresponding iso-compound could be modelled this seemed a bit strange. Being able to model the compounds in one trial but not the other could be explained by a matrix effect, as the two trials were carried out applying different substrates. However, this does not explain the success obtained for the iso-compound. Therefore, the three models were contemplated. For valeric acid the problem was not the precision of the model but the accuracy (slope), this was probably caused by the fact that the validation set didn't span the range of the calibration set – for undisclosed reasons.

For propanoic and butanoic acid the problems exist within the calibration-validation relation Prior to modelling it was decided that fermentation 1 and 3 should be applied as calibration

sets and fermentation 2 as a test-set for the models, which turned out not to be optimal for these two components. For butanoic acid the validation range was situated in the lower 20% of the calibration range, which might have caused the results. For propanoic acid a wider concentration span in the validation set than in the calibration set was obtained, and as a consequence the model was validated outside the calibration range. In an attempt to correct this, samples outside the calibration range were removed, which however did not improve the model performance, indicating that fermentation 2 was significantly different from fermentation 1 and 3.

5.4 Conclusion

The overall conclusion from the two trials was that the a.c. model for total and volatile solids had fair prediction abilities, but a low RPD. Therefore, further experiments should be carried out, preferably at industrial scale to test the real feasibility of a.c. for TS and VS monitoring. It is likely that a.c. sensor deployment also may have contributed to the relatively discouraging results (Esbensen, pers. com.).

The NIR models obtained from the two trials had good prediction abilities for total and volatile solids and glycerol. For VFA fair to good prediction abilities were established.

The VFA models obtained have been validated applying VFA concentrations that will hopefully never be encountered in a sound biogas process context. This is not necessarily a serious problem in practice, but it does pose a risk of obtaining more-or-less relevant calibration models, where samples with high concentrations have too high influence on the model. The specific VFA operating level in biogas plants is different for different locations, as it is very much substrate dependent or plant specific, and as a consequence some reactors have low VFA level < 1 g/L, whereas others are above 5 g/L. The reason that some reactors are able to run at high levels whereas others are not, is that the bacteria cultures have been accustomed to high VFA levels over a period of time. They have however not been accustomed to such a degree, that the methanogenic step can degrade the fatty acids formed in the acidogenic and acetogenic step, thus indicating that the reactors are running under suboptimal conditions. For the models to be implementable in an industrial biogas monitoring application, they have to be validated in the lower concentration range where the majority of plants are operating; 0.1 – 5 g/L[97].

6 2nd Generation Bioethanol

For production of 2nd generation bioethanol a wide variety of lignocellulosic materials such as straws, corn stover, bagasse and forest residues have been reported[40,94,128].

In Denmark 25.6 % of the total agricultural area is used for wheat production, resulting in substantial wheat straw production, making it the natural focus of attention in Danish 2nd generation bioethanol research.

6.1 Wheat Straw

Lignocellulosic materials, i.e., wheat straw, consist of three main components; cellulose, hemicelluloses, and lignin, with the average composition being 30-40 % cellulose, 20-50% hemicelluloses, and 10-20% lignin[75,93,114,115].

Cellulose is a linear molecule consisting of up to 15,000 D-glucose sub-units linked by β-1,4 glycosid bonds. Every second glucose unit is rotated 180° relative to the neighbour molecule. The repeating unit is cellobiose. See Figure 6.1.

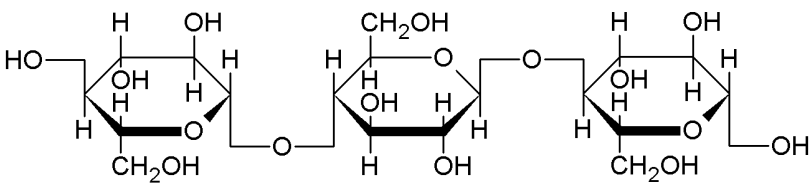

Figure 6.1: Cellulose chain consisting of D-glucose units.

The long cellulose chains link to each other by hydrogen bonds and Van der Waals forces forming micro fibrils containing highly ordered crystalline regions combined with less ordered amorphous regions. As a consequence cellulose is a crystalline, strong molecule, which is not easily hydrolysable.

Hemicelluloses are heterogeneous polysaccharides consisting of different pentoses, hexoses and sugar acids, the specific composition being raw material dependent. The main pentoses in wheat straw are xylose and arabinose. The carbohydrates are linked together mainly by β-1,4 glycosid bonds, but β-1,3 glycosid bonds are seen as well. To these chains, branches of short chains (di or trisaccharides, acetyl groups) are linked making it non-crystalline and as a consequence more easily hydrolysable. In plant cells the hemicelluloses form a protective coat around the cellulose fibrils and acts as the connecting link between cellulose and lignin.

Lignin's are present in the plant cell wall. It is an amorphous complex molecule consisting of aromatic polymers of phenylpropane molecules linked together in a three dimensional structure. In wheat straw the primary building blocks are aromatic alcohols; p-coumaryl alcohol, coniferyl alcohol, sinapyl alcohol[15,78,101]. See Figure 6.2. The proportion between these in wheat straw is 5, 49, and 46% respectively[15].

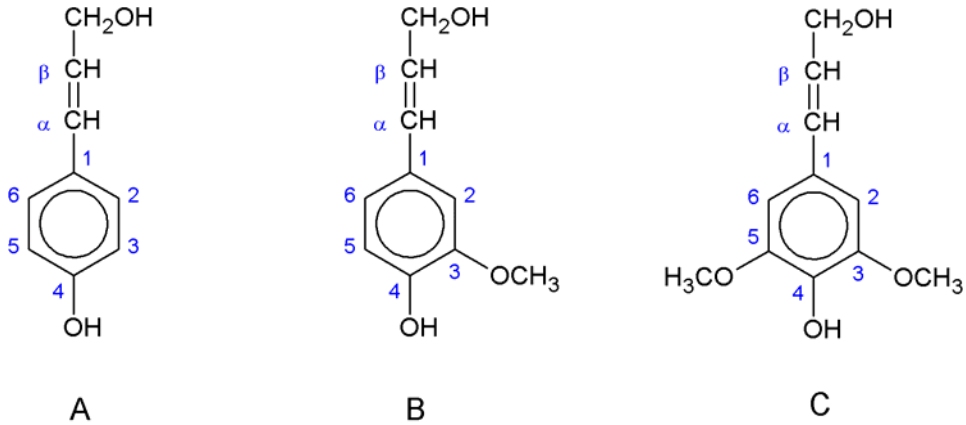

Figure 6.2: Primary building blocks in lignin. A) p-coumaryl alcohol, B) coniferyl alcohol, C) sinapyl alcohol (adapted from reference 15).

During lignifications these building blocks are linked together in a three dimensional structure through a variety of different linkages, α-O-4, β-O-4, β-5, β-1, 5-5, 4-O-5, and β-β. Furthermore the lignin structure is linked to the hemicelluloses structure via covalent bonds from the α-carbon and C-4 in the benzene ring.

Lignins presence in the cell wall provides rigidity, internal transport of water, and nutrients and protection against microbial degradation. It is extremely resistant to both chemical and enzymatic hydrolysis[15,78,101].

6.2 Ethanol Production

A general outline of the bioethanol production is given in Figure 6.3.

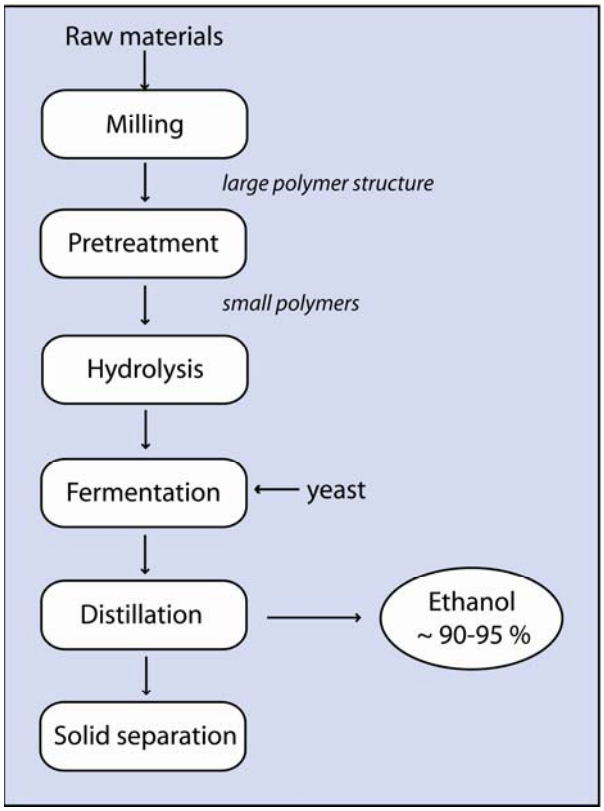

Figure 6.3: Outline of 2nd generation bioethanol production.(adapted from reference 75)

6.2.1 Pretreatment

The objective of having a pretreatment process is to speed up the following hydrolysis by opening up the biomass structure making the material more accessible to hydrolysis. A number of different pretreatments processes have been developed in laboratories; physical, physiochemical, and chemical. Examples of the different techniques are seen in Table 6.1.

Table 6.1:Different pretreatment techniques applied for lignocellulosic material.

Pretreatment methods	
Physical:	mechanical comminution, pyrolysis and irradiation
Physiochemical:	steam explosion, ammonia fiber explosion (AFEX), SO_2 explosion and CO_2 explosion
Chemical:	Dilute acid hydrolysis, alkaline hydrolysis, organosolvent process and oxidative delignification
Biological:	Fungal pretreatment

More information regarding the different techniques is given in reference 28, 94, 110, 114, & 125. In a given process one or several of the pretreatments methods may be applied. The choice of pre-treatment method is raw material dependent.

6.2.2 Hydrolysis

The hydrolysis step may be carried out either as a chemical hydrolysis or an enzymatic hydrolysis. During chemical hydrolysis the lignocellulosic material is subjected to acid at a specific temperature and for a period of time. As a result cellulose and hemicelluloses are degraded into sugar monomers. The enzymatic hydrolysis is carried out applying highly specific cellulose and hemicelluloses enzymes, the group containing at least 15 protein families. Different enzymes; Endo-endo, exo-exo, and exo-endo are applied simultaneously, obtaining a synergetic effect that reduces the time needed for the hydrolysis. As a result of the enzyme hydrolysis the hemicelluloses and cellulose is degraded to sugar monomers (glucose, fructose, arabinose ect.)[90,112,113].

6.2.3 Fermentation

From the hydrolysis step both hexoses and pentoses are obtained, which should both be converted to bioethanol in a fermentation step. For the hexoses the conversion may be carried out under anerobic conditions applying a yeast; *Saccharomyces cerevisiae*.

$$C_6H_{12}0_6 \rightarrow 2C_2H_5OH + CO_2$$

This process has a theoretical yield of 0.51 g ethanol/ g sugar, which in practice never is obtained, as energy is used for biomass and by product (glycerol, carboxylic acids) formation as well. The pentoses are not converted by *S. cerevisiae*, so to improve the yield they should either be transformed to a component that may be metabolized by yeast or another organism should be used for fermentation. The major pentose constituent xylose may be isomerised into xylulose, which is converted by *S. cerevisiae* with the same theoretical yield as hexoses.

$$3C_5H_{10}0_5 \rightarrow 5C_2H_5OH + 5CO_2$$

The metabolism of xylulose is a slow process making it less attractive for industrial application. As a consequence research is being carried out to identify or develop an organism able to metabolise other sugars in addition to glucose. This should be done with a high rate and yield[28,114]. In some processes the enzymatic hydrolysis and fermentation step has been integrated into a process termed simultaneous saccharification and fermentation (SSF).

The broth from the fermentation step contains ethanol (2-12 %), biomass, fusel oil, volatile components, and stillage. Typically ethanol is removed by distillation. Before doing so the biomass is allowed to settle or removed by centrifugation.

At each process step a number of technologies have been investigated, and the choice of technology for a given production is, amongst others, raw material dependent. For production from wheat straw a production termed the Integrated Biomass Utilization System (IBUS) is close to commercial reality. IBUS consists of the following process steps; hydrothermal pretreatment, prehydrolysis (enzymatic liquefaction), SSF, and distillation. By applying this process 1 ton of wheat straw (dm 86%) 143 kg bioethanol, 353 kg solid biofuel, and 420 kg animal feed is obtained[63].

6.3 Monitoring

The focus of attention in the development of the 2nd generation ethanol from wheat straw has been primarily on the conditions for the different steps, thus obtaining knowledge about the optimal parameter settings for this specific feedstock[115,116]. These optimal settings are estimated by applying material with a certain composition, and the process is subsequently being run without taking into account the different feedstock compositions or the effect of the process prior to the one being carried out. The feedstock composition varies as a function of harvest time, year, and storage condition and is by no means stationary. As a consequence, the parameters are at best estimated from an average composition of the feedstock and subsequently applied in the production.

The consequence of applying pre-fixed parameter settings may be a suboptimal utilisation of the feedstock and of the utilisation of each process step, e.g., the addition of a too low or high enzyme dose to the enzymatic hydrolysis, both making the process less economical feasible. In the prior case due to a too low hydrolysis of the feedstock leading to a reduced utilisation of the feedstock, and the latter due to the extra enzyme cost. As bioethanol is a low value product, it is important not to apply such suboptimal conditions, as it may contribute to making the process non-feasible.

As a solution to this process analytical technologies may be applied for process monitoring throughout the process. In the bioethanol production knowledge of the sugar concentration through the different process steps and also the presence of inhibitors before carrying out the

fermentation would be of great value. In this thesis it was chosen to investigate the feasibility of applying NIR for such process monitoring starting at the beginning of the production line with the quantification of sugar components in the feedstock. NIR was chosen as it had successfully been applied for quantification of sugar concentration in others lignocellulosic feedstocks, see Table 6.2.

Table 6.2: Different application fields of NIR on various biomasses reported in the literature. DTG: Derivative thermogravimetric analysis, ADF: Acid detergent fibre, NDF natural detergent fibre.

Raw material	Application	Components
Rice straw[53]	Quantitative determination	cellulose, hemicellulose, lignin, moisture, total ash, and acid insoluble ash
Ground barley[105]	Analysis for fuel ethanol production	moisture, starch, β-glucan, protein, oil and ash
Foliar[88]	Alternative to wet lab analysis	mineral nutrients, carbon, fibre constituents (ADF, NDF, cellulose and lignin)
Corn stover[47]	Alternative to wet lab analysis	glucan, xylan, arabinan, lignin, protein, acetyl
Flax Fibres[35]	Assessing physical and chemical characteristics	AIRflow, DTG, fibre strength, NDF, ADF and hemicelluloses
Forages and byproducts[87]	Quantitative determination	fibres and protein, nitrobenzene oxidation products of lignin, six measures for lignin content
Silage[86]	Analysis of different components	dry matter, crude protein, different acids, pH

Ultimately what is of most interest to the bioethanol industry would be instantaneous compositional characterisation directly on the feedstock (straw bales and similar) for example by applying a handheld NIR spectrophotometer or by a suitable “at-line” facility in the field. Measuring directly at the bale arises a question regarding representativeness of the spectra being measured and the material collected for reference analysis. Unless the situation is handled in compliance with TOS a risk of obtaining non-corresponding samples is high. Once incorporated such a mismatch would follow the samples through the subsequent sample handling stages. Furthermore, in each stage a contribution from the stage itself, will be incorporated: this only increases the overall total sampling error (TSE). The TSE will subsequently be built into the analysis result and, hence, finally the modelling stage. As a consequence this could ultimately lead to false conclusions regarding the feasibility of the NIR

analysis itself. The question of obtaining such a primary sample is of course by no means irrelevant; ultimately it is a critical success factor for the future implementation of NIR in the process and should be dealt with.

In the current context, however, it was decided to focus exclusively on the feasibility of NIR for quantification of cellulose (glucan), hemicelluloses (xylan and arabinan) and lignin in wheat straw starting the sampling delineation from the individual primary bags collected. It is not meaningful to investigate the full field sampling variability before it is known with certainty that the subsequent analytical procedure in fact can deliver the results it is supposed to. Consequently a matching experimental design was delineated.

6.4 Experimental Design

For the NIR model on wheat straw a total of 95 samples were collected. Sixty-five of these were collected from fields and barns and had an overall straw size from a few cm to approximately 25 to 30 cm. These samples were called **field** samples. The remaining 30 samples were obtained from Køge Bio pellets factory (Denmark). They were sampled from the bio pellets production line over a 30 day period, and had a cut straw size of 0.5 to 2 cm. These were called **industrial** samples. The complex sampling procedure and handling is outlined in Figure 6.4.

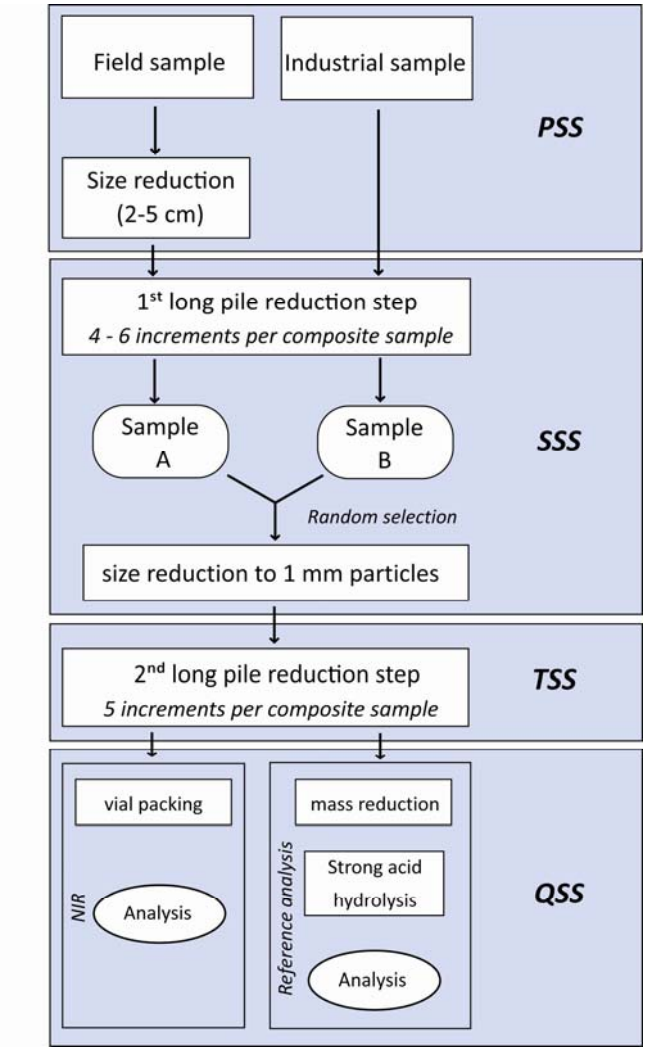

Figure 6.4: Sampling procedure and handling during wheat straw trial. PPS: Primary sampling step, SSS: Secondary sampling step, TSS: Tertiary sampling step, QSS: quaternary sampling step

The second sampling step applied the long pile method. Due to the physical properties of the material (static electricity) special care had to be taken to avoid loss of material (Incorrect Preparation Error, IPE). Therefore, a right-angle gutter with a length of 1 m was fabricated, Figure 6.5, and applied for splitting the sample in two approximately even sizes; an A and a B sample respectively. For further analysis sample A or B was chosen at random.

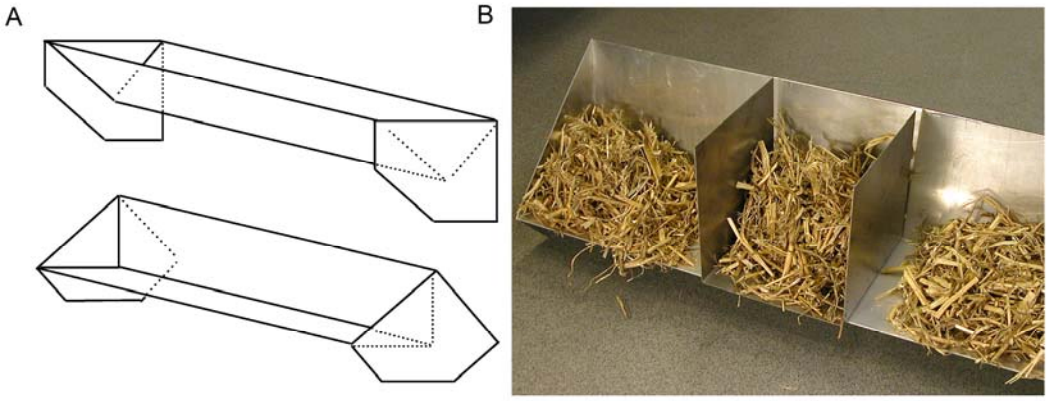

Figure 6.5: Right-angle gutter fabricated for mass reduction of primary wheat straw samples. B illustrates a specific long-pile cross-cut being taken, emphasizing attention to eliminate IDE.

The resulting sub-sample was now further size reduced to a particle size of 1 mm, see Figure 6.6.

Figure 6.6: Wheat straw size after size reduction in first and second mass reduction (and comminution) steps.

In the tertiary sampling step the material for NIR and strong acid hydrolysis was further sub-sampled by applying the long pile method once again extracting approximately 2 x 5 g wheat straw for NIR analysis and 2 g for reference analysis.

In the quaternary sampling step the material was packed in plastic bags and vials. The vials were scanned with NIR spectroscopy and the spectra applied as **X** in the subsequent PLSR. Furthermore 4 x 0.16 g of material was extracted from the plastic bags directly for reference analysis by composite sampling. The 0.16 g was subjected to strong acid hydrolysis and the carbohydrate concentration estimated by HPLC and Klason lignin as the dryweight after strong acid hydrolysis subtracted the weight of the ash. The obtained results were used as **y** parameters in a PLSR. Detailed information regarding the sampling procedure is given in paper 3.

Different pre-processing techniques were investigated for the spectra, and a 2nd order derivative with five smoothing points was found optimal for the carbohydrate, and a full MSC for lignin. The best models obtained are summarised in Table 6.3. The details regarding the pre-processing's are given in paper 3 and the discussion of the models in paper 3 and 4.

Table 6.3: PLS-1 models for sugar components and lignin. "Components": Number of PLS components applied in the model, numbers in brackets are the explained calibration and the residual validation variance, "Outliers": Number of outliers excluded from the set during calibration, numbers in brackets are the corresponding percentage, "RMSEP": Root Mean Square Error of Prediction

	Components	Outliers	Slope	r^2	RMSEP	% RMSEP
Glucan	5 (80% X, 83% Y)	5 (11%)	0.88	0.83	0.60	11
Xylan	5 (92% X, 83% Y)	1 (2%)	0.90	0.82	0.43	11
Arabinan	4 (97% X, 77% Y)	6 (14%)	0.76	0.77	0.12	13
Lignin	7(100% X, 72% Y)	8 (18%)	0.84	0.72	0.38	12

As seen the models have a relatively constant, intermediate % RMSEP. This may be due a significant material heterogeneity, not handled properly during the sample stages. To investigate this, a heterogeneity estimation of the material should be carried out, but for a material like wheat straw this is not straight forward. Furthermore, it is only possible theoretically to reduce the sampling error to a contribution from the heterogeneity alone, thus from a practical point of view making an empirical estimation of the TSE more interesting.

6.4.1 Estimation of Laboratory TSE

Estimation of TSE for the laboratory procedure was carried out by a replication experiment. This was carried by independently extracting and handling ten samples from a single primary field sample according to the procedure delineated in Figure 6.4. From these replicates TSE was estimated for each component, see Table 6.4. As seen the TSE is generally small and much below the 16 % set by *Pitard 2009*[85] as the maximum allowable error, meaning that the sampling procedure applied was in full control. Incorporation of an error in this range into the multivariate model is acceptable. For arabinan the error was a bit higher, than for glucan and xylan, which may only be explained by a larger contribution from the analytical procedure; expectably the HPLC analysis. As a consequence the models obtained for arabinan would have worse prediction abilities than for glucan and arabinan, which was also found to be the case.

Table 6.4: Mean [g·(100g DM)$^{-1}$], standard deviation [g·(100g DM)$^{-1}$] and coefficient of variation [%] for sample replicate experiment.

	Glucan	Xylan	Arabinan	Lignin
Mean	39.0	23.0	3.4	20
Standard deviation	0.76	0.41	0.16	0.26
Coefficient of variation	1.9	1.8	4.6	1.3

6.5 Conclusion

From the results delineated the feasibility of applying NIR spectroscopy for quantification of the selected sugars in wheat straw is substantiated. The models obtained for all carbohydrates have relative high % RMSEP, which means that they are not yet directly implementable in a professional monitoring context. This is not surprising as a minimum number of samples were used in the pilot study. In order to improve model performances, more samples should be collected, appropriately mass reduced and analyzed. The model obtained from lignin had distinctly less good accuracy and precision. Therefore, further samples need to be incorporated into the data set before any final conclusion can be made on the feasibility of applying NIR spectroscopy for lignin quantification in wheat straw. On the whole the feasibility of NIR characterisation in the context investigated appears sufficiently substantiated.

From the estimation of the total sampling error, it is evident that the procedure for handling this type of sample in the laboratory was in acceptable control.

Upon this basis, the next meaningful task remaining is incorporating the primary field sampling into the replication validation procedure applied above. This may very well relate to quite different orders of sampling errors.

7 Aerobic Bioconversion

A known important aroma constituent in Roquefort cheeses is 2-heptanone and it is, therefore, added to products like salad dressings, soups and crackers to simulate blue cheese flavour.

Currently, industrial production of 2-heptanone is carried out as a complex multiphase fermentation involving an aqueous phase, a hydrophobic phase, and a solid phase to which an enzyme and a filamentous fungi is added. During fermentation 2-heptanone is produced from β-oxidation of the octanoic acid by *Penicillium roqueforti*[73]. Octanoic acid is released from the organic phase by the added enzyme (lipase). See Figure 7.1.

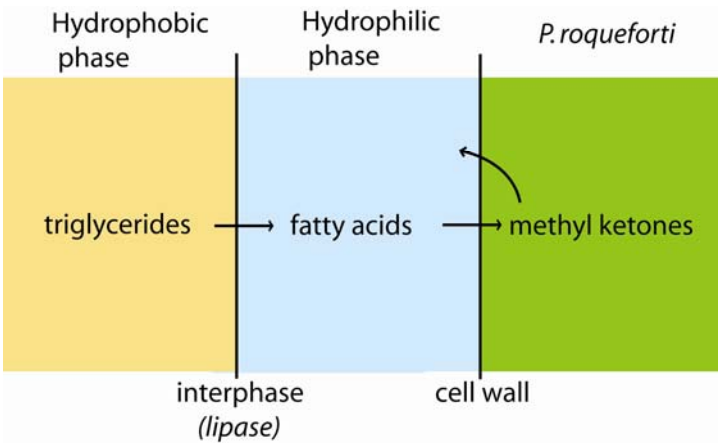

Figure 7.1: Schematic representation of the system applied for methyl ketone production.

Only a small part of these phases are used for product formation, and thus a substantial amount of waste is generated. Attempts have been made to optimise the production with respect to classical parameters such as stirring rate, pH, temperature, and enzyme load. Only minor improvements have been obtained however. An alternative solution would be a total redefinition of the process in terms of substrate compositions, running conditions etc.

In the literature different methods have been reported for production of 2-heptanone[24,61,77], one being a fed-batch technique, where octanoic acid is added in the feed[23]. This technique allows for 2-heptanone production from a well defined aqueous system, where only the amount of precursor needed (octanoic acid) is added.

To enable fed-batch fermentation, a process parameter suitable for monitoring the process is needed; in the current case octanoic acid in the broth could be such parameter. For monitoring this parameter an at- or on-line method is needed.

7.1 Monitoring

A production of 2-heptanone from a defined substrate should be run with the concentration level of octanoic acid between 5 to 10 mM at all times to avoid inhibition of the organism, see paper 5 for details. To enable this, process analytical technologies may be applied. The method needed is one that may be applied on- or at-line and is sensitive to the octanoic acid concentration in the water phase within 2.5 to 20 mM.

Different methods have been suggested in the literature, for example a pH state, which counteracts the decrease in H^+ concentration, as octanoic acid is degraded, by adding the octanoic acid as the feed, thereby enabling control of the fed-batch[23]. However, in a later paper by *Larroche et al.*[60] this method failed due to a parallel H^+ generating reaction taking place, making complex processes difficult to monitor using this method. Other methods reported are chromatographic method; either GC with different detectors or HPLC, see Table 7.1. Of the methods reported it is only the one reported by *Kellerhals et al.*[55] which is fast enough for monitoring and control.

Table 7.1: Methods reported for octanoic acid determination.

Subject	Method
Characterisation of moulds by measuring the fatty acids content[11]	HP 5890 series Gas Chromatograph with FID, capillary column; fused silica (30 m x 0.25 mm x 0.25 μm) HP 5890 series Gas Chromatograph equipped with Finnigan model 4023 mass spectrometer system
Method development for determination of carboxylic acids[51]	Hitachi HPLC separating system including an L-6300 intelligent pump, F-1080 fluorescent detector
Development of on-line GC, to maintain continuously fed substrate at a desired level, during the production of mcl-PHAs[1] by *Pseudomonas putida* KT2442[55]	Online determination of substrate concentration: Fermenter equipped with recirculation loop containing a crossflow filtration module. The permeate was analysed at a HP 5890 Gas Chromatograph with FID, and Permabond FFAP-0.35 column
Synthesis and characterisation of structured lipids in bench scale reactor[59]	Shimadzu 14A GC with FID
Investigation of the internal substrate concentration during the biotransformation of octanoic acid into 2-heptanone by *Penicillium roqueforti*[60]	Gas Chromatograph with FID, Capillary column; Supelcowax (30 m x 0.32 mm x 0.5μm)

A new method, facilitating reliable monitoring of low concentrations of octanoic acid during fed-batch production was needed. Based on the work by *Creuly et al.*[23] who showed that an ion

selective electrode works - but also taking into account the failure of this method reported by *Larroche et al.*[62] it was decided for the current thesis work to investigate the feasibility of applying the electronic tongue (ET) for octanoic acid determination in fermentation broth complete from a new beginning.

7.2 Experimental Design

Before carrying out the actual evaluation of the electronic tongue, a preliminary substrate and process settings for the fermentation has to be defined. This substrate was mainly defined contemplating the micro and macro nutritional requirements for the filamentous fungi, but also the experiences reported by others were contemplated (see paper 5 for details). For definitions of parameters settings the results reported in the literature with respect to pH and temperature were contemplated. A pH of 6.5 was chosen as earlier work had determined that optimum pH falls in the range from 5.5 to 7, the specific optimum being octanoic acid concentration dependent[41]. The temperature was fixed at 27°C based on work reported by Lawrence[64].

For investigating the feasibility of the electronic tongue at first an array consisting of 20 “most suitable” electrodes (from general a priori ET experience) were tested for their response to octanoic acid. From these six electrodes were selected with a cross sensitivity to octanoic acid.

The sensitivity of the electrodes was determined before hand, and as such the array should be applicable for the intended purpose; monitoring of the octanoic acid concentration. A fermentation broth consists of many constituents like sugar, proteins, and an amount of salt ions, which may affect the sensor membrane. To test the actual array performance in a fermentation broth and to incorporate the natural variation in to the model building phase the broth from four fermentations were applied as the aqueous matrix in the measurements.

After each fermentation run the substrate was heated and filtered to inactivate and remove mycelium. The aqueous phase was subsequently spiked with nine different octanoic acid volumes obtaining a calibration span from 0.65 to 20 mmol l^{-1}. A total of 36 samples were used in establishing the calibration relation. For detailed information regarding experimental conditions see paper 5.

From these experiments four individual models and a global model, all with good prediction abilities, were obtained, see Table 7.2.

Table 7.2: ET models obtained applying broth from the four independent fermentations. Y-data were linearised by taking the natural logarithm (ln) to the octanoic acid concentration and additionally the data were centred and scaled in the multivariate calibrations.

No	Slope	r^2	RMSEP	% RMSEP	RPD
1	0.92	0.96	0.13	5.5	5.1
2	0.95	0.98	0.10	4.2	6.7
3	0.89	0.96	0.15	6.3	4.3
4	0.84	0.90	0.21	8.8	3.1
Global	0.95	0.97	0.21	5.1	5.5

Besides the four fermentations a special fifth run was executed, to which octanoic acid was spiked. From this fermentation samples were withdrawn and handled according to sampling procedure outlined in Figure 7.2.

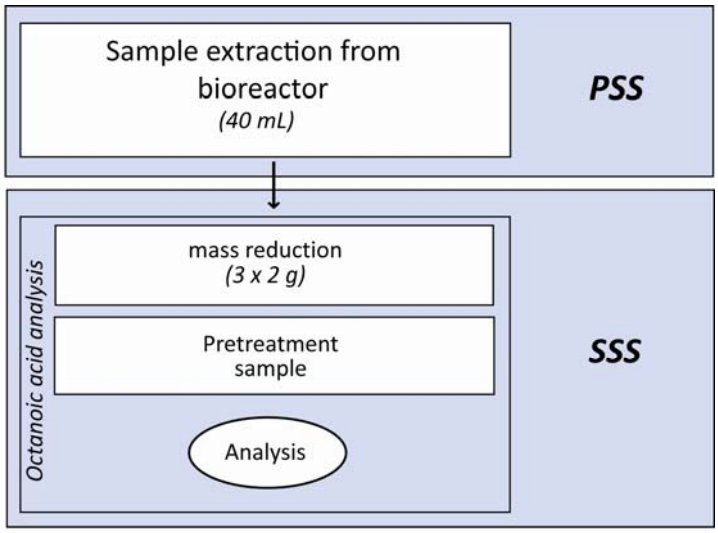

Figure 7.2: Illustration of the sampling extraction and handling from a 5L Applikon bioreactor.

The primary sampling was carried out from a 5 L Applikon bioreactor. In practice the sample must delimit a full-height column as the systems heterogeneity in the current system was assessed as being caused mainly by a density difference between cells and substrate, delineating a sample completely from top to bottom would be preferable. For the current process, it was critical to keep the environment aseptic, and this procedure could thus not be implemented. As a consequence it was chosen to apply the standard sampling equipment available for Applikon bioreactors, see Figure 7.3.

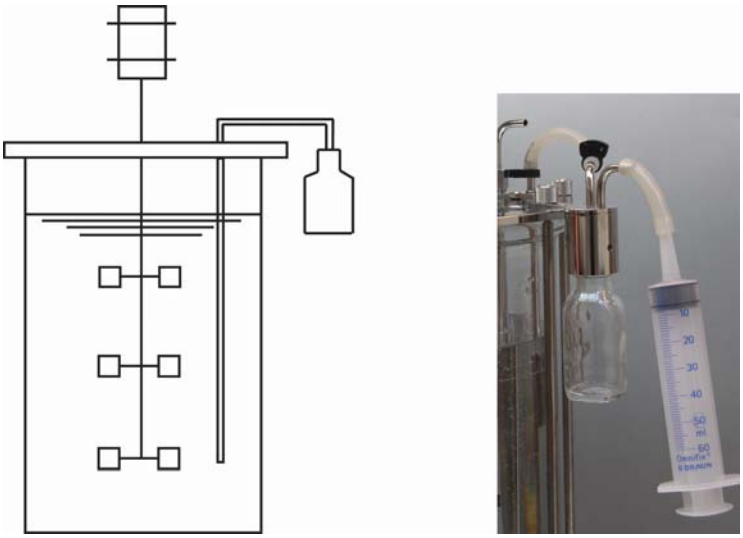

Figure 7.3: Primary sample extraction from 5 L Bioreactor. Shown to the right is the standard sampling equipment supplied- fermentation broth is sucked from the fermentor into the glass bottle by the syringe.

Applying the commercially available equipment resulted in a stationary sampling extraction, which was non-optimal, as it at the very least induced an incorrect sampling error IEE. As the main purpose of sampling was to obtain samples with different concentrations of octanoic acid, and not so much to relate the concentration to the process and the vessel at hand, this was deemed acceptable. It should be noted that applying a stationary sampling extraction point in a bioreactor calls for a total homogenisation of the broth at all times, as this is the only way of obtaining a representative sample, otherwise this standard sampling approach will simply yield a grab sample. Given this stipulation, the experiments continued.

After sample extraction (secondary sampling), the sample was placed on a magnetic stirrer and 2 g sample (analytical mass) was extracted applying the recommendations of *Holm-Nielsen et al.*[46] and pretreated. The solution was filtered and the octanoic acid concentration measured applying GC-FID. See paper 5 for details regarding the analytical procedure.

The samples withdrawn were used for testing the array on new, authentic samples. The deviations (RMSEP) obtained for these samples were generally in the range 0.10 – 0.15 (except one sample which had a value of 0.19), which was in the range of the RMSEP in Table 7.2. This indicated that the models perform equally well on new samples. More details regarding models are seen in paper 5.

7.3 Conclusion

The model obtained was characterised by a satisfactory prediction validation, although only leave-one-concentration-level-out cross-validation was applied (the model was very strong *w.r.t.* X-Y correlation, thus any validation would lead to approximately the same, satisfactory validation results in this particular case). Only one PLS-component was needed to span a

realistic compositional range with good validation characteristics. The global model was also able to predict samples from an octanoic acid degradation matrix, a much more stringent, realistic context.

The average future prediction error level, RMSEP% (test set validation), corresponded to 5.1% (ln-units). The RPD statistic was 5.5, also signifying a good prediction precision. The slope (accuracy) of the global model was 0.96. These statistics signified a highly satisfactory development result, allowing the conclusion that feasibility of an ET prediction model for octanoic acid prediction in realistic cultivation broths for industrial 2-heptanone production has been successfully demonstrated.

Establishing the feasibility of ET, the definition of the fermentation with respect to the optimal substrate composition, parameters settings, and precursor concentration, can now begin. In carrying out these investigations it would be beneficial to incorporate samples obtained during the different conditions into the ET model, thereby obtaining a more robust model. Another task is investigation of the on-line implementation of the method. It would be beneficial to be able to implement the ET on-line or in-line. Before doing so important questions regarding the maintenance of aseptic conditions of the process and sensor fouling should be resolved.

The sensors applied in the current study were made of plastic casing onto which the membranes were glued. These could not be autoclaved, a problem that also has to be solved before on-line implementation. A solution might be the implementation of a sterile barrier allowing only exit of the broth from the fermenter. Furthermore, the sensors applied all needed extensive washing after each measurement, to remove octanoic acid from the membrane; the washing time being proportional to the octanoic acid concentration. In an industrial application, this would evidently cause a delay. The consequence of such delay could be a fluctuation in the octanoic acid concentration in the fermentation broth. Whether this would have any real effect on the overall product performance would have to be evaluated for the industrial process implementation scenario.

8 References

1 Ahring, B.K., Sandberg, M. and Angelidaki, I., 1995. Volatile fatty acids as indicators of process imbalance in anaerobic digestors. Appl Microbiol Biotechnol 43, 559-565.

2 Andree, H., Dolud, M. and Hügle, T., 2005. Tenirs – NIR analysis of heterogeneous bioslurries in *Aalborg University Esbjerg, Baltic Biorefienry Symposium, Denmark, 26-28 May, (2005)*.

3 Angelidaki, I. and Ahring, B.K., 1995. Isomerization of n- and i-butyrate in anaerobic methanogenic systems. Antonie van Leeuwenhoek 68, 285-291.

4 Angelidaki, I., 2002. Environmental Biotechnology 12133, Environment & Resources DTU. Technical University of Denmark, Lyngby, Denmark.

5 ASTM 2008. Standard definitions of terms and symbols relating to spectroscopy, ASTM Vol. 14.01, Standard E131-90, cited p. 167 in *Handbook of Near-Infrared Analysis, 3rd ed*. Burns, D.A. and Ciurczak, E.W. (eds.), CRC Press, New Work, USA. ISBN: 0-8493-7393-X.

6 Bakeev, K.A. (ed.), 2005. Process Analytical Technology - spectroscopic tools and implementation for the chemical and pharmaceutical industries: 2nd ed. Blackwell Publishing, Oxford, UK. ISBN: 1-4051-2103-3.

7 Baptista, P., Felizardo, P., Menezes, J.C. and Correia, J.N., 2008. Monitoring the quality of oils for biodiesel production using multivariate near infrared spectroscopy models. J Near Infrared Spectrosc 16, 445-454.

8 Beebe, K.R., Pell, R.J. and Seasholtz, M.B., 1998. Chemometrics: A practical guide. John Wiley & Sons, Inc., New York, USA. ISBN: 0-471-12451-6.

9 Benito, M.T.J., Ojeda, C.B. and Rojas, F.S., 2008. Process analytical chemistry: Applications of near infrared spectrometry in environmental and food analysis: An overview. Appl Spectrosc Rev 43, 452-484.

10 Björnsson, L., Murto, M. and Mattiasson, B., 2000. Evaluation of parameters for monitoring an anaerobic co-digestion process. Appl Microbiol Biotechnol 54, 844-849.

11 Blomquist, G., Andersson, B., Andersson, K. and Brondz, I., 1992. Analysis of fatty-acids - a new method for characterization of molds. J Microbiol Meth 16, 59-68.

12 Boe, K., 2006. Online monitoring and control of the biogas process. Ph.D. dissertation, Institute of Environments & Resources, Technical University of Denmark, Lyngby, Denmark.

13 Boland, L., 2008. Evaluation of Sampling Facilities and Process Analytical Technologies in Applied Biotechnology. M.Sc. project, Esbjerg Institute of Technology, Aalborg University, Esbjerg, Denmark.

14 Bro, R., 1996. Håndbog i multivariable kalibrering - Kemometri: 1. udgave 1. oplag. Levnedsmiddelteknologi, MLI, KVL, Copenhagen, Denmark.

15 Buranov, A.U. and Mazza, G., 2008. Lignin in straw of herbaceous crops. Ind Crops Prod 28, 237-259.

16 Burns, D.A. and Ciurczak, E.W. (eds.), 2008. Handbook of Near-Infrared Analysis: 3rd ed. CRC Press, New Work, USA. ISBN: 0-8493-7393-X.

17 Callis, J., Illman, D.L. and Kowalski, B.R., 1987. Process analytical chemistry. Anal Chem 59, 624-637.

18 Chemometrics & Intelligent Laboratory Systems: http://www.elsevier.com/wps/find/journaldescription.cws_home/502682/description#description.

19 Chen, Q., Zhao, J. and Vittayapadung, S., 2008a. Identification of the green tea grade level using electronic tongue and pattern recognition. Food Res Int 41, 500-504.

20 Chen, Y., Cheng, J.J. and Creamer, K.S., 2008b. Inhibition of anaerobic digestion process: A review. Bioresour Technol 99, 4044-4064.

21 Ciosek, P. and Wróblewski, W., 2007. Sensor arrays for liquid sensing - electronic tongue systems. Analyst 132, 963-978.

22 Clementschitsch, F. and Bayer, K., 2006. Improvement of bioprocess monitoring: development of novel concepts. Microbial Cell Fact 5,

23 Creuly, C., Larroche, C. and Gros, J.B., 1990. A fed-batch technique for 2-heptanone production by spores of Penicillium roquefortii. Appl Microbiol Biotechnol 34, 20-25.

24 Creuly, C., Larroche, C. and Gros, J.B., 1992. Bioconversion of fatty acids into methyl ketones by spores of penicillium roqueforti in a water-organic solvent, two-phase system. Enzyme Microb Technol 14, 669-678.

25 Crowley, J., Arnold, S.A., Wood, N., Harvey, L.M. and McNeil, B., 2005. Monitoring a high cell density recombinant Pichia pastoris fed-batch bioprocess using transmission and reflectance near infrared spectroscopy. Enzyme Microb Technol 36, 621-628.

26 de Bok, F.A.M., Plugge, C.M. and Stams, A.J.M., 2004. Interspecies electron transfer in methanogenic propionate degrading consortia. Water Res 38, 1368-1375.

27 Diamantis, V., Melidis, P. and Aivasidis, A., 2006. Continuous determination of volatile products in anaerobic fermenters by on-line capillary gas chromatography. Anal Chim Acta 573-574, 189-194.

28 Drapcho, C.M., Nhuan, N.P. and Walker, T.H., 2008. Biofuels Engineering Process Technology. Mc Graw Hill, New York, USA. ISBN: 978-0-07-148749-8.

29 Eriksson, L., Johansson, E., Kettaneh-Wold, N. and Wold, S., 2001. Multi- and Megavariate Data analysis principles and applications. Umetrics, Umeå, Sweden. ISBN: 91-973730-1-X.

30 Esbensen, K.H., 2002. Multivariate analysis in practice: an introduction to multivariate data analysis and experimental design: 5th ed. CAMO Press, Norway, ISBN: 82-993330-3-2.

31 Esbensen, K.H. and Julius, L.P., 2009. Representative sampling: Data quality validation - a necessary trinity in chemometrics in *Comprehensive chemometrics: Chemical and biochemical data analysis*. Brown, S., Tauler, R. and Walczak, R. (eds.), Elsevier, Oxford, UK.

32 Esbensen, K.H. and Paasch-Mortensen, P., 2009. Process sampling: Theory of sampling - the missing link in Process Analytical Technologies (PAT) in *Process analytical technology: 2nd ed*. Bakeev, K.A. (ed.), Blackwell publishing, Oxford, UK. ISBN: 1-4051-2103-3 .

33 Esbensen, K., Kirsanov, D., Legin, A., Rudnitskaya, A., Mortensen, J., Pedersen, J., Vognsen, L., Makarychev-Mikhailov, S. and Vlasov, Y., 2004. Fermentation monitoring using mulitsensor systems: feasibility study of the electronic tongue. Anal Bioanal Chem 378, 391-395.

34 Esbensen, K.H., Friis-Petersen, H.H., Petersen, L., Holm-Nielsen, J.B. and Mortensen, P.P., 2007. Representative process sampling - in practice: Variographic analysis and estimation of total sampling errors (TSE). Chemom Intell Lab Syst 88, 41-59.

35 Faughey, G.J. and Sharma, H.S.S., 2000. A preliminary evaluation of near infrared spectroscopy for assessing physical and chemical characteristics of flax fibre. J Near Infrared Spectrosc 8, 61-69.

36 FDA PAT Initiative, 2005. Guidance for Industry PAT – A framework for Innivative Pharmaceutical devolpment, Manufacturing, and Quality Assurance. U.S. Department of Health and Human Services, Food and Drug Administration, Pharmaceutical CGMPs.

37 Fearn, T., 2002. Assessing calibrations: SEP, RPD, RER and R2. NIR News 13, 12-14.

38 Fearn, T., 2003. Are two pretreatments better than one?. NIR News 14, 9-11.

39 Galand, P.E., Fritze, H., Conrad, R. and Yrjälä, K., 2005. Pathways for methanogenesis and diversity of methanogenic archaea in three boreal peatland ecosystems. Appl Environ Microbiol 71, 2195-2198.

40 Galbe, M. and Zacchi, G., 2002. A review of the production of ethanol from softwood. Appl Microbiol Biotechnol 59, 618-628.

41 Gehrig, R.F. and Knight, S.G., 1963. Fatty acid oxidation by spores of Penicillium roqueforti. Appl Microbiol 11, 166-170.

42 Geladi, P., McDougall, D. and Martens H., 1985. Linearization and scatter-correction for near-infrared reflectance spectra of meat. Appl Spectroscop 39, 491-500.

43 Gy, P., 1998. Sampling for analytical purposes. John Wiley & Sons, Chichester, UK. ISBN: 0-471-97956-2.

44 Halstensen, M. and Esbensen, K., 2000. New developments in acoustic chemometric prediction of particle size distribution - 'the problem is the solution'. J Chemom 14, 463-481.

45 Halstensen, M., 2001. Acoustic Chemometrics - Experimental multivariate sensor technology and development of system prototypes for industrial multi-phase characterisation: selected forays. Ph.D. dissertation, Telemark University College, Norway.

46 Halstensen, M., de Bakker, P. and Esbensen, K.H., 2006. Acoustic chemometric monitoring of an industrial granulation process- a PAT feasibility study. Chemom Intell Lab Syst 84, 88-97.

47 Hames, B.R., Thomas, S.R., Sluiter, A.D., Roth, C.J. and Templeton, D.W., 2003. Rapid biomass analysis. Appl Biochem Biotechnol 105-108, 5-16.

48 Holm-Nielsen, J.B., Andree, H., Lindorfer, H. and Esbensen, K.H., 2007. Transflexive embedded near infrared monitoring for key process intermediates in anaerobic digestion/biogas production. J Near Infrared Spectrosc 15, 123-135.

49 Holm-Nielsen, J.B., Dahl, C.K., and Esbensen, K.H., 2006. Representative sampling for process analytical charatirization of heterogeneous bioslurry systems - a reference study of sampling issues in PAT. Chemom Intell Lab Syst 83, 114-126.

50 Hou, R., Hunt, A. and Williams, R.A., 1999. Acoustic monitoring of pipeline flows: particulate slurries. Powder Technol 106, 30-36.

51 Huang, C.Z., Santa, T. and Imai, K., 2002. Development of 7-(N,N-dimethylaminosulfonyl)-5-N-(4-N-aminoethyl)piperazino-2,1,3-benzoxadiazole as a water-soluble fluorogenic regent for the sensitive liquid chromatographic determination of saturated carboxylic acids. Analyst 127, 741-747.

52 Huang, H., Yu, H., Xu, H. and Ying, Y., 2008. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. J Food Eng 87, 303-313.

53 Jin, S. and Chen, H., 2007. Near-infrared analysis of the chemical composition of rice straw. Ind Crops Prod 26, 207-211.

54 Jöreskog, K. and Wold, H., 1981. The ML and PLS Techniques for Modeling with Latent variables: Historical and Comparative Aspects in *Systems under indirect observation: Causality, Structure, Prediction vol. 1.* Wold, H. and Jöreskog, K. (eds), Amsterdam Holland.

55 Kellerhals, M.B., Kessler, B. and Witholt, B., 1999. Closed-loop control of bacterial high-cell-density fed-batch cultures: Production of mcl-PHAs by Pseudomonas putida KT2442 under single-substrate and cofeeding conditions. Biotechnol Bioeng 65, 306-315.

56 Kim, N., Park, K., Park, I-S., Cho, Y-J. and Bae, Y.M., 2005. Application of a taste evaluation system to the monitoring of Kimchi fermentation. Biosens Bioelectron 20, 2283-2291.

57 Kourti, T., 2006. Process analytical technology beyond real-time analyzers: The role of multivariate analysis. Crit Rev Anal Chem 36, 257-278.

58 Känsäkoski, M., Kurkinen, M., von Weymarn, N., Niemelä, P., Neubauer, P., Juuso, E., Eerikäinen, T., Turunen, S., Aho, S. and Suhonen, P., 2006. Process analytical technology (PAT) nedds and applications in the bioprocess industry. VTT, Vuorimiehentie, Finland. ISBN: 951-38-6612-2.

59 Lai, O.M., Low, C.T. and Akoh, C.C., 2004. Lipase-catalyzed acidolysis of palm olein and caprylic acid in a continuous bench-scale packed bed bioreactor. Food Chem 92, 527-533.

60 Larroche, C., 1996. Internal substrate concentrations during biotransformation of octanoic acid into 2-heptanone by spores of Penicillium roquefortii. J Ind Microbiol 16, 29-35.

61 Larroche, C., Arpah, M. and Gros, J.B., 1989. Methyl-ketone production by Ca-alginate/Eudragit RL entrapped spores of Penicillium roqueforti. Enzyme Microb Technol 11, 106-112.

62 Larroche, C., Besson, I. and Gros J.B., 1994. Behavior of spores of Penicillium roqueforti during fed-batch bioconversion of octanoic acid into 2-heptanone. Biotechnol Bioeng 44, 699-709.

63 Larsen, J., Petersen, M.Ø., Thirup, L., Li, H.W. and Iversen, F.K., 2008. The IBUS process - lignocellulosic bioethanol close to a commercial reality. Chem Eng Technol 31, 765-772.

64 Lawrence, R.C., 1966. The oxidation of fatty acids by spores of Penicillium roquefortii. J Gen Microbiol 44, 393-405.

65 Legin, A. and Vlasov, Y. (eds.), 2008. Sensors and Actuators B Chemical - Special Issue Vol. B131 (1): Selected papers from the 12th International Symposium on Olfaction and Electronic Noses. Elsevier.

66 Legin, A., Rudnitskaya, A. and Vlasov, Y., 2003. Electronic Tongues: New analytical Perspective for chemical sensors in *Comprehensive Analytical Chemistry, XXXIX.* Alegret, S. (ed.), Elsevier, UK. ISBN 978-0-444-51037-2.

67 Legin, A.V., Rudnitskaya, A.M., Vlasov, Y.G., Di Natale, C. and D'Amico, A., 1999. The features of the electronic tongue in comparison with the characteristics of the discrete ion-selective sensors. Sens Actuators B 58, 464-468.

68 Lestander, T.A. and Rhén, C., 2005. Multivariate NIR spectroscopy models for moisture, ash and calorific content in biofuels using bi-orthogonal partial least squares regression. Analyst 130, 1182-1189.

69 Lindorfer, H., Pérez López, C., Resch, C., Braun, R. and Kirchmayr, R., 2007. The impact of increasing energy crop addition on process performance and residual methane potential in anaerobic digestion. Water Sci Technol 56, 55-63.

70 Martens, H. and Næs, T., 1991. Multivariate Calibration. Wiley & Sons, Inc., Chichester, UK. ISBN: 0-471-93047-4.

71 Mata-Alvarez, J., Macé, S. and Llabrés, P., 2000. Anaerobic digestion of organic solid wastes. An overview of research achievements and perspectives. Bioresour Technol 74, 3-16.

72 McLennan, F., 1995. Process analytical chemistry in perspective in *Process analytical chemistry.* McLennan, F. and Kowalski, B.R., Blackie Academic & Professional, London, UK. ISBN: 0-7514-0038-6.

73 McSweeney, P.L.H. and Sousa, M.J., 2000. Biochemical pathways for the production of flavour compounds in cheeses during ripening: A review. Lait 80, 293-324.

74 Mortensen, P.P., 2006. Process analytical chemistry - opportunities and problems for bioindustrial implementation. Pd.D. dissertation, Aalborg University, Esbjerg Institute of Technology, Aalborg, Denmark. ISBN: 87-7606-017-9.

75 Nag, A. (ed.), 2008. Biofuels and performance. McGraw-Hill, New York, USA. ISBN: 987-0-07-148970-6.

76 Nordberg, Å., Hansson, M., Sundh, I., Nordkvist, E., Carlsson, H. and Mathisen, B., 2000. Monitoring of a biogas process using electronic gas sensors and near-infrared spectroscopy (NIR). Water Sci Technol 41, 1-8.

77 Park, O.J., Holland, H.L., Khan, J.A. and Vulfson, E.N., 2000. Production of flavour ketones in aqueous-organic two-phase systems by using free and microencapsulated fungal spores as biocatalysts. Enzyme Microb Technol 26, 235-242.

78 Pérez, J., Muñoz-Dorado, J., de la Rubia, T. and Martinez, J., 2002. Biodegradation and biological treatments of cellulose, hemicellulose and lignin: An overview. Int Microbiol 5, 53-63.

79 Petersen, L. and Esbensen, K.H., 2005. Representative process sampling for reliable data analysis - a tutorial. J Chemom 19, 625-647.

80 Petersen, L., 2005. Pierre Gy's Theory of Sampling (TOS) – in practice: Laboratory and industrial didactics including – A first foray into image analytical sampling. Ph.D. dissertation, Aalborg University Esbjerg, Aalborg, Denmark; ISBN: 87-7606-004-7.

81 Petersen, L., Dahl, C.K., and Esbensen K.H., 2004. Representative mass reduction in sampling - a critical survey of techniques and hardware. Chemom Intell Lab Syst 74, 95-114.

82 Petersen, L., Minkkinen, P., and Esbensen, K.H., 2005. Representative sampling for reliable data analysis: Theory of sampling. Chemom Intell Lab Syst 77, 261-277.

83 Pind, P.F., Angelidaki, I. and Ahring, B.K., 2003. A new VFA sensor technique for anaerobic reactor systems. Biotechnol Bioeng 82, 54-61.

84 Pitard, F.F., 1993. Pierre Gy's sampling theory and sampling practice – heterogeneity, sampling correctness, and statistical process control: 2nd ed. CRC Press, Boca Raton, New York, USA. ISBN: 0-8493-8917-8.

85 Pitard, F.F., 2009. Pierre Gy's theory of sampling and C.O. Ingamells' poisson process approach. Doctoral thesis (Dr. Tech.), Aalborg University, Campus Esbjerg, Denmark. ISBN: 978-87-7606-032-9.

86 Reeves, J.B., Blosser, T.H. and Colenbrander, V.F., 1991. Analysis of silage composition by near infrared reflectance spectroscopy. Opt Agricul 1379, 28-38.

87 Reeves III, J.B., 1994. Near- versus mid-infrared diffuse reflectance spectroscopy for the quantitative determination of the composition of forages and by-products. J Near Infrared Spectrosc 2, 49-57.

88 Richardson, A.D. and Reeves III, J.B., 2005. Quantitative reflectance spectroscopy as an alternative to traditional wet lab analysis of foliar chemistry: Near-infrared and mid-infrared calibrations compared. Can J Forest Res 35, 1122-1130.

89 Riebe, M.T. and Eustace, D.J., 1990. Process analytical chemistry - an industrial perspective. Anal Chem 62, A65-A71.

90 Rodriguez, J., Kleerebezem, R., Lema, J.M. and van Loosdrecht, M.C.M., 2006. Modeling product formation in anaerobic mixed culture fermentations. Biotechnol Bioeng 93, 592-606.

91 Rudnitskaya, A. and Legin, A., 2008. Sensor systems, electronic tongues and electronic noses, for the monitoring of biotechnological processes. J Ind Microbiol Biotechnol 35, 443-451.

92 Rudnitskaya, A., Delgadillo, I., Legin, A., Rocha, S.M., Costa, A. and Simões, T., 2007. Prediction of the port wine age using an electronic tongue. Chemom Intell Lab Syst 88, 125-131.

93 Saha, B., 2003. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 30, 279-291.

94 Sánchez, Ó. And Cardona, C.A., 2008. Trends in biotechnological production of fuel ethanol from different feedstocks. Bioresour Technol 99, 5270-5295.

95 Scarff, M., Arnold, S.A., Harvey, L.M. and McNeil, B., 2006. Near infrared spectroscopy for bioprocess monitoring and control: Current status and future trends. Crit Rev Biotechnol 26, 17-39.

96 Schink, B., 1997. Energetics of syntrophic cooperation in methanogenic degradation. Microbiol Mol Biol Rev 61, 262-280.

97 Schöftner, R., Valentin, K., Schmiedinger, B., Trogisch, S., Haberbauer, M., Katzlinger, K., Schnitzhofer, W. and Weran, N., 2007. Best Biogas Practise. Bundesministerium für Verkehr, Innovation und Technologie, Vienna, Austria.

98 Serridge, M. and Licht, T., 1987. Piezoelectric accelerometer ad vibration preamplifier handbook - theory and application handbook. Brüel & Kjær, Nærum, Denmark.

99 Siesler, H.W., 2008. Basic Principles of Near-Infrared Spectroscopy in *Handbook of Near-Infrared Analysis 3rd ed*. Burns, D.A. and Ciurczak, E.W. (eds.), CRC Press, New Work, USA. ISBN: 0-8493-7393-X.

100 Simpson, M.B., 2005. Near- infrared spectroscopy for process analytical chemistry: Theory, technology and implementation in *Process analytical technology*. Bakeev, K.A. (ed.), Blackwell Publishing, Oxford, UK. ISBN: 1-4051-2103-3.

101 Sjöström, E., 1993. Wood polysaccharides in Wood chemistry; fundementals and applications. Academic Press Inc., San Diego, USA. ISBN: 0-1264-7481-8.

102 Skibsted, E., 2005. PAT and Beyond. Ph.D. dissertation, University of Amsterdam, Amsterdam, the Netherlands.

103 Skoog, D.A., Holler, E.J. and Nieman, T.A., 1998. Principles of Instrumental Analysis, 5th ed. Brooks Cole, USA. ISBN: 0-03-002078-6.

104 Smilde, A., Bro, R. and Geladi, P., 2004. Multi-way analysis applications in the chemical science. Wiley & Sons, Inc., Chichester, UK. ISBN: 0-471-98691-7.

105 Sohn, M., Himmelsbach, D.S., Barton II, F.E., Griffey, C.A., Books, W., and Hicks, K.B., 2007. Near-infrared analysis of ground barley for use as a feedstock for fuel ethanol production. Appl Spectrosc 61,

106 Solomons, T.W.G. and Fyhle, C.B., 2000. Organic Chemistry. Wiley & Sons, Inc., New York, USA. ISBN: 0-471-19095-0.

107 Spanjers, H., Bouvier, J.C., Steenweg, P., Bisschops, I., van Gils, W. and Versprille, B., 2006. Implementation of in-line infrared monitor in full-scale anaerobic digestion process. Water Sci Technol 53, 55-61.

108 Stams, A.J.M., Plugge, C.M., de Bok, F.A.M., van Houten, B.H..G.W., Lens, P., Dijkman, H. and Weijma, J., 2005. Metabolic interactions in methanogenic and sulfate-reducing bioreactors. Water Sci Technol 52, 13-20.

109 Steyer, J.P., Bouvier, J.C., Conte, T., Gras, P., Harmand, J. and Deigenes, J.P., 2002. On-line measurements of COD, TOC, VFA, total and partial alkalinity in anaerobic digestion processes using infra-red spectrometry. Water Sci Technol 45, 133-138.

110 Sun, Y. and Cheng, J.Y., 2002. Hydrolysis of lignocellulosic materials for ethanol production: A review. Bioresour Technol 83, 1-11.

111 Switzenbaum, M.S., Giraldo-Gomez, E. and Hickey, R.F., 1990. Monitoring of the anaerobic methane fermentation process. Enzyme Microb Technol 12, 722-730.

112 Taherzadeh, M. and Karimi, K., 2007a. Enzyme-based hydrolysis processes for ethanol from lignocellulosic materials: a review. Bioresour 2, 707-738.

113 Taherzadeh, M. and Karimi, K., 2007b. Acid-based hydrolysis processes for ethanol from lignocellulosic materials: a review. Bioresour 2, 472-499.

114 Taherzadeh, M.J. and Karimi, K., 2008. Bioethanol: Market and production processes in *Biofuels refining and performance*. Nag, A. (ed.), McGraw-Hill, New York, USA. ISBN: 987-0-07-148970-6.

115 Thomsen, M.H., Thygesen, A., and Thomsen, A.B., 2008. Hydrothermal treatment of wheat straw at pilot plant scale using a three-step reactor system aiming at high hemicellulose recovery, high cellulose digestibility and low lignin hydrolysis. Bioresour Technol 99, 4221-4228.

116 Thomsen, M.H., Thygesen, A., Jørgensen, H., Larsen, J., Christensen, B.H. and Thomsen, A.B., 2006. Preliminary results on optimization of pilot scale pretreatment of wheat straw used in coproduction of bioethanol and electricity. Appl Biochem Biotechnol 129-132, 448-460.

117 Turner, C., Rudnitskaya, A. and Legin, A., 2003. Monitoring batch fermentations with an electronic tongue. J Biotechnol 103, 87-91.

118 Vaidyanathan, S., Harvey, L.M. and McNeil, B., 2001. Deconvolution of near-infrared spectral information for monitoring mycelial biomass and other key analytes in a submerged fungal bioprocess. Anal Chim Acta 428, 41-59.

119 Vaidyanathan, S., White, S., Harvey, L.M. and McNeil, B., 2003. Influence of morphology on the near-infrared spectra of mycelial biomass and its implications in bioprocess monitoring. Biotechnol Bioeng 82, 715-724.

120 Vlasov, Y., Legin, A. and Rudnitskaya, A., 2002. Electronic tongues and their analytical application. Anal Bioanal Chem 373, 136-146.

121 Vojinovic, V., Cabral, J.M.S. and Fonseca, L.P., 2006. Real-time bioprocess monitoring - Part I: In situ sensors. Sens Actuators B 114, 1083-1091.

122 Williams, P. and Norris, K. (eds.), 2001. Near-Infrared Technology in the agricultural and food industries: 2nd ed. AACC Inc, Minnesota, USA. ISBN: 1-891127-24-1.

123 Workman, J.J. and Burns, D.A., 2008. Commercial NIR Instrumentation in *Handbook of Near-Infrared Analysis, 3rd ed*. Burns, D.A. and Ciurczak, E.W. (eds.), CRC Press, New Work, USA. ISBN: 0-8493-7393-X.

124 Workman, Jr. , J., Koch, M. and Veltkamp, D., 2007. Process analytical chemistry. Anal Chem 79, 4345-4364.

125 Wyman, C.E., 1996. Handbook on bioethanol: Production and utilization. Taylor & Francis, Washington DC, USA. ISBN: 1560325534.

126 Zhang, P., Zeng, G., Zhang, G., Li, Y., Zhang, B. and Fan, M., 2008. Anaerobic co-digestion of biosolids and organic fraction of municipal solid waste by sequencing batch process. Fuel Process Technol 89, 485-489.

127 Zhang, R., El-Mashad, H.M., Hartman, K., Wang, F., Liu, G., Choate, C. and Gamble, P., 2007. Characterization of food waste as feedstock for anaerobic digestion. Bioresour Technol 98, 929-935.

128 Öhgren, K., 2006. Fuel Ethanol Production from Corn Stover optimization of steam pretreatment and improvement of simultaneous saccharification and fermentation. Doctoral thesis, Lund university, Lund, Sweden. ISBN: 91-7422-124-8.

I

ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

On-Line Near Infrared Monitoring of Glycerol-Boosted Anaerobic Digestion Processes: Evaluation of Process Analytical Technologies

Jens Bo Holm-Nielsen, Carina Juel Lomborg, Piotr Oleskowicz-Popiel, Kim H. Esbensen

Applied Chemometrics, Analytical Chemistry, Acoustic Chemometrics, Applied Biotechnology, Bioenergy and Sampling Research Group, Esbjerg Institute of Technology, Aalborg University, Niels Bohrs Vej 8, 6700 Esbjerg, Denmark; telephone: +45-7912-7715; fax: +45-6650-1091; e-mail: jhn@aaue.dk

Received 29 March 2007; revision received 7 June 2007; accepted 26 June 2007

Published online 11 July 2007 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21571

ABSTRACT: A study of NIR as a tool for process monitoring of thermophilic anaerobic digestion boosted by glycerol has been carried out, aiming at developing simple and robust Process Analytical Technology modalities for on-line surveillance in full scale biogas plants. Three 5 L laboratory fermenters equipped with on-line NIR sensor and special sampling stations were used as a basis for chemometric multivariate calibration. NIR characterisation using Transflexive Embedded Near Infra-Red Sensor (TENIRS) equipment integrated into an external recurrent loop on the fermentation reactors, allows for representative sampling, of the highly heterogeneous fermentation bio slurries. Glycerol is an important by-product from the increasing European bio-diesel production. Glycerol addition can boost biogas yields, if not exceeding a limiting 5–7 $g\,L^{-1}$ concentration inside the fermenter—further increase can cause strong imbalance in the anaerobic digestion process. A secondary objective was to evaluate the effect of addition of glycerol, in a spiking experiment which introduced increasing organic overloading as monitored by volatile fatty acids (VFA) levels. High correlation between on-line NIR determinations of glycerol and VFA contents has been documented. Chemometric regression models (PLS) between glycerol and NIR spectra needed no outlier removals and only one PLS-component was required. Test set validation resulted in excellent measures of prediction performance, precision: $r^2 = 0.96$ and accuracy = 1.04, slope of predicted versus reference fitting. Similar prediction statistics for acetic acid, iso-butanoic acid and total VFA proves that process NIR spectroscopy is able to quantify all pertinent levels of both volatile fatty acids and glycerol.

Biotechnol. Bioeng. 2008;99: 302–313.

KEYWORDS: anaerobic digestion (AD); near infrared spectroscopy (NIR); on-line measurement; volatile fatty acids (VFA); glycerol; process analytical technologies (PAT)

Introduction

Bioethanol and biodiesel are the main biofuels utilized in the transportation sector world-wide, comprising of a total volume of 23.3 million tonnes in 2003, of which 6.4% was biodiesel (Mandil, 2004). The largest contributors of biofuels Brazil and the US produce mainly bioethanol. In Europe the production of biodiesel in the same year comprised 78% of the total amount of biofuels produced, resulting in 65% of the total world capacity. Further significant increase is planned for biofuels in Europe. A likely vision for year 2030 will be to exchange one fourth of the EU transportation fossil fuels by biofuels according to the Biofuels Research Advisory Council the predicted energy demand for transport is estimated to be 440 million tonnes of oil equivalents.

Biodiesel is produced by a direct trans-esterification of vegetable oils: rapeseed, palm, soybean, sunflower and other organic oil products or by-products. The main by-product from biodiesel production is glycerol with varying purity. When the conventional catalytic processes using caustic soda or sodium methylate process, known as the Fatty Acid Methyl Ester (FAME), is applied, purity falls between 80% and 95%. When applying a new heterogeneous process, consisting of mixtures of oxides of zinc and aluminium, a purity of 98% can be reached (Bournay et al., 2005). The glycerol helps to make the biodiesel production more economically feasible as it is a valuable by-product used by the cosmetic, medicine, and food industries. The industrial glycerol market is limited, so when the biodiesel production capacity is expanding glycerol ends up as a relatively low value energy by-product raising the cost of the biodiesel production (Mandil, 2004).

Correspondence to: J.B. Holm-Nielsen

Therefore, there is a potential need for a new application, as an integrated biodiesel and biogas production in an integrated biorefinery concept. The primary by-product from anaerobic digestion, digested bioslurry, can be applied as primary macro and micro nutrients for growing energy crops needed for biodiesel production. At the same time the by-product from the biodiesel plant, glycerol can be applied as an extra C-source in co-fermentation with manure and other organic by-products in the anaerobic digestion process, thereby increasing the biogas yield. This paper is studying one part of the biorefinery loop, the application of glycerol for anaerobic co-fermentation with manure and other organic by-product feedstocks.

The co-fermentation effect is well known, especially in the Danish Biogas sector, which has been documenting the effect since the 1990s, showing doubling or tripling of the methane yield when quality guaranteed food waste and similar types of organic waste were combined with cow and pig slurries at biogas plants (Nielsen et al., 2002). From the biogas plant documentations it was found that high biogas production was positively correlated with addition of high concentrate organic by-products like vegetable oils, fish oil, animal lipids or similar compounds. However, at a certain level organic overload was reached resulting in process imbalances and inhibition conditions. In a few cases there was a complete stop in the process that necessitated re-inoculum addition and new feedstock supply. To avoid the risk of inducing this type of imbalance well performing biogas plants are today running with a medium biogas and methane yield. The biogas plants are monitoring selected control parameters to avoid organic overloading. Recent experiments with co-fermentations applying glycerol with mixtures of pig manure, maize silage, and rapeseed meal have shown a significant increase in the methane yield. To maintain a stable digestion process, the amount of glycerol should not exceed 6% (Amon et al., 2006). From these results it is evident that the glycerol can be advantageously applied, but a strict control strategy of the intermediate glycerol and fatty acid concentrations in the biogas reactor is needed to avoid the risk of organic overloading.

The most common measurement techniques applied in the biogas plant are simple on-line pH and temperature monitoring, as well as biogas flow measurements combined with off-line quantifications using high performance liquid chromatography (HPLC), gas chromatography (GC), and flow injection analysis (FIA). Different methods and equipment available for anaerobic digestion are presented in (Boe, 2006; Vanrolleghem and Lee, 2003). The simple on-line methods are however not able to predict process imbalances, and the off-line analyses are inefficient. Sample preparation is often time consuming with the result not being reached within a suitable timeframe and furthermore the analysis requires expensive equipment and skilled laboratory technicians. Hence the methods currently applied are inadequate for this complex task.

As means of obtaining better control and also to improve understanding of the process, process analytical technologies (PAT) may be invoked. PAT has been applied to a variety of applications within refineries, the food processing industries, and the pharmaceutical industries. It covers a wide array of process analytical methods (Bakeev, 2005), for example, on-line GC determination of volatile components (Boe, 2006; Diamantis et al., 2006), spectroscopy coupled to a flow system containing glycerol dehydrogenase for determining glycerol in wine (Fernandes et al., 2004) and near infrared spectroscopy for monitoring of acetate, glycerol, ammonium, and biomass in a recombinant *E. coli* production (Macaloney et al., 1997). Of the established methods, near infrared spectroscopy (NIR) seem to be the most promising solution as it is rapid and non-invasive. Furthermore, it requires no chemical addition and once calibrated by the help of multivariate data analysis it needs little maintenance except for system drift.

The main objective of this paper is to investigate the possibility of measuring the amount of glycerol and volatile fatty acids: acetic, propionic, iso-butanoic, butanoic, iso-valeric, and valeric acid and total-volatile fatty acids (VFA) in anaerobic fermentation bio slurry by on-line NIR. VFA can accumulate during the anaerobic digestion process and result in an increase that directly reflects the process behaviour and/or imbalances. The VFA concentration has been the intermediary compound suggested most often for monitoring anaerobic digestion processes (Ahring et al., 1995; Angelidaki et al., 2003; Hill and Holmberg, 1988). Several studies have pointed out that in addition to monitoring the total VFA behaviour, individual volatile fatty acids will also contribute towards improved understanding of anaerobic digestion (AD). The ratio between acetic acid and propanoic acid in the process can provide valuable information as an early warning before a process failure would occur (Boe, 2006; Hill and Holmberg, 1988). VFA are excellent compounds for indicating organic overload and the toxic condition where acid consumers and methanogens are inhibited; however, the VFA response is still unclear under toxic stress levels where acid producers are also inhibited—as under a high concentration of long chain fatty acids (Boe, 2006; Mladenovska et al., 2003).

A secondary objective is to obtain knowledge about the effect of different imbalances in the anaerobic digestion process, caused by the addition of glycerol. The concentration level of the added glycerol was studied, and how to balance or manage the process when it includes glycerol by-products from the biodiesel refinery processes.

Materials and Methods

Raw Materials

Manure and Digested Manure

The raw feedstock is mainly composed of manure and organic food industrial waste. Together with the inoculum needed for the AD process, both were collected from the

Ribe biogas plant in Southern Denmark. The inoculum biomass—the digested manure, was sampled at the main pipeline outlet from the thermophilic AD process. The digested manure was screened through a 3 mm sieve to eliminate all larger solid particles, necessary to avoid clogging in the on-line flow cell system in the trials reported below. The digestate was transported in insulated containers to avoid temperature decrease and microorganism disturbances. It was immediately inserted in the fermenters after delivery. The feedstock consisted of a typical substrate mixture, which was divided into 1 L portions and kept in the freezer before being fed into the bioreactors. It was defrosted before application and then heated up to 53°C—the specific thermophilic temperature level.

Glycerol

Pure glycerol, was produced for laboratory purpose (J.T. Baker, purity: min. 99.5%), was injected regularly during the fermentation trials. It was decided not to use glycerol from biodiesel production plants with a varying content of water and impurities, in order to keep these factors neutral.

Fermentation Trials

The anaerobic digestion process was carried out in three 5 L bioreactors (Simax, Sázava, Czech Republic) each with a working volume 4 L. Each of the fermenters could be connected to the on-line recurrent loop system, thereby enabling on-line analysis with the TENIRS measuring system, described in section: on-line NIR. A schematic view of the fermenter setup is shown in Figure 1. Stirring and heating were monitored and maintained by a program written in LabVIEW 8.0 (National Instruments, Austin, TX) software. Agitation was set to 30 rpm and the process was operated under stable thermophillic conditions, 53°C.

The trial lasted 33 days. The AD process worked initially for 13 days with systematic addition of standard feedstock in order to achieve a stable biogas production. The first measurements were carried out after this period; day number 1 thus corresponds to the first day of taking samples from the reactor. The reactor was operated in a semi-continuous mode – two times each day sampling and feeding with standard feedstock was carried out. The amount of substituted biomass was equal to 150 mL at each feeding time. The hydraulic retention time (HRT) was at 13.5 days. After 20 days the first portion of glycerol, mixed with the standard feedstock, was added. Table I shows the planned amount of glycerol and the theoretical concentration development in the fermenters. The purpose of increasing the addition of glycerol was first of all to increase the methane production, and later to provoke organic overloading in which the introduced amount of substrate was greater than consumed by the microorganisms. This would lead to a process imbalance and subsequently a final collapse of the fermentation process.

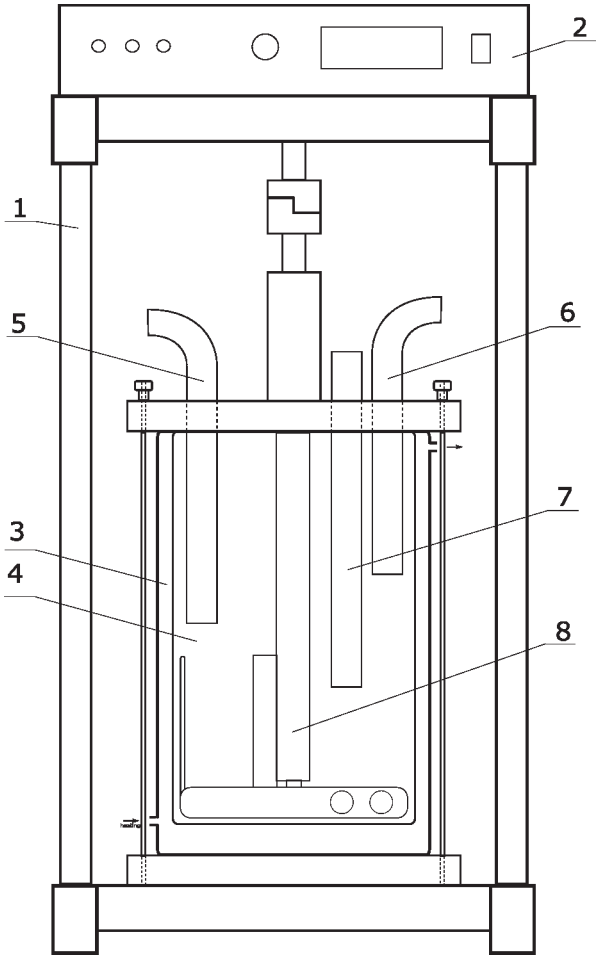

Figure 1. Fermentation setup: (1) holding frame; (2) stirring controller; (3) heating jacket with water; (4) 5 L glass fermenter; (5) outlet to TENIRS system; (6) inlet from TENIRS system; (7) feeding probe; (8) stirrer.

On-Line NIR

The Transflexive Embedded Near InfraRed Sensor (TENIRS) system is a prospective at- and on-line measurement system utilized in flowing heterogeneous bioslurry systems It was used in this trial as a semi on-line system, as it was moved from one fermenter to another during the trials. The TENIRS facility was developed by the Institute of Agricultural Process Engineering (ILV), at the University of Kiel for at-line measurements of bio-slurry and manure applications (Andree et al., 2005) to document that NIR spectroscopy can be a highly valuable tool for at- or on-line analysis of heterogeneous bioslurries such as pig and cattle slurry.

In this trial, the TENIRS loop was connected to each one of the three simultaneously running fermenters by the time. The loop was attached to the fermenter for approximately 30 min for each NIR measurements two times per day. During each NIR measurements, representative sampling of biomass were simultaneously performed for chemical

DOI 10.1002/bit

Table I. Changes of glycerol concentration during conducted trials.

Day of trials	Amount of glycerol added to each fermenter (mL/day)	Theoretical change in concentration level in the fermenter (V/V%)
1–7	0	—
8–12	50	0.0 → 1.0
13–15	114	1.0 → 3.5
16–19	188	3.5 → 6.5

analysis. The *semi on-line attachment* was chosen because only one prototype of the TENIRS equipment and the sampling construction existed.

The loop was connected to the centre part of the fermenter by individual outlet and inlet pipes (see Figs. 1 and 2). From the outlet pipe the biomass was pumped through the loop by an integrated impeller pump and through the flow through cell for transflexive NIR spectroscopic measurements, then the loop continued passing through a special sampling device, see section: sampling, and finally back to the fermenter by the inlet pipe diametrically opposite the outlet pipe. The total length of the external piping loop was 230 cm with a pipe diameter of 2 cm. The loop was kept thermophilic, at constant 53°C, throughout the trial, by using insulation materials around the pipeline in the loop. In the flow through cell the height was adjusted to 3 mm, enabling NIR measurements where the cell width expanded proportionally obtaining a constant pressure and flow in the pipe loop system. NIR data handling was organized with the software Aspect Plus, which allows communicating with the near infrared spectrophotometer Zeiss CORONA 45 NIR, with a scan range from 960 to 1,600 nm.

Sampling

To obtain primary samples in agreement with the TOS; theory of sampling (Gy, 1998), a pilot device facilitating sampling from the on-line stream flowing through the TENIRS loop was developed and implemented directly after the TENIRS measuring cell (Fig. 2). By applying this configuration the samples were obtained from a one dimensional pipeline flow of the biomass compared to the full three

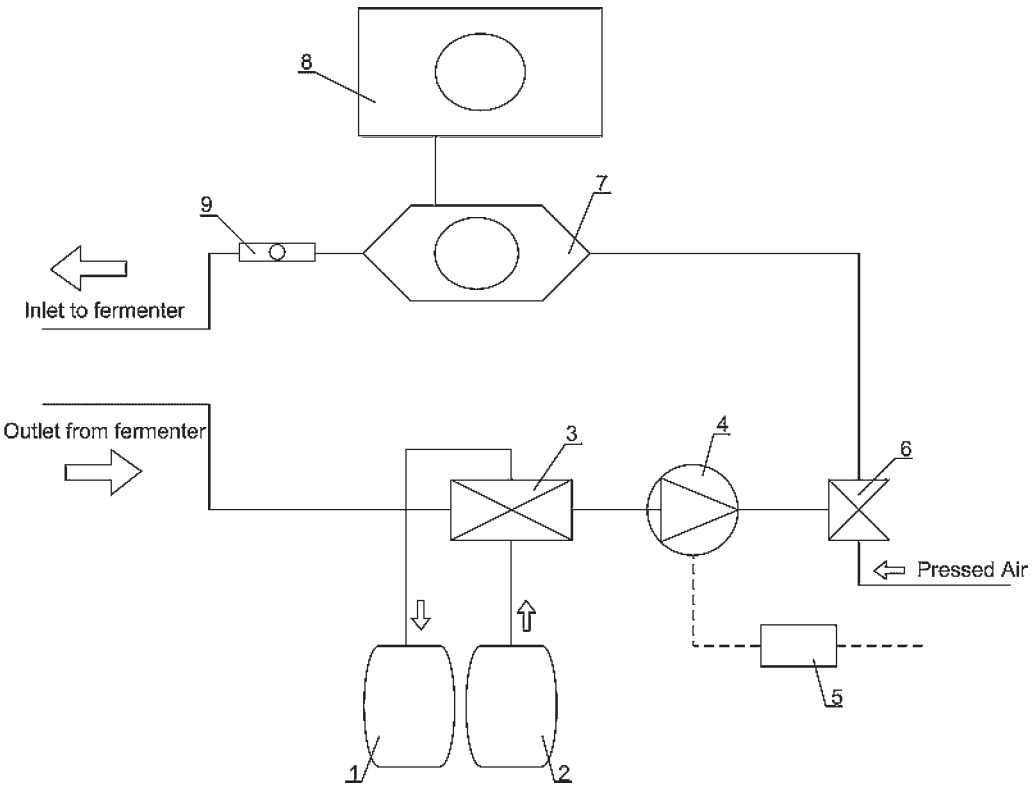

Figure 2. Recurrent measuring loop connected to fermenters: (1) outlet cleaning water; (2) inlet cleaning water; (3) multiway valve; (4) impeller pump; (5) frequency controller; (6) air valve, for drying after cleaning; (7) NIR flow-through cell; (8) Zeiss Corona NIR instrument; (9) sampling device.

dimensional section of the fermenter volume. The device (Fig. 3) consisted of a 10 mL bottle placed so, when a stainless steel disc was moved from side to side an incremental sample from the loop was sampled. Ten incremental samples were taken resulting in a primary 100 mL bio-slurry sample, insuring a composite sample (Gy, 1998; Mortensen, 2006; Pitard, 1993; Petersen, 2005).

This configuration was not completely in accordance with TOS as a total cross section of the flow in the pipeline was not completely ensured. This can only be obtained from an up-stream vertically flowing piping system. As a result the samples were slightly biased. However, this is a step in the direction for the development of a closed recurrent fermenter loop system.

Chemical Analysis

Total Solids/Volatile Solids

Total solids are a fraction of the total wet weight of a sample from which the water has been evaporated in an oven at 105°C for a period of 24 h. Total volatile solids were determined as the difference between total solids and the weight of the ashes from the sample kept in the oven for 2 h at 550°C.

Glycerol Concentration

Glycerol concentration was determined using Ion Chromatography (Dionex Series 4500i with microinjection valve at 100–120 psi) with a CarboPac PA10 (4 × 250 mm) column. All samples were diluted in order to ensure a concentration in the area of 20–100 mg glycerol per litre and carefully filtered before injecting the 0.2–0.3 mL sample in the loop. Two eluents with a concentration equal to 140 and 200 mM NaOH were used.

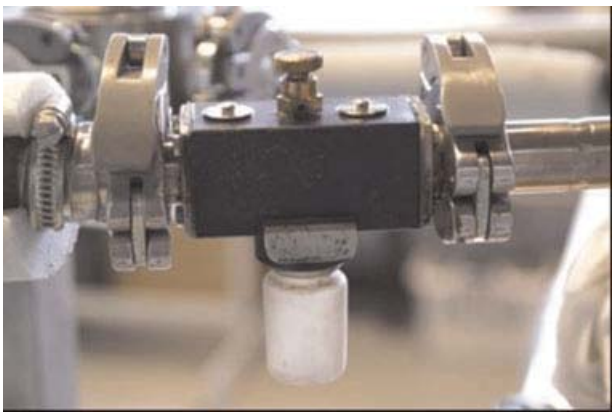

Figure 3. Prototype sampling device, 10 increments each of 10 mL were sampled during a period of 10 min. [Color figure can be seen in the online version of this article, available at www.interscience. wiley.com.]

Volatile Fatty Acids (VFA)

For all samples volatile fatty acids content was measured employing a GC/Varian 3800 chromatograph with a Varian 25 m 0.32 mm ID (Cat. No. CP 7488 WCOP fused silica coating FFAP-CB) column. The carrier gas had a constant heat pressure of 6 p.s.i., 2 mL of each sample was centrifuged, 200 μL HCl 1 M and 2.0 mL Internal Standard solution: 2,2-dimethylbutanoic were added. Samples were separated in a Mediafuge at 5,000 rpm for 5–8 min. The separated sample was filtered through a 45 μm acetate filter and injected into the GC. This method determined the individual concentrations of acetic, propanoic, isobutanoic, butanoic, isovaleric and valeric acid, and thereby also the accumulated total-VFA concentration.

Data Processing

Chemometric data analysis gives an overview of the state of the chemical and/or biological processes based on analytical measurements. The idea of chemometrics is to let the process or the data structures unfold their relations themselves (Mortensen, 2006). From chemometric data analysis it is possible to decrease the required number of variables in order to describe the investigated phenomena.

In *multivariate calibration and validation* a large number of *X*-variables are used to correlate to a given reference (chemical or physical) variable (Y), for example, X: NIR spectra; Y: chemical reference variables. Partial Least Squares (PLS) regression was used with the scope to predict future Y-data directly from X-data. PLS-Regression (PLS-R) consists of three stages: model training—calibration, validation, and prediction (Esbensen, 2001; Martens and Næs, 1991).

A test set validation is a procedure for validation of a model by a completely independent data set—the test set. This is the strongest validation of any model, which was a real advantage in this study where trials were performed in three parallel independent fermenter systems. The results from the test set validation are much more reliable than any internally re-sampling of the training data set alone, such as cross validation.

Data obtained from the TENIRS system were modelled using the UNSCRAMBLER software (CAMO, ver. 9.5). Multivariate calibrations using Partial Least Square regression (PLS-1) were performed for quantitative determination of glycerol and total-VFA as well as for the individual acids as acetic, propionic, iso-butyric, butyric, iso-valeric, and valeric acids. For each chemical component a separate model was set up. In all cases the models were evaluated by a comprehensive test set validation applying fermentation No. 1 and 3 as the calibration set respectively and No. 2 as the test set. The analysis includes best models obtained by the lowest numbers of principal components (PC), and by carefully removal of a minimum of outliers in the models (Esbensen, 2001).

DOI 10.1002/bit

Results and Discussion

The anaerobic digestion trial was performed in triplicate, lasting for 33 days; during the first 13 the process was stabilized. These days were excluded from further analysis, not included in the glycerol spiking trials. Sampling and chemical analysis was performed twice a day for the rest of the trial period (days 1–19). Each sample was analyzed for glycerol, VFA, and TS/VS content. Additionally, NIR spectra of the flowing biomass were acquired for the PLS-1 models (X). All results are presented in the sections below.

Chemical Analysis

The changes in the glycerol concentration in the bioreactor in the full experiment are presented in Figure 4. In the beginning of the glycerol addition, no accumulation was observed, which indicated that the microorganisms were able to degrade it to biogas. The increase in the feeding level at day 12 however, resulted in a slow accumulation, which became more pronounced when the feeding level was increased further at day 16. The overall accumulation of glycerol from the 16th to the 19th day of the trial, was from approximately 5 to more than 30 g L^{-1}, indicating that a severe organic overload might be taking place. The organic overload was coursed by a bottleneck in the system, where the slowest process in the degradation chain, was determining the overall speed of the process, consequently the substrate was accumulated in the reactor.

The overall tendency in the VFA concentrations in the fermenter increased simultaneously and even faster than the increase in the glycerol concentration (Fig. 5), thus indicating that the organic overload taking place was due to an inhibition of the methanogenic step. A slow accumulation of the VFAs was observed already at the 8th day. After the 12th day the VFA-acids content grew rapidly achieving the highest level: 36 g L^{-1} in fermenter no. 2. From days 12 to 16 organic overloading occurred, and after day 16 the process was strongly imbalanced and no biogas production was noticed. The acetic acid concentration, seen in Figure 6, was the most important contributor to the total VFA concentration, which means that exactly the same tendencies were seen in the acetic acid concentration behaviour during the trial (Fig. 6).

When VFA concentration significantly exceeds 5.0 g L^{-1}, the AD process is no longer stable and organic overloading is likely (Amon et al., 2006). Due to medium to high daily doses of glycerol since day 13, increasing VFA contents continued to imbalance the process ending by complete collapse. This can be seen when VFA concentration exceed 10–12 g L^{-1} (Fig. 5). Exactly the same tendencies can be seen in the acetic acid concentration development (Fig. 6).

Contemplating both the glycerol and the VFA and acetic acid results, it was seen that the VFA and acetic acid concentration increased earlier, compared to the concentration of glycerol, thus giving an early warning of the coming process instabilities. For this reason the VFA can be applied in the context of process monitoring and controlling.

According to Angelidaki et al. (2002) the VFA concentrations in biogas production do not necessarily have to be directly toxic to the process, but it indicates imbalance of

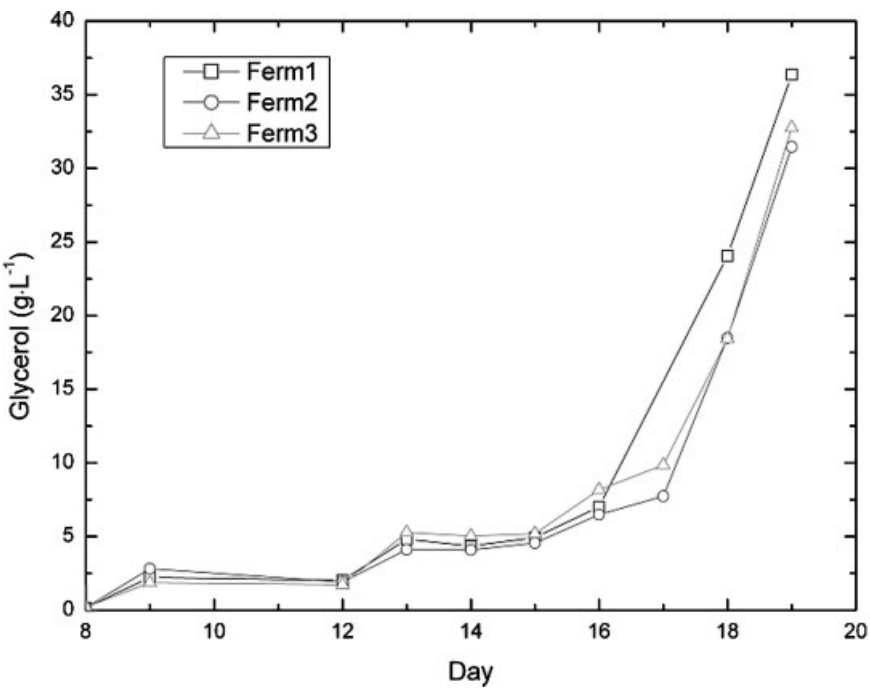

Figure 4. Glycerol development during the AD process trials, shown for each individual fermenter.

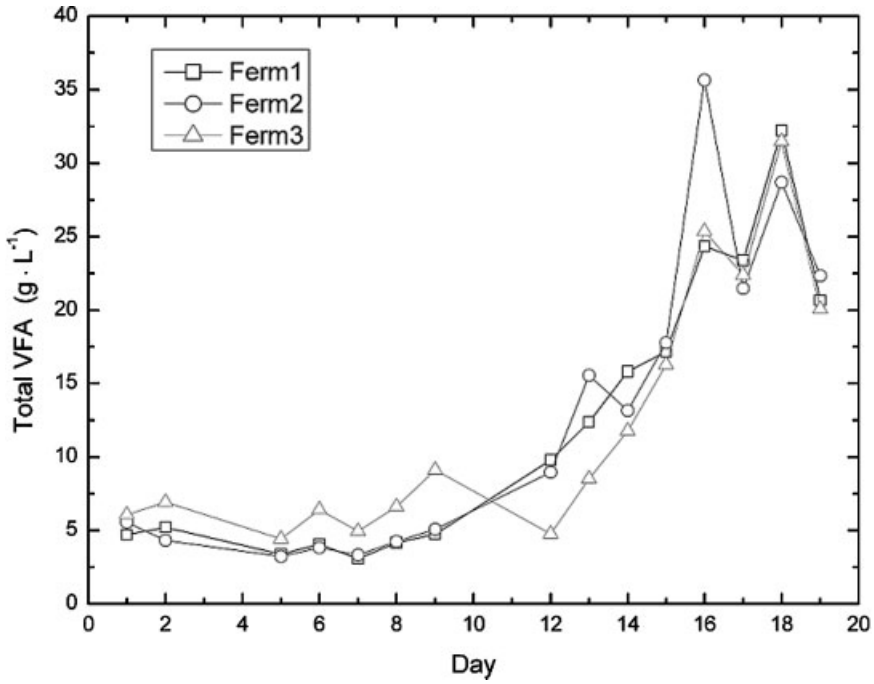

Figure 5. Volatile fatty acids contents. Development for each fermenter.

the process. Therefore, the increase of VFA is usually the result of process imbalances, not the reason, but can never-the-less give an important early warning of imbalances in progress.

Glycerol is a very promising feedstock for increasing biogas yields when the concentration does not exceed 5–7 g/L inside the digesters. Above this concentration level organic overloading occurs and methane is produced at a significantly

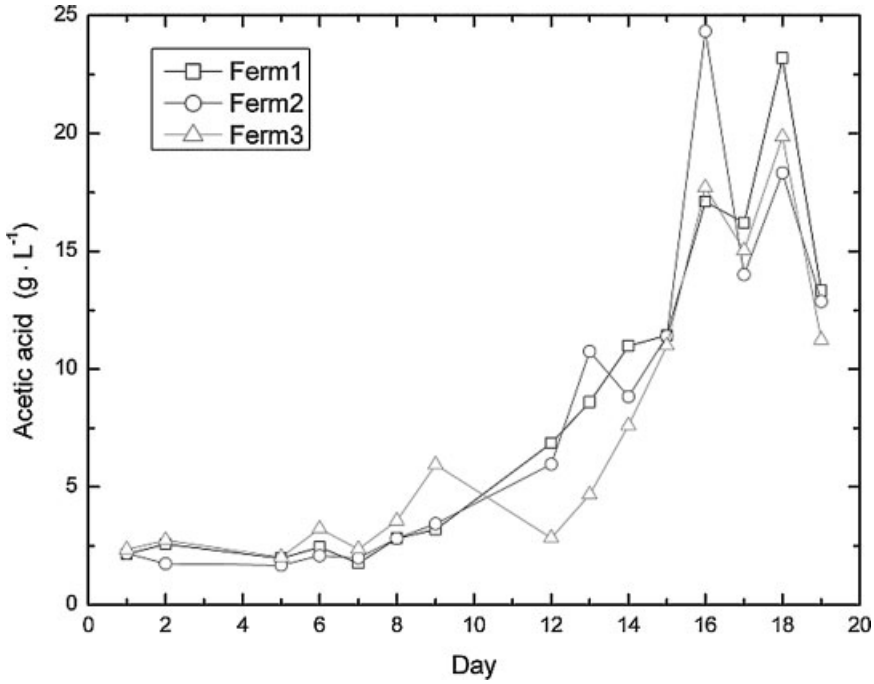

Figure 6. Acetic acid concentration changes during the investigated trial.

DOI 10.1002/bit

lower rate or could even be stopped. A longer adaptation period for the microorganisms may sometimes allow a higher feeding rate of glycerol in the anaerobic digestion process. In a full scale mesophilic biogas plant operation, the Hashoej biogas plant in Denmark, up to 9 vol.% glycerol has been registered for the feedstock blend, producing very high biogas yields.

At thermophilic conditions at 53°C ± 3 vol.% has been registered as a glycerol concentration which is fairly easy to manage. Above this concentration level a decrease in biogas production can occur or other measures shows organic overloading by increasing the VFA concentration during the fermentation. Similar results have been documented by Amon, (2006), indicating that a significant increase in the methane yield can be achieved with the addition of up to 6% of glycerol. However, further addition inhibits the process. This study has indicated fully that such a feeding situation can occur if no on-line process measures or proactive feeding management plan has been set in operation.

Several volatile acids were monitored in the present study. VFA changes constitute the most important indicators of early warning of process imbalances. High concentration is a sign of severe stress of the microorganisms and most likely organic overloading. An indication of VFA imbalances started between days 12 and 13 (Fig. 5). Severe organic overloading occurred on the 16th day of the trial when the concentration exceeded approximately 11.5 g/L and 18.0 g/L of acetic acid and total VFA, respectively. In the present study the organic overloading was however caused by purpose in order to achieve as wide a span as possible of acid concentrations allowing for PLS-1 models with high validations in all relevant ranges of volatile fatty acid concentrations.

Multivariate Calibration and Validation of On-Line NIR spectra

A PLS-1 model of glycerol concentration in the fermentation bio slurry, measured two times daily during the fermentation trial, showed very good performance (Fig. 7). The linear regression relationship needed primarily two PLS components to model the glycerol *Y*-variables measured at the Ion Chromatography equipment from the *X*-variables (NIR).

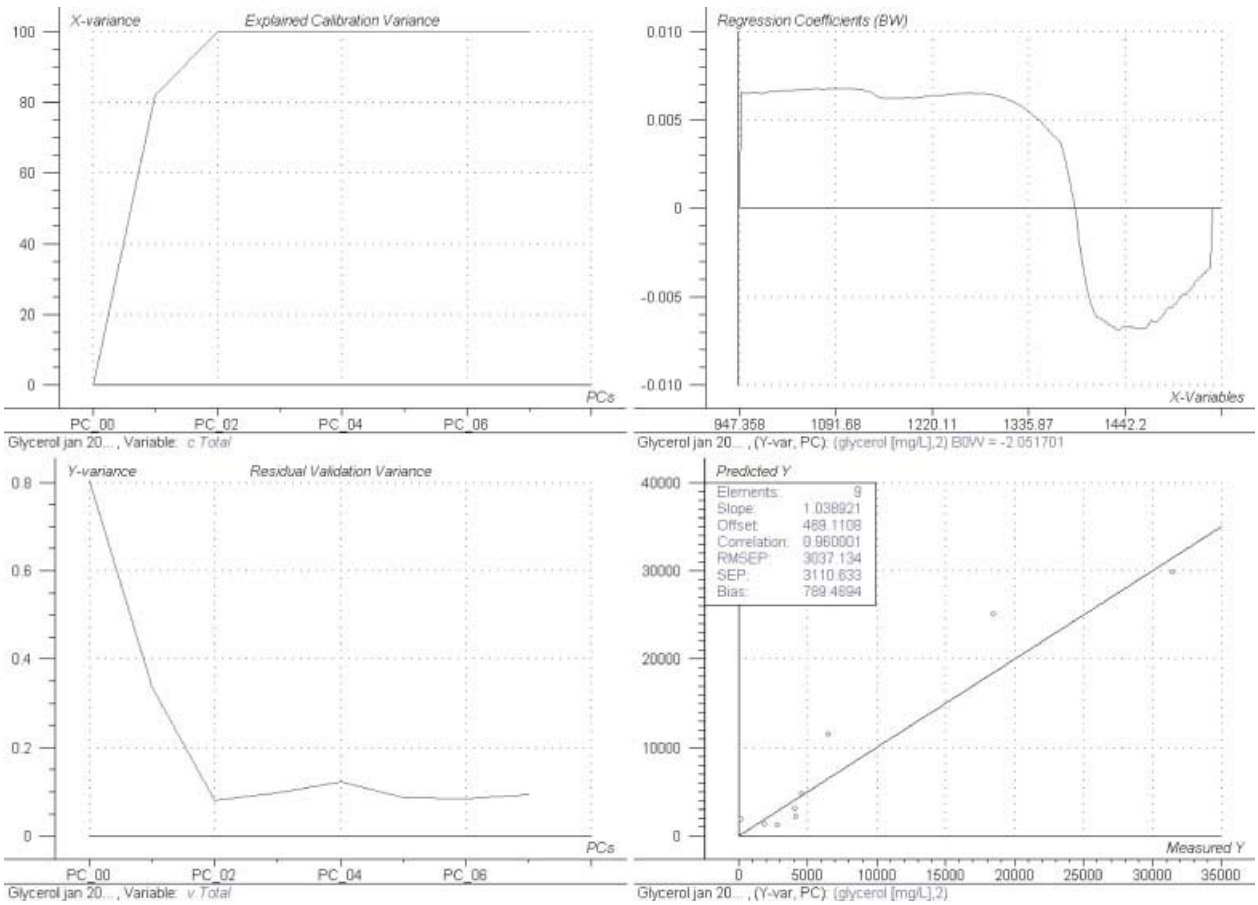

Figure 7. Glycerol PLS-1 model; number of required PLS components = 2; One outlier was removed; Test set validation was made from data obtained in fermenter No.2 and tested against data from fermenter No.1 and No.3; Measures of precisions $r^2 = 0.96$ and slope 1.04. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

Table II. Overview of PLS-1 models for volatile fatty acids; root square error of prediction (RMSEP).

	Acetic	Propanoic	Iso-butanoic	Butanoic	Iso-valeric	Valeric	Total VFA
No. of PCs	3	1	4	5	6	3	4
Explained validation variance (%)	93.1	54.8	89.3	92.7	92.8	86.3	94.0
No. of outliers	2	1	4	4	2	4	2
RMSEP	1,476	1,364	26.46	271.9	55.83	24.31	2,095
RMSEP/mean measurement level (%)	14.8	32.2	12.1	22.2	19.2	13.1	12.9
Correlation (r)	0.98	0.37	0.95	0.74	0.97	0.93	0.98

Definitions according to Esbensen (2001): PCs—are composite variables, that is, linear functions of the original variables, estimated to contain, in decreasing order, the main structured information in the data; Outlier—an observation (outlying sample) or variable (outlying variable) which is abnormal compared to the major part of the data; RMSEP—root mean square error of prediction—a measurement of the average difference between predicted and measured response values, at the prediction or validation stage; correlation—a unitless measure of the amount of linear relationship between two variables; the correlation is computed as the square root of the covariance between the two variables divided by the product of their variances. It varies from −1 to +1.

All the fatty acids were modelled by PLS-1 analysis in Unscrambler like the above model of glycerol. The results are summarized in Table II. As examples two models are shown below; acetic acid (Fig. 8), as a main contributor of the total VFA, and iso-butanoic acid (Fig. 9) as a minor fatty acid analogue, presumably more difficult to model due to its very low concentration levels. Finally a model of total VFA was shown (Fig. 10). This was done to document the strength in the application possibilities of NIR measurements.

The PLS-1 prediction model for acetic acid showed that the validation fermenter No.2 provided a realistic validation basis. The linear regression relationship needed 3 PLS components to model the Y-variable: acetic acid concentration analysed at the HPLC, from the X-variables (NIR). The overall model for acetic acid showed a very strong

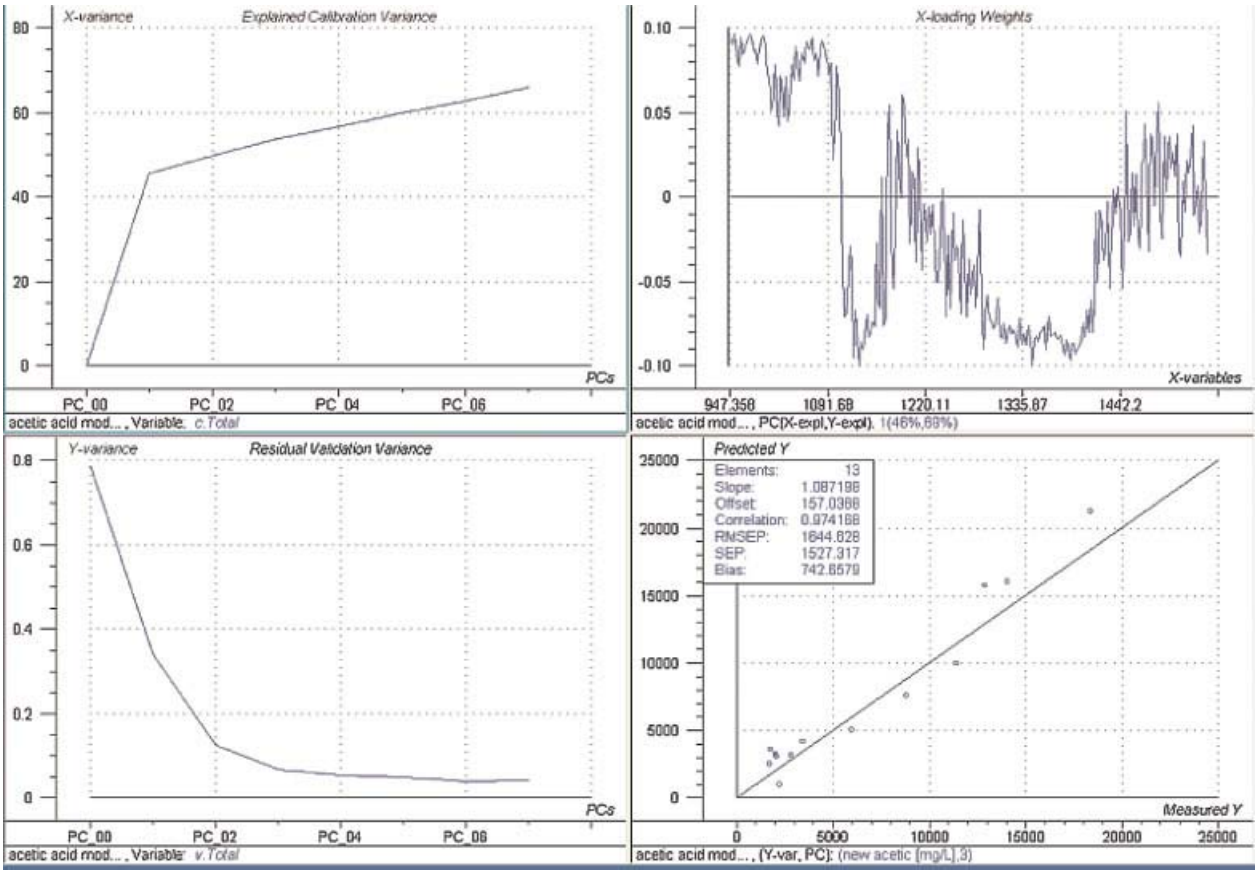

Figure 8. Acetic acid PLS-1 model; number of PLS-components required = 3; Two outliers were removed; Test set validation where data from the fermenter No.2 were tested against calibration data from fermenter No.1 and No. 3; measures of precision: $r^2 = 0.95$, and accuracy; slope = 1.09. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

DOI 10.1002/bit

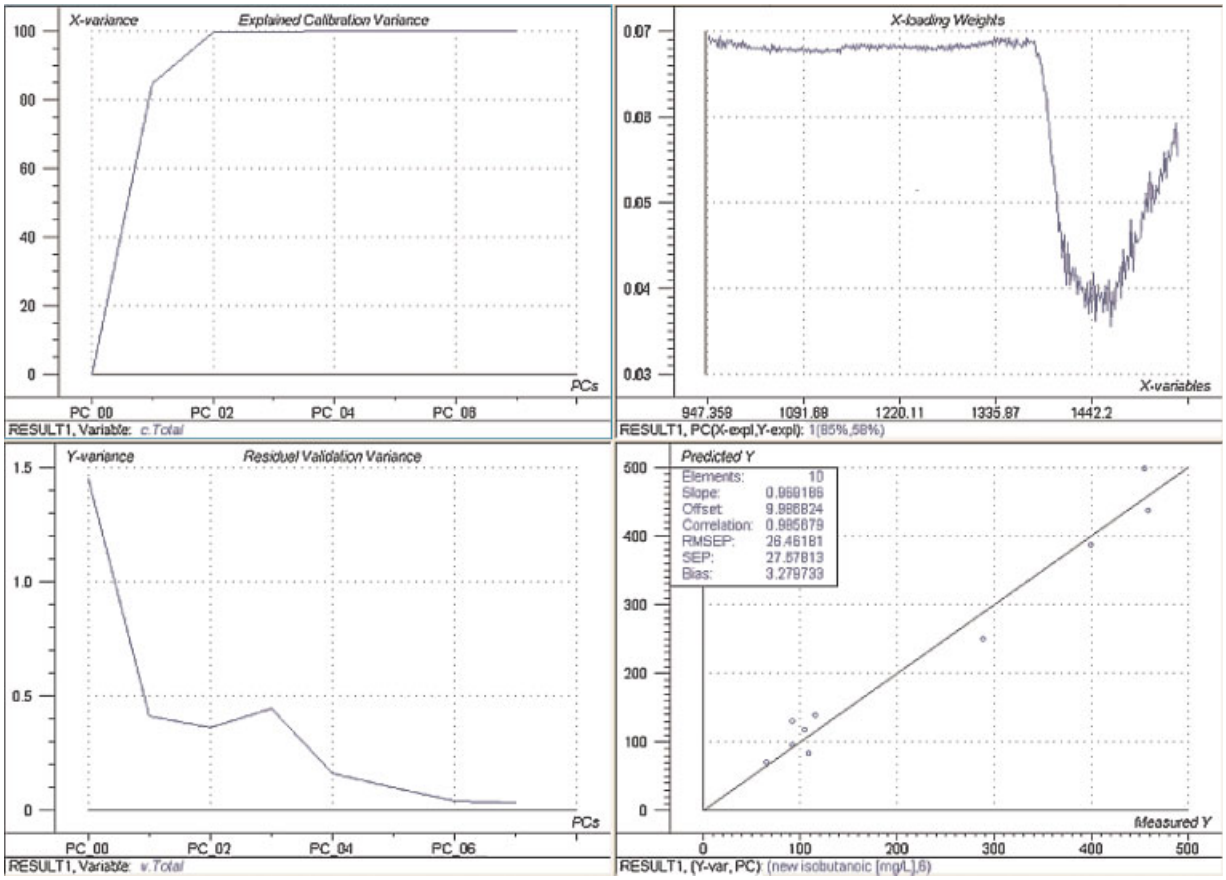

Figure 9. Iso-butanoic acid PLS-1 model; number of required PLS components = 6; Four outliers were removed; Test set validation of data from the fermenter no. 2; measures of precision: $r^2 = 0.97$ and accuracy; slope = 0.97. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

correlation between the HPLC measurements and the on-line NIR measurements. As acetic acid was the main contributor to the total VFA, the model obtained for total VFA showed, not surprisingly, the same trends with respect to the loading weight plot and the calibration and validation variance as the model for acetic acid (Table II).

The PLS-1 prediction model for iso-butanoic acid in the three bio-slurry fermenters spiked with glycerol showed a good validation model. For fermenter No.2 test set validation showed good statistic precisions; $r^2 = 0.97$ and the accuracy of predicting iso-butanoic acid by NIR had a slope of 0.97.

The PLS-1 prediction model for total VFA in the three bio-slurry fermenters spiked with glycerol provided a very good validation model. Test set validation showed good statistic precisions, see Figure 10. The on-line trials have been useful to document the possibilities to measure total VFA conditions during increased feeding of concentrated feedstock's like the glycerol in the anaerobic digestion systems to optimize the biogas production.

Table II gives an overview of all PLS-1 models computed for the main compounds of volatile fatty acids. The statistics show that the weakest model was developed for propanoic acid which was not acceptably, while all other VFA could be statistically modelled, *Y*-variance modelled span 86–94%.

Biogas production yields in volume and quality will always be some of the most important parameters to indicate overall process and reactor performance, but one weakness is that these parameters cannot give an early warning and indication of the stress status of the fermentation behaviour. The future combination with on-line monitoring of the VFA and other important intermediates like ammonium will provide extremely valuable information for improved process control as is documented in this study.

Conclusion

Due to high growth rates in the biodiesel production capacities in Europe and world-wide, glycerol is increasing in amounts year by year. The availability of glycerol for many industrial purposes is increasing. One of the sectors that have had increasing production and utilisation capacity growth rates is the European biogas sector. The co-digestion

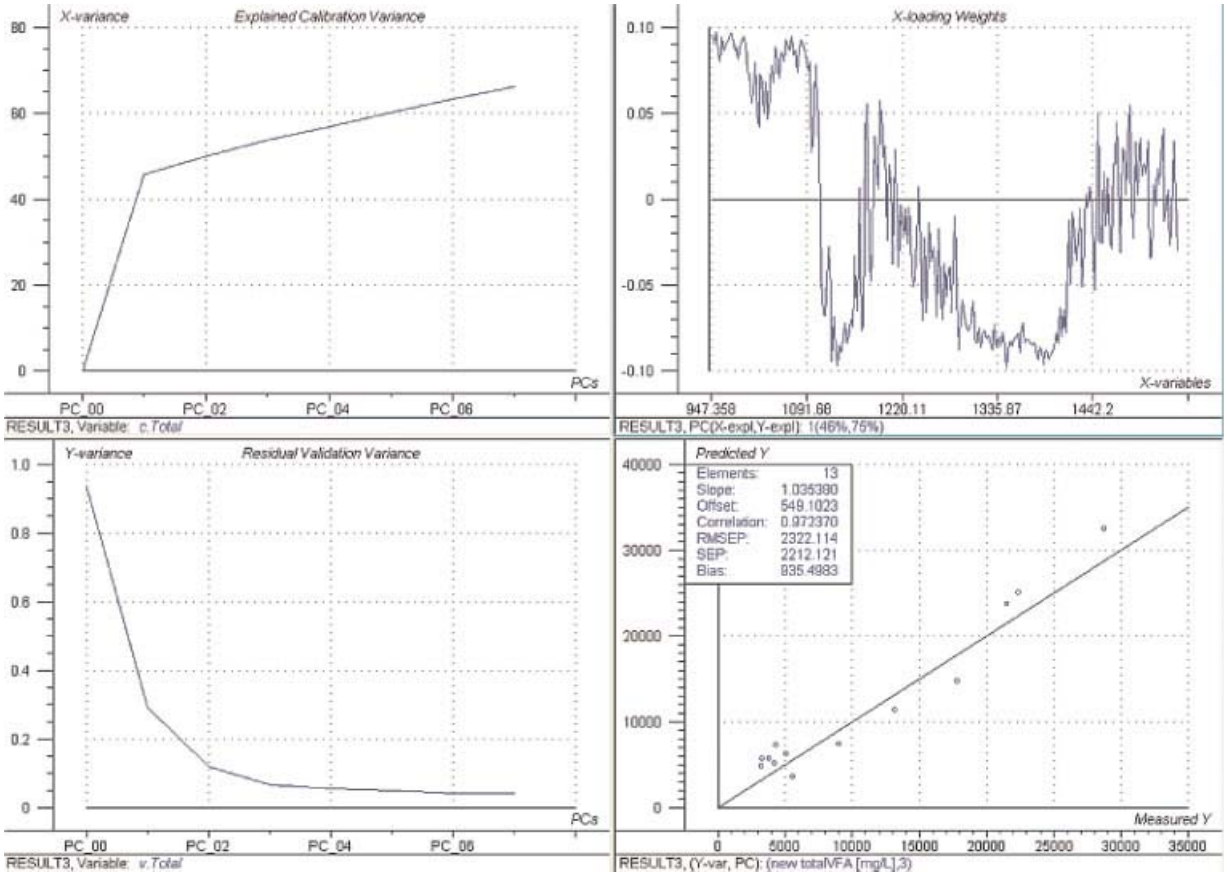

Figure 10. Total VFA PLS-1 model; number of required PLS-components = 3; Two outliers were removed; Test set validation where data from the fermenter No. 2 were tested against calibration data from fermenter No. 1 and No. 3; measures of precision: $r^2 = 0.98$, and accuracy; slope = 1.03. [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

of high yielding glycerol with a range of other feedstock's like manure and energy crops are rising in importance.

Spiking trials have shown that glycerol at concentration levels of 3–5 g L^{-1} makes the anaerobic digestion process run at continuously stable conditions. On-line measurement of the fermentation process has documented that with low concentrations of glycerol the VFA and the individual fatty acids show no sign of organic overloading. However, when rapidly increasing the content of glycerol, there are clear tendencies of organic overloading.

The results concluded a good correlation between on-line NIR measurement of glycerol and the VFA content in the anaerobic digestion process and analytical laboratory results. On-line process monitoring control is becoming more and more important when anaerobic digestion processes are spiked with concentrated feedstock, like glycerol. Analytical tools as applied in these experiments will bring forward the needed information of early warning about process imbalances.

The investment in the fermenters was made possible by the Esbjerg Seminarium Foundation and equal match funding by Aalborg University.

References

Ahring BK, Sandberg M, Angelidaki I. 1995. Volatile fatty acids as indicators of process imbalance in anaerobic digestors. Appl Microbiol Biotechnol 43:559–565.

Amon T, Amon B, Kryvoruchko V, Bodiroza V, Potsch E, Zollitsch W. 2006. Optimising methane yield from anaerobic digestion of manure: Effects of diary systems and of glycerine supplementation. Int Congr Ser 1293:217–220.

Andree H, Dolud M, Hügle T. 2005. TENIRS—NIR analysis of heterogeneous bioslurries. In: Madsen M et al., editors. Baltic Biorefinery Symposium, proceedings, p 141–148.

Angelidaki I, Ellegaard L, Sorensen AH, Schmidt JE. 2002. Anaerobic processes. In: Angelidaki I, editor. Environmental biotechnology. Environment & Resources. The Technical University of Denmark. p 1–114.

Angelidaki I, Ellegaard L, Ahring BK. 2003. Applications of the anaerobic digestion process. In: Ahring BK, editor. Biomethanisation II. Berlin, Germany: Springer. p 1–33.

Bakeev KA, editor. 2005. Process analytical technologies. Oxford: Blackwell Publishing. p 1–11.

Boe K. 2006. On-line monitoring and control of the biogas process. Institute of Environment and Resources, DTU. Ph.D. Thesis.

Bournay L, Casanave D, delfort B, Hillion G, Chodoroge JA. 2005. New heterogeneous process for biodiesel production: A way to improve the quality and the value of the crude glycerine produced by biodiesel plants. Catal Today 106:190–192.

DOI 10.1002/bit

Diamantis V, Melidis P, Aivasidis A. 2006. Continuous determination of volatile products in anaerobic fermenters by on-line capillary gas chromatography. Anal Chim Acta 573–574:189–194.

Esbensen KH. 2001. Multivariate data analysis—In practice. An introduction to multivariate data analysis and experimental design. Aalborg University Esbjerg. Camo Process AS, Norway.

Fernandes EN, Campos Moura MN, Lima JLFC, Reia BF. 2004. Automatic flow procedure for the determination of glycerol in wine using enzymatic reaction and spectrophotometry. Microchem J 77:107–112.

Gy P. 1998. Sampling for analytical purposes. Chichester: John Wiley & Sons.

Hill DT, Holmberg RD. 1988. Longchain volatile fatty acid relationships in anaerobic digestion of swine waste. Biol Waste 23:195–214.

Mandil C, editor. 2004. Biofuels for transport an international perspective. Chirat, France: International Energy Agency. p 25–47.

Macaloney G, Halls JW, Rollins MJ, Draper I, Anderson KB, Preston J, Thompson BB, McNeil B. 1997. The utility and performance of near-infrared spectroscopy in simultaneous monitoring of multiple components in a high cell density recombinant *Escherichia coli* production process. Bioprocess Eng 17:157–167.

Martens H, Næs T. 1991. Multivariate Calibration. Updated version. Chichester, UK: John Wiley and Sons.

Mladenovska Z, Dabrowski S, Ahring BK. 2003. Anaerobic digestion of manure and mixture of manure with lipids: Biogas reactor performance and microbial community analysis. Water Sci Technol 48:271–278.

Mortensen PP. 2006. Process Analytical Chemistry—Opportunities and Problems for Bio-industrial Implementation. Aalborg University Esbjerg. Ph.D. Thesis.

Nielsen LH, Hjort-Gregersen K, Thygesen P, Christensen J. 2002. Socio-economic Analysis of Cetralised Biogas Plants. Report no. 136. Danish Ministry of Food Agriculture and Fisheries.

Petersen L2005. Pierre Gy's Theory of Sampling (TOS) in Practice: Laboratory and Industrial Didactics. Aalborg University Esbjerg. Ph.D. Thesis.

Pitard FF. 1993. Pierre Gy's sampling theory and sampling practice, 2nd edition. Boca Raton, FL: CRC Press Ltd.

Vanrolleghem PA, Lee DS. 2003. On-line monitoring equipment for wastewater treatment processes: State of art. Water Sci Technol 47(2): 1–34.

II

Bioresource Technology 100 (2009) 1711–1719

Contents lists available at ScienceDirect

Bioresource Technology

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

Near infrared and acoustic chemometrics monitoring of volatile fatty acids and dry matter during co-digestion of manure and maize silage

Carina J. Lomborg [a,*], Jens Bo Holm-Nielsen [a], Piotr Oleskowicz-Popiel [b], Kim H. Esbensen [a]

[a] Applied Chemometrics, Analytical Chemistry, Acoustic Chemometrics, Applied Biotechnology, Bioenergy and Sampling (ACABS) Research Group, Aalborg University, Niels Bohrs Vej 8, DK-6700 Esbjerg, Denmark

[b] Biosystems Department, Risø DTU – National Laboratory for Sustainable Energy, Frederiksborgvej 399, DK-4000 Roskilde, Denmark

ARTICLE INFO

Article history:
Received 19 June 2008
Received in revised form 17 September 2008
Accepted 22 September 2008
Available online 8 November 2008

Keywords:
Near infrared spectroscopy
Acoustic chemometrics
Co-digestion
Maize silage
Biogas

ABSTRACT

In this study, two process analytical technologies, near infrared spectroscopy and acoustic chemometrics, were investigated as means of monitoring a maize silage spiked biogas process. A reactor recirculation loop which enables sampling concomitant with on-line near infrared characterisation was applied. Near infrared models resulted in multivariate models for total and volatile solids with ratio of standard error of performance to standard deviation (RPD) values of 5 and 5.1, indicating good on-line monitoring prospects. The volatile fatty acid models had slopes between 0.83 and 0.92 (good accuracy) and RPD between 2.8 and 3.6 (acceptable precision). A second experiment employed at-line monitoring with both near infrared spectroscopy and acoustic chemometrics. A larger calibration span was obtained for total solids by spiking. Both process analytical modalities were validated with respect to the total solids prediction. The near infrared model had an RPD equal to 5.7, while the acoustic chemometrics model resulted in a RPD of 2.6.

1. Introduction

Biomass has been used as a local low-tech energy source (fire, combustion) for thousands of years, but production of modern industrially produced biomass energy will gain a significant importance both locally as well as regionally in the near future. This increase can be facilitated because the cost of the conversion and production technologies is constantly being reduced, but more importantly because of increased societal and climate awareness concerning global warming and the pressing need for CO_2 neutral technologies. Several such predictions and calculations have been presented in Yamamoto et al. (1999), Sims (2001), Bull (2001). Hoogwijk et al. (2005) conducted a comprehensive simulation study on the geographical global energy crop potential. It was calculated that in the year 2050 the potential for abandoned agricultural land will be between 130 and 410 EJ/yr, and 35 and 245 EJ/yr for partially available area such as savannahs, shrublands or grasslands, depending on the different scenarios employed.

To significantly increase the energy supply from biomass, so-called advanced biofuels must be developed from dedicated energy crops (Koonin, 2006). In the present study, maize crop in the form of maize silage was used in a conventional biomass conversion scheme. Maize silage is the result of anaerobic ensilation of the whole maize crop (cob, stem, and leaves). The maize crop is harvested when the total solid (TS) content is around 30–33%. If the TS is below 30%, the harvested crops can be air dried for a few days. This low-technology pre-treatment usually increases the TS content between 30% and 35%. Afterwards, the whole maize crop is cut into small pieces, and degraded naturally by lactic acid bacteria under anaerobic conditions. The resulting silage has a very high biogas potential, and can be stored all year long, which is an important local logistic advantage.

The future energy market for biomass depends on different factors, such as biomass cost and availability, and the efficiency of the adopted conversion technologies, but it also depends on the availability of comprehensive, reliable, non-invasive, real-time process monitoring and control technologies. Even a slight variance in the operating conditions during fermentation can decrease the biogas production significantly. The parameters most commonly used for on-line monitoring of fermentations today are also the simplest to assess: acidity (pH), temperature, and headspace measurements of different gases, specifically methane, hydrogen sulphide, and ammonia. Application of these on-line analytical modalities will not detect process failure early enough however, and neither can they detect specific process imbalances. On the other hand, application of traditional analytical off-line methods like liquid chromatography (HPLC), gas chromatography (GC), and flow injection analysis (FIA) are difficult, because sample preparation is time consuming; moreover expensive equipment, maintenance and the

* Corresponding author. Tel.: +45 99407702.
E-mail address: cjp@aaue.dk (C.J. Lomborg).

0960-8524/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.biortech.2008.09.043

need for skilled technicians make them even less attractive in the bulk processing regimen of biomass conversion.

The situation appears different for the emerging process analytical technologies (PAT). Several experiments employing diverse sensors for anaerobic digestion process monitoring are currently under investigation, e.g., on-line continuous determination of volatile compounds by GC (Diamantis et al., 2006; Boe, 2006) or multi-wavelength fluorometry (Morel et al., 2004). A summary of the different methods and equipment available for anaerobic digestion is presented in Vanrolleghem and Lee (2003). Among all technologies, spectroscopic tools seem very favourable since these do not require chemicals to be added to the samples. More importantly, their assessment of the state of the process is non-invasive. The maintenance need for the multivariate calibration models is also less than other methods. The model however has to be updated when new feeds that have an influence on the spectrum, are added to the system. For on-line and in-line applications however a potential problem regarding sensor fouling due to microbial growth on the optics, is a disadvantage.

Of the spectroscopic methods, near infrared spectroscopy (NIR) for rapid and non-invasive, quantitative and qualitative analysis together with chemometric tools has shown a great potential for biological process control in the food, chemical, and pharmaceutical industries (Tosi et al., 2003). Additionally, it has been found applicable to monitor complex mixtures – one of which is fermentation broth from anaerobic digestion processes (Spanjers et al., 2006) and monitoring of the anaerobic digestion of a mixture of cellulose, albumin, and minerals (Nordberg et al., 2000).

A method which is less investigated but which may also have a great potential in monitoring these processes is acoustic chemometrics (a.c.): on-line signal acquisition via an acoustic sensor (piezoelectric accelerometer) placed strategically in the process, followed by multivariate calibration. In the biogas process it is optimal to run with as high an organic loading as possible, in processes running with maize silage addition this can be obtained by control of the total solids in the in-flow to the reactor. To be able to control this optimally reliable on-line monitoring is needed. The acoustic chemometrics could be such a method, as it is robust and well documented to work in industrially environments. The acoustic signal carries potential embedded information regarding the chemical and physical properties relating to the process, e.g., particle size distribution, flow, density, fibre length, chemical composition a.o. (Esbensen et al., 1999; Halstensen and Esbensen, 2000; Halstensen et al., 2006). An earlier investigation of particulate slurries showed that the sensor was indeed influenced by the hydrodynamics in the flow (Hou et al., 1999), which is closely related to the dry matter content – implying that perhaps acoustic signals could be easily calibrated for on-line monitoring of flow characteristics using indirect multivariate calibration (Martens and Næs, 1991; Esbensen, 2006).

The objective of the present study was to investigate the possibility of applying NIR spectroscopy for on-line monitoring of total/volatile solids and volatile fatty acids during co-fermentation of maize silage and manure. Furthermore, at-line application of acoustic chemometrics and NIR spectroscopy for total solids in manure was investigated for the first time.

2. Methods

Two co-digestions based on manure and maize silage were conducted. The first trial ran as an anaerobic digestion with addition of feed twice a day and was monitored by NIR spectroscopy solely. Based on the results obtained during this trial a second experiment was designed in which the total solid content span was artificially increased within an augmented range. This anaerobic digestion (constant manure substratum) was monitored by both NIR and a.c.

2.1. Trial 1

An anaerobic digestion trial was carried out in a 5 L bioreactor (Simax, CZ), with a working volume of 4 L. An external loop was connected to the fermentor, enabling on-line measurements with NIR (described in Section 2.3) and process sampling (Section 2.4).

The process was initially stabilised for a period of eight days during which only standard manure feedstock was added. This was followed by an addition of a feedstock containing two different levels of maize silage. Level 1 contained 59.5 g dried maize silage pr. 850.5 g manure and level 2, 119 g dried maize silage pr. 791 g manure. The entire trial lasted 42 days.

Stirring (set point = 30 rpm) and heating (set point = 53 °C) were both controlled by a program written in LabVIEW 8.0 software (National Instruments, TX, USA).

NIR spectra were acquired every hour, while every 12 h a representative 130 mL sample for reference analysis was collected. After each sampling operation the fermentor was fed 130 mL feedstock, resulting in a hydraulic retention time of approximately 15.4 days.

2.2. Trial 2

The objective of the second trial was to obtain a model covering a larger total solids (TS) calibration range than in trial 1, and specifically to investigate the possibilities for acoustic chemometric (a.c.) monitoring of this parameter. Therefore the second trial was carried out as a spiking anaerobic digestion (inactive manure substratum), where the total solid content was spanned artificially from 4.8% to 11.4% by adding maize silage to a constant matrix of manure and food waste, see Table 1. For each of the 14 levels triplicates were carried out, i.e., measuring on three separately prepared bottles.

The measurements were carried out applying a homemade reactor loop enabling simultaneous NIR and a.c. measurements. A bottle containing 1 L sample was mounted on the loop, turned up-side down, and the material circulated in the loop. Before each measurement the material was circulated in the loop for 1 min, after which NIR spectra and frequency spectra were collected every minute over a 4 min interval.

2.3. Equipment set-up

The equipment used in both trials was a TENIRS (Transflexive Embedded Near InfraRed Sensor) loop measurement system, developed by the Institute of Agricultural Process Engineering (ILV), University of Kiel for at-line measurements of bio-slurry and manure (Andree et al., 2005; Holm-Nielsen et al., 2007).

During the first anaerobic digestion trial the loop was modified to work in an on-line mode by connecting it directly to the

Table 1
Amount of added maize silage and manure to each level in the spiked anaerobic digestion (trial 2)

Level	Maize silage (g)	Manure (g)	TS obtained (%)
1	0	800	4.8
2	5	795	5.3
3	10	790	5.8
4	14	786	6.3
5	19	781	6.8
6	24	776	7.3
7	29	771	7.8
8	33	767	8.4
9	38	762	8.9
10	43	757	9.4
11	48	752	9.9
12	52	748	10.4
13	57	743	10.9
14	62	738	11.4

mid-height level of the fermentor. Broth was pumped to the flow-through cell for TENIRS measurements and immediately hereafter through a special sampling device (see Section 2.4), returning the broth to the fermentor via an impeller pump driving the entire loop flux. The pump ran for the entire duration of the trial (continuous loop circulation). The total length of the external loop was 2.30 m which was kept at thermophilic temperatures by appropriate insulation measures. The circular loop diameter was 2 cm, except the flow-through cell, which had a 6 mm planar parallel clearing. To keep the flow/pressure constant in the system, the geometry of the flow-through cell was designed commensurate to this decrease in inner duct width.

In the second anaerobic digestion trial only concerning TS, the TENIRS loop was applied in an at-line configuration by applying primary samples in 1 L bottles. The bottles were mounted on the loop, turned up-side down and the broth was pumped through the TENIRS loop measurement system for NIR spectroscopic measurements and a.c. characterisation. The samples was circulated in the loop for 4 min, at which time the bottle was turned downwards again, i.e., creating a sample flow back into the bottle. After measuring each sample, the loop was flushed with water and dried with compressed air to remove any material precipitated in the loop. The TENIRS loop functioned exactly as in trial 1, with only one difference namely that the broth was supplied directly from the reactor in trial 1 but from individual 1 L sampling bottles in trial 2. Great care was employed to ensure that the circulating flow rate was constant for the duration of each measurement.

2.3.1. NIR measurements

NIR data handling was done with the Aspect Plus software (Carl Zeiss Jena GmbH, Germany), allowing communication with the spectrophotometer CORONA (Carl Zeiss Jena GmbH, Germany). Measurements were carried out in the range from 947 to 1533 nm with a resolution of 6 nm.

2.3.2. Acoustic signal measurements

The acoustic signal was obtained via an uniaxial piezoelectric accelerometer Deltatron® 4396 (Brüel & Kjær A/S, Copenhagen, Denmark). The accelerometer was affixed (clamp-on) to a pipe section, specifically designed for a.c. measurements. This device was designed so as to facilitate a decrease of the inner diameter of the loop from 20 to 6 mm, resulting in a turbulent downstream flow pattern. From general acoustic chemometric experience, the acoustic signal is significantly influencing the hydrodynamics of the slurry; thus a change in solid content will show up as changes in the effective slurry density, and viscosity, thereby allowing for indirect multivariate calibration and parameter prediction (Esbensen et al., 1999; Halstensen et al., 2006).

The accelerometer was connected to a signal amplifying module that delivered a constant current to the accelerometer while also taking care of the signal adaption by amplifying and filtering the signal before it entered the computer. The module was a PSA 100i instrument (Detect, Porsgrunn, Norway). The recorded signal, which originally was in the time domain, was transformed via a fast Fourier transform to the frequency domain. Only the transformed FFT power spectra were stored for chemometric data treatment.

For each sample container, four a.c. spectra consisting of 1024 time series samples (corresponding to a frequency of 0–186 kHz) were recorded. Each spectrum was represented by an average of one hundred individual scans of 4 ms duration.

2.4. Sampling

During the fermentation monitoring trial a 130 mL sample (primary sample) was collected from the bioreactor for reference analysis twice a day.

A bioreactor is physically a three-dimensional lot, which means that special attention has to be paid to sample delineation and extraction to ensure a *representative* sampling process. For this to hold every part of the reactor volume has to have an identical, non-zero probability of ending up in the sample. In practice this either means that the sample must delimit a full cross section of the entire reactor or pipeline, or be produced as a full-height column (Pitard, 1993; Gy, 1998; Esbensen and Minkkinen, 2004). For physical three-dimensional bodies this is obviously not easily realisable in the general case, e.g., Holm-Nielsen et al. (2007). Instead of this impasse, TOS prescribes the regimen of one-dimensional sampling (process sampling), in which context it is relatively easy to employ representative sampling of increments and/or samples, see (Petersen et al., 2004; Petersen and Esbensen, 2005) for full details. Especially TOS is adamant that *composite sampling* (using as many increments as possible, together making up the sample) must be used for all two- and three-phase systems, which by their nature are significantly heterogeneous (Petersen et al., 2005; Petersen and Esbensen, 2005).

In this study, a device facilitating one-dimensional process sampling from the re-circulating flow-through the TENIRS loop was developed and placed immediately after the TENIRS measuring cell. For sampling to be "structurally correct" (a TOS term used in context to signify that all structural requirements for the physical sampling to be representative have been fully verified and documented), the sampling device, has to be implemented in an upwardly flowing stream, to avoid flow segregation (Holm-Nielsen et al., 2006). This could not be implemented in the current loop configuration however, where it had to be placed in the horizontally stream. To carry out effective composite sampling, the device's 10 mL increment bottle was effectuated 13 times for each composite sample. This also ensured that samples measured by the TENIRS and the volume withdrawn for reference analysis were pair-wise comparable, thereby allowing multivariate calibration with an absolute minimum of sample-spectrum mismatch. Recently this particular sampling device has been further characterized by Boland (2008), who showed that highly significant sampling bias will develop if circulation speed and solids load are unbalanced. More information regarding implementation of TOS' requirements for representative sampling in bio-slurry systems can be found in Mortensen (2006), Holm-Nielsen et al. (2006), Esbensen and Mortensen (in press).

The composite primary samples were thoroughly mixed before being subdivided into secondary samples having volumes suitable for reference analysis in the laboratory. Subsampling was achieved by vigorous shaking followed by fast pouring the amount needed for reference analysis, into a beaker. This technique was employed based on slurry sampling development work by Holm-Nielsen et al. (2006), who demonstrated that this particular procedure produced minimum contributions to the overall sampling error.

2.5. Raw materials

2.5.1. Manure

The raw material applied for starting up the anaerobic digestion was a substrate mixture obtained from the daily running anaerobic fermentors at Ribe centralised biogas plant (Ribe, Denmark), i.e., cattle and pig manure co-fermented with food waste, see Holm-Nielsen et al. (2007). Before transport to the laboratory, the material was screened through a 4 mm sieve and subsequently stored in thermally isolated containers to avoid any adverse decrease in temperature that might affect the microbial processes. The material was placed in the experimental fermentors immediately after delivery and the fermentation process was continued.

The raw material applied in trial 2 had the same origin as the material used in trial 1. After delivery of the material, 3×14 spiked

bottles were prepared containing the amounts of manure listed in Table 1. Maize silage was added and the solution was stored in the refrigerator until trial start-up. Before measuring, the bio-slurry was heated to 53 °C to fully simulate the exact same conditions as those pertaining to material coming directly from a thermophilic process.

2.5.2. Maize silage

The maize silage applied throughout both the active fermentation trial and the second experiment was obtained from a local farmer (Niels Tobiasen, Ribe, Denmark). For trial 1 it was ground into a particle size of 3 mm before being added to the manure. This mixture was stored in the freezer until approximately 24 h before usage, at which time it was defrosted and heated to the fermenting operating temperature (53 °C). For trial 2, the maize silage was first dried, obtaining a TS content of ca. 90%, and then ground into a uniform particle size of 2 mm before being added/spiked to the manure.

2.6. Chemical analysis

The following chemical analyses were applied in trial 1.

2.6.1. Volatile fatty acids (VFA)

The volatile fatty acids content was measured using a GC Varian 3800 gas chromatograph (Analytical Instruments, Denmark) with a Varian 25 m 0.32 mm ID column (Cat. No. CP 7488 WCOP fused silica coating FFAP-CB). The carrier gas was He at constant pressure: 6 psi. The temperature step-program was: 120 °C for 2 min, rate 20 until 165 °C, hold for 15 min followed by rate 20–200 °C and hold for 3 min.

Two milliliters of each sample was mixed with 250 μl HCl 4 M and 2.0 mL Internal Standard solution (2,2-dimethylbutanoic acid) and the mixture was centrifuged in a Medifuge (VWR & Bie & Berntsen, Denmark) at 5000 rpm for 5–8 min. The supernatent was withdrawn and filtered through a 0.45 μm acetate filter and injected to the GC. The concentrations of acetic, propanoic, iso-butanoic, butanoic, iso-valeric, and valeric acid were determined. Mass-reduction down to 2 mL can be considered representative following the procedures described in Holm-Nielsen et al. (2006).

2.6.2. Total solids/volatile solids

Total solids are the fraction of the sample total wet weight from which water has been evaporated in an oven at 105 °C until stable rest mass. The total volatile solids were determined as the difference between total solids and the weight of the ashes from the sample fully oxidised in a muffled furnace at 550 °C for 2 h.

2.7. Data processing

For both experimental trials, chemometric multivariate calibrations using partial least square regression (PLS-1) were performed for quantitative determination (prediction) of the relevant components.

In a PLS-1 model the goal is to obtain a multivariate model, where the structural part relevant for the *X*–*Y* regression relation and the "noise" part are separated. By doing this the effective dimension is often severely reduced (due to redundancies and correlations); the regression relationship is built between the independent *X*-matrix and the dependent *Y*-vector, based on this reduced number of *latent variables* (PLS components). In this case, the NIR spectra or a.c. frequencies formed the independent data matrixes (*X*) which were to be calibrated against the dependent reference analysis (*y*): the concentration levels for TS, VS, or VFA, respectively.

The PLS-1 algorithm projects the X-spectra onto the direction with the largest [*X*,*Y*] covariance, obtaining a vector containing so-called loading-weights, *w*, which represent this dominant PLS component directions in the *X* space. The original samples (objects) are projected on to this new set of axes thereby obtaining object *scores T* and *U* and variable *loadings W* and *Q*. The scores represent the original measurement data in the new PLS coordinate system, while the loadings-weights can be viewed as a transformation between the old and the new coordinate systems. After determination of the first PLS component, the second is calculated from *updated* version of the [*X*,*Y*] data, obtained after elimination of the first component. The second PLS component direction representing the direction of second largest covariance [*X*,*Y*] and similarly for the third. Full chemometric details of the PLS approach can be found in Martens and Næs (1991), Esbensen (2006).

For evaluating the prediction ability of the resulting regression models different validation methods were applied, test-set and leave-one-level-out segment cross validation following the paradigm delineated in Esbensen (2006). Test-set validation was carried out by building the model on a training data-set followed by validation on a completely new independent test-set, thereby obtaining the most realistic measure of the model's future prediction abilities. In the case, where a limited number of samples did not allow for separate calibration and validation data-sets, leave-one-level-out cross validation was applied, i.e., a complete concentration level was removed from the calibration set at a time and a model was created based on the remaining samples; subsequently this model was used to predict the concentrations of the left-out samples. This procedure was repeated for all of the available concentration levels; for the small sample case, leave-one-level-out cross validation is an acceptable, though not perfect approximation to the optimal test-set validation procedure, Esbensen (2006).

As a means of evaluating prediction performance, and simultaneously finding the optimal number of effective PLS components, a measure of the average prediction error; the root mean square error of prediction (RMSEP) is calculated

$$\mathrm{RMSEP} = \sqrt{\frac{\sum (y_{\mathrm{predicted}} - y_{\mathrm{reference}})^2}{n}}$$

where *y* is the dependent variable and *n* is the number of samples. Test-set RMSEP is a very reliable measure of the average error encountered when using the model for prediction of future, similar samples. The RMSEP statistic must be as small as possible for a model to be performing "optimally".

For model evaluation, also the "ratio of standard error of performance to standard deviation" (RPD) statistic was applied, which compares the RMSEP with the range of the calibrated parameter measured as, the standard deviation (SD) of the values obtained in the reference analysis

$$\mathrm{RPD} = \frac{\mathrm{SD}}{\mathrm{RMSEP}}$$

Generally, models with RPD above 2.5 are considered acceptable, while an RPD of 10 is excellent (Williams and Norris, 2001). These statistics are particularly useful in comparing the prediction abilities between alternative models.

In trial 1, total and volatile solids, total volatile fatty acids and the individual volatile fatty acids: acetic, propionic, iso-butyric, butyric, iso-valeric, valeric acid were calibrated for. The first model consisted of triplicate reference analysis for each spectrum and was used for the purpose of outlier detection only. After outlier exclusion the data were averaged, thereby creating a new data matrix for the final model containing only one spectrum per sample. For evaluating the model, test-set validation was applied, i.e., using

an independent second data-set: the samples of which were taken 12 h after each training data-set sample.

In trial 2, models for total solids based on the NIR and a.c. measurements were carried out. For each sample four spectra were recorded, which were averaged before modelling. For the NIR model no pre-processing was necessary, whereas a full (combined offset and amplification) multiple scatter correction (MSC) of the a.c. spectra (X) and logarithmic (ln) transformation was necessary to linearize the *y* data. MSC is applied to spectra to compensate for generic multiplicative and additive effects, for example caused by interference in solid–liquid mixtures and similar (Martens and Næs, 1991; Esbensen, 2006). In both cases in trial 2 leave-one-level-out segment cross validation was applied.

Data processing was carried out using the UNSCRAMBLER software version 9.7 (Camo, Oslo, Norway).

3. Results and discussion

3.1. Anaerobic digestion

The anaerobic digestion (trial 1) lasted for 42 days during which samples for reference analysis were collected twice a day. During the first 9 days the process was stabilised by feeding with manure only, after which a bio-slurry consisting of manure and maize silage was applied.

During this anaerobic digestion a gradual build up of a maize silage layer on the top of the liquid layer was observed. As the sample extraction took place only from the liquid layer at reactor mid-height, this resulted in a significant total sampling error with respect to the full bioreactor because these samples were not a measure of the complete system. Ordinarily this would disqualify this type of sampling, but because both the TENIRS and the sampling device were placed on the recirculation loop, the material measured on-line and the sampled reference material were in fact pair-wise strictly matched up and thus full comparable. For this reason these sample-pairs will thus facilitate multivariate calibration in a fashion acceptable for the present feasibility study. Mortensen and Bro (2006) discuss this situation in detail with a broad carrying over potential for all bio-conversion PAT monitoring scenarios.

For all samples the total and volatile solid and the fatty acid content was determined as described.

3.1.1. Total and volatile solids

The TS span obtained in the liquid phase during trial 1 was approximately 2.2%; from 4.4% to 6.6%. From Fig. 1 it is seen that the trial could be divided in three different intervals. From day 0 to 13 there was a steady increase followed by a constant level from day 13.5 to 24, after which an increase was observed again. On day 13.5, due to an overall decrease in pH from 8.5 to 7.4, it was decided to stop the maize silage feed, thereby trying to stabilise the process by feeding with manure only. This was done until day 15.5, where the normal feeding level was resumed, making this a possible explanation for the constant level. The constant level was maintained only until day 24, indicating that the maize silage added after day 15 was either degraded to biogas or did not stay in the liquid phase (swim layer). The steep increase in the total solid content from day 24 can likely be explained by the fact that the impeller and the swim layer were approaching each other, gradually releasing more solids into the liquid phase again.

In any event the sole purpose for the digestion experiment was to facilitate as wide a span in the concentrations of the target parameters (*y*) as possible in order to evaluate the potential for NIR monitoring.

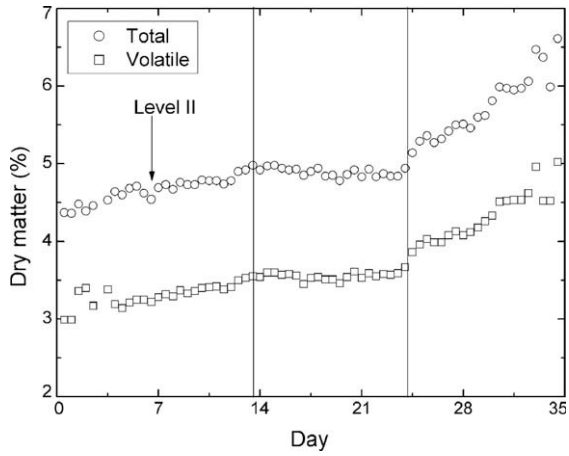

Fig. 1. Trend in total and volatile solid in trial 1.

The overall trend in the volatile solids (VS) followed the same trend as that of the total solids.

3.1.1.1. Multivariate calibration. For both the TS and VS in trial 1 multivariate calibrations with test-set validation were carried out. For calibration, the samples collected in the morning were applied, while samples collected in the evening constituted the test-set. Applying only one PLS component, good models ($RPD_{TS} = 5.0$ and $RPD_{VS} = 5.1$) were obtained for both TS and VS, see Fig. 2. For the TS model it was decided to exclude one sample from the test-set, as it was situated outside the calibration range. In the VS model two samples were excluded for the same reason. One of the samples excluded was the same as the one excluded from the TS model, due to a too low TS%, thus indicating an analytical error in the *y*-reference.

From these models it is evident that it is possible to monitor the TS content in a matrix as complex as manure and food waste within the range 4.6–6.5% TS and 3.2–4.5% VS, respectively. From Fig. 2 it is conspicuous that the VS model is almost gliding into the TS model, thus indicating that a global model spanning from 3.2% to 6.5% would be possible. The normal operating range in typical Danish biogas plants span 6–10% TS. The measurements therefore need to be extended to these higher levels and incorporated, to be able to use the model for future full-scale on-line monitoring. For a new TS model covering the entire range see the models obtained from trial 2.

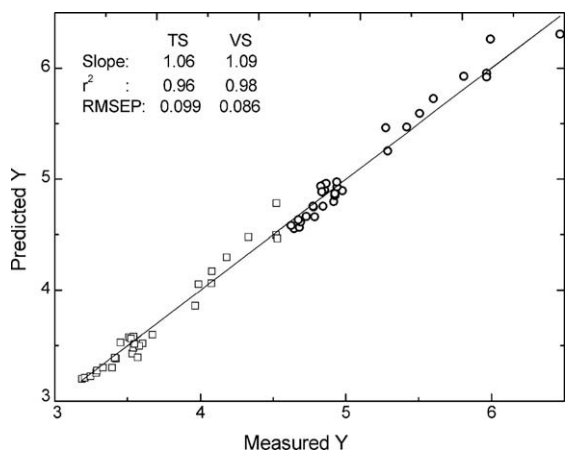

Fig. 2. PLS models (trial 1) applying one component: (○) total solids model and (□) volatile solids model. The line is the target line. In both models no pre-processing of the data was necessary.

3.1.2. Volatile fatty acids (VFA)

It has been debated which VFAs are the most important to monitor during the anaerobic digestion, and which is the best measure of process instabilities (Boe, 2006). In this anaerobic digestion (trial 1) it was decided to measure the largest contributors to the total VFA concentration; acetic and propionic acid as well as some smaller contributors; butanoic, iso-butanoic, valeric, and iso-valeric acid. The results are seen in Fig. 3.

In Fig. 3 one observes the significant large individual contributions to the analytical results from the total sampling errors, despite the care invested in the experimental sampling (cf. Section 2.4). It this context, it is gratifying how it is fully possible to delineate coherent overall trajectories for the individual process parameters (smoothed curves). Many biotechnological systems that do not address the inherent sampling issues in a similar strict fashion exhibit analogous, though much larger sampling errors still; for a comparison see e.g., Holm-Nielsen et al. (2006 and references herein).

From the beginning to the end of the anaerobic digestion an increase from 1.3 to 22.3 g L^{-1} in the total VFA was observed, Fig. 3A, indicating process instabilities due to uncoupling between the hydrolytic and fermentative step on one side and the acetogenic and methanogenic step on the other.

For the first 13 days the tendency in the total VFA concentration could be split in two. The first 8 days a logarithmic increase was observed followed by a linear increase until day 13. Looking at the individual fatty acids, this corresponds to the trend seen in the acetic acid concentration (Fig. 3B). The smaller increase in the total VFA after day 13 was a result of the complementary fatty acids that increased in concentration (see Fig. 3B and C) until day 21, where fluctuation around a stabilised level was observed. However, for the butanoic acid this stabilisation lasted until day 28 when an increase was observed.

Throughout the process the VFA levels observed were high for a biogas process. The normal operating concentration in Danish biogas plants for a normally functioning process is from 5 to 7 g L^{-1}. Above 7 g L^{-1} instabilities very soon start occurring. Elsewhere a limit as low as 1.5 g L^{-1} has also been found as an approximate limit above which instabilities occur (Angelidaki et al., 2005). In the present case, however, a gas production was observed throughout the experiment indicating that no process failure occurred (gas yield data only available for less frequent intervals during the trial, not shown). This was supported by the evaluation of the propanoic acid concentration, as this is known to be the most thermodynamically unfavourable process parameter, being affected by the hydrogen level before any other VFA, i.e., any process instability would be seen for this parameter first (Pauss and Guiot, 1993). Looking at the present curve, the concentration increased steadily during the complete trial, thereby not supporting any induction of an organic overload. This deviation from earlier experimental and published data indicates that the VFA level, where inhibition and process collapse occur, might be process dependent. In the present process no large addition of ammonium-containing material occurred, which resulted in a favourable C/N relationship throughout the entire process, thus allowing a gas production even under high VFA conditions.

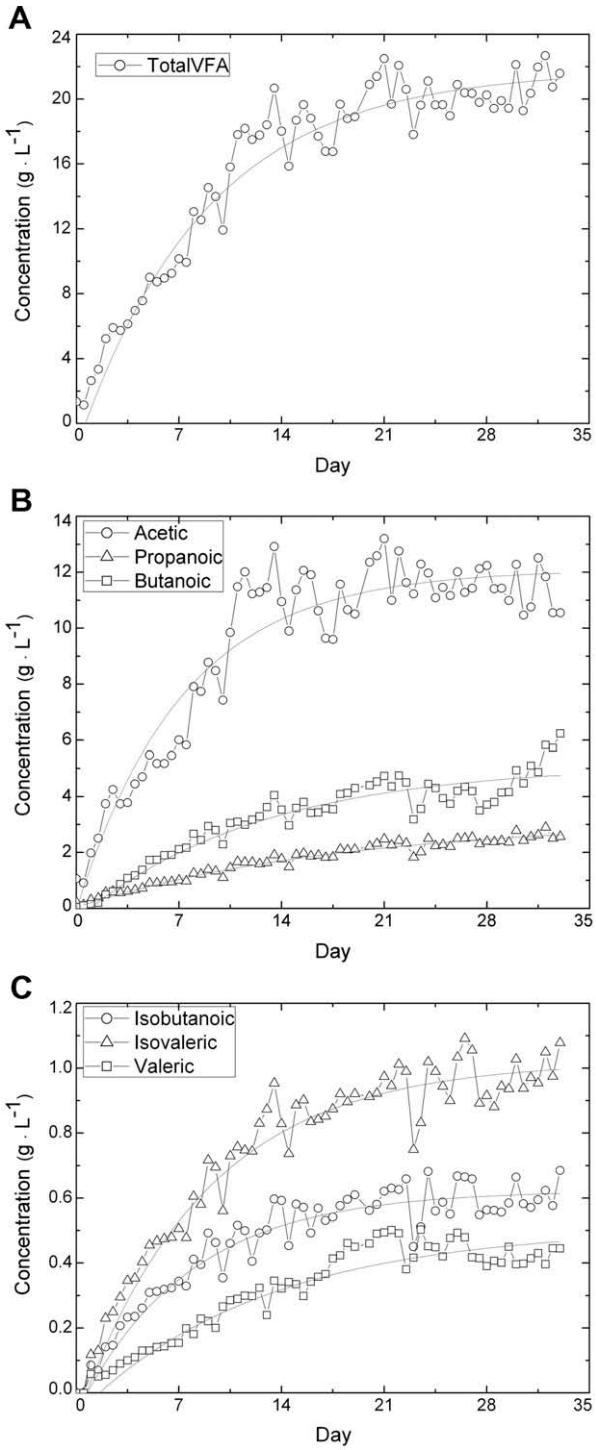

Fig. 3. Trial 1 biomass conversion process parameters: (A) overall VFA concentration, (B) acetic, propionic, and butyric acids and, (C) iso-butanoic, iso-valeric, and valeric acid.

3.1.2.1. Multivariate calibration. For all VFA models in trial 1 different pre-processing techniques of the NIR data matrix, see Fig. 4 for raw spectra, were tried out. A Savitzky-Golay with first order derivatives, employing a three-point smoothing window of a second order polynomial was found to give the best models.

In Table 2 the results of all VFA models are summarized. For these results, outlier exclusion from both the calibration and the validation set (test-set) was employed where necessary. Outliers excluded from the test-set were all values outside the limits of the calibration range, making their exclusion both necessary and acceptable. Calibration set outliers most likely represent samples with excessive total sampling errors, cf. Boland (2008). All outlier deletions represent only very low fractions of all available samples, Table 2.

The models obtained had slopes between 0.83 and 0.92, i.e., a fair to good prediction accuracy was obtained, thus indicating that these important parameters can be satisfactorily predicted by

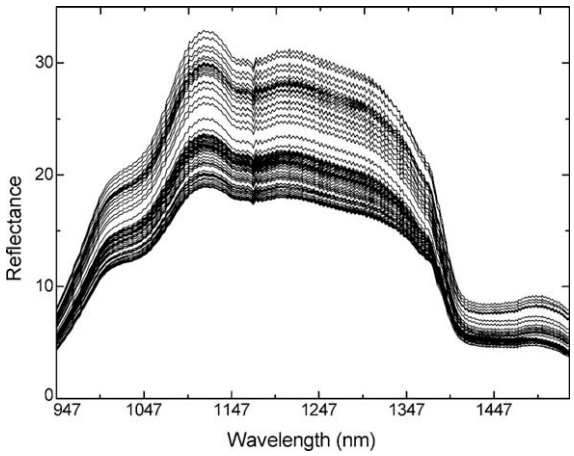

Fig. 4. Raw NIR spectra from trial 1.

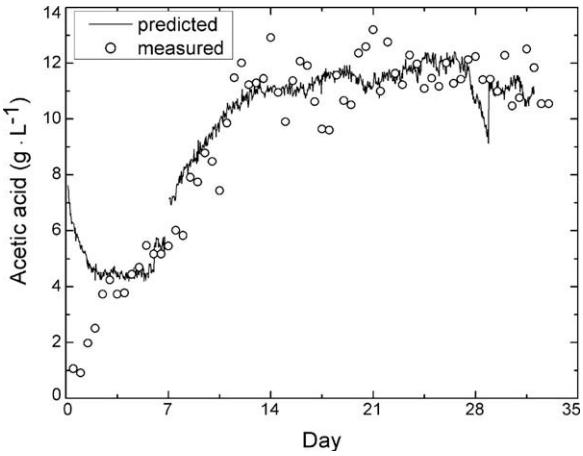

Fig. 5. Acetic acid trend during trial 1 predicted by the model (–) and the concentration measured by reference analysis (○). Predicted curvature breakage due to a computer malfunction.

on-line NIR spectroscopy. This was further substantiated by calculation of the RPD value (all RPD values are >2.5).

3.1.2.2. Discussion – future prediction possibilities. During anaerobic digestion in trial 1, NIR spectra were collected every hour, of which only two per 24 h were used in the modelling and validation. To illustrate the strength and possibility of future prediction applications of the models, as an example the acetic acid model was used for prediction of the acetic acid concentration in the broth for the remaining NIR spectra. The results are seen in Fig. 5.

From this illustration it can be seen that the dominant contribution to the total sampling error is associated with the *Y* (reference data). The obtained model was unable to estimate the concentration of acetic acid in the first four samples, i.e., samples with concentrations below approximately 4.2 g L^{-1}. This is fully understandable, however, as these were in fact excluded as outliers in the model, and thus situated outside the models validation range.

For the higher concentrations – on statistical average – the smoothed curve follows the expected process trend completely. Around day 27–28.5 a sharp decreasing tendency in the predicted curve was observed, which was not clearly observed from the samples collected. However, this cannot be explained by a failure in the model as the trend both before and after this excursion is correct, i.e., it is likely that a decrease in acetic acid actually took place. This illustrates the strength of NIR for monitoring and control purposes; as such a tendency would not have been noticed as quickly applying physical samples only.

The purpose of the present model feasibility studies is to *monitor* the overall process trend as accurately and precisely as possible. This has been amply demonstrated. To obtain stronger models, broadening the calibration range of the model by including more samples with the levels needed, will undoubtedly lead to improved results. It is also essential to improve on the sampling devices employed so as to reduce the prediction imprecision observed. While in many earlier studies such sampling errors were predominant, all necessary principles and experience needed are now available, Boland (2008), Petersen and Esbensen (2005), Holm-Nielsen (2008), Esbensen and Mortensen (in press).

3.2. Spiked anaerobic digestion

The spiked anaerobic digestion trial (trial 2) consisted of an at-line experiment containing 14 concentration levels spanning an augmented solid matter (TS) concentration range from 4.8% to 11.4%. For each level, three dilutions containing the same nominal solid content were prepared, totalling 42 dilutions, measured in a randomized order. For each sample container four a.c. and NIR spectra were recorded.

3.2.1. NIR model

For NIR measurements in trial 2, a PLS-1 model with segmented cross validation, leaving one concentration level out at a time was applied. For this model no pre-treatment of the data was necessary. The optimal number of PLS components for describing the relation between the NIR and the total solid was found to be one, where 100% of the information recorded by the NIR was describing 98% of the variation in the total solid, i.e., the total spectra from 947 to 1533 nm contained no noise and was important for modelling the total solids in the manure samples. A description of the model is seen in Table 3.

During data analysis a parallel displacement was observed in the spectra, which was originally reduced by a full MSC. This

Table 2
PLS-1 models for volatile fatty acids in trial 1

	Components	Out_{calib}	Out_{test}	Slope	r^2	RMSEP	RPD
Acetic acid	3 (92% *X*, 89% *Y*)	2 (6%)	5 (15%)	0.89	0.89	913	3.1
Propionic acid	3 (92% *X*, 91% *Y*)	1 (3%)	3 (9%)	0.89	0.91	206	3.4
Butanoic acid	3 (92% *X*, 92% *Y*)	0	3 (9%)	0.87	0.92	400	3.6
Iso-butanoic acid	3 (93% *X*, 88% *Y*)	3 (9%)	6 (18%)	0.83	0.89	43.8	2.8
Valeric acid	3 (93% *X*, 90% *Y*)	2 (6%)	3 (9%)	0.92	0.92	43.4	3.5
Iso-valeric acid	3 (92% *X*, 89% *Y*)	2 (6%)	2 (6%)	0.86	0.89	76.4	3.1
Total	3 (92% *X*, 90% *Y*)	2 (6%)	3 (9%)	0.87	0.90	1594	3.3

"Components": number of PLS components applied in the model, numbers in brackets are the explained calibration and the residual validation variance; "Out_{calib}": number of outliers excluded from the calibration set, numbers in brackets are the corresponding percentage; "Out_{test}": number of outliers excluded from the test-set, numbers in brackets are the corresponding percentage; "RMSEP": root mean square error of prediction; and RPD: ratio of standard error of performance to standard deviation.

Table 3
PLS-1 models for total solids for NIR and a.c. in trial 2

	Components	Out	Slope	r^2	RMSEP	RPD
NIR	1 (100% *X*, 98% *Y*)	0	0.98	0.97	0.36	5.7
a.c.	3 (79% *X*, 93% *Y*)	7 (17%)	0.85	0.85	0.077	2.6

"Components": number of PLS components applied in the model, numbers in brackets are the explained calibration and the residual validation variance; "Out": number of outliers excluded, numbers in brackets are the corresponding percentage; "RMSEP": root mean square error of prediction; and RPD: ratio of standard error of performance to standard deviation.

pre-treatment, however, only resulted in a less good model, i.e., it was evident that the original model (RPD = 5.7) was partially caused by the parallel displacement in the spectra that could be explained by the decreased transparency of the samples as a function of increasing total solid concentration.

3.2.2. Acoustic chemometric model

The a.c. spectra obtained during trial 2 were pre-processed by a full-spectrum MSC (both additive and multiplicative corrections were employed) and the *y*-variable (% TS) was ln-transformed before the model was obtained. Several other pre-treatments were tried, none of which resulted in better models.

Calibrating this model, 7 of the original 42 individual measurements had to be excluded as outliers. Two of the seven outliers can be explained by sample preparation deficiencies at the beginning of the experimental series. Of the remaining five, four belonged to high concentration levels which progressively showed more serious problems regarding a stable flow situation as the TS in the samples increased to well-nigh extreme levels. None of these outlier measurements represent a full concentration level, however, for which reason the total concentration range still spanned all 14 experimental levels; thus there was no significant impairment of the possibilities for multivariate calibration.

For describing the relation between the a.c. and the total solid a three component model applied 79% of the *X*-variance for explaining 92% of the total *Y*-variance, i.e., 21% acoustic vibration variance over 14 concentration levels was not correlated with the % TS range. From the loading-weights (not shown) it was evident that frequency range from 0 to 109 kHz was the most influential. The model obtained had reasonable good accuracy and precision results, see Table 3. The RPD, however, was only marginally on the acceptable side, indicating that the prediction ability must be improved before applying the model for real-world system predictions.

There are many optimisation options for acoustic chemometrics that could not be evaluated in this first pilot study, first and foremost a sub-optimal sensor deployment location (Esbensen et al., 1999). As a means of searching for improved results different sensor locations should be tried out. The most obvious alternative locations would be placing the sensor directly on the NIR measurement cell itself, as the associated complex cross section compensations between pipe and measurement cell, necessary for NIR to work, is bound to affect the flow regime of the liquid. Different locations along the pipe line, both before and after the NIR cell, should also be tested following Halstensen (2001). Furthermore the acoustic signal might also benefit from advanced *designed* changes in pumping speed. Kupyna (2008), in a recent thesis, explores many advanced options for zooming in on effective, operative X-signal characteristics related to *peak-shifts* as a function of the solids load in the slurry being forced through the acoustic chemometrics orifice.

4. Conclusions

To significantly increase energy supply from local biomass conversion facilities in the near future, so-called advanced biofuels applying dedicated energy crops are currently being developed. The addition of maize silage to conventional anaerobic digestion of cow and pig manure constitutes one of these new options. From the feasibility results obtained in this study it is evident that by spiking the anaerobic digestion with maize silage, a favourable C/N ratio can be obtained, allowing biogas production under unusual high fatty acid concentrations. Production under these circumstances is desirable, as it causes a high gas yield. However, this is very much a balancing act, where addition of maize silage has to be strictly monitored, as a too high concentration may result in process instabilities and system breakdowns if no countermeasures are taken immediately. As part of ongoing developments, the focus is therefore on optimisation via introduction of reliable process monitoring, especially the concept of process analytical technologies, PAT.

A spiking trial showed that NIR can be satisfactorily applied for chemical concentration monitoring of VFAs: acetic, propionic, butanoic, iso-butanoic, valeric, and iso-valeric acids in the complex matrix associated with conventional anaerobic digestion of biomass. This trial also showed a clear potential for effective monitoring of TS, a key physical process parameter.

A second trial, where both NIR and acoustic chemometrics were investigated confirmed that both methods can in fact be applied for TS monitoring in manure-based biomass conversion. Looking at the RPD values, the model obtained by NIR was clearly better predicting future samples than the current a.c. model. This pilot study outcome is likely to be altered by a change in the a.c. sensor location however. Only one location was included here, while from general acoustic chemometrics experience it is seldom the case that an optimized sensor location is to be found already in the pilot study stage. We were here restricted to report only the very first results within an acceptable range, while the work of optimizing acoustic chemometrics as a routine PAT modality in the bio-conversion industry has only begun. Much challenging work remains following these first positive forays.

These findings have recently been substantiated by subsequent studies a.o. regarding NIR and acoustic chemometrics monitoring from laboratory-, over pilot- to full-scale bioreactor systems, Holm-Nielsen (2008).

Acknowledgements

This study was financed by Aalborg University and Esbjerg Seminariefond, greatly appreciated. Thanks also to farmer Niels Tobiasen, who provided the maize silage, and to Ribe Biogas for providing the stock manure. Henry Enevoldsen and Carsten Andreasen are thanked for their extensive help setting up the fermentors and the laboratory technicians Dorte Spangsmark and Marie Georgette Susanne Roy are acknowledge for their kind help with the reference analysis. Finally Anne Cole and Michael Madsen are thanked for their linguistic contributions to this paper.

References

Andree, H., Dolud, M., Hügle, T., 2005. Tenirs – NIR Analysis of Heterogeneous Bioslurries. Baltic Biorefinery Symposium, 26–28 May, Aalborg University Esbjerg, Denmark, 2005, pp. 141–148.

Angelidaki, I., Boe, K., Ellegaard, L., 2005. Effect of operating conditions and reactor configuration on efficiency of full-scale biogas plants. Water Sci. Technol. 52 (1–2), 189–194.

Boe, K., 2006. Online Monitoring and Control of the Biogas Process, Ph.D. Thesis. Institute of Environments & Resources, Technical University of Denmark.

Boland, L., 2008. Evaluation of Sampling Facilities and Process Analytical Technologies in Applied Biotechnology, M.Sc. Thesis. Esbjerg Institute of Technology, Aalborg University.

Bull, S.R., 2001. Renewable energy today and tomorrow. Proc. IEEE 89 (8).

Diamantis, V., Melidis, P., Aivasidis, A., 2006. Continuous determination of volatile products in anaerobic fermenters by on-line capillary gas chromatography. Anal. Chim. Acta 573–574, 189–194.

Esbensen, K.H., 2006. Multivariate Analysis in Practice: An Introduction to Multivariate Data Analysis and Experimental Design. Reprint of 2001 edition, fifth ed. CAMO Press, Norway.
Esbensen, K.H., Hope, B., Lied, T.T., Halstensen, M., Vermoen, T.G., Sundberg, K., 1999. Acoustic chemometrics for fluid flow quantifications – II: a small constriction will go a long way. J. Chemometr. 13, 209–236.
Esbensen, K.H., Minkkinen, P. (Eds.), 2004. Special Issue: 50 Years of Pierre Gy's Theory of Sampling Proceedings: First World Conference on Sampling and Blending (WCSB1) Tutorials on Sampling: Theory and Practice. Chemom. Intell. Lab. Syst., vol. 74, pp. 1–236.
Esbensen, K.H., Mortensen, P., in press. Process sampling (theory of sampling, TOS) – the missing link in process analytical technology (PAT). In: Bakeev, K.A. (Ed.), Process Analytical Technology, second ed. Blackwell Publishing.
Gy, P., 1998. Sampling for Analytical Purposes. John Wiley & Sons, Chichester, England.
Halstensen, M., 2001. Acoustic Chemometrics – Experimental Multivariate Sensor Technology and Development of System Prototypes for Industrial Multi-phase Characterisation: Selected Forays, Ph.D. Thesis. Telemark University College, Norway.
Halstensen, M., de Bakker, P., Esbensen, K.H., 2006. Acoustic chemometric monitoring of an industrial granulation process – a PAT feasibility study. Chemometr. Intell. Lab. Syst. 84, 88–97.
Halstensen, M., Esbensen, K., 2000. New developments in acoustic chemometric prediction of particle size distribution – the problem is the solution. J. Chemometr. 14, 463–481.
Holm-Nielsen, J.B., 2008. Process Analytical Technologies for Anaerobic Digestion Systems – Robust Biomass Characterisation, Process Analytical Chemometrics, and Process Optimisation, Ph.D. Thesis. Esbjerg Institute of Technology, Aalborg University.
Holm-Nielsen, J.B., Andree, H., Lindorfer, H., Esbensen, K.H., 2007. Transflexive embedded near infrared monitoring for key process intermediates in anaerobic digestion/biogas production. J. Near Infrared Spectrosc. 15, 123–135.
Holm-Nielsen, J.B., Dahl, C.K., Esbensen, K.H., 2006. Representative sampling for process analytical characterization of heterogeneous bioslurry systems – a reference study of sampling issues in PAT. Chemometr. Intell. Lab. Syst. 83, 114–126.
Hoogwijk, M., Faaij, A., Eickhout, B., de Vries, B., Turkenburg, W., 2005. Potential of biomass energy out to 2100, for four IPCC SRES land-use scenarios. Biomass Bioenergy 29, 225–257.
Hou, R., Hunt, A., Williams, R.A., 1999. Acoustic monitoring of pipeline flows: particulate slurries. Powder Technol. 106, 30–36.
Koonin, S.E., 2006. Getting serious about biofuels. Science 311 (5760), 435.
Kupyna, A., 2008. Acoustic Chemometrics – A Potential Method for Determination of Characteristics of Liquid Flows, Ph.D. Thesis. Norwegian University of Life Sciences, Department of Chemistry, Biotechnology and Food Sciences, Norway.
Martens, H., Næs, T., 1991. Multivariate Calibration. Biddles Ltd., Great Britain.
Morel, E., Santamaria, K., Perrier, M., Guiot, S.R., Tartakovsky, B., 2004. Application of multi-wavelength fluorometry for on-line monitoring of and anaerobic digestion process. Water Res. 38, 3287–3296.
Mortensen, P.P., 2006. Process Analytical Chemistry-Opportunities and Problems for Bio-industrial Implementation, Ph.D. Thesis, Aalborg University Esbjerg.
Mortensen, P.P., Bro, R., 2006. Real monitoring and chemical profiling of a cultivation process. Chemometr. Intell. Lab. Syst. 84, 106–113.
Nordberg, Å., Hansson, M., Sundh, I., Nordkvist, E., Carlsson, H., Mathisen, B., 2000. Monitoring of a biogas process using electronic gas sensors and near-infrared spectroscopy (NIR). Water Sci. Technol. 41 (3), 1–8.
Pauss, A., Guiot, S.R., 1993. Hydrogen monitoring in anaerobic sludge bed reactors at various hydraulic regimes and loading rates. Water Environ. Res. 65 (3), 276–280.
Petersen, L., Dahl, C.K., Esbensen, K.H., 2004. Representative mass reduction in sampling: a critical survey of methods and devices. Chemometr. Intell. Lab. Syst. 74, 95–114.
Petersen, L., Esbensen, K.H., 2005. Representative process sampling for reliable data analysis – a tutorial. J. Chemometr. 19, 625–647.
Petersen, L., Minkkinen, P., Esbensen, K.H., 2005. Representative sampling for reliable data analysis: theory of sampling. Chemometr. Intell. Lab. Syst. 77, 261–277.
Pitard, F.F., 1993. Pierre Gy's Sampling Theory and Sampling Practice: Heterogeneity, Sampling Correctness, and Statistical Process Control, second ed. CRC Press Ltd., USA.
Sims, R.E.H., 2001. Bioenergy – a renewable carbon sink. Renew. Energy 22, 31–37.
Spanjers, H., Bouvier, J.C., Steenweg, P., Bisschops, I., van Gils, W., Versprille, B., 2006. Implementation of in-line infrared monitor in full-scale anaerobic digestion process. Water Sci. Technol. 53 (4–5), 55–61.
Tosi, S., Rossi, M., Tamburini, E., Vaccari, G., Amaretti, A., Matteuzzi, D., 2003. Assessment of in-line near-infrared spectroscopy for continuous monitoring of fermentation processes. Biotechnol. Prog. 19, 1816–1821.
Vanrolleghem, P.A., Lee, D.S., 2003. On-line monitoring equipment for wastewater treatment processes: state of art. Water Sci. Technol. 47 (2), 1–34.
Williams, P., Norris, K. (Eds.), 2001. Near-Infrared Technology in the Agricultural and Food Industries, second ed. American Association of Cereal Chemists Inc., pp. 145–169.
Yamamoto, H., Yamaji, K., Fujino, J., 1999. Evaluation of bioenergy resources with a global land use and energy model formulated with SD technique. Appl. Energy 63, 101–113.

III

NIR- characterization of (cut) wheat straw for bioethanol production – a feasibility study

Carina J. Lomborg[a], Erik Steen Jensen[b], Kim H. Esbensen[a]

[a] Applied chemometrics, analytical chemistry, acoustic chemometrics, applied biotechnology, bioenergy and sampling (ACABS) Research group, Aalborg University; Niels Bohrs Vej 8, DK-6700 Esbjerg, Denmark

[b] Biosystems department, Risø National Laboratory for Sustainable Energy, Technical University of Denmark, Fredriksborgsvej 399, DK – 4000 Roskilde, Denmark

Abstract

Fast, reliable field or intake estimates of carbohydrate and lignin compositions are necessary for optimal straw characterization at the first stage of 2G bioethanol production. Near Infrared spectroscopy was applied for quantifying the concentration of cellulose (glucan), hemicelluloses (xylan, arabinan) and lignin in chopped wheat straw. Field samples were collected from sites in Denmark, from which 44 were selected. Alternative pre-processing's were evaluated; 1st and 2nd order derivative spectra (Savitzky-Golay) and Multiplicative Scatter Correction (MSC) was found optimal for the carbohydrates and lignin respectively. Full spectrum PLS-1 regression models resulted in less good prediction abilities, than the models based on wavenumber variable selection applying a Jack-knife approach. The present first foray models were validated using 4-segment cross validation leading to fair to good accuracies (slope 0.76-0.90) and fair precisions (r^2 0.77-0.83) for the carbohydrates. The lignin model also showed a fair accuracy (0.84) but a distinctly less good precision (0.72). The carbohydrate models were furthermore test-set validated, resulting in an accuracy of 0.85-0.94 and precision estimates 0.86-0.87. This corresponded to a relative RMSEP in the interval 8 -10 % which is satisfactory for a feasibility study, but the models should be optimized further before application in a routine monitoring context.

Key-words Near infrared spectroscopy, multivariate analysis, wheat straw, 2nd generation bioethanol

Introduction

Due to increasing fuel demands and a need to decrease dependence on fossil fuels and green-house gas emissions, there is an increasing attention on biofuels and other renewable energies (solar, wind power etc.). As a case in point the target for incorporation of biofuels to replace fossil fuels in EU has recently been set to 10% by 2020.[1] Current bioethanol production facilities are therefore in need of both increased efficiency as well as significant quantitative growth. The plants that are mostly used are for 1st generation bioethanol, where biofuel is produced directly from sucrose or starch in cereal and tubers. However, the reduction of green-house gas emissions from 1st generation bioethanol is limited compared to the reduction from 2nd generation bioethanol. By producing 2nd generation bioethanol i.e., using lignocellulosic biomass for biofuel production it is also avoided to use biomass which can be use for food.[2] Agricultural residues are abundant around the world and are becoming important as raw materials for bioenergy production. Wheat straw, rice straw, and bagasse are the most well known examples.[3]

In Denmark wheat straw is mostly combusted in coal-fired Combined Heat and Power plants (CHP). However it is also suitable as a raw material for the 2nd generation bioethanol production. In a process called the Integrated Biomass Utilization System (IBUS), a combination of CHP and production of bioethanol, is currently being investigating. In this process the raw material is first converted to a fermentable substrate by applying steam and by adding selected enzymes. The C6 sugars from cellulose are now fermented to bioethanol, the C5 sugars from the hemicelluloses are used as a feed molasses, and the lignin residue is applied as a solid biofuel and used for generating process energy and a surplus of heat and power. The IBUS process is at present close to commercial reality.[2]

A way of optimizing this process is by a more comprehensive characterization at the CHP intake, as the efficiency and quality is very much raw material dependent. One promising way to optimize this process is thus by optimal control of the feedstock. Typically at a CHP the moisture content is being measured only, this however tells nothing about the content of fermentable sugars, and thus the real quality of the bales.

Different classical methods for carbohydrate determinations, for example high liquid performance chromatography (HPLC), have been reported.[4] HPLC analysis on wheat straw is laborious and slow, and therefore not a real solution for on-line monitoring of an industrial processes.

In the literature different methods that are principally applicable as at- or on-line monitoring modalities have been investigated for other types of biomasses, e.g. barley, foliar, flax. One of these modalities is near infrared spectroscopy (NIR). A selection of the raw material upon which NIR have been applied are listed in Table 1. For foliar and “forages and byproducts” a comparison between NIR and mid infrared spectroscopy was done. As seen in Table 1 the potential of applying NIR have been investigated for a number of different constituents in

lignocellulosic materials, amongst others cellulose, hemicellulose, and lignin. This implies that it may also be applicable for wheat straw characterization.

Table 1: Different application fields of NIR on various biomasses reported in the literature. *DTG: Derivative thermogravimetric analysis, ADF: Acid detergent fibre, NDF natural detergent fibre.*

Raw material	Application	Components
Rice straw[8]	Quantitative determination	Cellulose, hemicellulose, lignin, moisture, total ash, and acid insoluble ash
Ground barley[9]	Analysis for fuel ethanol production	Moisture, starch, β-Glucan, protein, oil and ash
Foliar[10]	Alternative to wet lab analysis	Mineral nutrients, carbon, fibre constituents (ADF, NDF, cellulose and lignin)
Corn stover[11]	Alternative to wet lab analysis	Glucan, xylan, arabinan, lignin, protein, acetyl
Flax Fibres[12]	Assessing physical and chemical characteristics	AIRflow, DTG, fibre strength, NDF, ADF and hemicelluloses
Forages and byproducts[13]	Quantitative determination	Fibres and protein, nitrobenzene oxidation products of lignin, six measures for lignin content
Silage[14]	Analysis of different components	Dry matter, crude protein, different acids, pH

The present work assesses the feasibility of developing multivariate models from NIR spectra to be applied for quantification of cellulose (glucan), hemicelluloses (xylan, arabinan) and lignin in wheat straw. Several alternative spectral pre-processing combined with multivariate calibration[5] are investigated, together with the possible merits of variable selection[6], while taking careful aim at proper validation of the feasible models.[7]

Material and Methods

Raw materials

A total of 95 samples were collected all over Denmark in 2006; 65 originated from south east/west Jutland and Funen while 30 originated from the island of Zealand.

The 65 sample set were from the harvesting period 2003 – 2006, with the 2006 samples constituted the majority. These samples were collected from fields and barns and hence called

field samples. All samples were accepted independently of the straw quality, resulting in freshly cut straw of the fields to almost spoiled straw that had been stored while it was still wet and thus subject to microbial degradation. This range in quality was deliberately accepted, as the straw quality is different from year to year. Furthermore it was hypothesized that samples subjected to microbial degradation may have lower sugar content, thus producing a potentially larger calibration span for the sugar concentrations. The overall straw size ranged from a few cm to approximately 25 to 30 cm.

The 30 Zealand samples were obtained from Køge Bio pellets factory (Denmark). They were sampled from the bio pellets production line over a 30 day period, hence called **industrial** samples. The material in the production line was from regional farmers, all harvested in 2006. The composition of these samples constituted a mix of wheat and barley straw; the more precise compositional distribution is unknown. Of the 30 samples eight distributed randomly over the time period were selected and further processed. The overall cut straw size was here 0.5 to 2 cm.

In all cases it was decided to apply the whole straw (stem and leaves)

Sampling

The overall sampling procedure applied, can be divided in four steps, see figure 1.

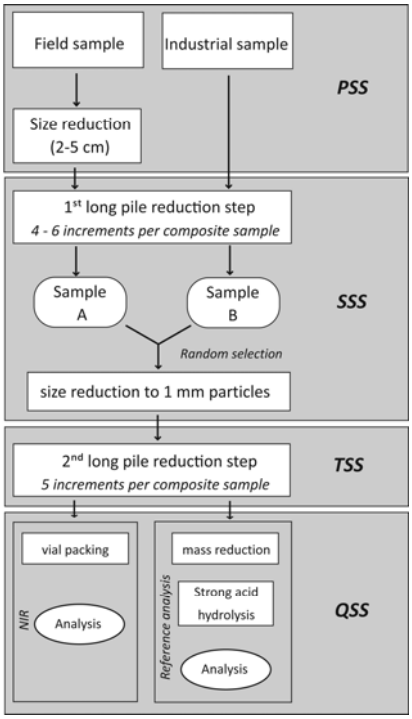

Figure 1. Sampling procedures as applied for mass reduction of samples. PSS: primary sampling step, SSS: secondary sampling step, TSS: tertiary sampling step, QSS: quaternary sampling step.

For the feasibility purpose a grab sampling approach was found acceptable at the primary sampling step, as maximal variation *between* samples was of prime interest (the primary goal being a maximally spanning multivariate calibration database). However all subsequent mass reduction and sample preparation steps were carried out in full compliance with the Theory of Sampling (TOS)[15,16,17] to ensure representative sub-samples for the different analytical steps delineated.

The first size-reduction step was only applied to the field samples, as it was carried out to obtain a uniform straw size of 2 cm. The industrial samples were already cut at the production line. The size reduction was done by a garden shredder (Bosch AXT 1600 HP). Between each sample the garden shredder was vacuum-cleaned and rinsed with compressed air, to avoid cross contamination.

The second sampling step applied the so-called *long pile* method. Due to the physical properties of the material (static electricity) special care had to be taken to avoid loss of material (Incorrect Preparation Error, IPE). Therefore a right-angle gutter with a length of 1 m was fabricated, figure 2. The material to be sub-sampled was placed into the gutter, while it was in an up-right position, by moving the delivering implement from one end to the other several times. This method is known as “bed-blending” and used by the industry in need of efficient pre-mixing.[15,18] The gutter was then tilted 90 degrees and a sub-sample was withdrawn with a specially designed scoop ensuring the withdrawal of a total, planar-parallel cross section of the pile. The length of the scoop is 13 cm and the width 15 cm, the latter corresponding to 15% of pile length. The material was poured back into the original storage container, and the procedure was repeated until the primary sample was split in two approximately uniform sub-samples; an A and a B sample respectively. For further analysis sample A or B was chosen at random.

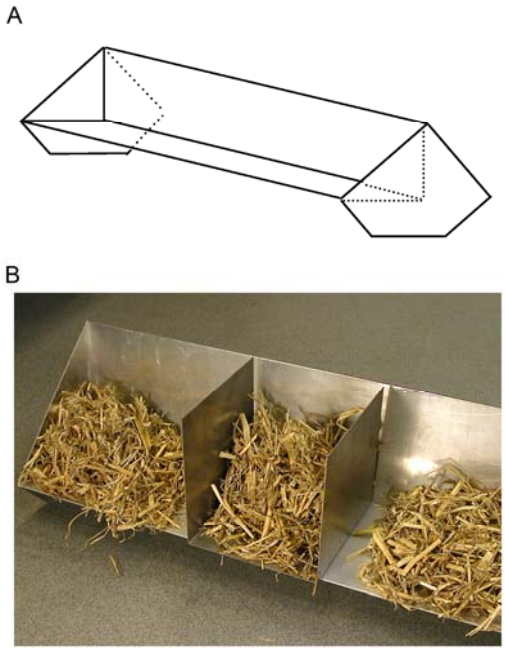

Figure 2. The gutter fabricated for mass reduction of primary field samples. A) Schematic view. B) Increment extraction from the gutter during secondary sampling step.

The resulting sub-sample was now further size reduced to a particle size of 1 mm, in a mill; Foss Cyclotec^{TM} 1093 (Foss, Denmark).

Finally the material for NIR and strong acid hydrolysis was further sub sampled by applying the long pile method once again, but this time at a distinctly smaller scale with a 50 % reduced gutter length. The scoop applied had a length of 13 cm and a width of 0.5 cm, the latter corresponding to 1 % of the total pile length.

For NIR approximately 2 x 5 g wheat straw was extracted as composite samples and packed directly in the NIR vials. For reference analysis 2 g samples was withdrawn, also applying composite sampling and packed in plastic bags. From these bags 4 x 0.16 g of material was extracted for reference analysis. Intentionally the material should have been poured out of the bags for further “long-pile” reduction, thereby obtaining the 0.16 g pr analysis. The material was however static electric and it was not possible to pour everything from the bag. Furthermore a fraction of the material poured out was lost. It was therefore decided to extract the 0.16 g of material directly from the bag, acknowledging that this procedure was not in complete compliance with TOS. However these 0.16 g increments were obtained by composite sampling, with thoroughly mixing between each increment. Any residual sampling error from this procedure will show up as an inflation of RMSEP in the final calibration models and will thus be fully accessible for quantitative assessment.

Dry matter and ash

The dry matter content was determined by weighing and drying in a Mettler Toledo moisture Analyzer HR83, Halogen (Mettler Toledo, Denmark).

For ash determination 0.5 g of material was weighed directly in a pre-weighed crucible and incinerated at 550°C for 3 hours. Doublet determinations were carried out.

Carbohydrate analysis

To quantify sugar polymers in wheat straw, a two step acid hydrolysis was carried out. In the first step 1.5 mL 72% (w/w) H_2SO_4 was added to 160 mg dry matter and incubated at 30°C for 1 hour. In the second step the samples were diluted with 42 mL milipore water and autoclaved at 121°C for 1 hour. The hydrolysates was filtered and analysed for sugars on HPLC. The Klason lignin content was determined as the weight of the dried filter cake subtracted the ash content. The concentration of sugar monomers were measured by HPLC.

The recovery of D-glucose, D-xylose, and L-arabinose was determined by standard addition of sugars to samples before autoclavation.

HPLC analysis

The amount of sugar monomers; glucose, xylose, and arabionse were determined after separation on a HPLC-system (Shimadzu) with a Rezex ROA column (Phenomenex) at 63°C with a flow rate of 0.6 $mL \cdot min^{-1}$. As an eluent 4 mM H_2SO_4 was applied. The detector used was a refractive index detector (Shimadzu Corp., Kyoto, Japan)

Near infrared spectroscopy

NIR spectroscopy was based on, a Quant FT-NIR instrument with an InAs detector (ABB, Q-interline, Denmark). The instrument was fitted with a powder sample holder, rotating the vial while collecting 128 scans, which was averaged as the final spectrum recording/storage for each sample. By this approach an average of the material placed peripherally in the vial was obtained. For each primary sample approximately 2 x 5 g of the 1 mm particles was packed, thereby basing the final analytical results on duplicate measurements. Alternative resolutions of 8 cm^{-1} and 32 cm^{-1} were investigated in this study based on previous experience. NIR spectroscopy was carried out in the range 4,000 – 10,000 cm^{-1}. Before measuring a background spectrum was recorded on a blank PTFe vial.

Data Processing

For quantifying concentrations of glucan, xylan, arabinan, and lignin from the NIR spectra, chemometric multivariate calibrations using Partial Least Square regression (PLS-1) were performed.

In PLS-1 modeling the goal is to separate the structural part of the spectral information which is relevant for describing the X-Y regression relation from the "noise" part. As the regression relationship is built directly between the independent X-matrix and the dependent Y-vector, the effective model dimension is often severely reduced compared to the initial number of wavelengths recorded, i.e., a model applying only a few latent variables (PLS components) is usually obtained. In the present study the NIR spectra formed the independent data matrix (X) which was calibrated against the dependent reference analysis (y): the concentration of the analytes. Full chemometric description of this method can be found in elsewhere.[5,7,19,20]

In the first modeling step the effect of alternative pre-processing techniques on the **X** matrix were investigated for all analytes:

Multiplicative Signal Correction (MSC)

- Full correction (both alfa and beta correction)
- Alfa correction
- Beta correction

Savitzky – Golay

- 1st derivative
- 2nd derivative

Description of these techniques can be found elsewhere. [5,6,21,22,23]

Esbensen 2006, recommends that for this type of restricted comparative purpose, cross-validation may be used; in fact, for this specific purpose (only), cross-validation is optimal. Due to the relative small number of samples in the feasibility study, full cross-validation was used in this context.

For each analyte a Jack-knife approach was applied for variable selection in the X-data, thereby estimating the wavenumbers relevant for optimizing the X spectral interval w.r.t. maximal correlation to the variation in the Y-vectors. [24,25]

For evaluating the prediction ability of the resulting regression models, different validation methods were compared; 4-segment cross validation vs "test-set" validation. Cross validation was applied both in model training and calibration, while the "test-set" was used in the final validations exclusively. The test-set, strictly speaking, was not a *true* test-set as the samples had also been applied in the X-variable selection step. It however served the purpose of the most stringent testing of the prediction ability of the final NIR models in a similar sense, and was the best possible approximation to a real test-set validation procedure. A true test set should have been based on new field and industrial samples exclusively.

For evaluating the models prediction performance and simultaneously finding the optimal number of effective PLS components, a measure of the average prediction error; the Root Mean Square Error of Prediction (RMSEP) was calculated:

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (y_{i\_predicted} - y_{i\_reference})^2}{n}}$$

where y is the dependent variable and n is the number of samples. Test set RMSEP is a reliable measure of the average error encountered when using the model for prediction of future, similar samples, Esbensen 2006. The RMSEP statistic must be as small as possible for a model to be performing "optimally".

Furthermore a relative % RMSEP measure was calculated allowing comparison of RMSEP to the calibration interval for the given analytes;

$$\%\,\mathrm{RMSEP} = \frac{\mathrm{RMSEP}}{y_{max} - y_{min}}$$

For all samples, duplicate measurements NIR were carried out. Before building the calibration models these were assessed in a Principal Component Analysis (PCA) and spectral outliers in the X-space deleted. In the calibration and validation model, all non-deleted duplicates were subsequently averaged.

Data processing was carried out using the UNSCRAMBLER software ver. 9.7 (Camo, Norway).

Results and discussion

Sample Selection

Based on a working hypothesis that straw sugar concentration to some extent may be correlated to colour (significant field and plant variations), the samples for reference analysis were selected via a PLS-1 model applying colour as a semi-quantitative y-variable. This hypothesis follows the rationale that when a straw sample lies on a field during a rainy period of increasing duration, the colour of the straw shifts from yellow to greyish brown. During this period a part of the water soluble fraction (hemicelluloses) may be washed out and the sample in general will become subject to microbial degradation, both resulting in a decrease in sugar concentration. If this hypothesis has merit, the set of reference samples selected will allow an approximate maximized span in sugar concentration (lignin following suit as per closed-array correlations).

The colour of each sample was evaluated visually and assigned a single digit value. This resulted in a range from 1 to 11, 1 being the darkest colour (grey/brown), and 11 the lightest (yellow), see Figure 3. From this PLS model, initially 18 samples were chosen for first foray reference

analysis. The samples chosen were the ones that had the largest contribution to the span of the first and second X – space scores.

Figure 3. The colour classification of the wheat straw samples. From left to right: grey to yellow

From these 18 samples a tentative first PLS-1 model for glucan was built and compared to the PLS-1 colour model. The two models were found to have very similar model features (loading weights), indicating that the information in the NIR spectra was indeed relevant for modelling the sugar concentration in the sample. Subsequently, a multi-step procedure was employed to pick out 5 to 6 additional samples at a time for reference analysis, with the objective of augmenting the relevant calibration concentration span as much as possible. This resulted in a final training set with a total of 44 samples for the carbohydrate and lignin models. The concentration intervals and average values for each analyte are shown in table 2. As seen, a relatively narrow spread in the analyte concentrations is on record, fully as expected however. The composition of wheat straw documented in the literature is 30-40 % cellulose, 20-50% hemicelluloses, and 10-20% lignin on average.[26,27,28] 50% as the upper limit was only reported by Nag 2008, and seems high compared to the levels else reported; 20-30% and 26.4%.

Table 2. Minimum, maximum, average concentration, and coefficient of variation for each analyte in the 44 feasibility study samples.

	Composition of straw [g·(100 g DM^{-1})]			**Coefficient of variation**
	min	**max**	**average**	
Cellulose (Glucan)	34.9	40.3	37.6	0.036
Hemicellulose	21.0	25.9	23.8	0.048
Xylan	18.9	22.8	21.2	0.047
Arabinan	2.1	3.1	2.6	0.093
Klason lignin	18.8	24.6	20.5	0.060
Ash	1.6	6.9	3.9	0.31
Residual	7.7	17.5	14.2	-

Before assessing the effect of alternative pre-processings, a PCA model on NIR spectra alone and PLS-1 regression models for each analyte: glucan, xylan, arabinan, and lignin were built for both 8 cm^{-1} and 32 cm^{-1} spectra. The sole purpose of these was to identify and delete any possible strong outliers in the data set (e.g. as possibly due to excessive microbial degradation, or due to unrecognized components of non-wheat straw, e.g. barley). This analysis was carried out on data that had not been pretreated.

From this PCA analysis no such spectra were identified. From the accompanying initial PLS-regression analysis however, a few outlying samples were indeed identified by applying modeling T-U plots and predicted vs. measured plot (no 49 for glucan and 14 for lignin). Including these led to a highly significant direction shift in the models; only these two samples were excluded from the data prior to investigating the effect of the different pre-processing techniques.

Pre-processing

NIR spectra obtained after applying the different pre-processing techniques on the X-data for 32 cm^{-1} are seen in Figure 4.

While the overall trend investigated was the same for all analytes, the detailed effect of pre-treatment was to some extent associated with the specific Y-variable, therefore it was evaluated separately for each analyte. Before investigating the effect of the pre-treatment, the duplicate measurements were averaged, reducing the total no. of samples to 42-44 depending on the analyte. Furthermore the spectral wavenumber intervals were reduced, spanning only the intervals from 4000 – 7180 cm^{-1} for 8 cm^{-1} and 4000 – 10000 cm^{-1} for 32 cm^{-1}. This reduction was carried out based on a visible judgement of the spectra alone, excluding severely degraded marginal spectral regions only.

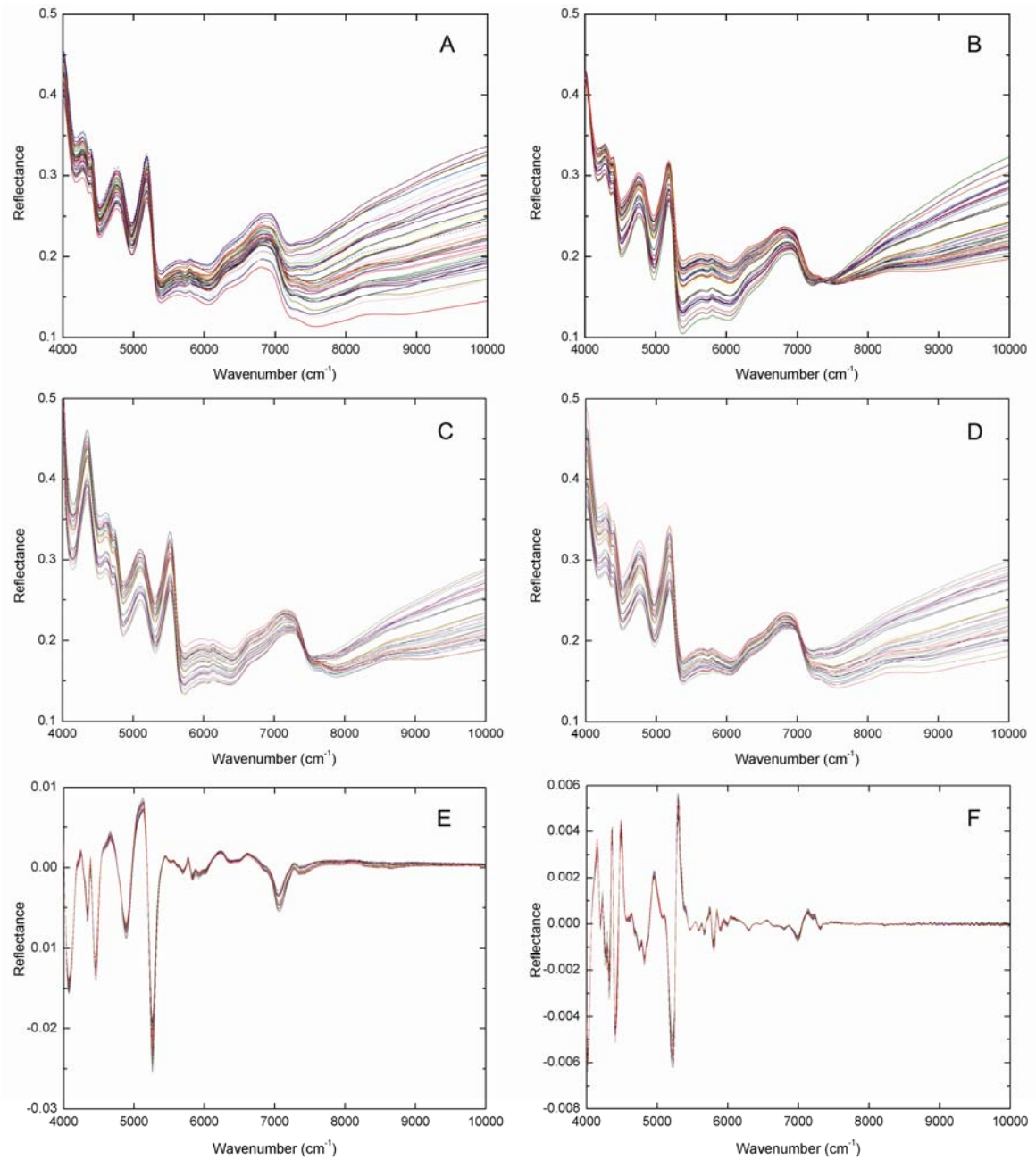

Figure 4. NIR spectra after applying all alternative pre-treatments on the X-data (32cm^{-1} resolution). A) No pre-treatment, B) MSC full, C) MSC alfa, D) MSC Beta, E) 1st derivative, and F) 2nd derivative.

The models for testing the pre-treatments were all validated with full cross validation. Outlier identification was carried out based on relevant T-U plots and the predicted vs. measured plots. The optimal results for both 8 and 32 cm^{-1} are seen in Table 3.

Table 3. Optimal calibration/pre-treatment assessment obtained for all four analytes. Validation: full cross validation. "predicted vs. measured" statistics: slope, correlation (r^2), RMSEP: root mean square error of prediction, no of PLS components, and fraction of outliers excluded. 1.der: S. Golay 1st derivative 2nd order polynomial, 2.der: S. Golay 2nd derivative 2nd order polynomial. **Bold**: Best of 1st order or 2nd order derived spectra.

	Pre-treatment	**slope**	**r^2**	**RMSEP**	**Components**	**Outliers**
8 cm^{-1}						
Glucan	**1. der**	0.67	0.57	0.91	7	5 (11%)
	2. der	0.41	0.46	1.01	6	5 (11%)
Xylan	1. der	0.80	0.69	0.53	10	1 (2%)
	2. der	0.56	0.58	0.61	5	3 (7%)
Arabinan	**1. der**	0.73	0.69	0.14	3	5 (11%)
	2. der	0.64	0.66	0.15	5	8 (18%)
Lignin	MSC full	0.75	0.61	0.46	10	9 (20%)
32 cm^{-1}						
Glucan	**1. der**	0.73	0.63	0.89	7	6 (14%)
	2. der	0.73	0.68	0.82	9	6 (14%)
Xylan	1. der	0.78	0.70	0.52	8	1 (2%)
	2. der	0.71	0.63	0.58	3	3 (7%)
Arabinan	**1. der**	0.74	0.71	0.13	3	5 (11%)
	2. der	0.77	0.73	0.12	4	8 (18%)
Lignin	MSC full	0.74	0.49	0.56	11	7 (16%)

In this first modelling step all outliers identified were excluded, in an attempt to obtain the best possible model. For all other analytes than xylan this resulted in a comparatively high outlier fraction. The different outlier types found were assessed as follows.

For xylan outliers were probably due to an analytical error, as a high relative difference between the duplicate measurements was observed for all. For glucan, arabinan, and lignin this was the case with only 30-40 % of the outliers found.

Different explanations for the remaining high outlier fractions were contemplated: 1. an error in the NIR spectra (X-data), 2. The error is raw material dependent, 3. A sampling error. The first two possible causes would expectably result in identical outlier identification for the different

analytes, which was not the case - making a physical sampling (TOS) error seem most likely. An error in the quaternary sampling stage in the procedure described in Figure 1 would lead to exactly this kind of deviation between the NIR spectra and corresponding reference values. [29,30,31] During increment delineation, special care was on the sampling, but due to the effects stemming from the static electric nature of the plastic material making up the sample bags, not all material had a uniform probability of ending up in the sample .i.e., the material handling of the reference sample was not in total compliance with the mass reduction principles in Theory of Sampling. This small deviation from TOS might have induced this particular error, thus emphasizing the importance of applying a correct TOS procedure from the beginning to the end.

Prediction evaluations were carried out employing the following "predicted vs. measured" statistics: slope, correlation (r^2), RMSEP, no of PLS components, and fraction of outliers excluded. In the case of Arabinan ($32cm^{-1}$) a model with a slope (0.76), correlation (0.68), and RMSEP (0.14) was obtained by applying full MSC. This was similar to the model obtained by applying a Savitzky-Golay 1st order derivative. The MSC model however applied three extra components and thus a Savitzky-Golay 1st order derivative was assessed as the best pre-treatment for arabinan. Similar reasoning was applied to the other analytes as well.

For all carbohydrates, the best models, on average, corresponded to applying a Savitzky-Golay pre-treatment. Based on the statistics and the number of components needed, applying a 1st derivative yield the best model for glucan and arabinan. This was also the case for xylan ($8cm^{-1}$), but here the 1st derivative model needed an extra five components, which may therefore very likely be the explanation for the improved prediction ability. Comparing models applying different resolutions (8 and 32 cm^{-1}); no effect was seen for xylan and arabinan. For glucan and lignin 32 cm^{-1} and 8 cm^{-1} spectra, yielded the best models respectively. Since no uniform trend is established both resolutions were further investigated.

Variable selection

As a means of optimising these models, an algorithmic variable selection approach applying the so-called Jack-knife approach ("Martens uncertainty test") was carried out.[24] This variable selection is expected to have a positive effect on the models in general. For completeness sake it was decided to investigate both the 1st and 2nd order derivative models.

The variable-reduced models were validated by a 4 segment cross validation. The outlier exclusion was done solely from the T-U plot of the components relevant for explaining Y. The best models obtained for each component are seen in table 4.

Table 4: Variable-reduced models for glucan, xylan, arabinan, and lignin by applying a 4 segment cross validation and Jack-knife pertubation. 1.Der: S. Golay 1st derivative 2nd order polynomial, 2.Der: S. Golay 2nd derivative 2nd order polynomial.

	Pre-treatment	**Smooth**	**slope**	**r^2**	**RMSEP**	**Components**	**No outliers**
8 cm^{-1}							
Glucan	2 der	5	0.88	0.83	0.60	5	5 (11%)
Xylan	2 der	5	0.90	0.82	0.43	5	1 (2%)
Arabinan	2 der	3	0.76	0.77	0.12	4	6 (14%)
Lignin	MSC full	-	0.84	0.72	0.38	7	8 (18%)
32 cm^{-1}							
Glucan	2 der	5	0.85	0.80	0.61	8	8 (18%)
Xylan	2 der	3	0.76	0.77	0.50	5	1 (2%)
Arabinan	2 der	7	0.75	0.75	0.13	4	6 (14%)
Lignin	MSC full	-	0.72	0.69	0.36	5	9 (20%)

Evaluating the accuracy (slope) and precision (r^2, RMSEP) for the carbohydrate models after variable selection, the best models were obtained with the high resolution spectra (8 cm^{-1}). The improvement seen in the glucan and arabinan models were only marginal however. A S. Golay 2nd order derivative with a three or five point smoothing window was optimal. For the glucan and arabinan model the number of outliers was high which might indicate overfitting and thus an unstable model.

The models obtained for glucan and xylan had good accuracies (0.88 and 0.90) and precisions (0.83 and 0.82). For arabinan the model accuracy and precision was in the lower end 0.76 and 0.77.

Comparing these results to the ones obtained prior to variable selection (Table 3), it is seen that for glucan and xylan the predicted vs. measured statistics was significantly improved applying less components even though the validation method here applied was less optimistic. For arabinan the model was only marginally improved applying an extra component. Furthermore it is seen that now a 2nd derivative is preferred over the 1st order derivative. Generally the 2nd order derivative is known to be more sensitive to noise and generate more artefacts[23], which without variable selection to some extend may be implemented in the model, thus yielding the worsened prediction ability recorded a prior to variable selection.

For glucan, xylan, and arabinan 5, 7, and 10 % of the original variable interval was selected respectively. The identified relevant wavenumbers were mainly placed in the following five regions:

a. 4000-4500 cm^{-1}: Combination band region (CH_3, CH_2, CH)
b. 4800-5000 cm^{-1}: OH combination band region
c. 5200-5500 cm^{-1}: C=O stretch 2nd overtone region
d. 5700-6300 cm^{-1}: CH 1st overtone region
e. 6800-7200 cm^{-1}: 1st OH overtone and CH 1st overtone combination region

These regions are depicted in the spectra seen in Figure 5.

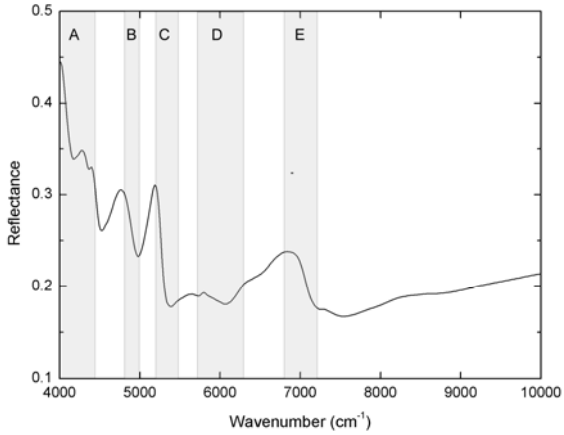

Figure 5. Relevant wavenumber bonds identified for xylan. A) Combination band region, B) OH combination band region, C) C=O stretch 2nd overtone region, D) CH 1st overtone region, and E) 1st OH overtone and CH 1st overtone combination region.

Besides these regions there were sporadically distributed wavenumbers identified for all carbohydrates. These may be artefacts generated by the 2nd order derivative and thus not essential for the model performance. Furthermore contemplating the 32cm^{-1} regression coefficients the same overall regions were identified but almost no sporadically distributed wavenumbers were identified, thus emphasising that these wavenumbers were indeed artefacts arisen from the combination of applying a high resolution and 2nd order derivative. This was tested by removal, and found to be true. In the model reported in Table 4 the sporadically distributed wavenumbers have been removed.

For lignin the precision obtained were broadly comparable applying both resolutions, but the highest accuracy was obtained applying 8 cm^{-1} spectra. This furthermore reduced the number of outliers by one, though two extra PLS components were needed. Overall it was concluded that applying the spectra with a resolution of 8 cm^{-1} yielded the best model. In this model 15% of the original dataset was applied, where the important wavenumbers identified were: 4250-4650, 4900-5000, 5640-6000, and 6100-6200 cm^{-1}.

For lignin the best model obtained only had a fair accuracy of 0.84 and a distinctly low precision of 0.72. The relative RMSEP for this model was 12%. Together with a high fraction of outliers, 18% lead to the overall conclusion; the risk of over-fitting is high for lignin. It might be possible to apply NIR for lignin monitoring, but with the current calibration sample set it is not possible to reach a stronger conclusion.

Validation of carbohydrate models

The final carbohydrate models were further evaluated by test-set validation. Having only 37-43 samples at disposal makes a test-set selection difficult, as the risk involved in reducing the data-set is critically high. Therefore the smallest test-set possible, still being realistic was decided on. Approximately 25% of the total data-sets were uniformly selected, obtaining test-sets that spanned the complete concentration intervals for each analyte. For evaluating the models, explained X-calibration variance, residual Y- validation variance, and predicted vs. measured plot were contemplated, see figure 6.

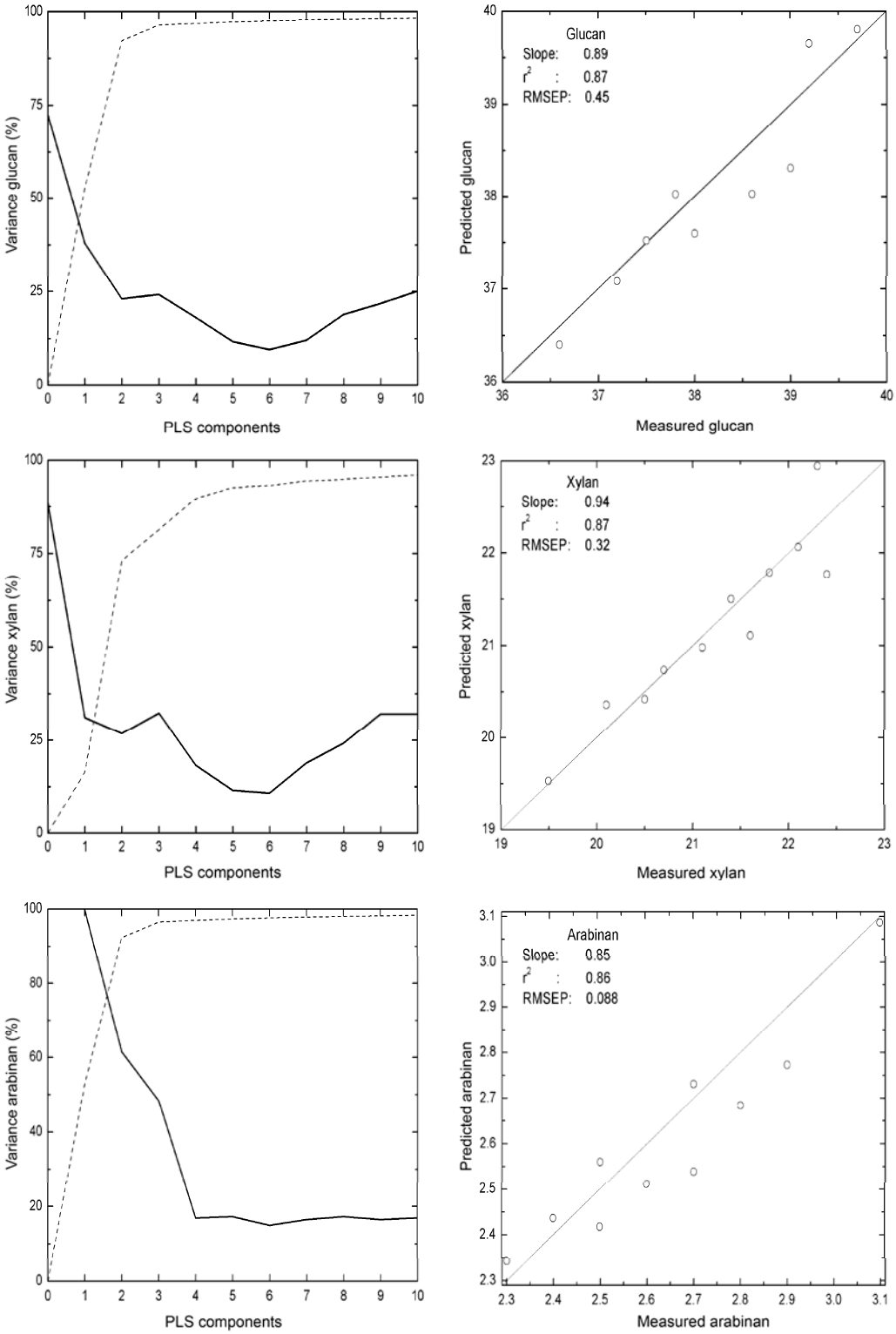

Figure 6. A) Explained X- calibration variance (--), residual Y-validation variance (-) and B) predicted Y vs measured Y, for models depicted in table 4 after applying a 25% test-set. RMSEP: Root mean square of prediction.

Contemplating the variance plot in Figure 6 between 80-97 % of the variable screened X-data could be used explaining 84%, 88%, and 86 % of the variation in glucan, xylan, and arabinan respectively. This strongly supports that Jack-knife variable selection did in fact select wavenumbers most relevant for modeling/predicting the different analytes.

Comparing to the cross validations, the accuracy for glucan decreased a bit, whereas for xylan and arabinan it increased. The increase for xylan was only minor however. For all models the precision increased, and the RMSEP decreased. This was obtained by applying the same number of components as for the cross validated models. For glucan and xylan the improvements in the models were only minor. For arabinan however the prediction ability improved much, which was unexpected. It may however be an overoptimistic result, that to some extend may be coursed by the presence of one extreme test-set sample. The sample was selected from one out of four samples in the region, thus representing 25% of the overall dataset and it was therefore chosen not to exclude it from the test-set. To avoid this, samples in the upper concentration range should be added.

The RMSEP for the model looked fair, comparing it to the calibration interval these corresponded to a relative % RMSEP of 8, 9, and 10 % for glucan, xylan, and arabinan respectively. These prediction errors are not fatally high, but the models do need further work before routine implementation.

In the literature no models have been found for wheat straw but similar models have been reported for corn stower.[11] In these models a SECV (which may be compared to our cross-validastion RMSEP) at 1.4, 0.95, and 0,20 wt% have been reported for glucan, xylan, and arabinan respectively. For comparative purposes these where converted into relative SECV% applying the min/max range also printed in the publication. The relative SECV % obtained were now 7, 9, and 5 % for glucan, xylan, and arabinan respectively. For glucan and xylan it is the same range as the one reported here, but for arabinan the results for corn stower was clearly better. The better result may be explained by a calibration range that was four times larger than the present one. For arabinan especially the concentration interval is very narrow 0.9 $g\cdot(100g\ DM)^{-1}$, and thus increasing the current calibration range would be beneficial. This demands positive identification of samples with a lower or high concentration of arabinan. A solution for this applied on corn stower[11] was to manually add or remove specific plant tissues such as leaves and nodes, thereby changing the content of the different sugars in the sample. Whether this would be possible for wheat straw has not been investigated in the present study.

Another solution may be to harvest the straw at different times during the ripening period as the carbohydrate content is changing as a function of ripening. This has also not been investigated here.

Conclusion

The objective of this work was to assess the feasibility of applying NIR spectroscopy for quantification of cellulose, hemicelluloses, and lignin in cut wheat straw. For this purpose wheat straw samples were collected all over Denmark and 44 samples were selected that gave a realistic span of the concentration intervals as can reasonably be expected in future samples. Alternative pre-treatment methods were investigate; applying the one found optimal for each analyte, test-set validated models with fair prediction abilities for glucan, xylan, and arabinan were obtained.

For lignin the number of outliers that had to be excluded to obtain a model was assessed as very high (18%) however and the % RMSEP was also relatively high (12%). From the present work it could thus not be concluded that the lignin content in the sample could be satisfactory quantified applying the reported method.

The models obtained for the carbohydrates all had intermediately high predictions errors (8-10 %), which however are in the same range as reported on corn stower. As a feasibility study these results are deemed fully acceptable. To our knowledge this is the first study dealing with NIR analysis on wheat straw.

Acknowledgements

This study would not have been possible without the help of many Danish farmers and “Halm 80” supplying the field samples applied, and Hans Møller, Vattenfall and operators at Køge Bio pellet factory supplying the plant samples. They are all warmly thanked for their contribution. Q-interline, Denmark is warmly thanked for their help with NIR measurements. The laboratory technicians Tomas Fernqvist, Ingelis Larsen, and Annette Eva Jensen are all acknowledge and thanked for their kind help with reference analysis.

References

1. EU Resolution. *Promotion of the use of Energy from renewable energy sources* (P6_TA(2008)0609)

2. J. Larsen, M.Ø. Petersen, L. Thirup, H.W. Li and F.K. Iversen, "The IBUS process - lignocellulosic bioethanol close to a commercial reality", *Chem. Eng. Technol*., 31, 765 (2008)

3. L. Oliveira, D. V. Evtuguin, N. Cordeiro, A.J. D. Silvestre, and A.M. Soares de Silva. "Chemical composition and lignin structural features of banana plant leaf sheath and rachis" in *Characterization of lignocellulosic materials*, Ed by T.Q. Hu., Wiley-Blackwell, (2008)

4. F.A. Agblevor, A. Murden and B.R. Hames, "Improved method of analysis of biomass sugars using high-performance liquid chromatography", *Biotechnol. Lett.*, 26, 1207 (2004)

5. H. Martens and T. Næs. *Multivariate Calibration.* Biddles Ltd. Great Britain (1991)

6. K.R. Beebe, R.J. Pell and M.B. Seascholtz, *Chemometrics a practical guide.* John Wiley & Sons. USA (1998)

7. K.H. Esbensen *Multivariate analysis in practice: an introduction to multivariate data analysis and experimental design*, 5th ed. CAMO Press. Norway (2006)

8. S. Jin and H. Chen "Near-infrared analysis of the chemical composition of rice straw", *Indust. crops and prod.*, 26, 207 (2007)

9. M. Sohn, D.S. Himmelsbach, F.E. Barton II, C.A. Griffey, W. Books and K.B. Hicks, "Near-infrared analysis of ground barley for use as a feedstock for fuel ethanol production", *Appl. Spectrosc.*, 61 (2007)

10. A.D. Richardson and J.B. Reeves III, "Quantitative reflectance spectroscopy as an alternative to traditional wet lab analysis of foliar chemistry: Near-infrared and mid-infrared calibrations compared", *Can. J. Forest Res.*, 35, 1122 (2005)

11. B.R. Hames, S.R. Thomas, A.D. Sluiter, C.J. Roth and D.W. Templeton, "Rapid biomass analysis, New tools for compositional analysis of corn stover feedstocks and process intermidiates from ethanol production", *Appl. Biochem. Biotechnol.*, 105-108, 5 (2003)

12. G.J. Faughey and H.S.S. Sharma, "A preliminary evaluation of near infrared spectroscopy for assessing physical and chemical characteristics of flax fibre", *J. Near Infrared Spectrosc.*, 8, 61 (2000)

13. J.B. Reeves III, "Near- versus mid-infrared diffuse reflectance spectroscopy for the quantitative determination of the composition of forages and by-products", *J. Near Infrared Spectrosc.*, 2, 49 (1994)

14. J.B. Reeves, T.H. Blosser and V.F. Colenbrander, "Analysis of silage composition by near infrared reflectance spectroscopy", *Opt. agricul.*, 1379, 28 (1991)

15. P. Gy *Sampling for analytical purposes*. John Wiley & Sons. Chichester, England (1998)

16. L. Petersen, C.K. Dahl and K.H. Esbensen, "Representative mass reduction in sampling - a critical survey of techniques and hardware", *Chemom. Intell. Lab. Syst.*, 74, 95 (2004)

17. L. Petersen, P. Minkkinen and K.H. Esbensen, "Representative sampling for reliable data analysis: theory of sampling", *Chemom. Intell. Lab. Syst*., 77, 261 (2005)

18. P. Gy, " Part IV: 50 years of sampling theory – a personal history", Chemom. Intell. Lab. Syst., 74, 49 (2004)

19. C.E. Miller, "Process analytical Technology", in *Process Analytical Technology*. Ed by K.A. Bakeev, Blackwell Publishing, UK p. 226 (2005)

20. H.R. Bjørsvik and H. Martens, " Data Analysis: Calibration of NIR Instruments by PLS Regression", in *Handbook of Near- Infrared Analysis* (3rd) , Ed by D.A. Burns and E.W. Ciurczak, CRC Press, Boca Raton p. 189 (2008)

21. A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data by simplified least squares procedure", *Anal chem*., 36, 1627 (1964)

22. M. Zeaiter, J.-M. Roger and V. Bellon-Maurel, "Robustness of models developed by multivariate calibration. Part II: the influence of pre-processings methods", *Trends anal. Chem*., 24, 437 (2005)

23. W.R. Hruschka, " Data Analysis: Wavelenght Selection Methods" in *Near-infrared Technology in the Agricultural and Food Industries (2nd ed)*, Ed by P. Williams and K. Norris American Association of Cereal Chemistry, Inc., St. Paul USA p. 39 (2001)

24. H. Martens and M. Martens, "Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR)", *Food qual. and Pref.*, 11, 5 (2000)

25. F. Westad and H. Martens, "Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression", *J. Near Infrared Spectrosc.*, 8, 117 (2000)

26 .M.H. Thomsen, A. Thygesen, H. Jørgensen, J. Larsen, B.H. Christensen and A.B. Thomsen, "Preliminary results on optimization of pilot scale pretreatment of wheat straw used in coproduction of bioethanol and electricity", *Appl. Biochem. Biotechnol.*, 129-132, 448 (2006)

27. M.J. Taherzadeh and K. Karimi, Bioethanol "Market and production Processes" in *Biofuels refining and performance*, Ed by A. Nag. Mc Graw Hill. USA p. 76 (2008)

28. C.M. Drapcho, N.P. Nhuan, T.H. Walker, *Biofuels engineering process technology*. McGraw Hill. USA (2008)

29. P.P. Mortensen and R. Bro, "Real-time monitoring and chemical profiling of a cultivation process", *Chemom. Intell. Lab. Syst.*, 84, 106 (2006)

30. J.B. Holm-Nielsen, C.K. Dahl and K.H. Esbensen, "Representative sampling for process analytical charatirization of heterogeneous bioslurry systems - a reference study of sampling issues in PAT", *Chemom. Intell. Lab. Syst.*, 83, 114 (2006)

31. K.H. Esbensen & P.P. Mortensen 2009, " Process Sampling (theory of Sampling, TOS) – the missing link in Process Analytical Technologies (PAT) in *Process Analytical Technology*. Ed by K.A. Bakeev, Blackwell Publishing, UK p. _ (2009)

IV

Power plant intake quantification of wheat straw composition for 2nd generation bioethanol optimization – a Near InfraRed Spectroscopy (NIRS) feasibility study

Carina J. Lomborg[a,*], Mette H. Thomsen[b], Kim H. Esbensen[a]

[a] Applied chemometrics, analytical chemistry, acoustic chemometrics, applied biotechnology, bioenergy and sampling (ACABS) Research group, Aalborg University; Niels Bohrs Vej 8, DK-6700 Esbjerg, Denmark,Telephone: (+45) 99407646; e-mail: cjp@aaue.dk

[b] Biosystems department, Risø National Laboratory for Sustainable Energy, Technical University of Denmark, Fredriksborgsvej 399, DK – 4000 Roskilde, Denmark

Abstract

Optimization of 2nd generation bioethanol production from wheat straw requires comprehensive knowledge of plant intake feedstock composition. Near infrared spectroscopy is evaluated as a potential method for instantaneous quantification of the salient fermentation wheat straw components: cellulose (glucan), hemicelluloses (xylan, arabinan) and lignin. Aiming at chemometric multivariate calibration, 44 pre-selected samples were subjected to spectroscopy and reference analysis. For glucan and xylan prediction accuracies (slope: 0.89, 0.94) and precisions (r^2: 0.87) were obtained, corresponding to error of prediction levels at 8-9 %. Models for arabinan and lignin were marginally less good, and especially for lignin a further expansion of the feasibility dataset was deemed necessary. The results are related to significant influences from sub-sampling/mass reduction errors in the laboratory regimen. A relative high proportion of outliers excluded from the present models (10-20%) may indicate that comminution sample preparation is most likely always needed. Different solutions to these issues are suggested.

Key words: Wheat straw, 2nd generation bioethanol, near infrared spectroscopy, sugar analysis, theory of sampling

1. Introduction

A resolution set by the European Union for 2020 is that 20 % of the EU energy consumption must originate from renewable energy sources (biomass, wind power etc.) and furthermore that biofuels should account for 10% of the total fuel consumption (EU resolution). To obtain the latter, both a quantitative growth and an increased efficiency of the existing biofuel production plants are needed. As a feed stock for bioethanol, agricultural lignocellulosic biomass such as wheat straw, rice straw, and bagasse may be used as feed stock in the 2nd generation context, as these are abundant around the world. (Oliveira et al. 2008)

In Denmark 25.6 % of the total agricultural area is cultivated with wheat. The average composition of wheat straw is 30-40 % cellulose, 20-50% hemicelluloses, and 10-20% lignin on average (Thomsen et al., 2006; Taherzadeh & Karimi 2008; Drapcho et al., 2008). Currently it is mostly combusted in coal-fired Combined Heat and Power plants (CHP), but research is carried out on combination of CHP and production of bioethanol from wheat straw, the so-called Integrated Biomass Utilization System (IBUS). According to Larsen et al. (2008) this is now close to commercial reality. As a mean of further optimizing this process, gaining more detailed process knowledge is highly valuable. It is of particular interest to be able to furnish bioreactor operators with a reliable feedstock composition already at the intake stage.

In a classical process technology context this could be done by extracting samples (at each point of interest in the process) which are sent for analysis at a centralized laboratory. This poses several problems; notably many samples, and a delay time of hours to days.

A much more desirable solution is implementation of the analytical facilities directly in the process, Process Analytical Technologies (PAT). This could either be done by placing the process analyses next to the production line (*at-line*) or in the process (*on-line*). With this approach it would be possible to monitor the production line from feed stock to the end product, with a much reduced delay time, effectively down to minutes – seconds. This allows obtaining knowledge of the chemical composition at different process stages in real-time, thus facilitating feed- forward improved regulation opportunities.

This is to some extent already seen at CHP facilities, where the feedstock quality is estimated from the moisture content in incomming straw bales. For straw to be used for bioethanol production this seems inadequate as this says nothing about the content of the fermentable sugars and lignin i.e., the relevant quality of the straw in this process context. Obtaining such knowledge has several advantages. For the feedstock on-line knowledge means that prices can be regulated according to sugar content and furthermore it can be used for feed-forward regulation of the pre-treatment processes and fermentation, thus improving both process efficiency as well as economics.

Several methods that are principally applicable as at- or on-line monitoring modalities have been investigated, one of which is Near InfraRed Spectroscopy (NIR). NIR spectroscopy is an

analytical technique that measures optic absorption spectral properties in organic molecules caused by combination bands and overtones in the region 12500 – 4000 cm^{-1} (800-2500 nm). The advantage of NIR is that the measurements are carried out in a fast non-destructive way. In the literature NIR has been applied on similar types of biomasses, e.g. barley, corn stover, foliar, and flax, (Jin & Chen 2007; Sohn M., 2007; Richardson A.D. and Reeves III J.B., 2005; Fauughey G.J. and Sharma H.S.S. 2000; Reeves III J.B. et al, 1991) thus it may also be applicable for direct wheat straw characterization.

Ultimately what is of highest interest to the bioethanol industry would be instantaneous compositional characterisation directly on the feedstock (straw bales and similar) for example by applying a handheld NIR spectrophotometer or a more permanent facility, see Fig.1. It would be only prudent to elucidate the limiting factors before such an undertaking, e.g.; how to compensate the field heterogeneity displayed by raw straw loads, bales or samples? How to assure a sufficiently robust sample illumination in the field?

Fig. 1. Handheld NIR on straw - the ultimate solution for compositional characterization of straw.

Before taking on this challenge, the present task is to carry out a feasibility study of the potential of NIR in this context. This study therefore investigates the potential of compositional analysis of cut-wheat straw by NIR spectroscopy, with respect to cellulose (glucan), hemicelluloses (xylan and arabinan) and lignin. As such the present study involves a brief focus on sampling (Theory of Sampling), representative mass reduction and analysis in the laboratory in order to *simulate* the procedures and intervention(s) necessary in order to reap the fullest benefit of field or on-line NIR analysis. Full technical details as to the many potential alternative spectral pre-processings, model optimization and validation has been described elsewhere (Lomborg et al. 2009). These are not addressed further here; below only the final optimized models are presented.

2. Methods

2.1. Samples

The first step was to collect samples as representative as possible with respect to the sugar and lignin concentration for the type of wheat straw on which the model has to be applied in the future. A minimum of 30 samples is needed for initial model evaluations; while as many as 100-300 are needed for calibration of robust methods for industrial use (Hames et al., 2003). The present feasibility study strikes a practical intermediate level.

In this case a total of 95 samples were collected in 2006 in Denmark; 65 originated from south east and west Jutland and Funen (field samples) while 30 originated from the island of Zealand (industrial samples). All material was harvested in the period from 2003 – 2006, the 2006 samples constituting the majority.

The 65 field sample set was collected in barns and on fields, whereas the 30 Zealand samples were sampled at the production line at Køge Bio pellets factory over a 30 day period. This was done in order to try to capture as large a compositional span as possible in the interest of securing the best possible calibration dataset. During field sampling, special effort was made to obtain samples of maximally different quality: from fresh straw directly as harvested in the field to spoiled straw that had been left on the field for an extended period of time or sometimes stored while it was still wet. This was considered an important experimental factor, as this could also be expected for a fraction of future feedstock samples. Furthermore it was hypothesized that samples that were almost totally spoiled would have become subject to microbial degradation and wash out of the water soluble fractions, thus producing a potentially enlarged span in sugar and lignin concentration. In all cases it was decided to apply the whole untreated straw (stem and leaves).

The industrial plant intake samples were a mix of wheat and barley straw; the compositional distribution was unknown. This choice was also a deliberate anticipation of possible future sampling settings (unavoidable mixtures of straw types, perhaps minor fraudulent feedstock declarations, other).

The overall original straw lengths ranged from 0.5 cm to approximately 25 to 30 cm, while all industrial samples were in the interval from 0.5 to 2 cm.

2.1.1. Sampling

In order to assess the feasibility of NIR-characterization vs. reference methods, the ultimate issue of representativeness with respect to starting material (the lot) can sometimes be left to be solved independently i.e., for this specific purpose it is only necessary to obtain representative splits from the field samples down to the sample being measured. Therefore the overall sampling procedure covered only the laboratory regime. As an efficient way of reducing the overall total

sampling error a procedure involving comminution, incremental sampling, mass reduction, and homogenization was designed, resulting in four steps from field sample to analysis:

1. Size reduction of primary field samples; stem and leaves were cut to approximately 2 cm
2. Long pile mass reduction of both field and industrial samples and subsequent size reduction to 1 mm, see Fig. 2.
3. Long Pile mass reduction of 1 mm comminute samples
4. Packing of vials for NIR analysis and plastic bags for reference analysis. Subsequent mass reduction for reference analysis

The procedures applied through all four steps were carried out in as full compliance with the Theory of Sampling (TOS) as possible to ensure representative sub-samples for the different analytical steps. In-depth treatment of the systematic of representative sampling can be found in Gy (1998), Petersen et al. (2004), Petersen et al. (2005), Esbensen et al. (2006).

Field samples had a harvesting cut size ranging from a few cm to 25 – 30 cm. In order to counteract heterogeneity mismatches between samples, they were in the first step chopped down to an equal particle size (approximately 2 cm) before being mass reduced (2nd step). This was done applying a commercial shredder of the type Bosch AXT 1600 HP. Between each sample the shredder was vacuum-cleaned and rinsed with compressed air, to avoid cross contamination.

For the second and tertiary sampling step incremental, composite sampling was applied. For this a right-angle gutter with a length of 1 m was fabricated. The sample was poured into this in a snake-like motion making sure to cover as many lengths as possible, thus ensuring maximal mixing while laying up the entire (comminuted) sample, see Fig 2. This procedure is known in bulk materials handling as “bed-blending” (Gy, 2004). Without going into further details, the long pile approach allows problem-specific, representative mass reduction, by fast manufacturing of fit-for-purpose implements as the one illustrated here. The long pile mass reduction principle is completely scale-invariant; it has been used on lot masses ranging from 1+ ton to 1 gram with commensurate differences in instrument size etc.

Fig. 2. Right angle gutter applied for long pile mass reduction of cut and particulate wheat straw. The gutter had two working positions: an upright (the one seen), applied when pouring the sample and a tilted (90°), applied for sample extraction.

After bed-blending, a complete, planar-parallel cross section (increment) of the pile was extracted with a scoop designed for the purpose. In the secondary sampling step the length of the scoop applied is 13 cm and the width 15 cm, the latter corresponding to 15% of pile length. After each increment withdrawal, the material was poured back into the container, and the procedure was repeated aiming at maximum further mixing before sub-sampling. This was carried out in repeated steps until the entire primary sample was split in two equal sizes; an A and a B sample. One of these samples (A or B) was selected at random. The particle size was now reduced to 1 mm in a comminution mill; Foss Cyclotec™ 1093 (Foss, Denmark).

In the tertiary sampling step only 50 % of the gutter length was applied, the scoop length now 13 cm and the width 0.5 cm, the latter corresponding to 1 % of the total pile length. Of the material extracted approximately 2 x 5 g wheat straw was packed directly in the NIR vials. For reference analysis a further 2 g sample was obtained and packed in a plastic bag. From the homogenized bag, the final 4 x 0.16 g material needed for reference analysis was extracted also by incremental sampling. The full procedure is outlined in Lomborg et al. (2009).

2.2. Dry matter and ash

The dry matter content was determined applying a Mettler Toledo moisture Analyzer HR83, Halogen (Mettler Toledo, Denmark). Double determinations were carried out.

Ash determination was carried out by weighing 0.5 g of the 1 mm material directly in a pre-weighed crucible and incinerated at 550°C for 3 hours. Doublet determinations were carried out.

2.3. Analysis of carbohydrates

A two-step acid hydrolysis was carried out to quantify the sugar polymers in the wheat straw. In the first step 1.5 mL 72% (w/w) H_2SO_4 was added to 160 mg dry matter and incubated at 30°C for 1 hour. In the second step the samples were diluted with 42 mL milipore water and autoclaved at 121°C for 1 hour. Subsequently the hydrolysates were filtered and the sugar monomer concentration determined applying High Liquid Performance Chromatography (HPLC). The HPLC system applied was a Shimadzu with a Rezex ROA column (Phenomenex) at 63°C with a flow rate of 0.6 $mL \cdot min^{-1}$. As an eluent 4 mM H_2SO_4 was applied. The detector used was a refractive index detector (Shimadzu Corp., Japan)

The recovery of D-glucose, D-xylose, and L-arabinose was determined by standard addition of sugars to two of the four samples before autoclavation.

Klason lignin content was determined as the weight of the dried filter cake subtracted the ash content.

2.4. Near Infrared spectroscopy

For Near Infrared Spectroscopy (NIR) spectroscopy, a Quant FT-NIR instrument with an InAs detector was applied (ABB, Q-interline, Denmark). For this purpose the instrument was fitted with a powder sample holder, rotating each sample while measuring, thereby obtaining an average of the material placed peripherally in the vial. For each sample a total of 128 scans were collected and averaged for the final spectrum recording/storage. For each primary sample, double determinations were carried out. Alternative resolutions of 8 cm^{-1} and 32 cm^{-1} were tested throughout all calibrations reported here. The 8 cm^{-1} spectra were found to be optimal. Spectroscopy was carried out in the range 4,000 – 10,000 cm^{-1}. Before measuring a background spectrum was recorded applying a PTFe vial.

2.5. Data analysis

For quantifying concentration of glucan, xylan, arabinan, and lignin from the NIR spectra, chemometric multivariate calibrations using Partial Least Square regression (PLS-1) were performed. In a PLS-1 model the regression relation is built directly between the independent X-matrix (NIR spectra) and the dependent y-vector (concentration of sugar or lignin), thereby splitting the spectral information in a part relevant for describing the sought for X-Y regression relation and a "noise" part (uncorrelated to concentration). As a consequence, the effective model dimension is often severely reduced compared to the original number of wavenumbers recorded, yielding a model applying only few latent variables (PLS components). Full chemometric description can be found in e.g. Martens & Næs (1991), Esbensen (2002), Miller (2005), Bjørsvik & Martens (2008).

To maximize correlations between the absorption NIR spectra and the concentration of glucan, xylan, arabinan, and lignin, different pre-processing techniques were tested: Multiplicative scatter correction and 1st and 2nd derivative (Savitzky – Golay). Description of these techniques can be found in Savitzky & Golay (1964), Martens & Næs (1991), Beebe et al. (1998), Hruschka (2001), Zeaiter et al. (2005). Furthermore a variable selection guided by a Jack-knife was applied. Description of this technique may be found in Martens & Martens (2000), Westad & Martens (2000).

For comparing these alternative pre-processing techniques <u>only</u>, a full cross validation was applied. Further model validation was carried out applying a four segment cross validation. The final models were evaluated applying a test-set, following the recommendations by Esbensen 2002.

Prediction evaluations were carried out employing "predicted vs. measured" statistics: slope, correlation (r^2), Root Mean Square Error of Prediction (RMSEP), % RMSEP, no of PLS components, and fraction of outliers excluded.

RMSEP evaluation was used for finding the optimal number of PLS components in the model and for evaluating the model's prediction performance (minimum RMSEP):

$$\mathrm{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n} (y_{i\,predicted} - y_{i\,reference})^2}{n}}$$

where y is the dependent variable and n is the number of samples. The RMSEP statistic must be as small as possible for a model to be performing "optimally".

To enable a fair comparison of the RMSEP to the calibration interval for the specific analyte across models, the relative RMSEP [%] was used, calculated as:

$$\%\,\mathrm{RMSEP} = \frac{\mathrm{RMSEP}}{y_{max} - y_{min}}$$

Data processing was carried out using the UNSCRAMBLER software ver. 9.7 (Camo, Norway)

3. Results & discussion

Of the original 95 samples collected, 44 were selected for training the multivariate calibration model a.o. using appropriate score plots from a principal component analysis (PCA). It was important to obtain samples displaying the largest possible span in the different concentrations, and which were simultaneously representative for future samples. All original samples were screened and selected by evaluating the color of the sample, as it was hypothesized that color is correlated to the leaching caused by either biodegradation or of washing the samples and therefore also to the sugar concentration in the samples. Yellow samples should thus have a higher concentration of xylan and arabinan than variously degraded grayish-brown samples. Furthermore, other types of information about each samples obtained during sample collecting was taken into account.

The span obtained for the different analytes is seen in table 1. The rather low compositional variability results from the fact that straw displays a very restricted biological variability at harvest time.

Table 1. Minimum, maximum, and coefficient of variation (C.V.) for each analyte in the 44 samples.

	Composition of straw [g·(100 g DM)$^{-1}$]		**C.V. [%]**
	Min	Max	
Glucan	34.9	40.3	3.6
Xylan	18.9	22.8	4.7
Arabinan	2.1	3.1	9.3
Lignin	18.8	24.6	6.0

After carrying out reference analysis, this working hypothesis (sugar concentration is correlated to overall straw color) was tested by plotting the average sugar concentration on each color level against the color assigned. It turned out that the samples situated in the high concentration range, were all samples collected right after harvest, whereas the lowest level all had been left on the field in a rainy period, i.e., the water soluble compounds was washed out. The overall correlation (r^2) obtained was 0.97, indicating that straw color may indeed be applied as a rough guide, a fact which will be useful in the field for selecting samples in the upper, intermediate and lower concentration ranges in the future.

3.1. Multivariate models

For obtaining multivariate models with optimal prediction abilities, different pre-treatments for the NIR spectra were evaluated, followed by an algorithmic variable selection approach (Jack-knife) to select the most influential wavenumbers (wavelengths). For sugar components a 2nd order derivative (employing a tree - five point smoothing window of a second order polynomial) was found to give the best model. For lignin a full MSC gave the best model. In Table 2 the models obtained applying a four-segment cross validation are summarized together with other relevant prediction model statistics.

Table 2: PLS-1 models for sugar components and lignin. “Components”: Number of PLS components applied in the model, numbers in brackets are the explained calibration and the residual validation variance, “Outliers”: Number of outliers excluded from the set during calibration, numbers in brackets are the corresponding percentages, “RMSEP”: Root Mean Square Error of Prediction.

	Components	**Outliers**	**slope**	**r^2**	**RMSEP**	**% RMSEP**
Glucan	5 (80% X, 83% Y)	5 (11%)	0.88	0.83	0.60	11
Xylan	5 (92% X, 83% Y)	1 (2%)	0.90	0.82	0.43	11
Arabinan	4 (97% X, 77% Y)	6 (14%)	0.76	0.77	0.12	13
Lignin	7(100% X, 72% Y)	8 (18%)	0.84	0.72	0.38	12

The models seen in Table 2 applied only 5, 7, 10 and 15 % of the original full-spectrum wavelengths regions for glucan, xylan, arabinan, and lignin respectively. The regions identified as most influential corresponded to expectations from general NIR experience and the chemical structure of the different analytes. For all models, except xylan, a relatively high outlier fraction (11-18%) had to be excluded from the models. Different explanations for these were contemplated and it was concluded that 30-40 % of these outliers could be explained by a reference analytical error. For the rest it was hypothesized that they were due to incorrect sampling errors during the complete procedure from field down to the final mass reduction step - from the field plastic bagging to the analytical determination.

The models obtained for glucan and xylan display slopes of 0.88 and 0.90, and correlations at 0.83 and 0.82 i.e., a good prediction accuracy and precision was obtained all applying five PLS components. For these models the relative RMSEP was 11%, fully acceptable for first model predictions in the present feasibility context.

For arabinan and lignin both the prediction accuracy (slopes of 0.76 and 0.84) and the precision (r^2: 0.77 and 0.72) were slightly lower. In the arabinan model 97% of the spectra total variance explained 83 % of the compositional variation applying only four PLS components. Contemplating the y-validation variance plot, Fig 3., it can be concluded that the model is quite acceptable. The prediction and accuracy statistics must be viewed in the light of the narrow calibration span for arabinan. Widening the span would very likely lead to improved prediction ability.

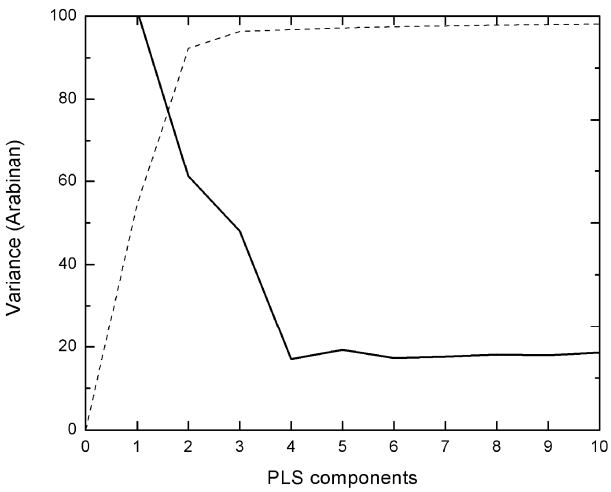

Fig.3. Residual Y validation variance plot for arabinan.

For lignin 100% X explained 72% of y using 7 components, indicating a significantly more complex model. Combined with the highest outlier fraction (18%), lead to the conclusion that the lignin model obtained is not fully acceptable yet, and that more samples spanning a wider compositional range must be included into the model before concluding more firmly as to the possibilities of routine lignin prediction by NIR in wheat straw.

3.2. Realistic validation

Employing only a fraction of the spectra (5, 7, and 10 %) always poses a risk of model overfitting, especially if based on cross-validation (Esbensen, 2002), and it was therefore decide to cary out a fully realistic validation of the prediction abilities of the best models obtained; glucan and xylan. This was done applying “test-set” procedure, *ibid*. Normally a test-set is obtained by collecting and analysing completely new samples. Because of the logistics of the present feasibility study, a slightly modified procedure was applied, in which approximately 25 % of the data set was extracted, and used in testing the new model which was established on the remaining 75%. While in the strictest sense this is not a true “test-set”, as it has been involved in the pre-treatment evaluation and the variable selection step, in the current context it is fully acceptable. The test-set was selected so as to completely span the concentration interval of the given analytes. This ensures the fullest, most realistic validation. Predicted vs. measured plots for the test-set validated sugar models are seen in Fig. 4.

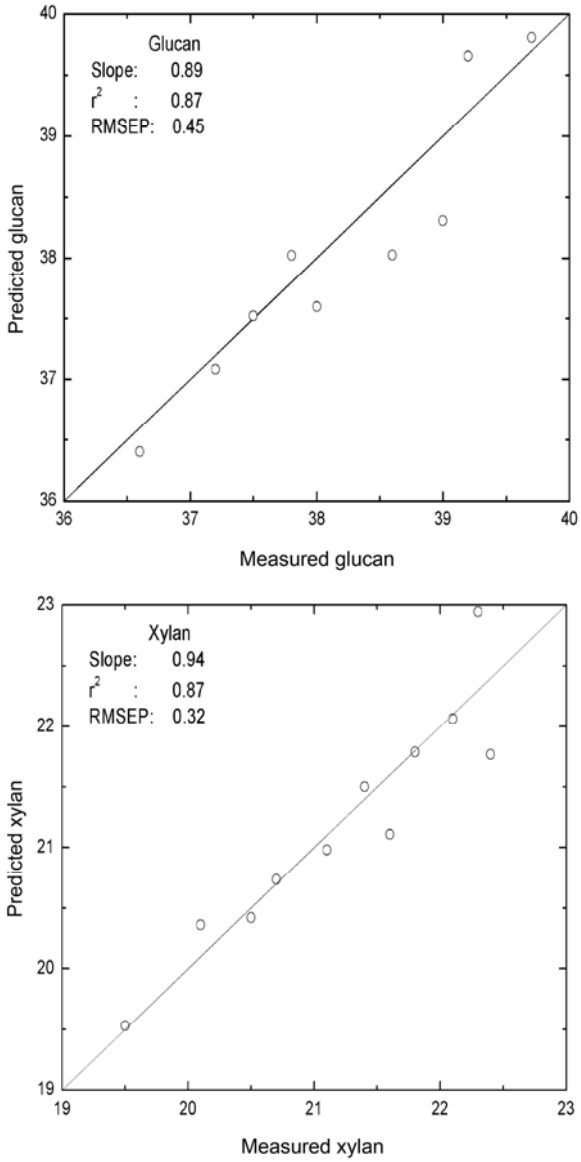

Fig 4. Predicted vs. measured plot for test-set validated models. A) Glucan, B) Xylan

These test-set validations for glucan and xylan showed good accuracies and precisions, indeed slightly better than what was assessed by the cross validated model (the differences reflecting a minor sub-setting sampling variability). Also an improvement in the relative % RMSEP was seen, which dropped to 8 and 9 % for glucan and xylan respectively. This indicates that the models obtained are stable, and can be employed for predicting the concentration in “new” samples. The models thus demonstrated the feasibility of NIR-characterization of glucan, xylan and to some extend arabinan in cut wheat straw. But all models reported here are only first foray attempts, and in no means ready for implementation in a real at-line monitoring context. For such a purpose the % RMSEP is still high.

Incorporating more samples in the entire concentration range in the training data-set is highly likely to contribute towards a further reduction. One should be prepared that it may not be enough for all analytes however, as especially for arabinan an extensive broadening of the calibration interval is necessary, but the feasibility of this undertaking appears strongly indicated. There may well also be further benefits associated with even more diligence in the various sampling steps, even though this was part of the experimental scope.

3.4. Obtaining new samples

As a guiding line for obtaining more samples in the upper and lower concentration interval of the C-5 sugars, the color of the samples can again be employed. For obtaining samples outside the calibration range this may not be enough, as the range that can be obtained in natural samples is limited. Therefore other solution needs also to be contemplated, three are suggested here:

1. New primary sampling strategy
2. Physical manipulation of primary samples
3. “Designed samples”

1) A new primary strategy may be employed. One solution could be setting-up an experiment where samples are collected in a predefined interval from the early reproductive state to well after the normal harvesting period. This is expected to influence the sugar composition and concentrations in the straw significantly, as sugars are remobilized from vegetative to reproductive organs during the ripening period (Yang et al. 2000). All that is needed for this to be successful is careful logistical planning. Whether this would in fact have a positive effect on the model is not certain, as the concurrent structural changes in the straw samples might be picked up in the NIR spectra as well, perhaps leading to other types of extreme samples (outliers) that cannot be fitted into the model, as they are too different. Only further experimentation can tell.

2) Physical manipulation of the sample may be beneficial for modeling. The positive correlation between the color of the straw and the C-5 sugar concentration showed that C-5 sugars are washed out of the samples. To obtain a lower concentration of xylan and arabinan, the solution might simply be to wash the straw with hot water. If successful this would yield the desired higher concentration of glucan and lignin, due to closed array correlation.

3) Finally it might also be possible to “design” calibration samples from collected wheat straw with the concentrations needed as suggested by Hames et al. (2003). In this experiment the entire straw including stem, leaves and nodes was applied, thus giving an average sugar content of the straw. By separating the component parts and mixing these in different proportions like done by Hames et al. (2003) with corn stover, it is expected to be able to obtain samples spanning an even broader calibration range.

None of the three suggested solution could be investigated as a part of the present feasibility study however. We here focused only on assessing the analytical possibility for measuring the relevant sugar concentration in cut wheat straw by NIR spectroscopy.

3.5. Direct on-line NIR analysis of untreated samples?

All models presented here are based on NIR measurements on materials comminuted to an average 1 mm particle size. This means that the present method can at best be incorporated at-line, where samples are collected and can be cut down before measuring. This would give a small delay in obtaining the desired results, but in no way comparable to the delay caused by e.g. wet chemical analysis or similar. Such a minor logistical infraction is likely still acceptable in most situations, but has to be evaluated for each specific process setting in question. This study did not address the issue of NIR analysis of completely untreated straw samples. After the present results, further experimentation on this issue would appear well justified however.

Together with broadening the interval of the existing data-set, the next step on the road to obtaining a model that can be fully implemented as an at- or on-line production facility is therefore to ensure new samples are covering the full compositional range met in nature, and to start elucidating the somewhat daunting issues regarding “as is” field samples vs. robust, reproducible NIR illumination and spectral acquisition.

4. Conclusions

From the results delineated above it is concluded that NIR spectroscopy is a viable modality for quantification of the selected bioethanol-related sugars in comminuted wheat straw. The models obtained for three carbohydrates have relative % RMSEP at the level of 10%. While broadly satisfactory for a feasibility study, this level also signifies that the models are not implementable in a professional monitoring context. This is not surprising considering the restricted number of available samples. In order to improve model performances, it would be beneficial to obtain more samples in the peripheral concentration intervals, as especially the arabinan model suffered from a very narrow calibration range. Unfortunately the compositional span is not a directly observable feature in the field, although this study demonstrated a useful indirect correlation between straw color and sugar content.

This study also reflected upon the critical success factor of representative sampling and mass reduction, which must be considered in all calibrations of multi-channel instrumental analytical techniques, NIR no exception. This issue does not always receive its proper attention. Interestingly were it possible to apply NIR-analysis on completely untreated materials (completely untreated wheat straw in the present case) the sample preparation issues encountered would all be eliminated. The primary field issues were kept out of the present study, and will pursued elsewhere.

Perhaps the relative high proportion of outliers excluded from the present models indicates that a minimum of comminution sample preparation is most likely always needed.

Acknowledgements

This study would not have been possible without the help of many Danish farmers and “Halm 80” supplying the field samples applied, and Hans Møller, Vattenfall and operators at Køge Bio pellet factory supplying the plant samples. The people are all warmly thanked for their contribution. The people at Q-interline, Denmark are warmly thanked for their help with NIR measurements. The laboratory technicians Tomas Fernqvist, Ingelis Larsen, and Annette Eva Jensen are all acknowledge and thanked for their kind help with reference analysis.

References

Beebe, K.R., Pell, R.J., and Seascholtz, M.B., 1998. Chemometrics a practical guide. John Wiley & Sons. USA

Bjørsvik, H.R. and Martens, H., 2008. Data Analysis: Calibration of NIR Instruments by PLS Regression, in: Burns, D.A. and Ciurczak, E.W. (Eds), Handbook of Near- Infrared Analysis (3rd), CRC Press, Boca Raton p. 189

Drapcho C.M., Nhuan N.P., Walker T.H., 2008. Biofuels engineering process technology. McGraw Hill. USA

Esbensen, K.H. 2002. Multivariate analysis in practice: an introduction to multivariate data analysis and experimental design, 5th ed. CAMO Press. Norway. Reprint of book from 2001.

Esbensen, K.H., Friis-Petersen H.H., Petersen L., Holm-Nielsen J.B., and Mortensen P.P. 2006. Representative process sampling-in practice: Variographic analysis and estimation of total sampling errors (TSE). Chemom. Intell. Lab. Syst., 88, 41

EU Resolution. Promotion of the use of Energy from renewable energy sources (P6_TA(2008)0609)

Faughey, G.J. and Sharma, H.S.S., 2000. A preliminary evaluation of near infrared spectroscopy for assessing physical and chemical characteristics of flax fibre. J. Near Infrared Spectrosc. 8, 61-69.

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

Gy, P, 2004. Part IV: 50 years of sampling theory – a personal history. Chemom. Intell. Lab. Syst., 74, 49

Hames, B.R., Thomas, S.R., Sluiter, A.D., Roth, C.J., and Templeton, D.W., 2003. Rapid biomass analysis, New tools for compositional analysis of corn stover feedstocks and process intermidiates from ethanol production. Appl. Biochem. Biotechnol. 105-108, 5-16.

Hruschka, W.R., 2001. Data Analysis: Wavelenght Selection Methods. In: Williams P. and Norris K. (Eds) Near-infrared Technology in the Agricultural and Food Industries (2nd ed), by American Association of Cereal Chemistry, Inc., St. Paul USA **p. 39**.

Jin, S. and Chen, H., 2007. Near-infrared analysis of the chemical composition of rice straw. Indust. crops and prod., 26, 207-211.

Larsen, J., Petersen, M.Ø., Thirup, L., Li, H.W. and Iversen, F.K., 2008. The IBUS process - lignocellulosic bioethanol close to a commercial reality. Chem. Eng. Technol. 31, 765-772.

Lomborg, C.J., Jensen, E.S., and Esbensen, K.H., 2009. NIR- characterization of (cut) wheat straw for bioethanol production – a feasibility study. Submitted to J. Near Infrared Spectrosc.

Martens, H. and Martens, M., 2000. Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). Food qual. and Pref . 11, 5-16.

Martens, H. and Næs, T., 1991. Multivariate Calibration. Biddles Ltd. Great Britain

Miller, C.E., 2005. Process analytical Technology. In: Bakeev, K.A. (Ed) Process Analytical Technology. Blackwell Publishing, UK p. 226

Oliveira, L., Evtuguin, DV, Cordeiro, N., Silvestre, A.J.D., and Soares de Silva, A.M., 2008. Chemical composition and lignin structural features of banana plant leaf sheath and rachis, in: (Eds) T.Q. Hu, Characterization of lignocellulosic materials, Wiley-Blackwell

Petersen, L., Dahl, C.K., and Esbensen K.H., 2004. Representative mass reduction in sampling - a critical survey of techniques and hardware. Chemom. Intell. Lab. Syst. 74, 95-114.

Petersen, L., Minkkinen, P, and Esbensen K.M., 2005. Representative sampling for reliable data analysis: theory of sampling. Chemom. Intell. Lab. Syst. 77, 261-277.

Reeves, J.B., Blosser, T.H., and Colenbrander, V.F., 1991. Analysis of silage composition by near infrared reflectance spectroscopy. Opt. agricul. 1379, 28-38.

Richardson, A.D. and Reeves III, J.B., 2005. Quantitative reflectance spectroscopy as an alternative to traditional wet lab analysis of foliar chemistry: Near-infrared and mid-infrared calibrations compared. Canadian Journal of Forest Research. 35, 1122-1130.

Savitzky, A. and Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedure. Anal. chem. 36, 1627-1639.

Sohn, M., Himmelsbach, D.S., Barton II, F.E., Griffey, C.A., Books, W., and Hicks, K.B., 2007. Near-infrared analysis of ground barley for use as a feedstock for fuel ethanol production. Appl. Spectrosc. 61,

Taherzadeh, M.J. and Karimi, K. 2008. Bioethanol: Market and production Processes. in Nag, A. (Ed) Biofuels refining and performance. Mc Graw Hill. USA p. 76

Thomsen, M.H., Thygesen, A., Jørgensen, H., Larsen, J., Christensen, B.H. and Thomsen, A.B., 2006. Preliminary results on optimization of pilot scale pretreatment of wheat straw used in coproduction of bioethanol and electricity. Appl. Biochem. Biotechnol. 129-132, 448-460.

Westad, F. and Martens, H., 2000. Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression. J. Near Infrared Spectrosc. 8, 117-124.

Yang, J., Zhang, J., Huang, Z., Zhu, O., and Wang, L., 2000. Remobilization of Carbon Reserves Is Improved by Controlled Soil-Drying. Crop Sci. 40, 1645–1655

Zeaiter, M., Roger, J.-M., and Bellon-Maurel, V., 2005. Robustness of models developed by multivariate calibration. Part II: the influence of pre-processings methods. Trends Anal. Chem. 24, 437-445.

V

Available online at www.sciencedirect.com
ScienceDirect
Journal of Biotechnology 133 (2008) 162–169
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

At-line determination of octanoic acid in cultivation broth-An electronic tongue (ET) feasibility study

Carina Juel Lomborg [a,*], Lars Wiebe [b], Kim H. Esbensen [a]

[a] *ACABS (Applied Chemometrics, Analytical Chemistry, Applied Biotechnology, Bioenergy and Sampling research group), Aalborg University Esbjerg, Denmark*
[b] *Danisco A/S, Tårnvej 25 DK-7200 Grindsted, Denmark*

Received 1 March 2007; received in revised form 26 August 2007; accepted 20 September 2007

Abstract

Production monitoring of "natural" 2-heptanone from octanoic acid in an industrial fed-batch cultivation based on *Penicillium roqueforti* requires development of a method for determination of octanoic acid dissolved in the water phase. An electronic tongue array using six non-specific potentiometric sensors with solid inner contact, and a pH electrode, has been introduced by spiking octanoic acid to a substrate obtained from four different cultivations, representing variations in the relevant industrial matrix. Multivariate calibration was performed on acid concentrations spanning 0.65–20 mmol l^{-1}. Excluding the lowest concentration a global Partial Least Square regression model with a predicted versus measured correlation of 0.98 and a relative root mean square error of prediction of 5.1% (ln units) (RPD = 5.5) signifies a highly acceptable prediction facility. This model was further tested by subjecting it to undiluted as well as diluted samples obtained from a cultivation process in which octanoic acid was catabolized; this led to acceptable prediction errors within the same range as for the global model. It is concluded that the ET sensor array can be applied for determination of octanoic acid in cultivation systems of the general *P. roqueforti* type.

Keywords: Electronic tongue; Potentiometric sensors; *Penicillium roqueforti*; At-line monitoring; Cultivation broth; Multivariate calibration

1. Introduction

Octanoic acid can be used for production of the "natural" aroma 2-heptanone by addition to a cultivation broth containing the filamentous fungi *Penicillium roqueforti*. 2-Heptanone is a known important aroma constituent in Roquefort cheeses and is therefore used in products like salad dressings, soups and crackers to simulate blue cheese flavour. Different methods have been reported for production of this aroma (Creuly et al., 1992; Larroche et al., 1989; Park et al., 2000), one being a fed-batch technique, where octanoic acid is added in the feed (Creuly et al., 1990). When added to the broth, octanoic acid is catabolized in the beta oxidation pathway by *P. roqueforti* (McSweeney and Sousa, 2000). If concentrations are too low, the octanoic acid is utilized solely as a carbon source i.e. turned into carbon dioxide only at the expense of the desired product: 2-heptanone. According to (Gehrig and Knight, 1963) a concentration of 1 μM leads only to carbon dioxide, whereas 2-heptanone production is registered based on an initial octanoic acid concentration at 20 μM. For high concentrations different results have been reported, e.g. inhibition of the microorganism (lag phase) followed by a degradation of the acid according to (Lawrence, 1966), whereas (Larroche et al., 1994) have reported total inhibition of the microorganism at concentrations above 5 mM. These variations and conflicting results may be due to interacting factors, e.g. different treatments of the spores applied during cultivation, as storage time and temperature is known to effect the subsequent methyl ketone formation (Gehrig and Knight, 1963), or they may be due to small differences in the strain applied (Larroche et al., 1989).

Even though production at high concentrations may be possible, it is severely disadvantaged by a prolonged lag phase; therefore a production at an intermediate concentration in the proximity of 5–10 mM is desirable from an industrial point of

Abbreviations: ET, electronic tongue; PDA, potato dextrose agar; PLS, partial least square; RPD, ratio of standard error of performance to standard deviation; RMSEP, root mean square error of prediction; TSB, tryptic soy broth.

* Corresponding author.
E-mail address: cjp@aaue.dk (C.J. Lomborg).

0168-1656/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2007.09.013

view. To enable this, an on-line or at-line analytical method that is sensitive to the octanoic acid concentration in the water phase within a relevant range, say from 2.5 to 20 mM is required.

A method suggested by (Creuly et al., 1990) is a pH state, which counteracts the decrease in H^+ concentration, as octanoic acid is degraded, by adding the octanoic acid as the feed, thereby enabling control of the fed-batch. In a later paper by (Larroche et al., 1994) this method fails however, due to a parallel H^+ generating reaction taking place, making complex processes difficult to monitor using this method.

Other methods reported in the literature are determination of octanoic acid by applying a chromatographic method; either GC with different detectors or HPLC, see Table 1. GC and HPLC are both two-step procedures, involving a sample pre-treatment before analysis. This particular pre-treatment is crucial for the analysis to work: the pre-treatment is finely adjusted to the specific sample and chromatographic equipment, thereby minimizing the analytical error on the result. In the methods reported in Table 1, derivatization and extraction are applied in all instances, except from the method by Kellerhals et al. (1999), where a filtration through a specific membrane was reported. The pre-treatment time for a derivatization or an extraction was typically in the range from 30 min to 12 h and the filtration was momentary. All these methods are commonly used in the laboratory as they are robust and precise, but for monitoring and control of a cultivation, only the method reported by Kellerhals et al. (1999) is fast enough.

From this ambiguous state of affairs a new process analytical method, facilitating reliable monitoring of low concentrations of octanoic acid during fed-batch production was needed. Based on the work by (Creuly et al., 1990) who showed that an ion selective electrode works—but also taking into account the failure of this method reported by (Larroche et al., 1994), a feasibility study of the electronic tongue (ET) used in an at-line configuration was carried out in the present study. The ET uses the basic principle of the ion selective electrode, but expands the number of sensors, solving the issue presented by (Larroche et al., 1994).

The Electronic Tongue consists of an array of non-specific electrodes which are applied directly to the sample. No pre-treatment is necessary, hence the method is fast, but often a simple filtration of the sample is recommended however, as this prolongs the life time of the electrodes significantly without any loss of analysis time. A large number of potential ET electrodes exist. Typically electrodes for an ET array are selected in a screening experiment, where the sensitivity of the electrodes towards a given compound in a specific matrix is determined. After selecting the electrodes with the largest *cross-sensitivity*, the array is calibrated. As the electrodes are non-specific, they are sensitive to both the relevant compound (octanoic acid) and the matrix. It is therefore necessary that chemometric multivariate calibration is carried out on samples that are as close to real samples as possible.

The advantages of applying the electronic tongue instead of HPLC or GC would be, once the electrode array is calibrated and properly validated, the determination of octanoic acid in a sample can be done within few minutes. In the industrial process monitoring scheme a short analysis time is a premium, as only then it is possible to keep the octanoic acid concentration in the process constant.

From earlier work reported on a different cultivation (Turner et al., 2003), it is further speculated that in addition to allow process monitoring the ET electrodes will also be able to monitor the catabolism of the octanoic acid.

2. Material and methods

2.1. Chemicals

All chemicals were of analytical grade. Ammonium nitrate, calcium chloride dehydrate, carbamide diamine, copper dichloride dihydrate, D-glucose, dipotassium dihydrogen phosphate, manganese dichloride tetrahydrate, octanoic acid, and zinc sulfate heptahydrate were obtained from Merck (Darmstadt, Germany). Polyethylene sorbitan monoorleate (Tween 80) was

Table 1
Reported analytical methods for determination of octanoic acid

Ref	Subject	Method
Blomquist et al. (1992)	Characterisation of moulds by measuring the fatty acids content	HP 5890 series Gas Chromatograph with FID, capillary column; fused silica (30 m × 0.25 mm × 0.25 μm) HP 5890 series Gas Chromatograph equipped with Finnigan model 4023 mass spectrometer system
Huang et al. (2002)	Method development for determination of carboxylic acids	Hitachi HPLC separating system including an L-6300 intelligent pump, F-1080 fluorescent detector
Kellerhals et al. (1999)	Development of on-line GC, to maintain continuously fed substrate at a desired level, during the production of mcl-PHAs by *Pseudomonas putida* KT2442	Online determination of substrate concentration: Fermenter equipped with recirculation loop containing a crossflow filtration module. The permeate was analysed at a HP 5890 Gas Chromatograph with FID, and Permabond FFAP-0.35 column
Lai et al. (2004)	Synthesis and characterisation of structured lipids in bench scale reactor.	Shimadzu 14A GC with FID
Larroche (1996)	Investigation of the internal substrate concentration during the biotransformation of octanoic acid into 2-heptanone by *Penicillium roqueforti*	Gas Chromatograph with FID, Capillary column; Supelcowax (30 m × 0.32 mm × 0.5 μm)

obtained from Struers (Denmark). Ammonium heptamolybdate tetrahydrate and 5-methyl-2-hexanole were obtained from Sigma-Aldrich (Denmark). Acetone, hydrochloric acid and sodium hydroxide were obtained from Bie and Berntsen (Denmark). Potato Dextrose Agar (Difco™) was obtained from Becton, Dickinson and Company while Tryptic Soy broth for microbiology was obtained from Merck (Darmstadt, Germany). Soy flour and parboiled rice was of food quality and obtained from the local supermarket.

Filters were 0.45 μm nylon filters from MicroLab (Aarhus, Denmark). Filter paper (diameter of 240 mm), was of the type Schleicher and Schuell MicroScience Ref No. 10311651.

A strain of *Penicillium roqueforti* used in industrial production of natural methyl ketones at *Danisco, Denmark* was used in this study. Spores were stored on silica-gel.

2.2. Culturing conditions

The pre-culture was prepared in three steps. In the first step a small amount of silica-gel containing spores was incubated at PDA-plates for 7 days at 25 °C. In the second step spores from the first incubation was transferred to new PDA plates and incubated again for 7 days at 25 °C. In the third step the spores were transferred from the PDA plate with a sterile spatula to a 1 l conical flask containing 225 g rice, 6.75 g soy flour, and 75 ml demineralized water, which had been sterilized at 121 °C for 60 min prior to inoculation.

After an incubation period of 5 days at 25 °C, 250 ml sterile 0.01% Tween 80 solution were added to the rice, followed by soaking and gently shaking to separate the spores from the rice grains. Thirty minute later the suspension was filtered through a piece of gaze, which after the filtration was washed with 100 ml 0.01% Tween 80.

After harvesting, 2 ml of the spore suspension was added to 25 ml TSB, which was incubated for 12 h in an environmental shaker at 30 °C and 200 rpm. This solution was investigated for contaminations, under a microscope before inoculation of the bioreactor. The spore solution was stored at 5 °C for no more than 24 h before use, to avoid a loss in spore activity.

2.3. Substrate

It was decided to apply a carbon source beside the octanoic acid in the substrate thereby enabling a batch without the presence of octanoic acid and subsequently directing as much of the octanoic acid towards product formation. The effect of different carbon sources has been reported, where D-glucose was shown also to facilitate the production of 2-heptanone (Lawrence and Bailey, 1970). Different inorganic and organic compounds have been investigated as nitrogen sources, where the organic ones were found to have the best effect on methyl ketone production (Meyers and Knight, 1958). Carbamide diamine was selected here, as it was reported to induce sporulation. The buffer applied was dipotassium dihydrogen phosphate, because phosphate is known to enhance the octanoic acid degradation ability (Lawrence and Hawke, 1968). In total the substrate contained: 100 g l^{-1} glucose, 2.16 g l^{-1} carbamide diamine, 7.17 g l^{-1} dipotassium dihydrogen phosphate and 2 ml of a trace element solution. 100 ml trace element solution contained: 15.5 mg $CuCl_2$, 175.6 mg $ZnSO_4 \cdot 7H_2O$, 36 mg $MnCl_2 \cdot 4H_2O$, 183.4 mg $CaCl_2 \cdot 2H_2O$, and 10.2 mg $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ dissolved in HCl and diluted with water, and was based on the results of (Meyers and Knight, 1958).

2.4. Cultivation

Five liter substrate was added to a 7.5 l bioreactor and heat treated at 100 °C for 60 min prior to inoculation. After heat treatment, the pH was adjusted to 6.5 applying 0.5 M HCl. The pH was kept constant throughout the cultivation, by adding 0.5 M HCl or 4 M NaOH. This pH value was chosen as earlier work had determined that optimum pH falls in the range from 5.5 to 7, the specific optimum being dependent on the octanoic acid concentration (Gehrig and Knight, 1963). The fermenter was inoculated with 3.6×10^9 spores l^{-1} and the process ran from two to six days under the following conditions: 0.042 vvm air, 200 rpm (tip speed: 0.63 m s^{-1}), and 27 °C. This temperature was fixed based on work reported in (Lawrence, 1966). The spore amount in the inoculum was counted by applying a counting chamber and a microscope. During the cultivation pH and O_2 tension were measured on-line. At the end of each cultivation, the broth was heated to inactivate the fungus, followed by filtration through a paper filter, thereby separating the hypha and spores from the aqueous phase. Subsequently the aqueous phase was applied in the electronic tongue measurements, thereby obtaining a realistic background matrix.

Finally a fifth cultivation was carried out under identical operating conditions, to which octanoic acid was added 15 h after oxygen limitation was reached. From this samples were withdrawn and analyzed at GC-FID.

The bioreactor applied was an Applikon Microbial Biobundle system with electrodes AppliSens pH^+ and AppliSens Dis Ox. Data collection was facilitated via a RS 232 connection in the BioBundle and a dedicated program written, for the purpose, in LabView.

2.5. Electronic tongue measurements

The electronic tongue (ET) consist of six non-specific potentiometric sensors that were obtained from the Laboratory of Chemical Sensors at St. Petersburg University and a standard pH electrode (Radiometer Analytical, DK). Prior to development of the present analytical procedure, an extensive pilot experiment applying 20 different electrodes was carried out to determine the minimum number of electrodes needed in the array for reliable determination of octanoic acid (results not shown). From these experiments it was concluded, that an array of six specific anion selective electrodes had an enhanced cross-sensitivity to octanoic acid in the applied substrate and was adequate. Each anion sensor consisted of an active material (transition metal doping) incorporated in a PVC membrane with a solid inner state contact. Technical details regarding these sensors may be found in (Legin et al., 1999; Legin et al., 2003). All measurements were

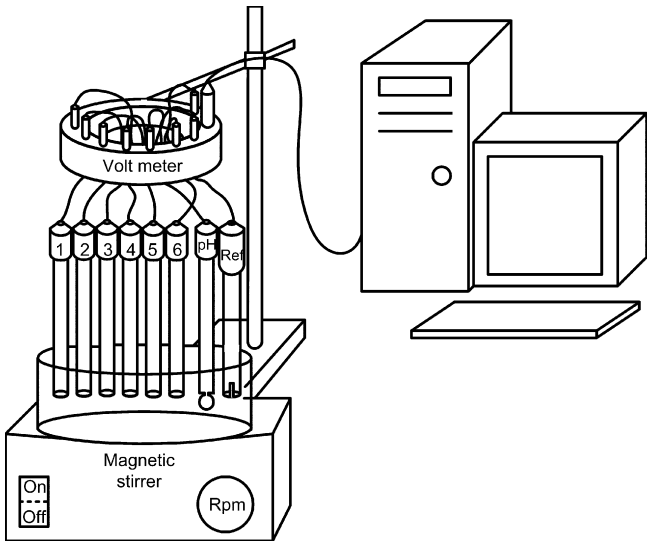

Fig. 1. Experimental set-up for the electronic tongue measurements.

made against a double junction electrode (Radiometer analytical, DK), with an inner reservoir of saturated potassium chloride and an outer of ammonium nitrate using a custom-made multichannel digital voltmeter with high input impedance. See Fig. 1 for experimental set-up.

For each of four cultivation broths, triplicate determination of nine different concentration levels: 0.65, 2.5, 5.0, 7.0, 8.5, 10, 12.5, 16, and 20 mmol l^{-1} were carried out. In addition, 13 samples with a concentration within this concentration range were obtained by dilution of samples from the fifth cultivation with broth from cultivation no. 2.

Each measurement had a 14 min equilibration period followed by electrode response recording for 1 min. Between measurements the sensors were washed thoroughly with distilled water.

2.6. Gas chromatography

For determination of the octanoic acid concentration, 2.00 g broth (analytical mass) was weighted directly into a 10 ml calibrated flask, diluted with 7 ml acetone and left for 5 min, after which 100 μl internal standard solution and acetone were added. The internal standard was 5-methyl-2-hexanole. The sample was filtered through a 0.45 μm filter directly into a vial and analyzed using a GC-FID with a Varian CP7485 WCOT Fused Silica column (25 m × 0.32 mm i.d., film thickness 0.30 μm) designed for free fatty acids. The temperature program applied was 115 °C for 2 min, 30 °C min^{-1} to 220 °C, and hold for 1.5 min.

2.7. Data processing

Processing of the electronic tongue data was carried out using the UNSCRAMBLER software (version 9.5). Multivariate calibrations using Partial Least Square regression (PLS) were performed for quantitative determination of octanoic acid, following the chemometric procedures outlined in (Esbensen, 2001). For each of the four cultivations a separate PLS model was

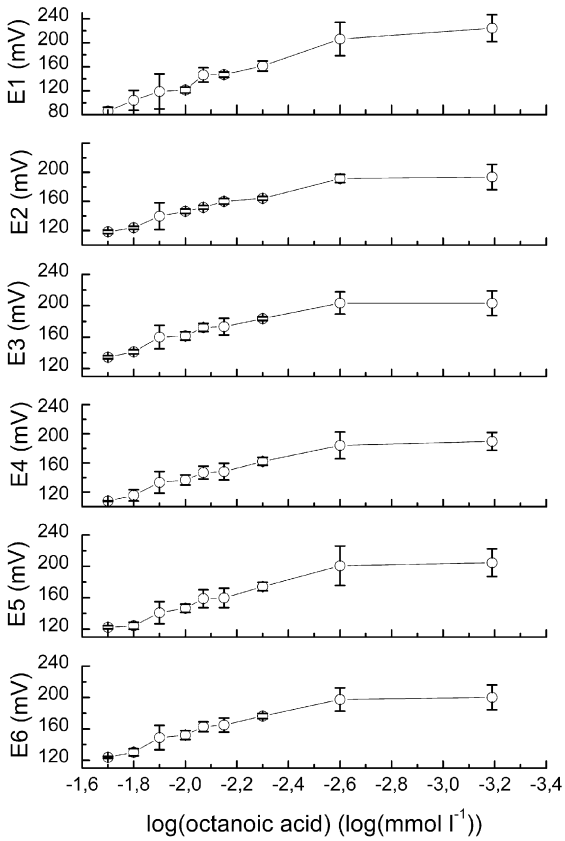

Fig. 2. Examples of electrode response curves for the six electrodes applied. Each response is based on triplicate determination of octanoic acid in broth no. 1. The bars indicated 1× standard deviation.

made, from which a few outliers were detected and excluded. After outlier exclusion the data were agglomerated, triplicate measurements were averaged and a global model containing data from all four cultivations was also calibrated. All models were evaluated by segmented cross-validation, leaving an entire concentration level out, one at a time. A random leave-one-measurement-out-at-the-time cross-validation would not test the model sufficiently, as the model would still contain measurements pertaining to the concentration level left out. Esbensen

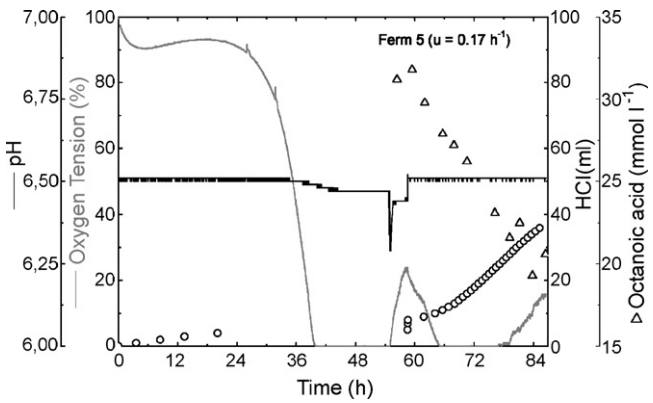

Fig. 3. Cultivation chart for test cultivation. Addition of octanoic acid is indicated by a sharp decrease in the pH and a concomitant increase in oxygen tension.

(2001) details the appropriate strategies for cross-validation, a.o. that this approach is mostly acceptable for internal comparison purposes only, e.g. between the models for the four cultivars employed in the present study.

The final global model was tested by prediction of octanoic acid concentration in samples originating from an independent degradation of octanoic acid i.e., in a proper test set validation context.

3. Results and discussion

To obtain a realistic evaluation of the electronic tongue performance, the measurements were carried out in real broth originating from an industrial cultivation analogue, to which known concentrations of octanoic acid was spiked. In total it was decided to apply four batch cultivations, with an average maximum specific growth rate (μ) of 0.15 h^{-1}. This was done in an attempt to catch and incorporate the natural variation that may exist between cultivations, into the model building phase, hence making the model i.e., electronic tongue more robust and capable of predicting the octanoic acid concentration in future samples. To facilitate an even broader, but also realistic variation, it was furthermore decided to terminate the cultivations at different times relative to the on-set of the oxygen limitation, see Table 2.

To 50 mL of each of these four cultivation broths different amounts of octanoic acid were added, hence a span in the concentration range from 0.65 to 20 mmol l^{-1} was obtained. For each concentration triplicate determinations were carried out with the electronic tongue.

In Fig. 2 an example of results obtained from measuring on the broth originating from cultivation 1 is seen. As seen the voltage was increasing as a function of log(octanoic acid) i.e. increasing as a function of an increasing octanoic acid concentration for all electrode applied. This was due to the nature of the sensors, as they were all anion selective. In general the electrodes had a linear response to the octanoic acid concentration in the interval from approximately 2.5–20 mmol l^{-1} (−2.3 to −1.7 log unites) i.e. they had a Nernstian behavior. At the lowest concentration; 0.65 mmol l^{-1} the slope of the curves were decreasing, which can be explained by the uncertainties encountered when measuring at low concentrations i.e. other ions in the solution effects the electrodes.

Table 2
Specific growth rate, duration of the cultivation before oxygen limitation, and time elapsed after on-set of oxygen limitation before the cultivation was ended for cultivation 1–4

Cultivation	μ (h^{-1})	O_2 limitation (h)	Stop relative to O_2 limitation (h)
1	0.15	40	80
2	0.15	44	72
3	0.13	44	0
4	0.17	36	36

As a mean of evaluating the model obtained, a fifth cultivation ($\mu = 0.17\ h^{-1}$) was carried out under the same conditions, see Fig. 3, to which 30 mmol l^{-1} octanoic acid was added. The fifth cultivation had a lag period of about 24 h after which the exponential growth of the organism started, which lasted 16 h. After additional 15 h of oxygen limitation, the octanoic acid was added, which resulted in a sharp increase in the oxygen tension curve and a decrease in the pH curve. The pH was adjusted back to 6.5 again manually, whereas the oxygen tension decreased after an adaption period. Simultaneously with the oxygen consumption, octanoic acid catabolisation was initiated and the total concentration in the substrate was decreased from approximately 32 to 20 mmol l^{-1}.

During the catabolisation of octanoic acid samples were collected, with the purpose of electronic tongue model evaluation. Of the samples collected it was however only the last

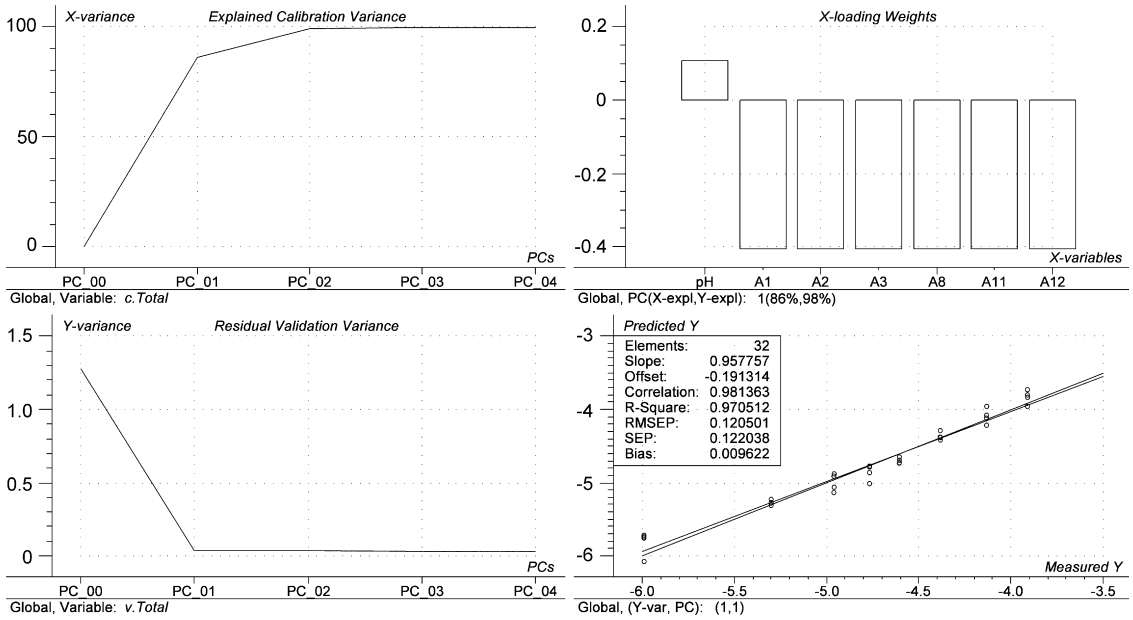

Fig. 4. Global model for octanoic acid measured by ET.

Table 3
Modeling and prediction statistics for the four individual models

No	Total	Component	# objects	Slope	r^2	RMSEP	%RMSEP	RPD
2	21	1	8	0.92	0.96	0.13	5.5	5.1
	24	1	8	0.95	0.98	0.10	4.2	6.7
	22	1	8	0.89	0.96	0.15	6.3	4.3
	24	1	8	0.84	0.90	0.21	8.8	3.1

The data was centered and scaled. "Total" refers to the number of determinations upon which the average for each model was based. "Component" is the number of PLS components applied in the model. "# objects" correspond to the number of concentration levels in the model.

two samples, that were situated within the calibration range of the model developed. Hence to test the prediction ability of the model in the lower concentrations interval as well, it was decided to *dilute* the true samples with broth from cultivation no. 2. Cultivation no. 2 was chosen, due to the high degree of similarity between the oxygen tension curves of the second and fifth cultivation. This compound set forms a relevant independent test set, which can be used for realistic prediction validation.

3.1. Individual models

From the non-linear tendencies observed in the electrode response plot, it was decided to linearise the Y-data, by taking ln to the octanoic acid concentration. In addition the data were centered and scaled. In Table 3 the pertinent statistics for the four individual PLS-1 models are given. Each model was based on averages over triplicate measurements after outlier exclusion.

From these it is seen that the best model obtained was for cultivation no. 2 which was characterized by growth at the average maximum specific growth rate followed by a somewhat slow death phase. During the death phase, lysis of the cells may occur; thereby releasing proteins into the environment, which can interfere with the membrane of the electronic tongue sensors. This may explain the better model obtained for this cultivation. This hypothesis is consistent with the somewhat poorer models obtained for cultivation no. 1 and 4, which both had a faster death phase. It is not consistent with obtaining the poorest model for cultivation no. 3 on the other hand, as this was stopped even before the death phase started. However other factors may also have influenced the latter cultivation, as the maximum specific growth rate here was lower than the average.

Eight of the full set of nine concentration levels were included in the models, as the lowest concentration 0.65 mmol l^{-1} consistently showed outlier tendencies (inclusion in the model was possible, but required two extra PLS-components, as compared to only one in the global model, see below). From the electrode response plots this was not totally unexpected, and it was decided to leave out this concentration level in the feasibility study. This could be done without loss of generality.

In general the individual PLS prediction models were characterized by satisfactory accuracy (as evidenced by the slope of a fitted regression line between predicted and reference values) between 0.84 and 0.95.

As a means of assessing the prediction performance of the model further, the relative RMSEP%, which is the average error associated with octanoic acid prediction, for all future samples with a similar compositional range (Esbensen, 2001), and the RPD which is a dimensionless ratio relating the reference span to the RMSEP (Fearn, 2002) were evaluated. In general it is the experience that an RPD in the interval 3.1–4.9 is fair, in the interval 5.0–6.4 is good and from 6.5 to 8 is very good. The prior may be applied for screening purposes whereas a model with an RPD in the interval above 6.5–8 may be applied in the process control (Williams and Norris, 2001).

For the models built, the relative prediction errors (precision) were of the order 4.2–8.8%. All acceptable as first attempts at evaluating the feasibility of octanoic prediction although the ET model 4 was marginal only. This was substantiated from the pertinent RPD values, which were in the interval 3.1–6.7. In the evaluation of the results it was further concluded that the total variation seen in the four cultivations, was relevant in the industrial cultivation context, as they all reflect a realistic matrix composition. Therefore general octanoic acid prediction model development would be based on aggregating all four cultivation runs.

3.2. Global model

The resulting global model, which was first evaluated by an 8-segmented cross-validation, again leaving out one whole concentration level at a time, is presented in Fig. 4.

Fig. 4 confirmed the results obtained from the individual models i.e., one PLS-component was statistically significant–no extra explanation of Y is obtained by adding an additional component. Therefore 86% of the variance in X was relevant and applied for explaining 98% of the Y-variance (octanoic acid

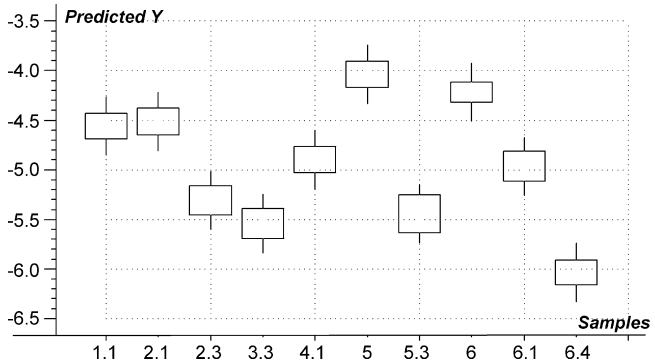

Fig. 5. Predicted levels of octanoic acid for samples from test set validation (run no. 5).

Table 4
Individual octanoic acid concentration predictions and deviation

	Sample number									
	6	5	6.4	6.1	5.3	4.1	3.3	2.3	2.1	1.1
Octanoic acid (mmol l^{-1})	20.3	19.7	2.5	10.1	4.92	11.2	5.85	6.7	13.4	15.1
	−3.90	*−3.93*	*−5.99*	*−4.59*	*−5.31*	*−4.49*	*−5.14*	*−5.01*	*−4.31*	*−4.19*
Predicted	−4.22	−4.04	−6.03	−4.97	−5.44	−4.90	−5.54	5.31	4.51	4.56
Deviation	0.10	0.13	0.13	0.15	0.19	0.13	0.15	0.15	0.13	0.13

Italics correspond to the linearized ln octanoic concentrations.

span). This signifies highly satisfactory prediction ability over the entire calibration range.

The standard measure RPD was found to be 5.5 and the RMSEP% level of 5.1% (ln units). If so desired the prediction precision can also be expressed by the asymmetric interval +12.7/−11.4% derived by the inverse ln-transform. For evaluating the performance of each electrode, the X-loading-weight w_1 can be evaluated. Here the X-loading weight shows how much each electrode contributes to explaining the response variation along each model component, hence a possibility to evaluate all six electrodes in the array. As each electrode shows nearly identical, high X loading-weight for PLS component number 1, they are equally important in building the octanoic acid model, and hence there will be no improvement of the model by any stepwise optimization approaches or similar. This observation is also evidenced by the pilot experiment, where the particular set of six electrodes were singled out simultaneously as being the optimal set of electrodes.

The global model broth was spiked with octanoic acid thereby obtaining realistic fed-batch composition of both matrix as well as analyte. This makes it likely that the model will also perform on real samples from catabolization of octanoic acid. To test this, the fifth cultivation described above was applied. The prediction results obtained are presented in Fig. 5; and individual prediction deviations are listed in Table 4.

Overall the prediction deviations in Table 4 were comparable with those from the calibration model RMSEP, 0.12 (ln units), except for sample no. 5.3, which means that the model performed equally well on completely new samples.

4. Conclusions and perspectives

The final global model was based on averaged triplicate ET recordings. It was characterized by a highly satisfactory prediction validation (leave-one-concentration level out cross-validation) which pointed to only one PLS-component needed to span a realistic compositional range. The global model was also able to predict samples from an octanoic acid degradation matrix.

The average future prediction error level, RMSEP% from test set validation, corresponded to 5.1% (ln-units). The RPD statistic was 5.5, also signifying a good prediction precision. The slope (accuracy) of the global model was 0.96. These statistics signified a satisfactory development result, allowing the conclusion that feasibility of an ET prediction model for octanoic acid prediction in realistic cultivation broths for industrial 2-heptanone production has been successfully demonstrated.

The present successful development of an Electronic Tongue analytical procedure for low concentration of an organic acid (octanoic acid) has a wide carrying-over potential for similar industrial on-line/at-line monitoring scenarios with comparable demands, i.e. a critical sensitivity for the analyte and a need for fast (minutes), reliable analytical result. There are numerous potential application areas in the general industrial fermentation-based process industries, aerobic as well as anaerobic.

References

Blomquist, G., Andersson, B., Andersson, K., Brondz, I., 1992. Analysis of fatty acids—a new method for characterization of molds. J. Microbiol. Methods 16, 59–68.

Creuly, C., Larroche, C., Gros, J.B., 1990. A fed-batch technique for 2-heptanone production by spores of *Penicillium roqueforti*. Appl. Microbiol. Biotechnol. 34, 20–25.

Creuly, C., Larroche, C., Gros, J.B., 1992. Bioconversion of fatty acids into methyl ketones by spores of *Penicillium roqueforti* in a water-organic solvent, two-phase system. Enzyme Microbiol. Technol. 14, 669–678.

Esbensen, K.H., 2001. Multivariate Analysis in Practice, fifth ed. CAMO Press.

Fearn, T., 2002. Assessing calibrations: SEP, RPD, RER and R^2. NIR news 13, 12–14.

Gehrig, R.F., Knight, S.G., 1963. Fatty acid oxidation by spores of *Penicillium roqueforti*. Appl. Microbiol. 11, 166–170.

Huang, C.Z., Santa, T., Imai, K., 2002. Development of 7-(*N*,*N*-dimethylaminosulfonyl)-5-*N*-(4-*N*-aminoethyl)piperazino-2,1,3-benzoxadiazole as a water-soluble fluorogenic regent for the sensitive liquid chromatographic determination of saturated carboxylic acids. Analyst 127, 741–747.

Kellerhals, M.B., Kessler, B., Witholt, B., 1999. Closed-loop control of bacterial high-cell-density fed-batch cultures: production of mcl-PHAs by Pseudomonas putida KT2442 under single-substrate and cofeeding conditions. Biotechnol. Bioeng. 65, 306–315.

Lai, O.M., Low, C.T., Akoh, C.C., 2004. Lipase-catalyzed acidolysis of palm olein and caprylic acid in a continuous bench-scale packed bed bioreactor. Food Chem. 92, 527–533.

Larroche, C., Arpah, M., Gros, J.B., 1989. Methyl-ketone production by Ca-alginate/Eudragit RL entrapped spores of *Penicillium roqueforti*. Enzyme Microbiol. Technol. 11, 106–112.

Larroche, C., Besson, I., Gros, J.B., 1994. Behavior of spores of *Penicillium roqueforti* during fed-batch bioconversion of octanoic acid into 2-heptanone. Biotechnol. Bioeng. 44, 699–709.

Larroche, C., 1996. Internal substrate concentrations during biotransformation of octanoic acid into 2-heptanone by spores of *Penicillium roqueforti*. J. Ind. Microbiol. 16, 29–35.

Lawrence, R.C., 1966. The oxidation of fatty acids by spores of *Penicillium roqueforti*. J. Gen. Microbiol. 44, 393–405.

Lawrence, R.C., Hawke, J.C., 1968. The oxidation of fatty acids by mycelium of *Penicillium roqueforti*. J. Gen. Microbiol. 51, 289–302.

Lawrence, R.C., Bailey, R.W., 1970. Evidence of the role of the citric acid cycle in the activation of spores of *Penicillium roqueforti*. Biochim. Biophys. Acta 208, 77–86.

Legin, A.V., Rudnitskaya, A.M., Vlasov, Y.G., Di Natale, C., D'Amico, A., 1999. The features of the electronic tongue in comparison with the characteristics of the discrete ion-selective sensors. Sens. Actuators B 58 (1–3), 464–468.

Legin, A., Rudnitskaya, A., Vlasov, Yu., 2003. Electronic tongues: new analytical perspective for chemical sensors. In: Alegret, S. (Ed.), Integrated Analytical Systems, Comprehensive Analytical Chemistry, vol. XXXIX. Elsevier, pp. 437–486.

McSweeney, P.L.H., Sousa, M.J., 2000. Biochemical pathways for the production of flavour compounds in cheeses during ripening: a review. Lait 80, 293–324.

Meyers, E., Knight, S.G., 1958. Studies on the nutrition of *Penicillium roqueforti*. Appl. Microbiol. 6, 174–178.

Park, O.J., Holland, H.L., Khan, J.A., Vulfson, E.N., 2000. Production of flavours ketones in aqueous-organic two-phase systems by using free and microencapsulated fungal spores as biocatalysts. Enzyme Microbiol. Technol. 26, 235–242.

Turner, C., Rudnitskaya, A., Legin, A., 2003. Monitoring batch fermentations with an electronic tongue. J. Biotechnol. 103, 87–91.

Williams, P., Norris, K., 2001. Near-Infrared Technology in the Agricultural and Food Industry. American Association of Cereal Chemists, Inc.