Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt

Lehrstuhl für Brau- und Getränketechnologie

Computationally aided reliability analysis of sensor data using multivariate data analysis and modelling approaches for bioprocesses

Daniel Krause

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. U. Kulozik

Prüfer der Dissertation:

1. Prof. Dr.-Ing. Th. Becker

2. Prof. Dr.-Ing. A. Kremling

Die Dissertation wurde am 27.06.2017 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 30.01.2018 angenommen.

Acknowledgements

First, I would like to thank my thesis supervisor, Prof. Dr.-Ing. Thomas Becker, for the possibility to develop this work at his chair, for his support and the chances he gave me beyond of the thesis. I would further like to thank for the constructive directions contributed to the successful completion as well as to the confidence, that nothing is too difficult. His support reached also beyond science by means of trust in times, when I almost gave up hope.

Further, my appreciation goes to the members of my thesis committee, namely Prof. Dr.-Ing. U. Kulozik for acting as chief examiner as well as Prof. Dr.-Ing. A. Kremling and Prof. Dr.-Ing. Th. Becker for their time and effort reviewing my thesis and taking part in my thesis defence.

Furthermore, one of the most important persons in the time of my research was my supervisor and good friend Mohamed A. Hussein. He deserves my special appreciations for his all-time full support without compromises and more than fruitful supervision. He was also one of the persons never gave up hope with me.

A lot of my gratitude I want to show to my colleagues, especially but not exclusively the members of the PAT group, and my students supporting me in my research and some of them even in private issues. This work contains allot of their input. Additionally, many of my thanks go to my friends – overall, I want to apologize for not naming you all, the list would be too big!

Last but not least and for sure more than equal appreciation go to my parents and Tine for everything - they showed elusive amount of patience without any time of complaint but always full support.

Thank you all! You possibly cannot imagine, how much!

Erfurt, February 2017

Daniel Krause

Publications

Peer reviewed publications

1. Hoche, S. Krause, D., Hussein, M. A., Becker, T., Ultrasound-based, in-line monitoring of anaerobe yeast fermentation: model, sensor design and process application. International Journal of Food Science & Technology, 2016. 51(3): p. 710-719.

2. **Krause, D., Hussein, M. A., Becker, T., Online Monitoring of Bioprocesses via Multivariate Sensor Prediction within Swarm Intelligence Decision Making. Chemometrics and Intelligent Laboratory Systems, 2015. 145(0): p. 48-59.**

3. **Krause, D., Holtz, C., Gastl, M., Hussein, M. A., Becker, T., NIR and PLS discriminant analysis for predicting the processability of malt during lautering. European Food Research and Technology, 2015. 240(4): p. 831-846.**

4. Holtz, C., Krause, D., Hussein, M. A., Gastl, M., Becker, T., Lautering Performance Prediction from Malt by Combining Whole Near-Infrared Spectral Information with Lautering Process Evaluation as Reference Values. Journal of the American Society of Brewing Chemists, 2014. 72(3): p. 214-219.

5. Schirmer, M., Zeller, J., Krause, D., Jekle, M., Becker, T., In situ monitoring of starch gelatinization with limited water content using confocal laser scanning microscopy. European Food Research and Technology, 2014: p. 1-11.

6. **Krause, D., Hussein, W. B., Hussein, M. A., Becker, T., Ultrasonic sensor for predicting sugar concentration using multivariate calibration. Ultrasonics, 2014. 54(6): p. 1703-1712.**

7. Procopio, S., Krause, D., Hofmann, T., Becker, T., Significant amino acids in aroma compound profiling during yeast fermentation analyzed by PLS regression. Food Science and Technology, 2013. 51(2): p. 423-432.

8. **Krause, D., Birle, S., Hussein, M. A., Becker, T., Bioprocess monitoring and control via adaptive sensor calibration. Engineering in Life Sciences, 2011. 11(4): p. 402-416.**

9. Krause, D., Schöck, T., Hussein, M. A., Becker, T., Ultrasonic Characterization of Aqueous Solutions with varying Sugar and Ethanol Content using Multivariate Regression Methods. Journal of Chemometrics, 2011. 25(4): p. 216-223.

Contents

Abstract

The process industry is confronted with a global competition and dynamic market. Therefore, the establishment and provision of innovative and future oriented process intelligence represents one solution for higher competitive capacity. Thus, the development and usage of innovative sensor principles in combination with intelligent linkage between sensor systems as well as process knowledge will lead to better process control in any life science area. Furthermore, food and beverage industry is confronted with the demand for large quantities of goods within constant quality corridors determined by the market, while at the same time only a small individual value added. Thus, a rising demand of novel sensing methodologies appears, implying the central necessity of sophisticated and adapted data modelling and analysis to establish robust measuring devices for bioprocess given surrounding conditions and analysing existing data pool to extract more relevant information and knowledge, much more linked to processing of food stuff. The very broad field of multivariate data analysis offers beneficial features to such demands. It contains a variety of tools which are typically convenient to apply and in the same time do not lead to over-complexified solutions to the challenge of interest. Additionally, this area provides a numerous amount of possibilities for a task and only few instructions and recommendations for applying which method for what challenge. Thus, with the focus on solutions "as simple as possible and as complex as necessary", the central working hypothesis of this thesis is termed "*Is there a standard way of applying multivariate data analysis on bioprocess sensor data?*"

This includes all aspects, such as data pre-processing, model generation, data post-processing, aspects of non-linearity, model robustness, and the transfer of models. The thesis shows the application of a confined selection of algorithms on three different fields: first, ultrasonic sensor calibration used for the online detection of biochemical fluid properties using ultrasonic features in different calibration model approaches. Amongst others, variable selection on features, external parameter orthogonalisation on temperature and kernel-PLS were applied, reaching a robustified result of ~0.9g/100g prediction error for binary maltose solutions and ~1.5g/100g for ternary maltose and ethanol solutions in a temperature range between 10 and 20°C. Second, qualification of raw material with respect to processability by using PLS-DA on NIR-spectra of malt kernels and expert knowledge reaching a maximum classification error of 76.6% between model output and expert classification on industrial data. Further, multivariate process control tools were applied on process data to reach data driven classification of process quality reaching 84% match between model output and expert classification. Third, intelligent sensor network inspection and failure compensation for robust online process monitoring applying multivariate process control together with swarm intelligence reaching stable monitoring even under single sensor failure in 100% of the tested cases. The additional benefit of this system is, that it is not restricted to the number sensors or any specific sensor reading. The presented approach can be used as combined multiple sensor analysis reaching universal process control with integrated sensor evaluation.

Even though, just a selection of possibilities are investigated, the answer to the hypothesis is twofold: yes, when coping all necessary aspects from data processing to model refinement and robustness issues and no, when it comes to fully automated solutions without respective knowledge of the user.

Zusammenfassung

Die Prozessindustrie ist einem globalen Wettbewerb und einem dynamischen Markt ausgesetzt. Eine zentrale Lösung zur Erhöhung der Wettbewerbsfähigkeit stellt dafür unter anderem die Bereitstellung von innovativer und zukunftsorientierter Prozessintelligenz dar. Dabei wird die Entwicklung und Nutzung neuartiger Sensorprinzipien in Kombination mit intelligenten Verknüpfungen von Sensorsystemen und Prozesswissen zu einem besseren Prozessverständnis und folglich besserer Prozesskontrolle im Life Science Bereich führen. Weiterhin wird die Nahrungsmittel- und Getränkeproduktion mit der Nachfrage von großen Mengen an Gütern innerhalb der durch den Markt bestimmten, gleichbleibenden Qualitätskorridore bei gleichzeitig geringer, individueller Wertschöpfung konfrontiert. Damit wird auch die steigende Nachfrage an neuartiger Messmethoden aufgezeigt, welche die zentrale Notwendigkeit an geeigneter Datenanalyse und –modellierung impliziert. Mit deren Hilfe können entweder robuste Messsysteme für die durch Bioprozesse gegebenen Umgebungsbedingungen erstellt oder bereits bestehende Datenbestände analysiert werden. Letzteres wird genutzt, um weitere, relevante Informationen und Wissen zu extrahieren, unter anderem mit mehr Verknüpfung zur Verarbeitung der Lebensmittel. Das breite Feld der multivariaten Datenanalyse bietet nützliche Funktionen, um derartigen Forderungen gerecht zu werden. Es enthält eine Vielzahl von Werkzeugen, welche typischerweise bequem angewendet werden können, gleichzeitig aber nicht zu über-komplexen Lösungen für die jeweilige Problemstellung und deren Herausforderungen führen. Dieser Bereich bietet eine zahlreiche Menge an Möglichkeiten, um eine Aufgabe zu lösen. Allerdings existieren nur wenige Leitfäden und Empfehlungen für die Anwendung welcher Methode und deren Nutzung für eine gegebene Herausforderung. Daher und mit dem Fokus auf Lösungen „so einfach wie möglich und so komplex wie nötig“ wird die zentrale Arbeitshypothese dieser Dissertation wie folgt benannt: *„Gibt es eine standardisierte Möglichkeit der Anwendung von multivariater Datenanalyse auf Sensordaten von Bioprozessen?“*

Dies umfasst alle Aspekte, von der Datenvorverarbeitung über Modellgenerierung, Datennachbearbeitung, Nichtlinearität von Daten und Modellrobustheit bis hin zu dem Transfer von Modellen. Die Arbeit zeigt die Anwendung einer begrenzten Auswahl von Algorithmen auf drei verschiedene Bereiche: erstens, Kalibrierung eines Ultraschallsensors für die Online-Erfassung der biochemischen Eigenschaften eines Fluides mittels berechneter Merkmale der Ultraschallsignale in verschiedenen Kalibrierungsmodellansätzen. Unter anderem wurden dabei Methoden zur Variablenauswahl von Ultraschallmerkmalen, externe Parameter-Orthogonalisierung auf Temperatur und Kernel-PLS angewendet. Damit konnte ein robusteres Ergebnis in einem Temperaturbereich von 10 bis 20°C mit einem Vorhersagefehler von ~0.9g/100g für binäre Maltose-Lösungen und ~1.5g/100g für ternäre Maltose- und Ethanol-Lösungen erreicht werden. Zweitens, Qualifikation von Rohstoffen im Hinblick auf deren Verarbeitbarkeit unter der Anwendung von PLS-DA auf NIR-Spektren von Malzkörnern und Expertenwissen. Hier konnte ein maximaler Klassifikationsfehler von 76.6% zwischen Modelresultat und den Experteneinteilungen bei industriellen Daten erreicht werden. Weiterhin wurden Werkzeuge der multivariaten Prozesssteuerung verwendet, um Prozessdaten datengetrieben zu qualifizieren. Dabei wurde eine 84%ige Übereinstimmung zwischen Modelresultat und Expertenmeinung erreicht. Drittens, intelligente Sensornetzwerkkontrolle und Fehlerkompensation für robuste Online-Prozessüberwachung unter der Verwendung von Algorithmen der multivariaten Prozesssteuerung zusammen mit Schwarmintelligenz. Der Ansatz führte zu stabilem Monitoring in allen getesteten Fällen, auch bei Ausfall einzelner Sensoren. Ein zusätzlicher Vorteil des Systems ist, dass dieses nicht abhängig von der Anzahl an Sensoren sowie spezifischer Sensordaten ist. Der präsentierte Ansatz kann zur kombinierten, multiplen Sensoranalyse genutzt werden, wobei eine umfassende Prozesssteuerung mit integrierter Sensorevaluierung erreicht wird. Auch wenn nur eine Auswahl an Möglichkeiten untersucht wurden, so ist die Antwort auf die Hypothese zweifältig: Ja, wenn alle notwendigen Aspekte von der Datenverarbeitung über Modellanpassung und -verfeinerung bis hin zu Robustheit bewältigt werden und nein, wenn es um vollautomatische Lösungserstellung ohne entsprechende Kenntnisse des Benutzers geht.

1. Introduction

Enhancing productivity of existing process units as well as increasing the knowledge in any life science area provoke the demand of novel sensor principles. Thus, extracting biochemical process leading parameters increase the productivity of existing process units. Such sensors and systems have to fulfill the robustness attributes of production (including hygienic design aspects), simplicity in service, maintenance as well as optimal cost-benefit ratio. Over the last decades, a variety of different principles were investigated measuring univariate as well as spectral process information [1]. Additionally, diverse sensors as well as sensor networks produce a big amount of data. Such data pool contain a variety of hidden information, which is mostly not fully used. Further, data pool are often superimposed with noise as well as corrupted information. This might lead to failure of process models, misleading interpretations and sometimes to total process failures.

Summing these aspects together with the mostly time invariant and non-linear behavior of bioprocesses, this area of industrial production phases a variety of difficulties. The dynamic behavior of such processes is in contrast to the widely applied static and recipe driven process control. Under this perspective, computationally aided techniques such as computational intelligence provide helpful methods in solving such challenges that are hard to solve with conventional possibilities [5]. Therefore, this implies, amongst others, the usage of multivariate data analysis for extracting the most relevant patterns out of large data pool [6]. Such techniques are supported by the fast development of computational capacities as well as the reduction in size of utilized hardware.

Such possibilities recently gave way for the rise of intelligent data handling and analysis. Nevertheless, there are still many challenges on the path to reach the full potential of intelligent sensing and data handling. Both areas of bioprocess and computational intelligence have steadily developed based on the rise of such techniques [5]. There is a need to emphasize the efforts in science of each single discipline together with a high exchange between them [7]. According to the PAT/QbD initiative of the FDA the future in process analysis will include the generation of knowledge through data in pharmaceutical, medical and food and feed industry. This implies the automation of non- automated systems, integration of quality assurance into production processes, enhanced product and process safety, efficiency as well as sustainability. It also includes the use of multivariate data analysis for extracting the most important information out of huge data sets [7]. Especially in the food and feed sector, the current practice of comparably low degree in automation makes this industrial branch attractive for future implementations [7].

This thesis exactly targets the handling of still existing challenges for intelligent sensing. This is achieved by following several, mostly data driven strategies and applications:

a. Ultrasonic sensor calibration using (a) physical and (b) statistical sound features used for detection of biochemical fluid properties
b. Raw material qualification with respect to multivariate full process behavior
c. Sensor network inspection and failure compensation

These topics include all relevant aspects of multivariate data analysis, such as data pre-processing, variable selection, model generation, outlier analysis, model validation, stability and robustness. Therefore, the core question of this thesis crystalizes as:

Is there a standard way of multivariate model generation addressing all the listed keywords?

Thus, the focus of this thesis is based on the point reported by Munck *et al.*: “The aim is to study complex processes as a whole in order to model interaction of the underlying latent functional factors which may later be defined more precisely by deductive methods” [8]. They showed in their feasibility study by using multivariate data analysis in food science applications, that explorative strategies can also lead to “fundamental scientific significance” [8]. Therefore, the aspects of bioprocess sensors and calibration, multivariate process data and influences of raw materials as well as sensor network inspection will be discussed in the following paragraphs, all under the aspect of data driven modelling.

1.1 Bioprocess sensors and calibration

In the process industry there is a variety of novel sensor technologies, developed over the last decades. High scientific effort is given to spectroscopic techniques in diverse types of electromagnetic waves. These contain the benefit of measuring a big amount of variables and information, therefore several physical as well as chemical process constituents at one time point. An overview on existing process sensors is given in various literature.

First of all, these sensors can be divided (next to their specific targets) amongst their complexity, their detection performance and selectivity (Figure 1.1), [see [1]]. Next to the sensor systems on the left lower part of Figure 1.1, which are already established in industry, optical systems over the whole bandwidth from UV to IR light including fluorescence and Raman spectroscopy are successfully applied for investigations in fluids and hard matter, also in beverage area [see [9] for example]. In the brewing sector, these applications are widely laboratory equipment. However, the usage for online measurements increased over last decades due to the fast developments with respect to boundaries given by the techniques. One limit for MIR measurements in absorption is the necessity of little layer thickness. Therefore, ATR-flow cells (attenuated total reflection) were established, which opens the possibility to measure MIR spectra of fluids without reducing the layer thickness [10]. Nevertheless, the MIRS for online usage is still limited by maximal fiber optics length of less than five meters due to the high absorption of the material. Single channel excitation for fluorescence measurements were enhanced by multichannel excitation reaching two-dimensional spectra (2D-fluorescence) and the fluorescence interference in Raman spectroscopy for example is partially eliminated by using excitation wavelength below 250 nm. These examples are just a short extract of developments. Nevertheless, it shows the importance of spectroscopic systems. To not reach beyond the scope of this thesis, a comprehensive overview on online monitoring of bioprocess using spectroscopic methods is given by Lourenço *et al.* [11], where standard methods for calibration as well as the measurement instrumentation, advantages, disadvantages and applications of NIR-, MIR-, UV-Vis-, Raman-, Fluorescence- and Terahertz spectroscopy are discussed.

The main benefit of such spectroscopic methods such as IR- spectroscopy is the possibility to measure a variety of analytes at the same time [10] due to their specific absorption bands. In case of for instance NIR, chemical bonds are interacting with radiation induced by NIR allowing investigations on (amongst others) organic systems [12]. Typically, spectroscopic methods have one feature in common: to cover a suitable area of calibration model validity of such indirect principles, a sufficient and more or less equally distributed number of samples within defined boundaries is necessary [12].

Additionally, multimodal optical spectroscopy extends the possibilities of singular usage. The three different implementation strategies are explained in Kessler, 2013 [7]. The usage of different wavelength ranges combines the benefits of single ranges and partially reduces their disadvantages (1), the combination of different optical configurations in one range opens the possibility to separate morphological scattering from chemical absorption (2) and the angular resolved measurement or line scans with an imaging system lead to different penetration depths and can be used, amongst others, to describe particle properties (size, distribution).

Spatial resolved spectroscopy is another quite new approach, which measures spectral and spatial information at approximately the same time. This technique is known as chemical imaging. The three main possibilities are Whiskbroom, Staring and Pushbroom imaging. Basics can be found in Kessler 2013 [7], who states this technology as important for the future. However, challenges for optical spectroscopy are measurements in aqueous systems at rather low concentrations of analytes together with a strong scatter of cell suspensions [7], which is one main characteristic for biotechnological, especially brewing fluids.

Nevertheless, online application for running bioprocesses, especially in industry, is still rare. In summary, there are two main reasons: first, sensitive and selective devices for analyzing multi-parametric systems via spectroscopy are mostly cost intensive. Second, the interpretation and analysis of such data is somewhat complex by means of calibration/modelling. However, for measurement of single analytes of known bandwidths, the use

of photometric sensors can be applied. For those using bandpass filters to measure specific wavelength, the advantage is the reduced cost of such systems compared to spectroscopic ones.

A key issue is to integrate first principles in any sensor calibration. However, this is not always applicable [7]. The known relations between ultrasound, density and temperature with solutes in fluids, especially ethanol and sugar are of major interest in the development of novel sensor strategies. Most commonly applied method for inline measurements is the combination of speed of sound together with separate density measurement. A detailed review on measuring both properties with ultrasound, its difficulties and different design possibilities is given by Hoche *et al.* [13]. Nevertheless, all possibilities phase several limits such as dissipation by gas bubbles or particles. Even though algorithms can be used to either detect distorted signals prior to analysis, such as Angle Based Outlier Factor (ABOF, see section 1.2.3.2 "Outlier Analysis"), the strategy of signal analysis can be different as well. Acoustic sensors are also used for holistically measured information [14]. The success is shown in several applications. The major disadvantage is the variability of sensor signals by changes of setup or the sensor itself. Nevertheless, data driven soft-sensors are widely applied, also for spectroscopic sensor information. Anyway, a causal relation has to be extracted to correlate measured values to the target.

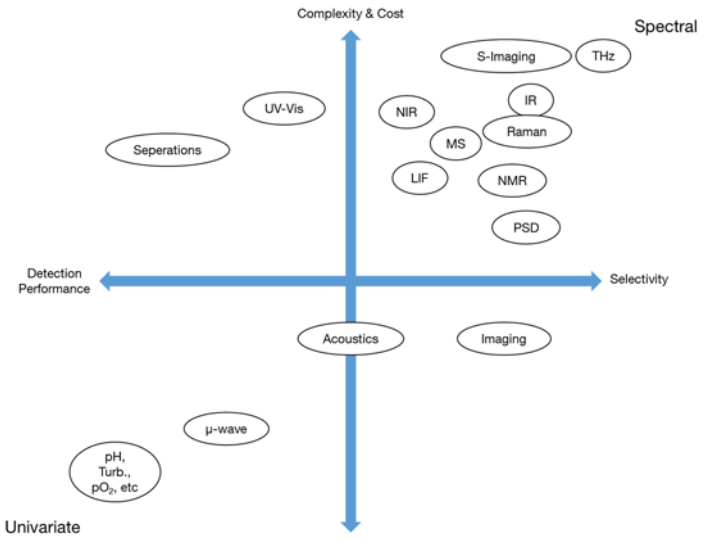


Figure 1.1: Classification of process instrumentation for online use, adapted from Bakeev, 2010 [1]; it shows a selection of process instruments classified according to their detection performance (sensitivity, precision, speed or quantification limits) as well as their selectivity on the one hand and their principle complexity and cost on the other (including e g. capital cost, training, maintenance and implementation); in contrast to Bakeev, only spectral versions of the more complex instruments are shown (IR, LIF, etc.); further, acoustics is moved more to selectivity, because recent developments reached higher resolutions and better detection accuracies from electronics and data mining point of view. Furthermore, spectral (S-) imaging is added including e. g. hyperspectral or pushbroom imaging; the figure is thought as an indication of diversity of existing process instrumentation (LIF…laser - induced fluorescence; NMR…nuclear magnetic resonance; MS…mass spectroscopy; PSD…particle size distribution)

Further, these systems can be distinguished according to their ability in measuring in aqueous solutions with respect to brewing conditions. Table 1.1 is summarizing major pros and cons, focus is given here on the more complex systems for measuring major substrates in fluids, more or less multivariate. The table shows, besides the difficulties mentioned for ultrasound measurements, there are obvious advantages. In addition to the same benefits like being non- invasive, inline applicability, rapid response time, low power consumption, excellent long term stability and high resolution and accuracy [15], there are low costs and advantages in areas of opaque medium due to its selectivity property (compared to optics in penetration and information depth [7]). Even though, it has to be mentioned, that opaqueness resulting from suspended particles may have negative impact due to attenuation and scattering of ultrasonic signal energy. Nevertheless, path length also does not play that important role as in transmission optics. Amongst all the listed possibilities and advantages, the light based systems are limited by non-scattered solutions, constant temperatures, and depending on the method used, on absorptions lower than a certain asymptotic value [10].

Table 1.1: selection criteria for different (assortment) measurement techniques for monitoring biological medium online [modified from [18]]; additionally, it has to be mentioned, that US is giving a bulk information over the path length (depending on the setup) and thus not spatially resolved in detail; furthermore, robustness against gas bubbles when using transmission mode for measuring is low and causes scattering

	UV/VIS/s-NIR	**NIR**	**MIR**	**Fluorescence**	**Raman**	**Ultrasound**
Selectivity	+	++	+++	++	+++	o
Sensitivity	+++	+(+)	+++	+++(+)	++(+)	+++
Sampling	+++	+++	+	++	+++	++
Working in aqueous medium	+++	+	+	++	+++	+++
Applicability	+++	++	+	+	+	+++
Process analytical tool	+++	+++	+	+	+++	+++
Length of fiber optics/coaxial cable	several	Up to 100 m	Few meters (Transmission)	several	several	~1 m, otherwise noise level increases
Signal	Absorption	Absorption	Absorption	Emission	Scattering	Absorption, Scattering
Acquisition mode	Transmission, Reflectance, Transflectance, Internal reflection	Transmission, Reflectance, Transflectance	ATR (Transmission), Internal reflection	Emission	Transmission	Transmission, Reflectance, Emission, Resonance
Relative costs	1	3-5	6-10	4-6	8-12	1
Robustness to:						
- **Temperature fluctuations**	Very high	Low	Very low	Low	Low	Low
- **turbidity**	Low	High	Very low	Very low	Very high	High*
Detection limits (ppm)	0.3	0.1 - 1	100 – 2000	25 - 150	< 0.1	1000**
References	[11, 18]	[7, 11, 18]	[7, 11, 18]	[11, 18]	[11, 18]	[13, 15]

*robustness of US against turbidity (color) is high, but particles cause energy dissipation

**approximated from a necessary density accuracy reported by Hoche *et al.*, 2013 [13]

1.1.1. Ultrasonic sensing

Even though acoustics, especially ultrasound is said to be not too specific (see Figure 1.1), it was shown earlier, that differences in various types of solutes (salt, sugar) as well as various sugar types are detectable [16]. Nevertheless, applications of ultrasound are rarely documented [7] and the general usage for process monitoring is low compared to its potential [15]. Numerous ultrasound implementations already exist, such as monitoring brewing process or concentration and particle distribution measurements [15], in which developments of hardware (electronics), software (computational effort of algorithms for analysis) over the past decade [e. g. [13]] and the comparably low cost of the necessary electronic components [17] show the capability for ultrasonic sensors. Furthermore, data driven strategies ("acoustic chemometrics", [1]) enlarges the measurable ultrasonic features by unspecific methods. Most commonly measured specific features are attenuation, sound velocity or impedance. These are dependent on three wave parameters that are phase, frequency and amplitude. Accuracy of ultrasonic measurements demands sampling rates in range of nanoseconds and picoseconds, the amplitude resolution is 12 bit and higher [13, 15], which prevails as challenges for electronics. Hauptmann *et al.*, 2002 and Hoche *et al.*, 2013 conclude, amongst others, that in general, all possibilities are temperature (gradient) dependent and accuracy limitations are caused by signal resolution (sampling rate as well as amplitude). Therefore, main enhancements for reaching the necessary accuracy should be achieved by correcting temperature gradients and maximizing signal to noise ratio.

Further, an ultrasonic setup is comparably simple and the costs are usually quite low. Furthermore, the field of US does cover first principle solutions, real physical parameters such as acoustical impedance, speed of sound, sound attenuation or density. Nonetheless, there are also unspecific parameters. Ultrasound based sensors and measurements are innovative and cost efficient and its potential is still higher than used, even with the technology available [e. g. Hauptmann *et al.*, 2002 [15]]. Taking together the fact of electronical and software developments with the possibilities of data driven modelling shown by Halstensen and Esbensen, 2010 [14], this field is of high interest in biotechnological applications.

1.1.1.1 Detection of substrate differences

In biotechnological fermentation medium, it is of most interest to monitor the substrates for biomass such as carbohydrates. It is known, that concentration changes of carbohydrates suspended in liquids is causing density changes. This leads to changing ultrasonic features such as speed of sound (c) or acoustical impedance (Z). Combining both under the assumption of small attenuation results in the bulk density as the ratio between impedance and speed of sound. For buffer rod setups, the prediction of impedance is possible using Equation 1. Therefore, the reflection coefficient has to be extracted out of the collected ultrasonic signal of the boundary between the setup specific buffer and the liquid sample of interest (subscript one and 2, respectively; for detailed theory refer to Hoche *et al.,* 2013 [13]).

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{1}$$

Since changes of concentration or density is of interest, just liquid phase related features need to be predicted. Whereas the prediction of speed of sound is mostly unproblematic in clear and bubble free liquids, it might fail in case of particles or gaseous inclusions, since the traveling wave is dissipated or attenuated.

Further, prediction of acoustical impedance needs highly accurate time and amplitude resolution [13]. Therefore, parts of this work are based on a data driven investigation of non-physical US features of buffer reflections to reduce the influence of suspended particles or gas bubbles and to overcome statistical inaccuracies by less accurate electronics.

1.1.1.2 Feature investigation

The following features have mostly statistical background. It is shown rudimentarily by following two demonstrations, that information such like acoustical impedance are covered by these features. This abstract is implemented to partially justify the usage of features just from buffer reflections of an ultrasonic pulse.

1. Energy balance

At planar interfaces, the following equation for pressure transmission takes effect, assuming an incident angle $\alpha=0°$ of the ultrasonic pressure wave and the definition of acoustic impedance:

$$T_p = \frac{2Z_2}{Z_1 + Z_2} \tag{2}$$

Combining Equation 1 as pressure reflection coefficient and Equation 2 together with the definition of acoustic intensity results in:

$$I_i = I_r + I_t \tag{3}$$

which is not surprisingly proving the law of conversion of energy (for detailed proof please be referred to Chapter 7.2 in David and Cheeke, 2002 [19]). Further, the transmitted intensity I_t is function in properties of the medium it is transmitted to. Therefore, it is clear, that also the reflected intensity is function in both, the properties of buffer and medium. This proves that properties of medium of interest are hidden inside of the buffer rod reflection of the ultrasonic pulse.

2. Interest on changing properties

In general it was shown, that concentration differences of the carbohydrates ethanol and sugar (dissolved in water) in the same time are detectable knowing density ρ, speed of sound c and acoustical impedance of the medium. However, if just one measurand changes, one parameter is sufficient. Therefore, just one property out of the signal is necessary, such as the reflection coefficient (Equation 1) as a function of acoustic impedance of both, the setup specific buffer rod and the liquid of interest. Under the assumption of constant temperature as well as c and ρ of the buffer rod, the reflection coefficient R is proportional to acoustic impedance Z_2 of the liquid. Further, the investigations presented in this thesis are dependent on changes in ultrasonic parameters. Hence, transforming Equation 1 and a derivation results in:

$$\frac{d}{dR} Z_2 = \frac{2Z_1}{(1-R)^2} \tag{4}$$

This proves, that changes in Z_2 are just dependent on R. Taking the theory of the so called “reference reflection method” (RRM) for prediction of R (for detailed theory please be referred to Hoche *et al.*, 2013 [13]) leads to:

$$R_{12,s} = R_{12,ref} * \exp(a_s - a_{ref}) \tag{5}$$

Whereas subscript 12 resembles the interface between buffer and liquid, s for sample and ref for a reference measurement in another medium (like water or air). The parameter a is the slope of the linearized equation

$$lnA_{rk} = a * k + b \tag{6}$$

whereas A_{rk} is the maximum amplitude of the k^{th} reflection (r). Under the assumption of steady surrounding conditions whilst investigating a sample (e.g. no changes in temperature, pressure or any other relevant physical environmental condition), $R_{12,s}$ is only function in slope a of Equation 6. This slope is dependent on the amplitudes of buffer reflections.

These relations prove that under the mentioned assumptions only signal changes inside the buffer are necessary for the prediction of one measurand. It further shows that the calculation of R is only function in the amplitudes of the reflection. The relevant temporal features shown in the investigation “Ultrasonic sensor for predicting sugar

concentration using multivariate calibration" and investigations related to the project "Prozesstaugliches Ultraschallmesssystem für die Überwachung und Regelung der Konzentrationen von relevanten Komponenten in industriellen Hefefermentationsprozessen" (AiF 16536 N, shown in Figure 1.2) are, amongst others, slope and entropy . The figure shows, that there are dependencies of features with respect to changing concentration and temperature visible after applying multivariate data analysis. Further, it indicates different importance of individual features selected.

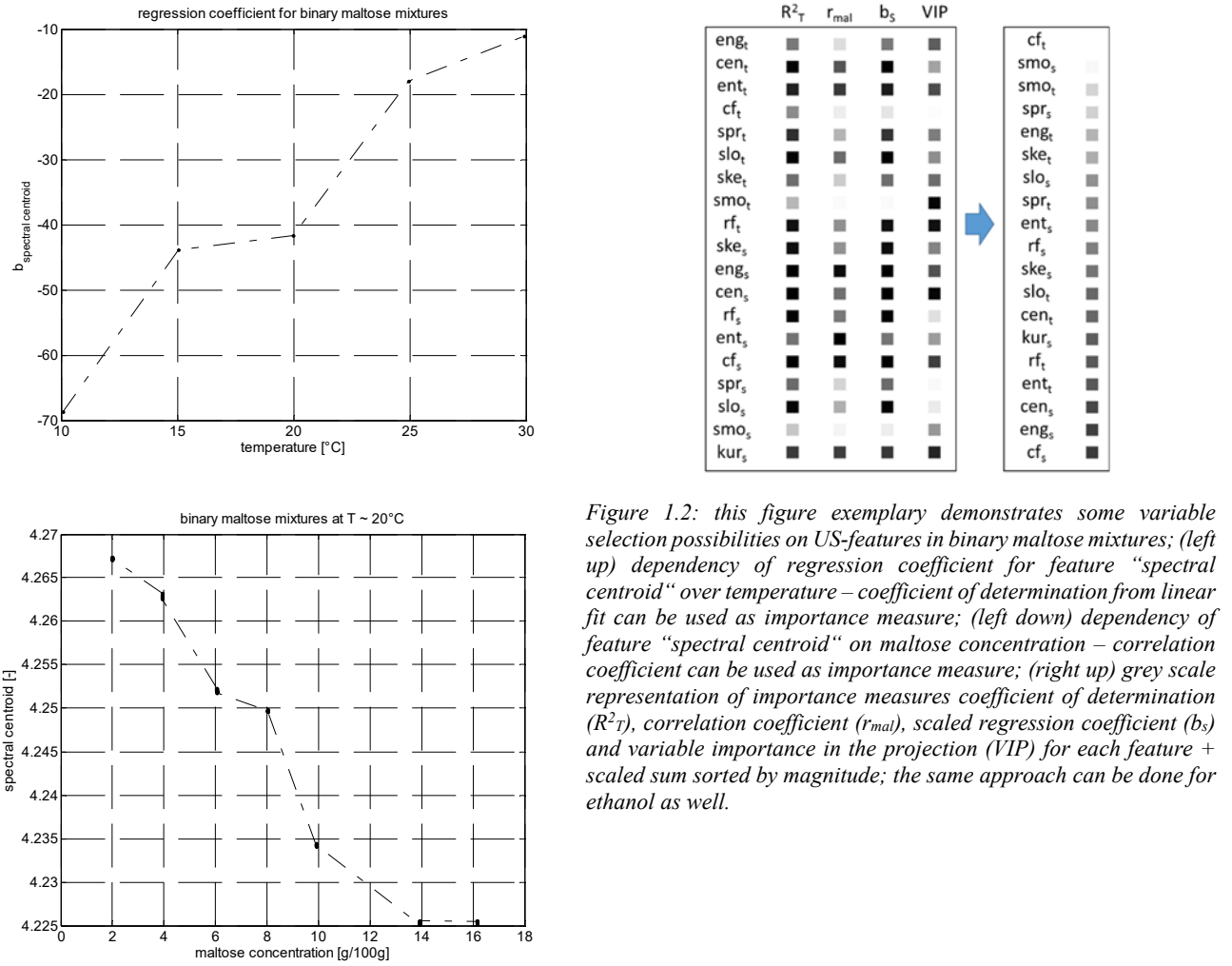


Figure 1.2: this figure exemplary demonstrates some variable selection possibilities on US-features in binary maltose mixtures; (left up) dependency of regression coefficient for feature "spectral centroid" over temperature – coefficient of determination from linear fit can be used as importance measure; (left down) dependency of feature "spectral centroid" on maltose concentration – correlation coefficient can be used as importance measure; (right up) grey scale representation of importance measures coefficient of determination (R^2_T), correlation coefficient (r_{mal}), scaled regression coefficient (b_S) and variable importance in the projection (VIP) for each feature + scaled sum sorted by magnitude; the same approach can be done for ethanol as well.

The slope (Equation 7) resembles similar meaning than the physical relation presented before (RRM method). The entropy (Equation 8) is a measure of disorder in the signal (higher values relate to noisy signals).

$$slope_t = \frac{N * \sum_{n=1}^{N}(n * A_n) - \sum_{n=1}^{N} n * \sum_{n=1}^{N} A_n}{\sum_{n=1}^{N} A_n * (\sum_{n=1}^{N} n^2 - (\sum_{n=1}^{N} n)^2)} \tag{7}$$

$$entropy_t = -\sum_{n=1}^{N} \left(\frac{|A_n|}{\sum_{n=1}^{N}|A_n|}\right)^2 ln\left(\frac{|A_n|}{\sum_{n=1}^{N}|A_n|}\right)^2 \tag{8}$$

The importance of the latter does not directly mean that the signals of investigation are noisy, but a numerical change of entropy of independent signals is linked to concentrations changes. However, both equations show that the values are function of amplitudes, too. Thus, it is most likely, that any of those features or a combination keep the information related to reflection coefficient. The temporal features used in the investigations are namely energy (as reported in [20] (short time energy), [21] and [22] (temporal energy envelope)), entropy (as reported in [20] and [23] (entropy based features)) and crest factor (as reported in [24]).

The relevant spectral features in the mentioned investigations are maximal magnitude [25], bandwidth [26-28], kurtosis [21, 22], skewness [21, 22], crest factor [21, 22, 24], energy [22, 29], entropy [24, 25], centroid [20-22, 29, 30] and spread [21, 22]. Those features are calculated similar to the equations for time domain but using the power spectrum. This is achieved by applying a Fourier transform on the time domain representation. The feature bandwidth is the range of frequencies over a certain amplitude. In this study, the change of bandwidth around center frequency was taken. Park *et al.* (1994a, b) [26, 27] as well as Mörlein *et al.*, 2005 [28] used this feature for investigations on fat content in meat samples. Skewness and kurtosis are statistical origin and describe the shape of a distribution. The crest factor resembles the ratio of the maximum magnitude to the average of the signal. Therefore, it represents the singularity of this property. The centroid resembles the point, where half of the energy of the signal is covered, the spread resembles the spread (variance) around the mean value (centroid) of the signal. The entropy is a measure of disorder in the signal information. Those features are described in more detail in Krause et al., 2014 [25]. Additionally, other features were taken into account but not further studied. Those are temporal smoothness [24], skewness [29], slope [22], centroid [21, 22, 29], spread [24], rolloff [24] and spectral smoothness [24], slope [21] and rolloff [20-22, 29, 30]. A detailed description of these features can be found in Hussein, 2013 [24]. Nevertheless, most mentioned references are from audio signal processing and it is known, that signals are linear time invariant. Still, the usage of features applied on a certain time frame of buffer reflections extracted out of US signals is possible under the assumption of quasi-stationary behavior in an appropriate time segment out of the full signal [24].

However, the successful usage of frequency domain representation of acoustic signals is shown in several references. The raising importance of acoustic sensors in general can be assumed from review of Halstensen and Esbensen, 2010 [14], Rathore *et al.*, 2010 [31] and Pomerantsev *et al.*, 2012 [32]. Even though the sensor setups are significantly different, the possibility of acoustic chemometrics on frequency domain representation of acoustic signals used for predicting particles sizes [33] or ammonia concentration [14] show the power of this relative simple and cost effective measuring principle. The detection of sodium chloride concentration in aqueous solution using multivariate data analysis and ultrasonic signals was shown by Schäfer *et al.* [34]. It could be shown, that detection of salt concentration using magnitude and phase in frequency domain together with temperature as predictors is possible via PLS regression. Even though the presented results look quite promising, some details are missing. First, the temperature dependence is not clear, since the range investigated is not presented. Further, discussion about the choice of final calibration model or the influence of chosen predictors is missing, respectively. From the chapter “material and methods” it might be assumed, that the used input data for calibration is built on the spectral representation of magnitude and phase. The temperature was entered to the predictor matrix as last column. Under the presented circumstances, it might have been beneficial to use PLS2 and temperature as second target value.

In summary, the last section proves the potential of feature investigation with respect to changing properties such as temperature or carbohydrates. This area of feature extraction for ultrasonic signals penetrating aqueous solutions opens a new option investigating such data pool. In combination with US setup investigations and the theory on impedance (see Hoche et al., 2013 [13]), this area show a promising path to novel information generation. This field is still quite new and shows various future possibilities.

One obvious disadvantage is the dependence on the measurement setup. However, investigations in this area of acoustics is quite low. Additionally, the coherences are not fully known. In such cases, data driven models can be used to overcome difficulties of such systems, where physical relations are partially missing. Such techniques are known to work superior in the field of optics, the application in acoustics for fluid inspection is quite rare. Nevertheless, calibration for optical systems are also setup specific. Therefore, the disadvantage in comparison is acceptable.

1.1.2. Biotechnological aspects

Another aspect in the aim of the investigated setup for online usage are the already mentioned difficulties for biotechnological fluids, such as gas bubbles or suspended particles. Both phenomena distort signals with respect to loss of energy. Thus, online application is often limited due to outliers (see chapter "discussion", Figure 3.11). In this investigation, features could help to overcome those challenges.

Nevertheless, there are two points to address when sensing a specific problem of interest.

(1) Is the sensing principle capable of sensing the relevant properties of interest?

(2) What sampling frequency has to be taken with respect to the progress of the investigated property in the process of interest?

The first point should be considered by determining the relevant time constants in the process under investigation. In this work, the temporal most critical process is the aerobic production of yeast under brewing conditions. This process is typically approached in a propagation tank, similar to the description in the fourth thesis publication, page 49. The sensor for investigating the progress of substrate or product is based on ultrasound and mounted in the pipe of the mentioned plant. Therefore, the following physical phenomena are assumed most relevant: mass flow of the fluid, which carries influencing particles and serves as driving force for substrate and temperature flux. The critical time constant T is given by a characteristic length, say 10^{-2} m divided by the velocity of the mass element. Knowing the volume flow in these processes investigated (around 1100 L/h) and the dimension of the pipe (diameter of $5*10^{-2}$ m) results in 0.153 m/s and therefore a time constant of 0.0654 s, thus in the range of 10^{-2}. Since ultrasound is in the range of 1500 m/s in aqueous fluids, its time constant is around 10^{-5}, which is lower than the time constant of the process (principle explained schematically in Figure 1.3). Therefore, it is legit to assume, that temperature or substrate gradients can be considered as controllable effects with respect to a property measured in average.

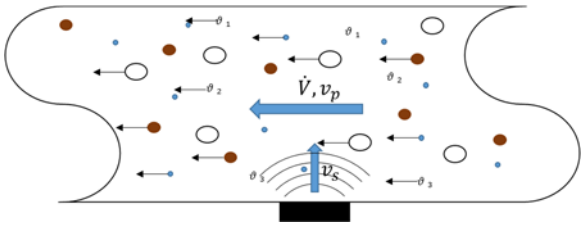


Figure 1.3: gas bubbles, particles, substrate and local temperatures with the velocity of product flow (subscript p) versus sensor information propagation velocity (subscript s)

Another influencing factor is the surface tension at the buffer of the sensor, where gas bubbles could agglomerate and therefore influence the measured signal. Visual inspections proved, that this effect is not influencing to a critical extent. Further, the influences are hard to estimate. Anyway, the material properties with respect to adhesion can be customized by production of the sensor setup.

The second point is considered with the following assumptions. Assuming a sufficiently high sampling frequency and the knowledge of the time constant for fermentation progress, it is valid to assume, that a number of temporal aligned signals should have similar properties (no significant change of any concentration, no major temperature change). Nevertheless, time constants of microbial systems are hard to estimate, in general caused by their complexity [35]. A respond might be very slow in case of a minor environmental change in any growth factor. A comparably lower reaction time of genetic adaption might be visible by changing temperature according to the used organism. One possibility presented by Szita *et al.*, 2005 [36] is simply taking the doubling time of the organism to estimate the critical process time (Equation 9).

$$t_{d,min} = \frac{ln2}{\mu_{max}} \tag{9}$$

In the applications presented in this thesis, the most frequently used microorganisms are from species *Saccharomyces spp.*, mainly *S. cerevisiae* or *S. pastorianus* var *carlbergensis*. The temporal process investigated under brewing conditions is the yeast propagation (growth of biomass). This process is usually never run at

temperatures higher than 28 °C. Since both organisms have similar properties, a maximal possible growth rate μ of 0.47 h^{-1} (see Sonnleitner and Käppeli, 1986 [37]) can be assumed.

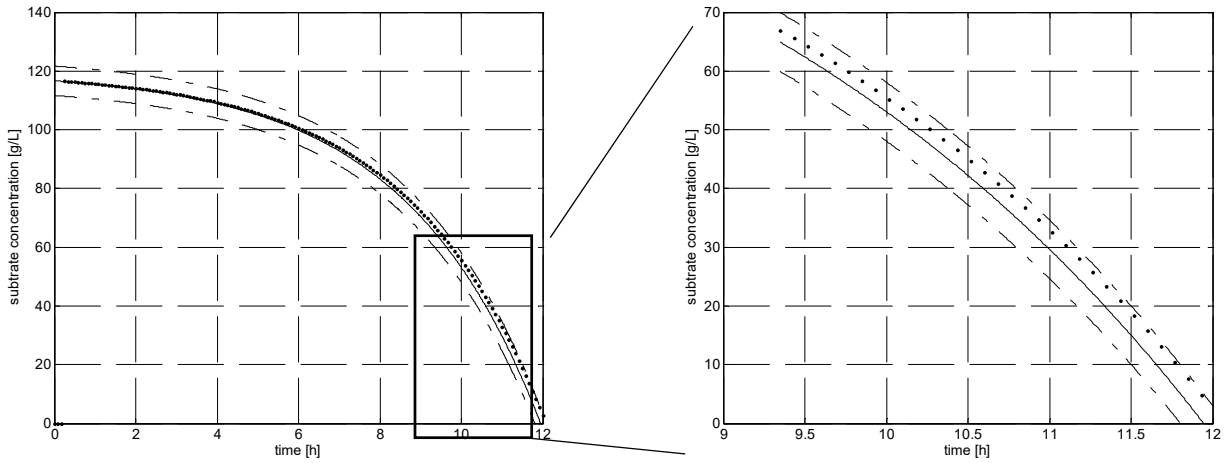


Figure 1.4: estimating critical process time – graphical illustration of the theoretical online trend deviation using moving average; the maximum acceptable error will be just reached at the end of the curve, which resembles areas of concentration never reached in industrial propagation process; black line - substrate decline; dashed black line(s) upper and lower Acceptable online error limit; black dots - theoretical moving average result of online sugar sensor – sampling rate 30 sec

This growth rate will just be reached, if a substrate concentration of approximately 0.1 g/L (no Crabtree effect, for details see Sonnleitner and Käppeli, 1986 [37]) and a temperature of 30 °C are given. Thus, using Equation 9 results in a critical process time t_{crit} of 1.5 h. Further, assuming a minimum total process time of 12 h and a more suitable maximum growth rate of 0.3 h^{-1} under mentioned conditions, results in the graphical representation shown in Figure 1.4. Both assumptions are never reached in industry under the normal brewing conditions. Additionally, a maximum acceptable measuring error of 0.5 g/100g for online sensor estimation is assumed. Thus, a time delay caused by an average filter of 30 signals with a sampling rate of 30 sec causes a "delay" of 0.5 g/100g at the end of this curve. Nevertheless, a maximal decrease of the initial substrate of 120 g/L down to 60 g/L is intended in real processes of this type. In conclusion, this delay would not influence in online applications.

Table 1.2: size of moving average window for online detection (n_{points}); standard deviation calculation (σ) of data set including temporal slope of US signal buffer reflections; 190 sample points at one temperature, no concentration change and the corresponding time delay (t_{delay}) in minutes (sampling time 30 sec.)

n_{points}	σ	t_{delay} [min]
20	6.96	10
40	2.55	20
60	2.5	30
80	1.4	40
100	1.31	50

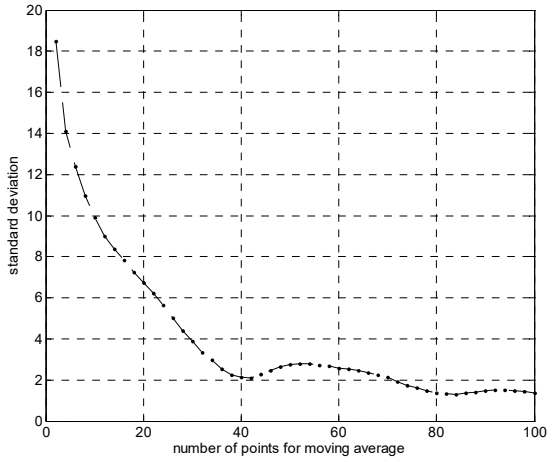


Figure 1.5: Investigation on the number of points for a moving average filter to reduce the standard deviation of US features – example temporal slope; data background are signals (buffer reflections) taken at approx. constant temperature on water (no change of concentration).

This information is necessary for the following two points. (1) Outlier detection of online signals using a number of temporal consecutive signals and (2) the effect on an average filter to reduce the statistical deviation of extracted US-features. The first issue will be discussed in detail in chapter "discussion" by using a buffer of signals for feature extraction. The second issue is shown in Table 1.2 and Figure 1.5. Those two representations show that 40 signals would cause an acceptable time delay of 20 min.

In addition, the applied feature analysis in this thesis was restricted to buffer reflections. Therefore, sampling rate (originally applied to collect a full signal travelling once through the setup being reflected once at the reflector; see first thesis publication for setup and explanation) could be reduced from 30 to 10 sec.

1.2 Multivariate data analysis

With the involvement of complex sensor data such as spectroscopy, computational data fusion techniques are necessary. Next to hard models, which use strict scientific relations, these soft models are data driven [7]. The latter can either be used as soft-sensor indirectly extracting leading parameters or as holistic process qualification. The biggest task in any of these systems is that the data processing reaches a predictive model for online use. Therefore, the field of multivariate data analysis, often referred to as chemometrics, provides plenty of possibilities for such tasks. In literature already a variety of reviews and tutorials are presented covering the basic methods used in multivariate data analysis. In general, one can divide between two tasks: (1) "quantitative model building", where the collected data is used to establish a prediction model for an unknown leading parameter (e.g. concentration of a chemical component) and (2) "qualitative model building", where data is used to classify samples measured (e.g. quality of samples). Figure 1.6 tries to give an overview on classifying methods and further subdivisions in multivariate data analysis. The methods used in this thesis are marked blue, the number of algorithms on each subclass should be seen as examples for numerous other possibilities. Three of the used methods, namely Support Vector Machines (SVM, originally used for classification), Partial Least Squares – Discriminant Analysis (PLS-DA, modified for classification) and unfold-PLS (Wold *et al.*, 1998 [38], for three dimensional data problems) are overlapping with other subdivisions (indicated by the red dashed arrows).

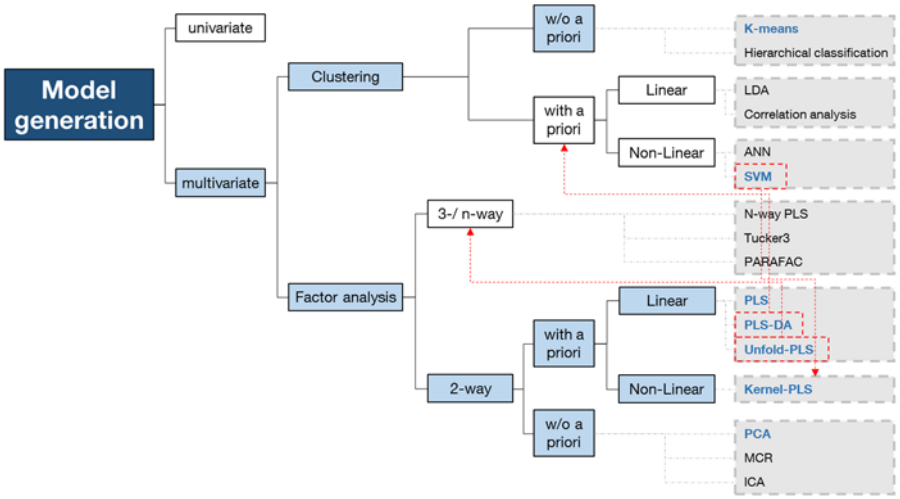


Figure 1.6: Methods (assortment) used for model generation in chemometrics based on different background; without (w/o) a priori knowledge is also referred to as unsupervised, with a priori as supervised, respectively; methods used in this thesis are coloured in blue, whereas Unfold-PLS is used for multivariate statistical process control (MSPC) (thus linked to 3-way methods) and PLS-Discriminant Analysis (PLS-DA) for classification (here referred to as with a priori knowledge clustering); SVM is used for regression and thus linked to nonlinear regression methods such as kernel-PLS; LDA – Linear Discriminant Analysis, ANN – Artificial Neural Networks, SVM – Support Vector Machines, PARAFAC – Parallel Factor Analysis, PLS – Partial Least Squares, PCA – Principal Component Analysis, MCR – Multivariate Curve Resolution (here referred to w/o a priori knowledge, even though the methods initial estimation as well as constraints can be seen as knowledge inclusion), ICA – Independent Component Analysis; this figure is adapted from Gendrin et al., 2008 [4]

Figure 1.6 indicates that multivariate data analysis is a brought scientific field. Next to the mentioned methods for model generation, it implies amongst others investigations in inputs (importance), non-linearity, and model dimension.

In general, multivariate data analysis can be divided into the following steps:

(1) Data pre-processing

(2) Model generation

(3) Data post-processing

Thus, Figure 1.7 gives an overview on the methods and algorithms applied in this thesis with respect to their functionality. Furthermore, some subdivisions are overlapping with other steps mentioned before. A selection of methods for each subclass will be explained in the corresponding following section.

1.2.1 Data Pre-processing

Figure 1.6 provides an overview of the used algorithms in this thesis with respect to their functionality or subclass, respectively. Each of the mentioned pre-processing subclass and its corresponding methods will be explained, a good overview including a variety of other algorithms is given by Axelson, 2012 [39].

1.2.1.1 Offset

Eliminating an offset can be accomplished by column wise **mean centering** of data. Therefore, differences in the data are stronger visible rather than similarities. Mean centering is used in this thesis and is one of the most common applied pre-treatment methods, almost independently from the source of data [39]. Nonetheless, Axelson also mentions, that it is not advised for data sets where the responses change linearly with the targets or have no baseline. The first might be a reason for errors reported in US-investigations in this work and thus need to be investigated in further studies.

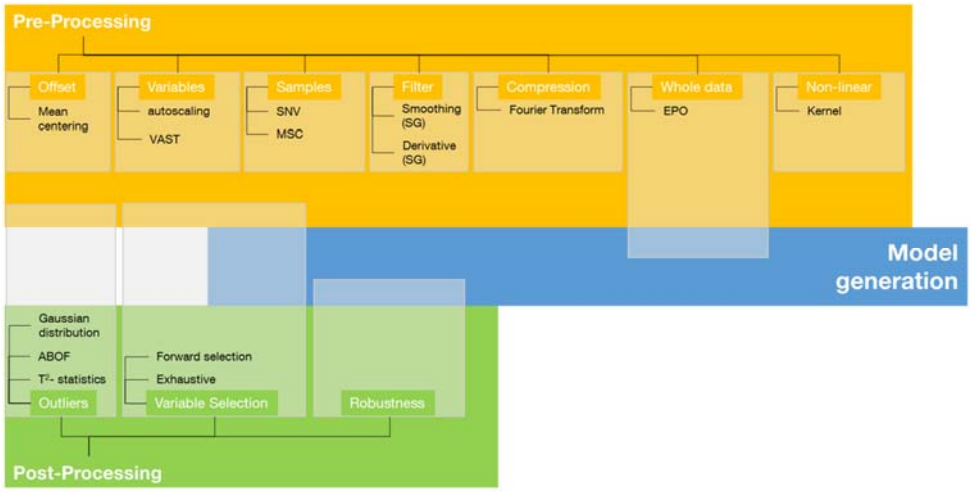


Figure 1.7: Methods used in this work for pre-processing and post-processing of data; some of the methods typically cannot be strictly classified as indicated – thus, overlapping with the respective group is graphically highlighted; VAST- variable stability scaling, SNV – standard normal variate, MSC – multiplicative scatter correction, SG – Savitzky-Golay-Filter, EPO – external parameter orthogonalisation, ABOF – angle based outlier factor; this figure is partially adapted from Miller, 2010 [3].

Baseline correction methods also belong to the group of offset, usually applied for spectral pretreatment [3]. The latter is partially included in next section "sample scaling".

1.2.1.2 Variable scaling

One very common applied method is called **auto-scaling** or **z-transform**. Here, mean-centered variables are additionally divided by their standard deviation to remove the total variance inside the data. For this methodology, data pre-knowledge is necessary. Firstly, to prevent amplification of noise in case of variables with low variance with respect to a given target of interest. Secondly, for handling variables with comparably low variance but high informative content to the respective aim. Thirdly, once variables do not have the same range of magnitude the variables with largest absolute will dominate the rest [3].

In case of high signal to noise ratio, the variance of some variables is artificially increased. This effect mostly leads to false correlations, which would even be enlarged by auto-scaling. Thus, a supporting scaling factor for additional stabilization was introduced by Keun *et al.,* 2003. This so-called **variable stability scaling** (**VAST**) can support data analysis if there is spurious variation appearing. In this thesis, auto-scaled data is divided by this scaling factor called coefficient of variation (ratio between standard deviation and mean). Thus, spectral areas of higher information content were stressed in contrary to areas with more noise related variation.

This method can also be used including *a priori* knowledge, such as class membership (**sVAST**). Thus, treatment is applied for each subgroup in calibration data individually [39, 40].

1.2.1.3 Sample scaling

This kind of pre-processing might be necessary for correcting multiplicative variations between samples caused by e.g. light scattering in spectral optical signals due to path length differences (caused by e.g. particle size or

thickness differences). It is not possible to remove such multiplicative variations by derivatives or variable pre-processing. One possibility to correct light scattering is using the **standard normal variate** (**SNV**) on each signal [41]. Here, the mean value of each signal will be subtracted from each signal value and divided by the signals standard deviation. Thus, gradient differences of spectra will be removed. These variations could be somehow convoluted with the expected ones related to the problem of interest [3].

In this thesis **multiplicative scatter correction** (**MSC**) is also used. It is based on the idea to correct the level of the base scatter of all spectra to the niveau of an ideal spectrum. In most cases, the mean spectrum is taken as ideal or reference spectrum. Each spectrum will be adapted to the mean spectrum using least squares [3]. The corrected spectrum will be calculated using the fitted constants a_i (intercept, additive correction factor) and b_i (slope, multiplicative correction factor) [3]. Thus, MSC is a data set dependent transformation [42].

1.2.1.4 Filtering

Pre-processing by filtering can be applied in case of variables being presented as continuous numerical representation (discrete, digitized data). This case is existent in optical or acoustic data, where the variables are represented as continuous function. One of the famous filtering method in chemometrics, especially for spectral data, is the **Savitzky Golay** (**SG**) **filter**. This method can be used either for smoothing or for derivatives of individual spectra. Derivatives remove baseline shifts between spectra as well as improvement of resolution between overlapping features. Huang *et al.*, 2008 for example report the application of second-derivative transformation together with extended multiplicative spectral signal correction in order to improve band resolution and remove physical and quantitative spectral variations [43]. Anyway, the derivative order, which is one of the three parameters of SG filter (window width, polynomial and derivative order), could be zero for just smoothing of data, one to effectively remove baseline shifts and two for additionally removing baseline slope differences between spectra. The window width determines the degree of smoothing as well as the deterioration of resolution improvements in an opposing manner. The SG filter itself is a set of coefficients defined by the three parameters and sequentially multiplied to local windows in a moving window manner. The content of this paragraph can be found in more detail in Miller *et al.*, 2010 [3].

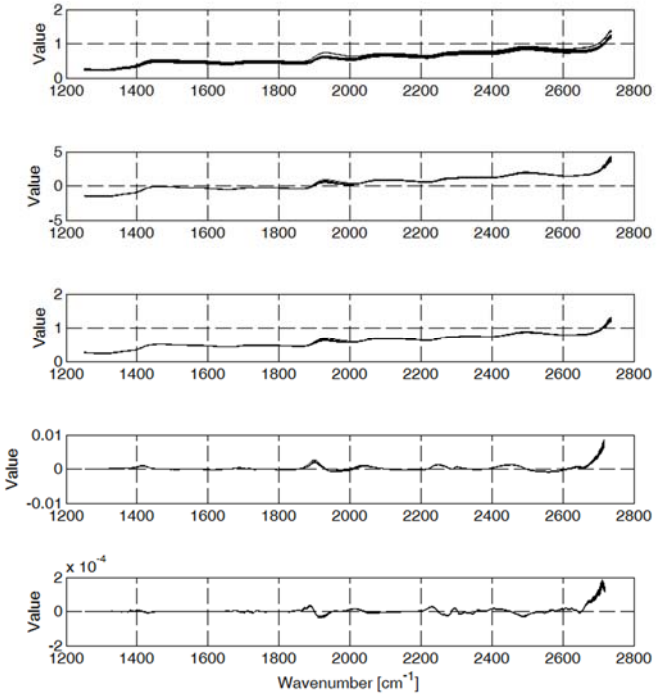


Figure 1.8: sample scaling/filtering (selection); from top to bottom: original data, standard normal variate (SNV), multiplicative scatter correction (MSC), first and second derivative using Savitzky Golay algorithm (number of sampling points N=21, polynomial degree p=5).

The methods mentioned under “sample scaling” and “filtering” were applied on spectra used in this thesis. Figure 1.8 presents these original spectra, deviating in their absolute values from each other, almost at each wavelength. Obvious changes by sample scaling are reduced differences in absolute values between spectra, caused by the aforementioned influences. Applying SG filter for first derivative highlighted areas, where changes in the slope of the spectra are appearing, most likely caused by changed sample composition. The same counts for the second derivative, which highlights changes in the absorption peaks.

1.2.1.5 Linearization – Kernel Matrix

Several different ways for linearization of potential non-linear data are reported in literature, such as non-linear extension of data matrix by **polynomial extensions** [44] or orthogonal signal correction [39]. Another common way is the construction of a **kernel matrix**. A certain sample set of input data **X** [n x m] is restructured using the direct dependence of two samples to each other. This means, the samples of **X** are transformed into a new feature space using nonlinear mapping [45]. This results in a new input matrix **K** [n x n] (Equation 10).

$$K = \begin{bmatrix} k_{1,1} & \cdots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{n,1} & \dots & k_{n,n} \end{bmatrix} \tag{10}$$

Where $k_{i,j}$ can take the form of several different functions. The most famous are the linear or covariance, the polynomial and the radial basis function kernel (RBF, Equation 11). The description of those functions can be found in literature, for instance Nicolai *et al.* 2007 [45].

$$k_{i,j} = e^{\frac{-\left\|x_i^T - x_j^T\right\|^2}{2*\sigma^2}} \tag{11}$$

The Kernel width parameter σ is linked to the reliability or the SNR of the data. If this parameter is higher, the solution of the model becomes more linear. Over all, the value of $k_{i,j}$ becomes one in case of samples that are more similar and zero in case of less similar ones.

Further, it is recommended to always perform a centering of the Kernel matrix prior to analysis. Therefore, an approach reported by Bennett and Embrechts was applied [46].

1.2.1.6 External Parameter Orthogonalisation

This methodology used in this thesis can be found in detail in Roger *et al.*, 2003 [47] and in appendix A1. They present a data matrix as follows:

$$\boldsymbol{X} = \boldsymbol{XP} + \boldsymbol{XQ} + \boldsymbol{R} \tag{12}$$

Where **P** contain the loadings of the projection onto the relevant target information (relevant subspace $\vec{\mathbf{C}}$) and **Q** onto the external parameter influence (subspace $\vec{\mathbf{G}}$, containing influence of external parameter); **R** resembles a residual matrix. After some intermediate steps the subspace $\widehat{\mathbf{G}}$ is estimated by PCA (see appendix A1). The original data matrix is modified by subtracting the influence of the external parameter following Equation 13:

$$\boldsymbol{X^{0*}} = \boldsymbol{X^0}\left(\boldsymbol{I} - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{G}}^{\boldsymbol{T}}\right) \tag{13}$$

Finally, a calibration can be calculated between $\mathbf{X}^{0*}$ and $\mathbf{Y}^0$. Two of several possibilities are presented in Roger *et al.*, 2003 [47]: a k-fold cross validation on the different $\mathbf{X}^i$ resulting in an error as a function of EPO component and PLS latent variables number. The second approach is based on an analysis of variance measured by Wilk's ratio between the inner group and the total variance [47]. The approach used in the presented thesis is slightly different. The choice is done on the final error (similar to the approach one above). In contrast to the mentioned approach, the error is calculated on a validation data set. This way was chosen to combine the approach with kernel PLS and the method of robust calibration.

1.2.2 Model generation

Aim of scientific research in PAT is to better understand the processes and backgrounds of the challenges investigated. This implies the option of increasing complexity with respect to knowledge integration. In most PAT tasks, not the individually measured values are of interest but more importantly functional features of raw material or product are required [7]. One of the major benefits of using process analysis is to simultaneously optimize

process behavior and product functionality [7]. This is reached, amongst others, by trying to use all the information existing and a joint analysis. Amongst others, a review about mathematical procedures for model building on bioprocess data is given by Becker and Krause, 2010 [48]. Most popular modelling strategies start from low complexity and knowledge, such as data driven modelling, until process modelling using physico-chemical relations via differential equations (deterministic models) or knowledge based approaches such as fuzzy logic or artificial neural networks. Even though the latter include most knowledge about important relations in a biochemical process surrounding, the application of such systems in most cases is of rather low success. For deterministic models it can be explained by the huge complexity of bioprocesses resulting in a variety of equations and unknown parameters [44]. For knowledge based approaches it is often necessary to have enough experience providing sinful relations and connections [44]. If such background information is not available, the usage of data driven modelling becomes beneficial. Even in such approaches, it is achievable to include knowledge. It was shown by using multivariate curve resolution (MCR) [7] to integrate biochemical knowledge into data driven analysis. Altogether, the following three guiding principles given by Miller, 2010 should be considered [3]:

“

1. When building a method for on - line use, keep it simple! Strive for simplicity, but be wary of complexity.
2. Do your best to cover the relevant analyzer response space in your calibration data. If this cannot be achieved, then at least know the limitations in your calibration data.
3. Regardless of your background, strive to use both chemical and statistical thinking
    a. Use your prior knowledge to guide you (chemical thinking),
    b. But still ‘listen’ to the data – it might tell you something new! (statistical thinking),,

The background of the presented solutions in this thesis is majorly the data driven and well-known Partial Least squares (PLS) method. This method is used quite often in the field of chemometrics model generation. Further, a variety of different algorithms for PLS is existing. The used ones are namely Non-linear Iterative Partial Least Squares and kernel PLS. Here it is important to mention, that literature presents two very different kernel PLS approaches. On the one hand, kernel is referred to as data pre-processing, thus dealing with non-linearity. The subsequent data transformation is based on adapted NIPALS algorithm. On the other hand, kernel-PLS is a different algorithm for decomposing the data matrix based on the kernel X^TYY^TX, proposed by Lindgren *et al.*, 1993 [49] and used in the work of Whitehead, 2012 [50] and in the third thesis publication.

Further, one can divide between the way of decomposition (kernel-PLS, NIPALS, SVD, and further methods reported by Burnham *et al.* 1996, Lavine and Workman, 2010 or Miller, 2010 [3, 51, 52]) and the adaption of PLS in principle to the usage of interest (e.g. PLS-DA, unfold-PLS, kernel-PLS).

However, several other approaches for model generation, namely Support Vector Machines (SVM), K-means clustering and Principal Component Analysis (PCA) are used in this thesis. In the following subsections, the algorithms shown in Figure 1.6 including the marked ones used in this thesis are explained under their subdivision with respect to their aimed usage.

1.2.2.1 Pattern Recognition

This topic describes methods for automatic detection of different groups, patterns or clusters out of an n-dimensional space of variables. Cluster analysis, which is based on the dissimilarity measure between objects [53], is one of the most known methods for unsupervised pattern recognition [54]. These methods without *a priori* knowledge include algorithms such as **K-means**, which is one of the most widely applied algorithm [53]. Important for these methods is the consideration of pre-processing, in particular variable scaling. Hastie *et al.,* 2008 shows, that the differences of clusters might disappear after applying this step (see Hastie *et al.,* 2008 [53], Figure 14.5). However, next to K-means there are varieties of other methods used for clustering, such as

hierarchical clustering or proximity matrices. These methods include (amongst others) PCA (Principal Component Analysis, mentioned in section 1.2.2.2 "Multivariate data analysis for regression") as well as Independent Component Analysis (ICA). The major difference in the latter two is the identification of unique components in ICA, which makes ICA applicable when PCA is limited [53]. The area of unsupervised as well as the following supervised algorithms are briefly discussed (amongst others) in Brereton, 2009 [55] and more application oriented in Hastie *et al.* 2008 [53].

Methods with *a priori* knowledge (supervised) applied for linear challenges include Linear **Discriminant Analysis** (LDA) and correlation analysis, for non-linear cases **Support Vector Machines** (SVM) and Artificial Neural Networks (ANN) are widely known. The most famous approach, mainly known for calibration of multivariate data sets is Partial Least Squares (PLS, explained in section 1.2.2.2 "Multivariate data analysis for regression"). A slight adaption of this algorithm allows classification as well. This method is referred to as PLS-DA, which is discussed in the third thesis publication as well as in Barker and Rayens, 2003 [56].

Whereas PLS is used for linear and slight non-linear cases, challenges of non-linear background are treated with methods like support vector machines (SVM, also used for regression [SVR]) The development of general non-linear SV algorithms in its present form was mainly done by Vapnik and co-workers in the early 20th century, even though it roots back to the framework of statistical learning theory developed in the 60ties and 70ties [57, 58]. In 1997, Vapnik *et al.* extended the framework to non-linear regression [59]. Tutorials on classifiers and regression are given by Bruges, 1998 [60] and Smola and Schölkopf, 2004 [58], respectively.

Further, the objective function for calculating the regression coefficients is different to that known from PLSR or MLR. Instead of minimizing the sum of the squared residual error, an alternative formulation serves as optimization background:

$$\min\left(\frac{\sum_{i=1}^{n}\delta_i}{2}+\theta\frac{b^T b}{2}\right) \tag{14}$$

Where θ is a parameter for regularization. Increased values of θ forces more on the root mean square magnitude of the regression coefficients. Instead of the classical least squares criterion, δ_i is defined by a significance threshold ε:

$$\delta_i=\begin{cases}0, & if\ |y_i-\hat{y}_i|<\varepsilon\\ |y_i-\hat{y}_i|-\varepsilon, & otherwise\end{cases} \tag{15}$$

Therefore, a sample with low residual will get a regression coefficient equal to zero. This means, that the others can easily describe those samples. The other samples with coefficients bigger than zero are referred to as support vectors.

Another formulation for SVR is given by Least Squares- SVM (LS-SVM) [61]. The difference to the objective function described above is usage of the general squared loss function instead of Equation 15. Therefore, all regression coefficients will be nonzero. Further, instead of three tuning parameters (θ, σ, ε) only two have to be tuned (θ, σ). The training of the model is achieved by solving the linear Karush-Kuhn-Tucker system (Equation 16).

$$\begin{bmatrix}0 & \boldsymbol{I}_n^T\\ \boldsymbol{I}_n & \boldsymbol{K}+\boldsymbol{I}/\gamma_i\end{bmatrix}\begin{bmatrix}b_0\\ \boldsymbol{b}\end{bmatrix}=\begin{bmatrix}0\\ \boldsymbol{y}\end{bmatrix} \tag{16}$$

Where **I** is a [n x n] identity matrix, $\mathbf{I}_n$ a column vector of ones, **y** the vector of reference values and 1/γ corresponds to θ [57]. This algorithm is implemented in this work, results are shown in chapter "discussion".

Another famous, non-linear algorithm used for pattern recognition and regression is named artificial neural networks (ANN). ANN have an adjustable mathematical background capable of building a non-linear relation between many inputs and multiple outputs with features such as on-line application and learning ability but more

or less empirical modeling capacity and less extrapolation capability (see Goyal, 2013 [62]). They are more dependent on the data than on comprehension of any relation between regressor and target of interest [62]. Nevertheless, a major disadvantage of ANN is their complexity and thus not easy interpretability.

However, the majority of data-driven models (including the ones mentioned in this thesis) will not be able to replace any of the laboratory measurements totally. They can be seen as powerful addition and reduction in effort. Further algorithms are mentioned in literature, e. g. Becker and Krause, 2010, Hastie *et al.*, 2008 or Brereton, 2009 [48, 53, 55].

1.2.2.2 Multivariate data analysis for regression

Two-way models

One of the major goals of multivariate data analysis is to find a relation between an n-dimensional data matrix and multiple samples with respect to corresponding observations. The two basic methodologies are named factor analysis and **principal component analysis** (**PCA**). There are similarities but also differences between these two methods [48, 53, 55], both analyzing one data matrix. These algorithms are also called component models [63]. Factor analysis was not applied in this work. PCA is basis for the widely applied **Partial Least Squares** algorithm (**PLS,** also referred to as Projection to latent structures). PLS is a method established for analysis of two data matrices simultaneously in a regression point of view, maximizing the covariance between x and y-data. The final model describes the approximated coherences between measured x-data and the desired target y [12]. PLS is the main method applied in this thesis, the algorithm of PLS is described in the second thesis publication. All those algorithms assume errors, noise or uninformative parts in the measured data. Thus, the problem of interest can be described by a condensed or reduced dimensional space of the data origin [64]. This is achieved by projecting the higher dimensional data onto latent variables, which reduces the dimensionality and simplifies any further analysis (calibration, monitoring and control, etc.) [64]. Other variants are Reduced rank regression (RRR), Canonical variate analysis (CVA), or canonical correlation regression (CCR) [51, 64]. The choice of method depends on the objectives. A detailed description of these algorithms is not given, the reader is referred to literature. Another distinction between these mentioned algorithms can be made based upon the knowledge included, whereas PCA, Multivariate Curve Resolution (MCR) and Independent Component Analysis (ICA) are categorized by Gendrin *et al.*, 2008 without and PLS with *a priori* knowledge [4] (see Figure 1.6).

Another objective is the consideration of non-linearity. Such issues can be handled with corresponding pre-processing (as mentioned in the paragraph before). However, there are also varieties of algorithms for model generation facing this challenge. Two methods are used is this thesis, namely **Support Vector Regression** (see paragraph pattern recognition) and **Kernel-PLS**. This algorithm is modelling non-linearity by PLS-regression on a kernel matrix like presented above. This feature space might have a more suitable content than the original variable space. However, the used PLS algorithm (for instance NIPALS) has to be rearranged to avoid an unreasonable computational effort due to the possible high dimension of the kernel feature space. Examples are given for instance in Bennett and Embrechts, 2003 [46], Nicolaï *et al.*, 2007 [45] or Krämer and Braun, 2007 [65]. The algorithm applied in this thesis was taken from Nicolaï et al., 2007 [45]. After extracting the latent variables, the regression model is built as follows:

$$\hat{\boldsymbol{y}}_{\boldsymbol{c}} = \boldsymbol{T}\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{y}_{\boldsymbol{c}} \tag{17}$$

for the calibration data and

$$\hat{\boldsymbol{y}}_{\boldsymbol{v}} = \boldsymbol{K}_{\boldsymbol{v}}\boldsymbol{U}\left(\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{K}\boldsymbol{U}\right)^{-1}\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{y}_{\boldsymbol{c}} \tag{18}$$

for validation data. All the presented equations were implemented and used on the same data set as presented for LS-SVM model building. SVM Regression & Kernel PLS results are presented in chapter “discussion”.

Three or n-way models

The last paragraph describes the basic techniques for multivariate analysis applied on two dimensional data structures. However, some data structures are of three or more dimensional kind, also referred to as multiway. A famous example is 2D fluorescence spectroscopy, where for each excitation wavelength the full spectrum of emitted fluorescence is measured. Applying this method on different samples results in three dimensions. Taking location and time into account, such data structures could have five dimensions. Other examples are chromatographic data or data measured under different chemical (pH) or physical (temperature or pressure) surrounding conditions [66]. Methods for multiway analysis are amongst others PARAFAC (PARAllel FACtor analysis), Tucker3 (both extensions of two-way PCA [63]) and its combination as well as PLSR for nPLS models. An overview of possibilities is given by Bro, 1998 [66]. Detailed description of algorithms is given in Bro, 1998 [66] or Smilde *et al.* 2005 [63], a brief comparison of methods is given in Smilde *et al.* 2005 [63]. However, Smilde *et al.* also state, that an *a priori* choice of model is mostly not possible [63]. This fact is supported by the statement of Bro in using constraints to improve a model and usually a model cannot be optimal in all cases [66].

There are two major advantages of using multi-way techniques. The first is uniqueness [66], which allows calibration in presence of unknown impurities. The second is better structural models [66], which is beneficial for robustness, economic models and most often better prediction accuracy.

Although these methods are obviously powerful, in this work the method presented by Wold *et al.*, 1998 [38], unfolding the three-way structure into a two-dimensional matrix combined with more or less standard PLS, was applied (similar to Tucker1 [63]). As presented by Whitehead, 2012, this method was adopted and applied for multivariate statistical process control, even though being not a truly tri-linear decomposition method [67].

Multivariate statistical process control

It is possible to distinguish between the usage of sensor data on the calibration of a specific target (usually a chemical component) or the complete process behavior. The latter is known as multivariate statistical process control (MSPC), which is mostly used for provision of multivariate trajectories for process inspection and control. The successful application of multivariate data analysis used even for industrial process monitoring, control and fault diagnosis has increased, especially in the last decades [64].

Kourti reviews methodologies and transfer possibilities for industrial usage on the emerging applications of latent variable extraction applied for monitoring and control and based on image analysis as inexpensive sensor for such occasions, for example [64]. It is also mentioned, that any data treatment has to be utilized with care to preserve multivariate structure and integrity [64].

One of the major benefits in using multivariate analysis is, that among predicting nominal changes in single variables changes, the covariance between them is also detectable. This means, that also minor changes in any variable causing major changes of the whole process trajectory are detectable, earlier than each variable deviating significantly on its own [64]. Grassi *et al.* (2014) give an example on the process of beer brewing. The main aim of the study is to extract the relevant process leading parameters online using NIR. Nevertheless, they also point out the difficulties in univariate quality assurance in industrial processes [68]. More specific, it is hard to find the error source for the abnormality of a certain batch [64], since a misbehaving variable could be originated by a combination of others. Thus, a broader view on the holistic information given is recommended.

One of the aims of multivariate process control is the real time release of products (rather than time-consuming laboratory proved release). This aim is even emphasized by the FDA Guidance released in 2004 [69]. Thus, a product will be released, if the process was inside its boundaries and just analysed in laboratory, if problems occurred.

1.2.2.3 Model Validation

One of the major task in latent variable methods is the final model validation. This step is quite important in choosing the correct amount of necessary components rather than over- or under- fitting the problem of interest.

Thus, the chosen criterion has to have the background which allows comparison between models, also of different origin [70]. Estimating the dimensionality can be accomplished using the predicting error (see below). Other possibilities, which are helpful to choose more appropriate model sizes with respect to over fitting are, amongst others, so called information criteria. Such criteria are manifold including the Bayesian and the Akaide Information Criterion (BIC and AIC, respectively) [70]. Amongst others, the AIC was used for analyzing the model size in this work.

$$AIC = n * \log(SSr) + 2 * df \tag{19}$$

Where n is the number of samples, SS_r the sum-of-squares of residuals and df the degrees of freedom. Choosing the correct degrees of freedom is one of the discussion points of several publications (e.g. Faber and Bro, 2002 for the prediction error [71], Bruns *et al.*, 2006, Chapter 5, p. 227 for ANOVA of least-squares fit (amongst others) [72] or Krämer and Braun, 2007 for the model selection on information criteria [65]). In contrary to Krämer and Braun, in the presented work the "naïve" approach df(PLS) = number of components was used.

However, literature presents a variety of criteria and methods for error propagation but no overall solution for all given cases is reported. Axelson reports some aspects of proper validation out of the work of Esbensen and Geladi. One of such aspects is, that test set validation should be the first choice for correct validation objectives [39, 73]. Axelson further heavily recommends to always use a completely independent validation set, never "in touch" with any model generation to not result in too optimistic model predictive quality [39]. However, this might not always be possible due to a limited sample size in the data set.

In conclusion, several different criteria were used for decision on best model size (number of latent variables). AIC as well as the RMSEV are complemented by the sum of squares for the model precision (SS_a, resembling the prediction accuracy of each level, e.g. concentration) and sum of squares for model validity (SS_r, resembling the overall regression accuracy).

$$SS_a = \sum_i^m \sum_j^{n_i} \left(\hat{y}_{ij} - mean(\hat{y}_i)\right)^2 \tag{20}$$

$$SS_r = \sum_i^m \sum_j^{n_i} \left(y_i - \hat{y}_{ij}\right)^2 \tag{21}$$

Both are calculated for calibration and validation. They are based on the theory of ANOVA for empirical model building, for instance shown by Bruns *et al.*, 2006, Chapter 5 [72]. Further, the mean sum of squares are calculated for comparison:

$$MS = \mathrm{SS}/df \tag{22}$$

where df is the degree of freedom; for "precision" (subscript a) df = m - p – 2 and for "validity" (subscript r) df = n - p – 2 with m as the number of levels, n the number of samples and p for the number of parameters/components and subtracted by two for pre-processing by auto-scaling. This mean sum of squares can be compared in several ways. Brereton ([74], section 5.7.2) reports on using latent variable methods for pattern recognition comparing the percentage of classification with cross-validation to "autoprediction" (error calculated between model and training set, also sometimes referred to as calibration error, in this work subscript c). If these two are similar, the model might be of good choice. If the accuracy for validation (cross-validation, subscript cv) decreases, the results should be further investigated [74]. When building a ratio (autoprediction/cross-validation) between these accuracies, it will increase with rising number of latent variables. This fact can be also used for regression, where the ratio would decrease. Further, the mean sum of squares can be taken in a ratio with similar conclusion, rather than using the final model error. This measure can be used as an indicator for over-fitting. Cross-validation is the better choice for this ratio, because in test set validation (subscript v) for regression values

above one are also possible. Nevertheless, it still can be used for analysis assuming that MS_{rc} decreases and MS_{rv} increases with too many number of latent variables in the model.

$$OF = MS_{rc}/MS_{rv} \tag{23}$$

Another way of comparing mean sum of squares is presented by Brereton, 2003, section 4.3.3.2 [75] where the MS with a certain number of latent variables p is compared to the one with p-1. If this value [$MS_{rv}(p\text{-}1)/MS_{rv}(p)$] (adapted from [75]) reaches below one, the model with p-1 latent variables should be taken [75]. This can be even further analysed by assuming, that any additional component does not affect the model output anymore, resulting in an asymptotic behaviour.

$$R_a = MS_{av}(p-1)/MS_{av}(p) \tag{24}$$

$$R_r = MS_{rv}(p-1)/MS_{rv}(p) \tag{25}$$

The last ratio investigated in this work is between precision and validity (Equation 26).

$$R_{P/V} = MS_{av}/MS_{rv} = Presicion/Validity \tag{26}$$

All ratios as well as the comparison of individual trend lines where analysed (see chapter "discussion" below). The reason in using several measures is, to support the choice of model size and accuracy as well as robustness in the same time (see section "Model Robustness").

1.2.3 Data Post-processing

Even though this naming is not usual in literature, the following topics are included under this header: model validation, outlier analysis and variable selection/inspection. All topics are necessary to stabilize the solution of a model.

1.2.3.1 Variable selection/inspection (for Partial Least Squares regression)

Multivariate calibration models consist usually out of several measured chemical or physical variables (predictors) and a few targets (responses). The number of predictors might range from a few (like presented in the second thesis publication) up to thousands for optical spectroscopic data (like presented in the third thesis publication). In any case, the information of those variables will be compressed to a few latent variables with most direct correlation to the targets of interest (e.g. PLS regression). Nevertheless, the contribution of each single variable to the final model structure and the targets of interest is varying. Further, regions of low single to noise ratio might be excluded. Thus, methods for feature selection play an important role in multivariate analysis (Lavine and Workmann, 2010 [52]). Several algorithms for variable selection are presented in literature. Some of them are based on moving windows, were length, position and the number of windows are tunable (see Chen *et al.*, 2014 [76]). Such approaches are usually unsupervised. Other methods are based on informative vectors also referred to as Filter Methods. These different vectors are calculated based on the created multivariate model and contain the influence of single variables on the targets. The final "window" investigated by iteration consist on most relevant predictors only. Therefore, such methods are supervised. The latter are preferable, since the least necessary information out of the dataset is used. Teófilo *et al., 2008* investigated the performance on a number of existing informative vectors for NIRS calibration [77]. The most promising vectors in their investigation were regression and net analyte signal (NAS) vector. However, they concluded that the other vectors analyzed might be more suitable for other data sets. Hence, *a priori* choice of the informative vector seems difficult.

Why is variable selection necessary?

It is known, that using all variables does not necessarily lead to best performance of the final model since uninformative variables could reduce dominance of informative ones [39]. Further, Nadler and Coifman, 2005 [78] state, that in multivariate models with large calibration sets the final prediction error is mainly influenced by noise divided by the length of the Net Analyte Signal vector ($\sigma/\mathrm{norm}(x^{nas})$). However, they also state that there

are additional error terms of the form $\sigma^2 m^2/n^2$ (n samples, m variables and σ noise level per variable). Thus, in case of m >> n, those terms can be quite large. Thus, the prediction error can be mainly dominated by those terms. In summary, the prediction error is influenced by the sensitivity ($\sigma/\mathrm{norm}(x^{nas})$), the degree of statistical correlation (the larger the correlation or interferences the worse the prediction error) and the number of variables (the more variables, the more noisy the estimates) [78].

Nevertheless, variable selection might avoid overfitting (more terms taken or more complicated than necessary) as well as improve the model performance. The major question is always, how much variables are necessary to optimally picture the given problem. This must not mean how much variables are significant. In general, optimal features for the specific algorithm do not necessarily equal to relevant features [79]. In most cases, highly correlated variables that are therefore redundant, do not add information, if used. In contrast, noise reduction might be also obtained by adding redundant variables. High correlation does also not mean that there is no complementary effect between variables. Finally, a variable with obviously no information alone might still support together with others. This count also for two uninformative variables, which might be useful together [80]. In conclusion, the topic of variable selection is of necessity but should be considered with care. For further reading about reasons please be referred to Axelson, 2012 [39].

In general, there are three possible ways in choosing a method out of several different ones used for the specific task: (1) "expert questionnaire", (2) mathematical comparison of algorithms, and (3) data driven (by iteration; therefore process or application oriented). In this work, the methods Variable Importance in the Projection (VIP), Net Analyte Signal (NAS) as well as the regression vector (Reg) where used.

(1) Expert Questionnaire by literature review

The method of acquisition of expert knowledge by expert questionnaire is used most often in the area of social science. This method does not completely belong to the close area of experimental design (DoE), but has to be taken into account in rather complex processes, for example [81]. Nevertheless, the way used in the present investigation was adapted to support the choice of algorithm usage for variable selection. Therefore, a literature investigation was performed on the number of times a single algorithm is mentioned in a period between 2000 and 2013. The detailed search settings are summarized in Table 1.3. The list contains just algorithms of supervised style.

A good overview for variable subset selection in general is given in Axelson, 2012 [39]. Further algorithms can be found in Teófilo *et al.*, 2009 [77], which are regression vector, correlation vector, residual vector, and covariance procedures vector. These could not be investigated in the questionnaire, since the naming is to general and therefore not comparable in the survey. The result of the investigation is shown in Figure 1.9.

Table 1.3: Search engine: Google Scholar, exclude patents and citations; advanced search with the exact phrase "Variable Selection" and with at least one of the words "Chemometrics", "Multivariate", "MVA" or "Calibration"

	Full name	**with all of the words**
VIP	Variable Importance in the Projection	Variable Importance Projection VIP PLS
NAS	Net analyte Signal	Net analyte Signal NAS PLS
SNR	Signal to Noise ratio	signal to noise ratio SNR PLS
UVE	Uninformative Variable Elimination	Uninformative Variable Elimination UVE PLS
ICA	Independent Component Analysis	Independent Component Analysis ICA PLS
GA	Genetic Algorithm	Genetic Algorithm GA PLS
PSO	Particle Swarm Optimization	Particle Swarm Optimization PSO PLS
ACO	Ant Colony Optimization	Ant Colony Optimization ACO PLS
NN	Neural Network	Neural Network NN PLS
SA	Simulated Annealing	Simulated Annealing SA PLS

The reviewed algorithms can be classified according their origin, which is either multivariate statistics (information criteria (single equations) such as NAS, VIP and SNR or algorithms such as UVE and ICA) or

evolutionary algorithms (optimization strategies such as PSO, ACO and SA). The latter ones have secondary importance in this investigation, since it was aimed at simple solutions without additional coding.

The algorithms based upon multivariate statistics can be again subdivided. The background of UVE and ICA are standalone algorithms whereas NAS, VIP and SNR are single equations based upon results from a prior accomplished PLS decomposition.

The algorithm for UVE is similar to the method presented in section "robust calibration". The background of ICA is the assumption of independency of components from original sources in an investigated, convoluted signal (for details in UVE and ICA refer to Axelson, 2012 [39]). Both algorithms are somewhat of more computational effort and were not considered in this thesis. Last, the reviewed single equation solutions contain the SNR, which can be calculated in many different ways. Thus, the result presented in Figure 1.9 is not unambiguous. Therefore, only NAS and VIP were taken into consideration. Additionally, the regression vector (vector of regression coefficients) was taken.

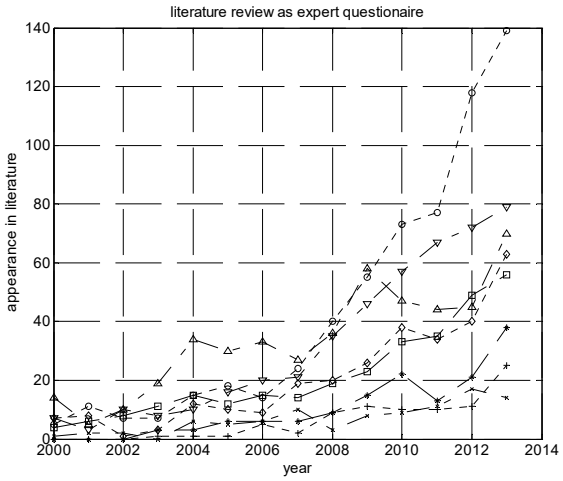


Figure 1.9: graphical illustration of the expert questionnaire; it is visible, that the algorithm VIP is having the highest impact in the field reviewed; dashed-dotted line and cross: Net Analyte Signal (NAS), dotted line and diamond: Uninformative Variable Elimination (UVE), dotted line and circle: Variable Importance in the Projection (VIP), dashed line and star: Particle Swarm Optimization (PSO), dotted line and plus: Ant Colony Optimization (ACO), dashed-dotted line and upper triangle: Simulated Annealing (SA), dashed-dotted line and lower triangle: Independent Component Analysis (ICA), dashed line and square: Signal to Noise Ratio (SNR)

(2) Mathematical Comparison

Mathematical representation of these algorithms can be compared to check similarity between the three chosen options. First it has to be mentioned, that NAS (Equation 27) is not comparable mathematically to the others, since it is calculated based on the regression vector.

Net Analyte Signal (NAS) [77]

In multivariate calibration, the NAS is representative for the part of a mixture signal, which is useful in prediction. The calibration model considers only the relevant part of the mixture spectrum, which is orthogonal to the interference spectrum and contributes to the target. To calculate this part of a signal, the equation presented by Teófilo *et al.*, 2009 can be used [77]:

$$\boldsymbol{x}_i^{nas} = \frac{\hat{y}_i}{\boldsymbol{b}^T\boldsymbol{b}} \cdot \boldsymbol{b} \tag{27}$$

where $\hat{y}$ represents the predicted target and **b** the regression vector. The NAS vector is calculated for each sample. Therefore, an average of those vectors result in an informative vector to be used for variable selection.

Variable Importance in the Projection (VIP) [2, 39, 82-84]

In VIP, Partial Least Squares reduce the input space **X** to the relevant subspace containing information for predicting **y**. To calculate the VIP parameters for each variable, the regression coefficients **β** in the variable space (also known as y-loadings **q**), the scores **T** and the weighted loadings **W** out of the PLS decomposition are used [83]:

$$VIP_j = \sqrt{\frac{M \cdot \sum_{a=1}^{A}\left(\left(w_{ja}/\|\boldsymbol{w}_a\|\right)^2 \beta_a^2 \boldsymbol{t}_a^T\boldsymbol{t}_a\right)}{\sum_{a=1}^{A} b_a^2 \boldsymbol{t}_a^T\boldsymbol{t}_a}} \tag{28}$$

where w_{ja} is the j^{th} element (belonging to j^{th} variable) of the a^{th} column of **W** (each component or latent variable vector is stored in one column), $\mathbf{t}_a$ is the a^{th} column of matrix **T**, β_a the a^{th} element of the (in case of a single target) row vector **β** (or **q**, respectively), and M the number of variables. The part $\sum_{a=1}^{A} \beta_a^2 \boldsymbol{t}_a^T \boldsymbol{t}_a$ resembles the percentage of explained **y** by the a^{th} latent variable. Interpreting Equation 28, the VIP value (which is a weighted sum of squares of the PLS weights **w**) of the j^{th} variable will be higher depending on the significance of:

- the weighted loadings w_{ja},
- the a^{th} element of the regression vector **β** (or **q**-loading),
- the scores for the a^{th} component ($\mathbf{t}_a$)

In general it is known, that variables with values higher than one should be informative, since the average of all $VIP^2 = 1$ [2, 82].

Regression Vector [2, 77, 83-85]

Calculation of the regression vector **b** (Reg) is accomplished by a relation proposed by Martens and Naes, 1991 [85]:

$$\boldsymbol{b}_S = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{q}^T \tag{29}$$

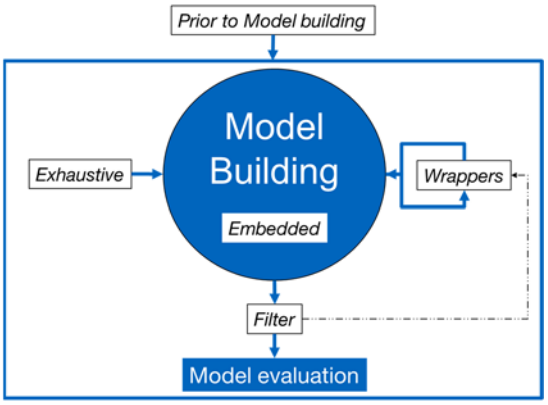


Figure 1.10: variable selection methods located around central model building/generation; Filter methods applied after a preliminary model generation and often used inside wrapper methods; the latter are most often integrated by an iterative scheme; exhaustive methods are located outside of the model generation algorithm investigation all possible variable combinations; embedded methods are integrated inside the model generation algorithm.

where the weighted loading matrix **W**, the x-loadings **P** and the y-loadings **q** are used. The subscript S counts for the scaled version of Reg, since in most of the PLS models original variables have to be preprocessed by mean centering or z-transform to reduce the influence of magnitude and make variables comparable to each other. Hence, the informative variables can be selected directly according to the magnitude of the absolute values of Reg [2, 83].

Another possibility in choosing variables by regression coefficients is given by a method called PLS-Bootstrap [86]. Here, bootstrap sampling is used to estimate a confidence interval for each regression coefficient. If any confidence interval includes the value zero, the corresponding variable can be sorted out [86]).

Nevertheless, calculation of NAS is based on regression coefficients. Further, information is included in regression coefficients as well as in VIP. It has to be mentioned additionally, that both filters described showed best performance in the investigations (see results section).

A schematic categorization of algorithms used for variable selection and there position in the model building strategy is given in Figure 1.10, an extract of algorithms belonging to these different groups is given by Figure 1.11.

Prior to model building

Meinshausen, 2008 [87] stated, that the efficiency of typical variable selection methods with connection to the model algorithm is low, if there is a high redundancy in the data. This fact is increasing if the multiplicity (many predictors lead to a reduction of the power to detect important variables) of the problem is taken into account [87]. This means, that typical variable selection techniques seems to be helpful for deleting non correlated variables with respect to the target but highly correlated variables get the same importance, which is not always beneficial. Nadler and Coifman, 2005 [78] reported that the noise level and the original dimension of the variable space lead to the same conclusion: Variable selection has to be performed before model building. Meinshausen, 2008

proposes Hierarchical Testing for variable selection. This method is based on statistically testing if clusters of variables have significant regression coefficients greater than zero.

Since the options are manifold, the whole topic is just mentioned for the sake of completeness but will be not discussed in detail.

Filter Methods

Those methods select according to a specific property of the data. They can be applied independent of the training algorithm as well as on the amount of variable inputs. Therefore, the model algorithm is performed on a number of variables and a certain chosen criteria is calculated to rank those variables accordingly. Thus, those methods provide a global selection, which is not influenced by the used algorithm. A filter method can be, amongst others, statistical test such as t- or F-test as well as Pearson's correlation coefficient. Further well-known methods are signal-to-noise, covariance procedures, residual vector or loading weights [2, 77, 88].

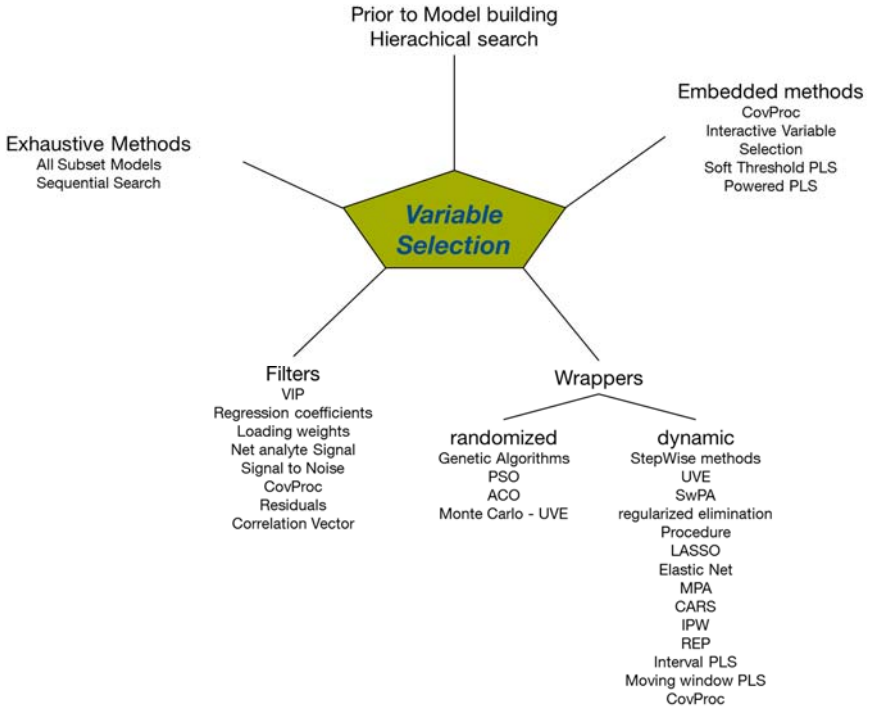


Figure 1.11: overview of some variable selection methods described in literature (adapted from Mehmood et al., 2012 [2]); they are all classified under their respective header

Wrapper Methods

In those kind of algorithms, the feature search is embedded in the model-building step, based on a search algorithm. Thus, the feature selection is performed together with the algorithm for model building and subsets are also tested on their predictive power by for instance cross-validation. Finally, the variable combinations with best performance are taken for retraining of the model structure. It should be noted, that often Wrappers are based on Filter Methods. Amongst others, Support Vector Machines and their variants or decision trees belong to the category of Wrapper Methods [39]. Further, StepWise methods (SW) [2, 39, 82, 83], which are either Forward Selection (FS) or Backward Elimination (BE), evolutionary algorithms such as Genetic Algorithms (GA) [2, 82, 84], Particle Swarm Optimization (PSO) [82, 84], Ant Colony Optimization (ACO) [82], Least Absolute Shrinkage and Selection Operator (LASSO) [53, 82, 83, 89], Elastic Net [53, 82], Uninformative variable elimination (UVE) [2, 39, 84], Monte-Carlo UVE (MC-UVE) [90], Sub-window permutation analysis (SwPA), Model Population Analysis (MPA) [91, 92], Competitive Adaptive Reweighing Sampling (CARS) [39, 91], Iterative predictor weighting PLS (IPW-PLS) [2], Regularized elimination procedure (REP) [2], Interval PLS (iPLS) [2, 39] and Moving window PLS (MW-PLS) [39] are mentioned under this topic.

Embedded Methods

Embedded variable selection procedures are coded to perform variable selection while model building. They are specific for a given algorithm [39]. In case of PLS for instance, each step for component extraction includes a variable selection procedure within the iterative algorithm. This is in contrast to the wrapper methods, which follow a double iterative procedure [2] with retraining after each variable subset. Methods in this category are amongst others Neural Networks or modified versions of standard PLS [39]. Also, Interactive variable selection

(IVS) [2, 39] for PLS, Soft-Threshold PLS (ST-PLS) and Sparse-PLS as well as Powered PLS (PPLS) [2] belong to this category.

Exhaustive Methods

Those methods test all possible combinations of variables [39]. The obvious disadvantage is the numerous iterations necessary, when the number of variables increase (see for instance Cassotti and Grisoni [82]). They mention All Subset Models (ASM) as well as Sequential Search (SS) under this header [82]. An overview of the listed algorithms is given in Figure 1.11.

Advantages/Disadvantages

In general it can be said, that wrappers based on data-driven iteration are preferable, since no parameter fitting has to be done (compared to randomized versions), require less computational effort (compared to exhaustive methods), and are model algorithm independent (with respect to the method chosen). The risk of overfitting can be partly avoided by an external dataset for proper final model validation. This is recommended in all mentioned applications (see for instance Mehmood *et al.*, 2012 [2] or Axelson *et al.*, 2012 [39]). They are usually slower than embedded since two iterations are needed, but more flexible in choice of filter with respect to the application. They are more guided to the problem if done with care. The only relevant disadvantage is the higher risk of local instead of global optima [2]. This can be partly solved by iterations as for robust calibration. Nevertheless, no algorithm can be taken as perfect or optimal. Even though it was tried many times to compare between methods, only a few neutral comparisons are given [2]. It is recommended by Mehmood et al., 2012, to test several methods on the given problem of interest [2]. Nadler and Coifman, 2005 [78] state under a different background, that even if algorithms have similar performance, they might be superior in special cases over others.

Table 1.4: Advantages and disadvantages of different strategies for variable selection; adapted from Saeys et al., 2007 [93]

	Advantages	**Disadvantages**
Prior to model building	- Completely model algorithm independent	- Might ignore feature dependencies - ignores interaction with the model algorithm - extra computational effort
Filter	- Fast and scalable - Independent from model algorithm - Better computational efficiency than Wrapper	- Ignores feature dependencies (if univariate techniques) - Ignores interaction with the algorithm - Threshold definition
Wrapper	- Interacts with model algorithm - Models feature dependencies - trying to minimize the number of errors directly - better prediction accuracy - more effective than Filter Methods	- slow - Risk of overfitting - features are less robust - local optima
Embedded	- Interacts with model algorithm - Better computational efficiency than Wrapper - Models feature dependencies	- Algorithm dependant selection

The result from the short literature review named “expert questionnaire” in the beginning of the chapter is therefore not surprising. Filters like VIP are often used since both wrapper and filter methods can be based on it – but it also shows the power with respect to simplicity and easy interpretability.

These mentioned points together with the simplicity support the choice of method within this work, which is based on a stepwise method with the filters VIP, regression coefficient or NAS or any of their combination. The reason

for the choice of several filters in comparison is the recommendation by for instance Chong and Jun, 2005 [83], which conclude in their study to use a complementary combination of VIP and regression coefficients. Another reason is given by the fact, that stepwise methods usually based on one filter are inefficient if multicollinearity (explainable by several variables doing the same job) is present in the data [83].

This also supports the choice of a wrapper method for flexible use of filters. Last it can be said, that for instance Chong and Jun, 2005 [83] concluded best performance for VIP and complementary power of regression coefficients, Mehmood *et al.*, 2011 [94] concluded best performance of VIP as filter and Teófilo *et al.*, 2009 [77] concluded best performance of regression coefficients and NAS in their investigation. The later also concluded, that this output counts for the data investigated, other filters might be more effective in other cases.

Finally, it has to be mentioned, that one of the most critical points amongst the choice of search algorithm is the performance prediction, which serves as final criterion in the choice of the best subset, discussed in section 1.2.2.3.

(3) Data driven or application oriented

The functionality of those methods (in this work limited to the three presented algorithms) can be tested on the data, whereas the final model quality is used as criterion for the choice. In the presented publications, most efficient algorithm was VIP (third thesis publication) as well as the combination with Reg (second thesis publication). The output of NIR investigations slightly shows that VIP seems to be enough for selecting different wavebands under the assumption, that those have low multicollinearity. The feature investigation in US signals show that a better performance is achieved combining two measures, since the features extracted are known to be highly multi-collinear. The aspect of performance under multicollinearity present is discussed by Chong *et al.,* 2005 [83].

1.2.3.2 Outlier analysis

Literature presents a variety of algorithms for outlier detection. Outlier analysis is another broad scientific field, which will not be discussed in detail here. Nevertheless, in multivariate statistics it is distinguished between x-sample outlier (on a complete sample profile), x-variable outlier (if just one variable behaves different in comparison to the others) and y-sample outlier (based on the samples characteristic response [regression only]) [3]. One of the most often applied outlier detection is based on Hotelling T^2 statistics compared to Q residuals. The first came up as an alternative to the standard method leverage over the past 15 years [95]. The relation between Hotelling T^2 and the leverage of a sample is given by [95]:

$$Hotelling's\ T^2 = (n-1)(h_i - 1/n) \tag{30}$$

The Q residuals are calculated on the residuals from the X matrix to investigate the sample outliers [96]. This can also be done by comparing the residual variances [97]. For calibration models it is also of interest to check y-outliers, which can be done by simply using the residuals received from predicted and original target value [6]. Nevertheless, predicting outliers and removing them is always a crucial topic.

Cao *et al.* [98] give a short review on outlier detection in statistics. They report two different approaches, namely diagnostics and robust estimators. The first starts with identification of outliers and subsequent regression of the remaining data by a variety of regression methods. One of the most common approaches is based on the mean standard deviation, which serves as basis for a statistic with a certain threshold. Such methods fail, if there are multiple outliers, since they effect the result to such an extent, that outliers are masked (masking effect, different to masking presented in the third thesis publication in discriminant analysis) [98]. Cao *et al.* also report several algorithms handling masking, but most of them are limited in just sample outliers [98]. In regression analysis, the targets are also of interest.

In robust estimators, the aim is to find a model fitting the majority of data and examine the outliers based on this solution. They are able to detect target outliers but mostly have lower effectiveness predicting sample outliers [98]. Further, their susceptibility to outliers is reported to be lower, but outliers found are varying between models

of different background [98]. Based upon those issues and the coexistence of outliers in both samples and targets, Cao *et al.* use Monte-Carlo cross-validation [98]. A similar approach is presented in different publications of Li *et al.* [70, 91, 92] named model population analysis. The Monte Carlo simulations are used for statistical analysis of sub-model outputs (regression coefficients or prediction errors), whereas prediction errors showed good results for effective outlier detection [70]. This method is also used for variable selection.

Another comparatively simple approach was presented by Mitzscherling, 2004, who used statistical variation of loadings during LOOCV analysis of data to discuss and interpret the impact of single variables (variable selection) and the source of variation (possible outliers) [44].

The statistical distribution of, for instance, residual errors indicate the quality of the sample data set. Cao *et al.* [98] use the background of Monte Carlo (MC) simulations on data set to create multiple models (as explained under MPA for variable selection) and investigate the residual errors of targets on their mean and standard deviation. They define outliers of three different categories – x-outliers, y or model outliers and a mixture of both [98]. Those different types are either visible in a histogram showing deviations from normal distribution (whereas y and model outliers show a single own distribution in the histogram, x-outliers influence the central peak by bigger mean and standard deviation) or in a plot of mean against standard deviation, similar to the leverage to residual variance plot mentioned above.

However, also variable outliers can affect the model output. Li *et al.* showed, that artificial outliers in variables results in skewed distribution of errors or slopes of the models [70]. Similarly, the errors in the second thesis publication where analyzed on single levels. It was visible, that in lower concentration levels the skewed distribution appeared. In contrast to Li *et al*., it cannot directly be linked to variables, since other levels showed only brought "normal" distributions possibly indicating sample outliers. Thus, such investigations should be further considered in the future.

A different way of outlier consideration was followed in detecting valid US signals. These signals represent a hyper-dimensional cluster in a feature space. Distorted signals would be geometrically outside of this cluster. Therefore, outlier detection using Angle Based Outlier Factor (ABOF) can be applied to sort out signals prior to further signal processing, saving time and space in an online application later. In multivariate cluster analysis, ABOF (Equation 31) detecting cluster outliers is presented by Kriegel *et al.* [99].

$$ABOF(S_P) = \mathrm{Var}_{B,C \in M/S_P}\left(\frac{\langle \overline{BS}_P, \overline{CS}_P}{\|\overline{BS}_P\|^2 \cdot \|\overline{CS}_P\|^2}\right) \tag{31}$$

This algorithm calculates the geometrical angles between a set of three data points and compares its variance. If a data point is inside of the cluster, the angles to other points vary quite a lot. A point outside the cluster will have almost the same angle to all of the points. In a preliminary investigation, some simple feature (integral, momentum and centroid, Equation 32-34) where used on signals from yeast propagation. It is worth mentioning, that choice of features was guided by minimal computational effort for online application.

$$F_{I,p} = \sum_{n=1}^{N} |A_n| \tag{32}$$

$$F_{C,p} = \sum_{n=1}^{N} (A_n^2 \cdot t_n) \Big/ \sum_{n=1}^{N} A_n^2 \tag{33}$$

$$F_{M,p} = \sum_{n=1}^{N} (|A_n| \cdot t_n) \tag{34}$$

Results on outlier detection procedures are shown in the publications included in this thesis (see results section) with the conclusion, that those might not always work satisfactorily and thus need intensive consideration in each special case.

1.2.3.3 Model Robustness

Another very important aspect of any kind of model generation is the robustness of the solution. This includes the ability of the corresponding algorithms used for model generation to cope with both, normal and experimental variability, to not result in unusual large deviating predictions [39]. Recommendation for such algorithms is also quite diverse. Amongst others, robustness and model stability as well as realistic prediction performance is influenced by the choice of validation sample. Thus, a realistic representative training set has to be found. Influences of sample subsets in model training as well as validation are discussed in Axelson, 2012 in chapter four [39]. This includes the influences reasoned by the presence of outliers such as masking (variations are not detected due to model distortion) or swamping (good data is identified as outliers) [39].

This fact is also supported by the points of Li *et al.* in 2012. They state, that a single quantity (e.g. RMSEV) for model assessment is highly dependent on the choice of the corresponding training and test set and thus lack in statistical substantiation.

Further, the number of samples for validation is of importance. Faber in 1999 concluded, that therefore the uncertainty of the errors have to be calculated additionally [100]. The proposed iterative approach on randomized sample sets results in a mean error and its standard deviation. The lower the error, the higher the accuracy and the lower the standard deviation, the higher the stability. Another, very similar approach is followed by model population analysis (MPA), presented by Li *et al.*, 2012 (see also sections variable selection and outlier detection in this work) [70].

A combination for model selection using model prediction performance and variable selection is shown by division into different data sets. The data presented in the appendix, Figure A.1 are light spectra of plant leaves collected in a grain study field to detect the nitrogen content. Typically, such investigations are analyzed using the REIP index, which showed low accuracy. Thus, multivariate data analysis was accomplished instead. The presented example (see appendix) aims at more certain predictive error calculation using a similar approach as presented by Chen *et al.*, 2014 [76]. A similar variant was accomplished for the special case of US feature investigation. Therefore, a data set of ~6600 samples from the setup presented in the first and second thesis publication on binary and ternary mixtures of sugar, ethanol and water over a certain temperature range was divided into 50 subsets. The calibration model was calculated on one data set extracted out of the data pool before iteration, the validation error (RMSEV) was calculated on each of the 50 subsets. Results of this investigation are shown below in the chapter “discussion”.

Other possibilities for robust calibration by weighting single samples with respect to outliers are proposed in literature (Martens and Naes, 1991 or Liebmann *et al.*, 2010 [85, 101]). However, the robustness of calibration models is dependent on the data (noise, outliers) or the application (laboratory, process). Therefore, it will always need special consideration.

1.3 Quality inspection of raw materials

Another aspect in food and beverage industrial production is the quality of raw material. This is varying from harvest to harvest, influenced by processing as well as storage. Again, literature presents a variety of different techniques for quality analysis and evaluation. Established methods such as gas chromatography (GC) for evaluating the sensorial appearance of products as quality indication are also possible (reviewed by Plutowska *et al.*, 2008 [102]). Even though, such methods are very necessary and are still not completely substitutable in quantitative analysis, they are comparably slow and need often extensive sample preparation with respect to quality evaluation of raw materials. Online application is possible but comparably sophisticated. Brosnan and Sun, 2004 give a review about the application of computer vision by image analysis such as X-ray, 3D and color vision on food products for quality analysis [103]. They report usage of those techniques for grain evaluation as well. The benefits of non-destructiveness and flexibility together with the rising speed of computational possibilities increase the online potential of computer vision for automated process analysis in food industry. One

important sensorial aspect and grading factor related to food quality and human appreciation is the color associating freshness or food safety [104]. This attribute could be estimated by spectrophotometers as well as by computer vision using image analysis. Wu and Chen, 2013 review the application of the latter on color but point out the potential for other attributes such as shape, size or defects [104]. They also report a high spatial resolution by image analysis. Even though, the informative content with respect to more detailed information in e.g. chemical composition might not reach the same level, it is an interesting field with respect to online industrial application. However, compared to hyperspectral imaging or comparable methods, the efficiency regarding costs seem much higher.

In general, methods for estimating quality or chemical composition of any raw material or product in food stuff should be fast, non-invasive and non-destructive, which is in contrast to UV and Vis spectroscopy provided by NIRS [12]. Even though those spectra contain a high amount of superimposed bands, the versatile information content is already widely used for chemical as well as physical qualification and quantification [105]. With respect to specific food stuff application, Ratcliffe and Pannozzo, 1999 applied NIR spectroscopy on wort samples to distinguish between different malt qualities based on multi-linear regression using four wavelengths [106]. Aim of the study by Ratcliffe and Pannozzo was to predict three major quality attributes, namely hot water extract, free alpha amino nitrogen and soluble protein [106]. However, their approach needs processing of raw material reaching liquid wort to be analysed. Even though this approach is applicable, the results achieved by using rapid laboratory processing might be too slow for industrial application. Another possibility is shown by Giovenzana *et al.,* 2014 presenting a rapid possibility to evaluate the beer quality on soluble solid content (extract) as well as pH- value during fermentation by analysing samples offline using vis/NIR spectral analysis [107]. Regardless of the presented accuracies, the approach shows chances for online monitoring using presented methods.

Sileoni *et al.*, 2010 report several online applications for meat, cheese, fruits and grain (e. g. whole malt kernel), but majorly on predicting single traits or quality attributes [12]. One reference is given to a quality index used as target for PLSR but based on a laboratory malt profile [12, 105].

Sileoni *et al.* further shows the potential of NIR spectroscopy applied for fast quality detection of whole malt grain and maize grits. Their application was investigated related to brewing relevant material features, namely moisture and total nitrogen content [12]. These results regarding full grain inspection support the finding from the third thesis publication. In another publication, Sileoni, 2011 reports several applications for quantitative use of NIR spectroscopy [105]. Even though, this method is not very sensitive for such aims [105], the benefits (reported amongst others in section 1.1.1 and 1.3) still make it a favourable option. Under restrictions, the authors are also stating the possibility of NIR used for qualitative analysis. One work reviewed by Sileoni proves NIR usage for physico-chemical fingerprinting and thus applicability for classification by applying MSC to partially separate physical and chemical spectral components [105]. Amongst predicting individual parameters of malt or beer for quality evaluation, Sileoni also showed the possibility to combine those parameters to a quality index and its prediction by applying PLSR on NIR spectra [105]. A similar approach was shown by Gianinetti *et al.*, 2005, who used different combinations of general malt parameters, PCA and discriminant analysis to group different classes of quality [108]. Here, focus lies on the differentiation of breeding programs for malting. The parameters were all received by laboratory analysis (e.g. hot wort extract [HWE] or wort viscosity). Munck and Möller, 2005 showed the possibility to evaluate malting quality by classification on vigour and vitality [109]. Fluorescence and image analysis were used for detecting germination, NIR spectroscopy was used to calibrate to vigour and vitality. Other methods in the field of quality inspection using multivariate information on barley or other grains are reported in the third thesis publication.

Lachenmeier, 2007 proposed a rapid quality control for spirit drinks and beer by applying NIR calibrated to a variety of spirit drink parameters (such as density or ethanol concentration) resulting in different accuracies [110]. Nevertheless, the majority of applications reported are oriented to one specific parameter of interest or the relation

of quality to single attribute, such as content of a specific chemical ingredient [12, 105]. This is mentioned in the third thesis publication to not always lead to exact predictions of the central target, material quality.

Further, Huang *et al.* in 2008 gave a brought review on the usage of online and inline applications of NIR spectral measurements on food stuff such like dairy products, beverages or meat [43]. Another review on NIR application with respect to food quality is given by Cen and He, 2007 [111], including more detailed reports on most popular algorithms for pre-processing, model building as well as the measurement devices. Majorly, the detection of interest in the reported grain or grain products were established quality parameters such as protein content, moisture or hardness, whereas the review of Cen and Hu contains, besides online and inline, a various number of general applications in a variety of food areas [43, 111]. But nevertheless, there are major constraints of NIR applied in food analysis such as cost, reliability of calibration (thus necessity of robust algorithms and non-conventional methods), and non-sensitivity to mineral content (thus combination with fluorescence spectroscopy or electronic nose) [43, 111] (some challenges applying NIR are also reported by Grassi *et al.* [68]). Georgieva *et al.*, 2014 applied NIR and multivariate data analysis to predict, identify, and qualify berry fruits. In contrast, they investigated fruit extracts, which implements at least one processing step to the raw material. Nevertheless, the aim of their study was focused on quality attributes directly, such as species or preparation procedure using classification and storage time applying regression. Even though, the multivariate methods used are quite simple and not always fully appropriate (PLSR on zero [no storage] and one [storage]), they could show the possibilities of the mentioned method [112].

Further, Nicolai *et al.*, 2007 give an extensive review on NIR used for fruit and vegetable quality. They state, that more research in modelling with incorporated knowledge based on real coherences, meaning rather than just simple and empirical more explorative or even inductive statistics are necessary [113]. Further, techniques including spatial or temporal resolved spectroscopy to separate absorption and scattering effects are quite interesting. These might lead to innovative prediction models on texture associated properties [113]. Such more advanced techniques include the aforementioned hyperspectral imaging (HSI). Gowen *et al.*, 2007 reviewed the potential in using these methods for food quality and safety analysis and give examples on hyperspectral reflectance, fluorescence and transmission imaging [114]. However, such methods are much higher in equipment costs than the “simple” NIR- or 2D-fluorescence spectroscopy and require much more sophisticated multivariate analysis together with image processing. Nevertheless, two separate features are used in HSI. The spatial feature can be used for investigating complex heterogeneous samples, the spectral feature again is used for characterizing different (chemical) components simultaneously on and underneath the surface of a sample [114]. Regardless of the cost issues and under the aspect of continuing emphasis on PAT, it is still most likely, that interest of such techniques for food analysis will rise [114].

Another possibility of quality estimation is given by combination of numerical and linguistic data using fuzzy logic. Perrot *et al.*, 2006 give a review on the application of fuzzy logic for control of food quality [115]. They address one of the challenges in the food area, namely non-existing sensors as well as monetary unattractive variants. They propose using knowledge combined with sensor data for indirect measurement or quality control instead. They also hint, that instead using classical fuzzy membership functions (e. g. Mamdani type), it is more efficient to estimate these functions by the theory of possibility [115].

Quality inspection and property prediction in cereals is a broad scientific and industrial field. One quite successful method over the past decades is artificial neural networks (ANN, see section 1.2.2.1 “Pattern Recognition”). Goyal reports, that the application of such networks for different tasks gained acceptance in the field of cereals over the last years [62]. Nevertheless, the setup of such networks are quite sensitive and need to be done with care [62]. Amongst others, Goyal reports the application of ANN for cereal grain quality, whole barley kernel identification and identification of physical parameters of grain quality with respect to malting barley. Further, he reports several novel techniques such as image analysis, classification, prediction and system modelling applied for grain analysis pointing out, that they are mostly in the development stage [62]. Nevertheless, Zapotoczny

reports that "The development of non-destructive methods for the evaluation of cereal grain varieties has significant implications for the food processing industry" (Zapotoczny, 2011; found in Goyal, 2013 [62]), which can be taken as fact for sensing techniques in other areas as well.

Although this chapter is reporting just selected publications from literature and does not give a comprehensive overview on the topic quality analysis in general, following points are most obvious:

- the application of new, multivariate measurements such as NIRS prove to be applicable
- multivariate data analysis with new computational possibilities and adapted to a specific challenge of interest are highly recommended
- limitation to existing knowledge in the standards of quality of a specific product investigated is not always leading to the target

Up to the knowledge of the author, none of the reported methods with respect to overall quality inspection in malt kernels was intensively applied in industrial scale, which lead to research and the third thesis publication.

1.4 Sensor network inspection

In general, bioprocess sensor networks are not monitored automatically with respect to failures. Therefore, models used assisting process monitoring based on some process theory as well as based on a fixed number of valuable sensors are either not grasping any changes induced by faults or fail themselves being exposed to any equipment failure. Thus, fault detection is a major task for stable and reliable progress of bioprocesses and its monitoring. Kourti defines fault detection with the "detection of sensor failure, equipment failure, presence of unusual disturbances, and any other situation that does not correspond to a good routine operation." and reports possibilities based on multivariate statistics including separated process state monitoring and handling missing data [64].

In contrast to many solutions presented in literature, the monitoring solution in this thesis is based not only on single sensor information but also on a combined treatment, additionally supported by swarm behavior. The evolutionary management of controlled swarms can result in robustness to sensor failures, its resilience in this regard is an emerging knowledge. Logically, sensors with redundant information should be of less importance in the state estimation. The identification of the importance of a sensor is not often an easy job; this is especially true in a complex swarm scenario. The importance together with the reliability can be measured by the magnitude of the sensor importance value. An example is shown by combining historic process data as cost function for statistical process model choices on different number of sensor inputs, shown in the fourth thesis publication (a simplified, schematic representation is shown in Figure 1.12). Controlled swarming can therefore be used to identify and explore effects of sensor failure. Swarms, as expected from nature, are able to adapt and survive with only a moderate loss in performance, however the prediction is limited. In some cases due sensor redundancy, sensor loss can result in a significantly modified swarm emergent behaviour.

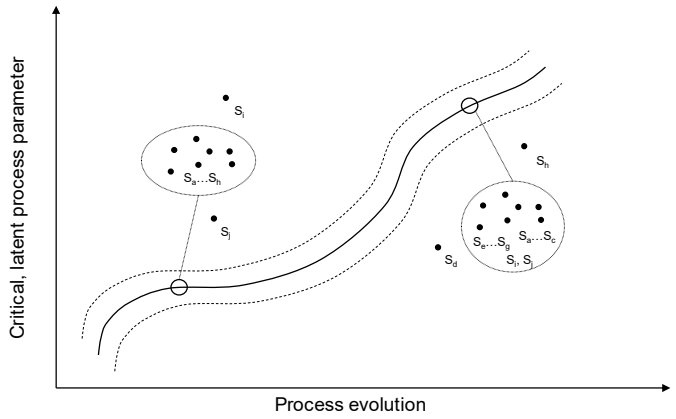


Figure 1.12: schematic representation of sensor swarm predicting latent process trajectory; the respective cloud of members succeed, sensors S_i and S_j (first case) as well as S_d and S_h (second case) are faulty and thus not taken into account

1.5 Thesis outline

Being capable of solving the aforementioned issues will lead to sustainable and resource efficient production. In the following, those different aspects are elaborated with the aim of exemplarily giving possibilities of solving such challenges with general, data driven approaches. This includes the possibility of data driven sensor calibration on innovative ultrasonic sensor system for fluid property inspection, the usage of near infrared spectroscopy for raw material quality inspection with the aid of multivariate process control strategies and finally, process sensor network inspection under swarm intelligent decision making.

The thesis gives a small abstract of the possibilities in data analysis applied to different challenges and with the respective adaptions. The core hypothesis of this work is:

Is there any general pathway though pre-processing, model generation and post-processing in data-driven bioprocess applications using multivariate data analysis?

This hypothesis implies the following key aspects:

- algorithms for:
    - variable pre-processing
    - sample pre-processing
    - model generation
    - variable selection
    - outlier analysis
- non-linearity
- discriminant analysis
- regression
- model robustness
- transfer of models
- intelligent sensor network evaluation
- process monitoring

The examples for application of PAT in this thesis follow the key aspect of process analysis to not only detect but also eliminate or minimize faults either in sensors or process [7]. Therefore, system integration by standardization (arbitrary data pool for MSPC and Swarm), knowledge creation with causality to the product (NIR fingerprint for processability) as well as the development of new instruments for in-line monitoring (US and combined with fuzzy control) is applied. To address these issues, the thesis is divided into three parts – calibration of ultrasonic sensor setups, quality inspection of raw material with respect to processing behavior, and sensor network evaluation together with full process monitoring.

Optimal product quality of bioprocesses can be ensured by intelligent control systems with integrated monitoring of key parameters. Quality of brewers' yeast is important to increase the efficiency of subsequent brewing processes.

The state-of-the art drawbacks are:

- lacks in online detection of yeast attributes or process guiding parameters
- un-flexible temporal control schemes typically used for industrial processes
- variations of raw products, such as brewing wort based on malt kernels
- lacks of flexibility in advanced modern monitoring and control models with multivariate background and sensor networks

Under the header of calibration of ultrasonic sensor setups the basic MVA algorithms for regression analysis PCA and PLS are used and discussed. The first publication is seen as an introductory guidance for pathways in the PAT topics, hardware and physics of Ultrasonic sensors in bioprocess fluids as well as full process considerations, sensor calibration, knowledge inclusion and its various aspects (handled individually, investigated by Krause *et al.*), focusing on the example of biomass growth under industrial, brewing relevant aspects. Birle et al. and Hoche et al. further investigate the first two parts, respectively. The latter point is addressed in this thesis. The outlook of this publication gives directions for future investigations on US signals by analyzing subdivisions of these. The part of sensor calibration in this publication serves as foundation for the second thesis publication as well as further results included in this thesis on ternary mixtures and temperature behavior.

In a preliminary publication (not included in this thesis) a spectral analysis on temperature spectra of speed of sound was applied, showing the possibility to use MVA in general to break down high dimensional data into simpler interpretable variants [116]. The outlook of this publication gives the path for following investigations by raising the issue of pattern recognition used for extracting new signal properties into a different MVA system combined with process knowledge.

The second thesis publication deals with the analysis of buffer reflections, US features and variable selection, presents a more realistic model than in the first thesis publication with respect to temperature inclusion and opens the aspects non-linearity, robustness and stability. It further shows preliminary investigations related to model population analysis reported by Li *et al.*, 2012 [70]. The novelty of this topic is the analysis of maltose-ethanol-water samples with varying temperatures using ultrasonic transducer mounted on a steel buffer, the combination of multivariate data analysis used on ultrasonic signal features from buffer reflections, variable selection to evaluate the strength of used features and, together with the additional results presented in the chapter “discussion”, external parameter orthogonalisation as well as considerations of robustness and non-linearities.

The second topic Quality inspection of raw material with respect to processing behavior is handled in the third thesis publication using discriminant analysis on NIR spectra for classification into groups of different processing quality as well as MSPC for process evaluation. Amongst others, it deals with the aspects, variables and sample pre-processing, masking, outlier detection, robustness and the transfer of models. The innovative aspects in this part of the thesis are NIR-spectra directly correlated to process quality, the direct application on industrial data and MSPC for support of expert classification.

The third topic sensor network evaluation together with full process monitoring is handled in the fourth thesis publication dealing with parameter settings of a particle swarm for sensor and model evaluation resulting in a combined multiple sensor investigation. This is capable of spotting sensor failures as well as holistic process system consideration by using MSPC as powerful control tool. In addition, the necessity and strength of variable importance detection and variable selection is underlined again. The improvement here amongst other existing solutions is, that the method is not restricted to any fixed number of multisensory inputs as well as the use of a specific sensor reading, it is capable of stable monitoring in case of sensor failures and thus combining process control, monitoring and sensor network inspection.

2. Summary of results (thesis publications)

2.1 Paper summary

Part 1: Bioprocess monitoring – brewing yeast propagation as example

Optimal product quality of bioprocesses can be ensured by intelligent control systems with integrated monitoring of key parameters. Quality of brewers' yeast is important to increase the efficiency of subsequent brewing processes. One solution is to first detect essential process parameters, second combine those with expert knowledge and third with linguistic control mechanisms. These can be fulfilled by fuzzy logic including process dynamics associated by accurate of sensing devices. Incipient stages for multivariate calibration of an ultrasound based device including temperature dependencies using temporal and spectral properties of ultrasonic waves are presented. Additionally, preliminary results of a mechanistic model for the temperature dependency of yeast growth adapted from literature is shown. The publication is thought as a perspective in combining new (ultrasonic) measurement devices for qualification of fermentation progress together with fuzzy logic control schemes enhanced by trend estimation using mathematical growth modelling. The results on mathematical growth modelling followed by further studies (McHardy, 2013 [117]) together with studies on fuzzy logic control and dynamics (Birle *et al.,* 2015 [118]) are used for flexible control of temperature and aeration resulting in vital yeast and enhanced transparency of propagation progress according to the demands. Preliminary results on these two parts are also included in this publication but left out in this thesis.

The contribution further shows the processing of ultrasonic signals estimating the ''apparent extract''. This value reflects the major components in solution (mainly carbohydrates such as sugar and ethanol). The calibration procedure on offline signals in frequency domain was absolved using PLS regression reaching a maximum prediction error of ~0.5 g/100 g by leave-one-line-out cross-validation (LOOCV). It is shown that using only ultrasonic characterization of such mixtures is enough to get insight into fluid properties without knowledge about progress of fermentation.

One of the most critical points considering online process application is the influence of temperature. This offline study showed that calibration of ultrasonic properties to apparent extract values at single temperatures is possible in a range of six to 22 °C. However, this analysis can only be regarded as perspective, since several influences are still not considered. Nevertheless, evaluating the connections between those independent temperature models, the presented results show a promising path for implementing the ultrasonic sensor online including accurate correlation to the process-related parameters to reach better process performance and understanding as well as a high quality end product.

Part 2: Data-driven calibration of new sensing devices

Following the aim of data-driven model building via multivariate regression, time and frequency domain of ultrasonic signals are analysed in order to predict maltose concentration in aqueous solutions. It is shown, that the prediction of concentrations at different temperatures is possible by using several multivariate regression models for individual temperature points. Combining these models by a linear approximation of each coefficient over temperature results in a unified solution, which takes temperature effects into account (temperature between 10 and 21 °C, fitting to brewing processes). The proposed methods have a low processing time required for analysing online signals and are based on non-invasive sensor setup, applicable in pipelines. In addition, ultrasonic signal sections used in the presented investigation were extracted out of buffer reflections, which remain primarily unaffected by bubble and particle interferences. Model calibration was performed in order to investigate the feasibility of online monitoring in fermentation processes. Processing of ultrasonic signals, model evaluation using features from time and frequency domain of ultrasonic pulse as well as input variable selection are discussed. The basic approach used for creating the final prediction solution was partial least squares (PLS) regression validated by cross validation. Feature selection was applied showing its power in choosing the required input features by their sensitivity towards the target of interest. The overall minimum prediction error was 0.64 g/100 g.

The applied approach highlights the strength of the methods used to detect less sensitive inputs in correlation to respective targets.

Part 3: Quality estimation of raw material by process discrimination

To handle the mentioned aspect of raw material variations such as malt kernels as basis for medium in brewing processes, a new strategy for quality analysis of brewing malt using near infrared (NIR) spectra taken from kernels in reflection as fingerprint to classify directly to processability of malt was established. Two main tasks are handled, namely discriminant analysis of NIR spectra to quality classes of malt kernels with the aid of partial least squares and automatic process evaluation classifying the different processes in the mentioned three categories. The accuracy achieved in the first task using pilot plant data in relation to the expert classification “good”, “normal” and “bad” was 90.6 and 92.7 % in validation and calibration, respectively. The second task was attempted by two numerical possibilities, one calculating the residual standard deviation of a process based on multivariate statistical process control (MSPC), and the second discretizing each process individually based on its single online trends. Both are finally compared to the expert opinion reaching a match of 85 % between single trend analyses using K-means clustering as well as 84 % between a RSD values from MSPC analysis and expert qualification, respectively. Furthermore, the results of the aforementioned calibration model were transferred to industrial scale, established via adjustment to corresponding system conditions reaching 93.6 and 76.6 % in calibration and validation, respectively. The authors investigated different data processing algorithms. The best possible algorithm combination was reached using either standard normal variate (SNV) or multiplicative scatter correction (MSC) combined with first derivative as spectral and variable stability scaling (VAST) as variable pre-processing. Further, studies on variable selection and outlier detection showed first positive results. Finally, possibilities of automatic qualification on lautering processes lead to reduction in expert efforts on qualifying processes.

Part 4: Sensor network evaluation

Typically, comprehensive process models are or at least should be based on online data provided by sensors or sensor networks. In MSPC models, even the full matrix of process data can be involved. In such cases, those models might fail, if one or more sensors are giving false information or are damaged. Thus, a methodology combining process knowledge with computational efforts aiming at a flexible sensor network for coping with sensor failures is presented. Therefore, multivariate linear and non-linear combinations of inputs (sensors) are utilized creating a search space based on the multisensory data pool. The raw data retrieved from several sensors is used for extracting multivariate statistical process control trajectories. Those different models are further scored by swarm intelligence (particle swarm optimization) leading to the optimal sensor/model combination at certain time step. The core of the presented approach is to determine the fermentation trajectory in combination with sensor output validation online by using a swarm intelligence based system. In addition, the network should be able to replace or ignore false sensor information caused by drift, wrong calibration, or a total sensor failure. Thus, the authors address adjustments of the basic algorithms, cost function, accuracy of output as well as the dynamic behaviour. The presented results on online data indicate the possibility of more robust online monitoring using the swarm sensing idea for biotechnological processes to insure optimal and timely effective processing. It is shown, that a discrete swarm with suitable parameter settings on a search space based on MSPC charts is able to overcome sensor failures including failure detection. MSPC trajectories in validation were supported over the whole process with 85 % decisions towards models with almost maximum inputs (representing correct progress of fermentation as well as fully functional sensors). Combined with historical similarity of sensors it was further possible to find false inputs in 100 %. Compared to MSPC and several other process models, the benefits of the presented methodology is not being restricted to any number of multisensory inputs as well as the use of an specific sensor reading.

2.2 Paper copies

2.2.1 Bioprocess monitoring and control via adaptive sensor calibration

Bioprocess monitoring and control via adaptive sensor calibration

To ensure optimal product quality of bioprocesses, it is necessary to develop intelligent control systems with integrated monitoring of key parameters. Having optimal yeast propagation in brewing technology is important to increase the efficiency of subsequent processes. Major drawbacks are: lacks in online detection of yeast attributes and temporal control schemes. One solution is to accurately detect essential process parameters combined with expert knowledge of linguistic control mechanisms. Those needs can be fulfilled by fuzzy logic or state observers including process dynamics associated with accurate multivariate calibration of sensing devices. Ultrasonic-based devices could monitor key parameter but their inline implementation is limited due to influences of the temperature and gas bubbles. Thus, incipient stages for calibration of the device including temperature dependencies using time and frequency properties of ultrasonic waves are presented. A multivariate model using offline measurements with a maximum prediction error of 0.48 g/100 g is reported in this study. Additionally, we show preliminary results of a mechanistic model for the temperature dependency of yeast growth adapted from the literature (biomass and ethanol production, substrate consumption). The results will lead to flexible control of temperature and aeration resulting in vital yeast and enhanced transparency of propagation progress according to the demands.

Keywords: Bioprocess monitoring / Fuzzy control / Multivariate statistics / Ultrasonic analysis

Received: December 2, 2010; *revised:* May 19, 2011; *accepted:* June 10, 2011

DOI: 10.1002/elsc.201000215

1 Introduction

Automatic monitoring and control of biotechnical fermentation processes represents a crucial aspect of various scientific studies. However, two of the biggest drawbacks are the online detection of relevant process parameters as well as the dynamical behavior of biological influenced reactions [1]. The fermentation of viable yeast cells (propagation) in the context of a modern yeast management in brewing industry is of essential importance with respect to the final product quality and its acceptance by the consumer. Such processes require careful, accurate monitoring and control as well as deep experience and knowledge of personnel according to their sensitivity in changes in physical leading parameters.

The PAT initiative, launched in 2004 by the FDA (Food and Drug Administration), forces a shift in the view on process validation and releases [2]. Therefore, instead of complex offline laboratory analysis, process-oriented validation as well as release of process parts is intended. This principle of the PAT initiative has opened a whole new perspective on the understanding of processes that were previously regarded as a black box. Owing to its complex living dynamics, the fermentation processes in the brewing industry present itself as ideal applications of PAT in order to explore its possibilities and limitations. Although significant process variables indicating efficiency as well as high quality of fermentations are well known, their time-efficient determination suitable to process demands is limited. The underlying requirement is thus to provide comprehensive process intelligence via innovative sensor concepts to improve process continuity, process safety and process efficiency. One major task is to combine

Correspondence: Daniel Krause (d.krause@wzw.tum.de), (Bio-) Process Technology and Process Analysis, Life Science Engineering, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany

Abbreviations: **FDA**, Food and Drug Administration; **LOOCV**, leave-one-line-out cross validation; **MISO**, multiple input single output; **PCA**, principal component analysis; **PLSR**, partial least-square regression; **RMSE**, root mean-square error; **RMSECV**, root mean-square error of cross-validation; **TOF**, time-of-flight

innovative sensor principles with modern methods of data analysis and modeling using process and product knowledge.

This contribution presents a perspective for combining ultrasonic measuring equipment used for detecting the relevant process parameters with soft sensing principles as well as intelligent control strategies using fuzzy logic. Therefore, every part in the contribution is divided into three subsections: ultrasonic measuring device including multivariate calibration, mechanistic growth model and fuzzy logic control.

1.1 Ultrasonic measuring device

Detection of relevant system parameters is very crucial regarding process monitoring and control. Among others, the importance of ultrasonic measuring devices for quality assurance in several fields increased over the past decades. Those sensor systems should have the ability to be robust, easy to use and non-invasive, especially in case of food-related applications. Typically used sensor equipment for analyzing fluid properties are in direct contact with the fluid of interest [3, 4] (http://www.iul-instruments.de/pdf/vitalsensors_2.pdf; http://www.anton-paar.com/Dichtesensoren/; http://www.sensotech.com/). Over the last decades, the importance of non-invasive, indirect measurement using ultrasonic sensors for such uses became more and more practical, since they provide rapid response, good long-term stability and high resolution as well as accuracy [5, 6]. The developed sensors for different applications mainly differ in the principle used, the sensor geometry and in the materials used and thus differ in the accuracy and performance [7–10]. Some research groups have studied the possibility of using ultrasonic devices for measuring sugar concentration in different variants [11, 12]. Furthermore, investigations in ultrasonic monitoring of certain fermentation processes including ternary mixtures containing sugar and ethanol were also reported [13–15]. Most of the presented devices in the literature are in contact with the medium and/or applied to an almost isothermal surrounding. Furthermore, investigations on ternary mixtures are usually made in direct relation with the process behavior [13–15]. Those approaches often lack the ability to fit dynamical bioprocesses with changing temperature as well as to fulfill hygienic conditions.

The applied sensor setup is measuring via pulse echo method according to McClements and Fairly, Bamberger and Greenwood and Schäfer [7, 9, 16]. This system is used to investigate changes in physical properties induced by different concentration of solutes, namely maltose and ethanol as well as temperature differences. Different fluid properties cause changes in the traveling velocity of the monitored ultrasonic wave as well as differences in the frequency spectra of respective signals [17]. Developing a physical relationship between pulse distortion and the fluid properties is quite complicated due to the variability (e.g. air bubbles, yeast cells) of the system.

Analyzing material properties as control variables for biotechnological applications in a physical modeling manner is often not possible. To overcome this challenge, the use of multivariate statistics (Chemometrics) is stated. Those statistical-based analysis techniques are used to find a relation between target values **Y**, which are not directly measurable, and some corresponding predictors **X** by means of multiple linear regression (MLR). The use of MLR is only possible, if a matrix **X** (estimators) does have full rank; otherwise, the inverse $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$ does not exist. Thus, data reduction using principal component analysis (PCA) could help solving this problem. PCA summarizes the matrix in new components due to the highest variance in the complete data set. The underlying assumption that large variation in **X** is of necessity when describing **Y** (targets) does not always fit when it comes to regression. It may happen that components with high predictability for certain target values are deleted. The alternative is partial least squares regression (PLSR), which calculates the PC due to highest covariance between estimators (**X**) and targets (**Y**) [18–20].

In this contribution, PLSR is used for modeling the variations in frequency spectra in combination with time-of-flight (TOF) of ultrasonic signals transmitted through aqueous solutions with varying maltose as well as ethanol concentrations using leave-one-line-out cross-validation (LOOCV).

1.2 Motivation for mathematical growth modeling

The control and monitoring of industrial bioprocesses faces special conditions. On the one hand, the living nature of organisms and substances involved in the process adds strong dynamics to the system. On the other hand, it often turns out to be difficult to grasp the process-related information and parameters such as substrate or product concentration online. The necessary long-term stable and robust online sensors for the measurement and monitoring of, for example, biomass, metabolites or nutrient concentrations is rarely available [21, 22]. Therefore, the background of monitoring and control strategies in industrial bioprocesses is laboratory offline analysis. These strategies are limited by the number of samples taken over a specific process. Furthermore, offline analyses include a time delay with respect to the fermentation progress. Therefore, a correct temporal reaction of any control mechanism cannot be ensured. This indicates the demand of sufficient sensor equipment for efficient online monitoring. Promising methods for online monitoring of concentration trends of the relevant metabolites, nutrients and biomass are optical or acoustical spectroscopy [23–25]. However, the industrial use of these techniques in biotechnological processes is rare due to high costs and operation-related adaption. Furthermore, an extensive analysis of the signals on the corresponding target variable is required (metabolites, biomass, etc.). These challenges forced an increase in the interest in the field of indirect measurement and monitoring using "software sensors" over the last decades [26–31]. SoftSensors used in biological-based fermentation processes are typically based on mathematical growth models. Therefore, the living phase in the reaction of interest and the corresponding metabolites will be simulated via differential equations.

The aerobic growth of the yeast strain *Saccharomyces cerevisiae* is well known and studied in detail by several groups [32–35]. Modeling investigations in growth of *Saccharomyces pastorianus* var. *carlsbergensis* is included in the work of Kurz

[36]. Yeast metabolism in general follows two different pathways for energy production when exposed to glucose concentration higher than a critical value, even under the presence of oxygen. This effect is typically known as respiratory shunt or Crabtree effect [32, 37–39]. It describes an overflow metabolism in which ethanol is produced by yeast even under full aerobic conditions. Several different explanations are discussed in the literature; a short overview is given by Kurz [36].

One of the major drawbacks of such models is the implementation of temperature dependence of certain growth parameters. Since control of yeast propagation under brewing relevant conditions is mainly influenced by the temperature, this part has to be considered. In Kurz [36], the temperature influence is investigated; this work was further used as the basis for the model used in the present contribution.

In this article, preliminary investigations on simple kinetics resembling the behavior of aerobic, temperature-dependent growth are shown and compared with the literature. Those equations will be used for further enhancing the monitoring and control issues of the proposed fuzzy propagation system.

1.3 Fuzzy logic control

The control of yeast fermentation in breweries is normally a static, recipe-driven process with isothermal management and continuous or intermittent ventilation. The settings are commonly based on the values obtained by experience. The process settings therefore do not take into account the current yeast requirements. State of the art of process monitoring in this field is achieved by manual sample taking and lab analysis. Therefore, interruptions or changing process conditions will be eliminated only with time-delayed response of the system operator. Owing to the inherent complexity (time-varying behavior and nonlinearity) of biological processes they set up high challenges to the control system. Since the introduction of fuzzy logic control by Zadeh [40], this technique has evolved into an established practice for the control of biotechnological processes [41–44]. Unlike the crisp set theory, the fuzzy set theory allows the transition from the classic bivalent concept of truth to the gradual and multivalent truth concept. Instead of crisp values the fuzzy theory deals with linguistic variables (fuzzified numerical values) that are expressed by corresponding fuzzy sets. In its classical configuration, a fuzzy controller consists of three parts:

(i) Fuzzification
(ii) Inference engine
(iii) Defuzzification.

In the fuzzification part, the distinct, numerical input values read out of the process (received from the measuring devices) get fuzzificated, are assigned to linguistic terms and evaluated by distinct rule bases representing the expert knowledge of the process in the form of "if⋯then" rules. In how far a crisp value belongs to a certain fuzzy set (grade of membership) is described by a membership function. The crisp values are mapped with their membership functions and the membership degree to their fuzzy set, respectively linguistic variable is calculated. In the case of applying multiple input single output (MISO) controllers and the MAXMIN-Inference method, the determination of the resulting membership degree of the premise part is achieved by the minimum (MIN) operator. The degree of performance of the corresponding set in the conclusion part is then calculated by cutting it to the resulting value of the premise part. The retransformation of the linguistic output values into numerical values is accomplished by defuzzification. Applying the widely used Center of Area (COA) (Eq. 1) defuzzification, the conclusion parts of all rules are conjoint by a maximum operator and interpreted as a geometric area. The calculation of the centroid finally delivers the new crisp value, respectively set point for the corresponding actuator.

$$y_{\mathrm{akt}} = \frac{\sum_{i=1}^{n} G_i \cdot y_i}{\sum_{i=1}^{n} G_i} \tag{1}$$

Practical applications of fuzzy logic control of fed-batch yeast fermentations are presented by Besli et al., Mahjoub et al. as well as Miśkiewicz and Kasperski [45–47].

The present work focuses on the analysis of frequency spectra as well as on TOF predictions in combination with temperature measurements for the prediction of the combined effect of ethanol and sugar on the ultrasonic pulse behavior. Multivariate calibration is applied using PLSR; the model size was estimated by the most commonly used cross-validation. The model output is given in terms of an extract value, which resembles the total amount of dissolved carbon sources. The measured ternary mixtures are presented with a value called "apparent extract," which is a common representation of a typical yeast propagation progress in the brewing industry [48]. This value was applied earlier for fuzzy control and indicated that temporal control of yeast propagation in the given boundary conditions is possible [49, 50]. Furthermore, investigations are made on modeling the temperature-dependent growth of yeast with simple model equations adapted from the literature. Those mathematical functions are thought to be implemented as objective functions in the fuzzy control scheme to enhance the controller accuracy. This work is thought as a perspective in combining new ultrasonic measurement devices for the qualification of fermentation progress together with a fuzzy logic control scheme, which is supposed to be enhanced by trend estimation using mechanistic modeling.

2 Materials and methods

2.1 Ultrasonic measuring device

2.1.1 Experimental setup

The experimental setup based on a VARINLINE® clutch as well as the measuring principle is explained schematically in Fig. 1. This figure includes a photograph of the setup as well (middle). The clutch made of stainless steel is widely used as a typical process access for inline sensors. The setup pipe diameter used was 50 mm (DN50). The indirect heating/

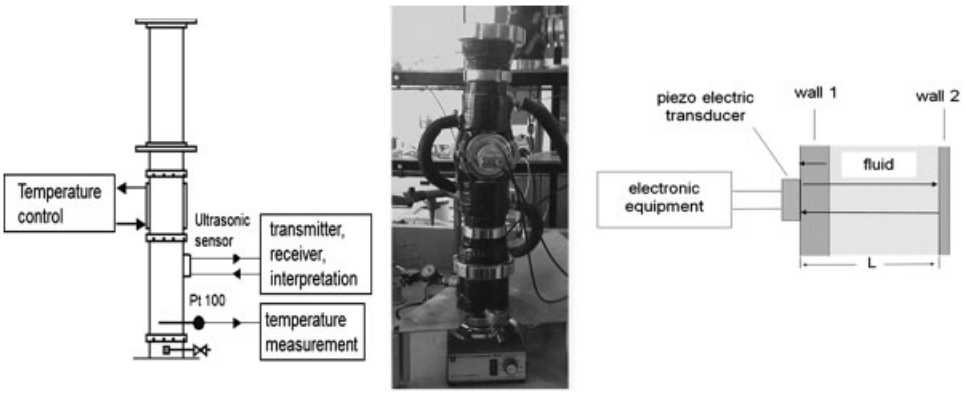


Figure 1. On the left hand side, a schematic drawing of the experimental setup is shown; it includes a tempering mantle provided with tempering fluid by an external thermostat, the piezo electric transducer (coupled to a microcontroller) and a Pt100 resistance thermometer for monitoring the temperature inside the chamber. This setup is shown also in the photography (middle); the right hand side shows the schematic principle (pulse echo method) with piezo electric transducer, excited via the electronic equipment; after passing the first wall (steel), the pulse travels through the sample and is reflected at the backside (wall 2); transducer works as receiver at the same time, signal is collected via the electronic equipment (microcontroller).

cooling of the container was established using an external thermostat. Investigations on ternary mixtures (ethanol, maltose as sugar equivalent, water) are made by filling the chamber with respective solutions, continuous heat supply and stirring by magnetic stir bar for homogeneous temperature distribution. Ultrasonic signals are recorded over a temperature range reaching from 6 to 22°C at constant container pressure over a time frame of around 3 h. The temperature points used in this study for calibration were extracted in steps of 0.5 K ($\pm$0.02 K) out of this continuous spectrum.

The in-house-produced piezo electric transducer (built using a piezo ceramic with a center frequency of 2 MHz) is used for creating an ultrasonic pulse by excitation with a rectangular electrical pulse (width of 200 ns, amplitude of 5 V). After passing container wall (wall 1) and fluid, the pulse is reflected at the backside (wall 2) and caught by the transducer, which works as a receiver in the same time (pulse-echo method).

The signal is recorded via a microcontroller connected to the measuring device. The temperature of the probe fluid is measured by Pt 100 temperature sensor (maximum accuracy of $\pm$0.1°C). The experiments were carried out for each solution mixture separately.

Several mixtures of maltose–water and maltose–ethanol–water in a range of 0–12% (per weight) maltose as well as 0–3% (per weight) ethanol for measuring the ultrasonic pulse behavior were prepared (Fig. 2). Homogenization was reached with a magnetic stir bar. For each investigation, a sample volume of 3 L was prepared by dissolving a known mass of crystalline maltose (D(+)-Maltose Monohydrate, Roth®) and ethanol (HPLC Gradient Grade, Roth®) using a Sartorius Laboratory® weight (L 2200 S) in demineralized water reaching a defined final weight.

The reference concentration for control was measured using an Anton Paar® Density Meter (DMA 4500). Signals for analysis were extracted from the total set of continuous measured temperature spectrum resulting in data sets of 40–60 objects with varying concentrations for each temperature point, respectively.

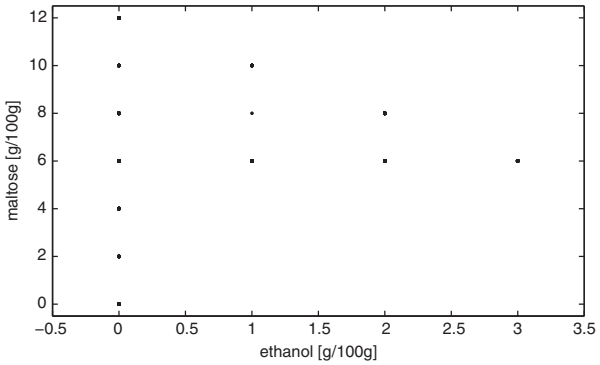


Figure 2. This diagram shows the experimental design of the measured samples (maltose–ethanol–water mixtures, chosen in relation to propagation processes for calibration.

2.1.2 Signal processing

2.1.2.1 TOF prediction

The setup-specific TOF prediction of the ultrasonic wave traveling through the fluid was achieved by comparing a reference reflection insight the buffer material with a reflection of the echo using the cross-correlation method.

A typical signal shape of the used setup with schematic explanation is shown in Fig. 3. This method and the used algorithm is explained in detail by Hoche [51].

2.1.2.2 Frequency analysis

Extracting the corresponding frequency spectra every signal is analyzed by Fast Fourier transform. The achieved magnitude spectrum ($P(f)$) is analyzed for phase changes ($\varphi(f)$). The

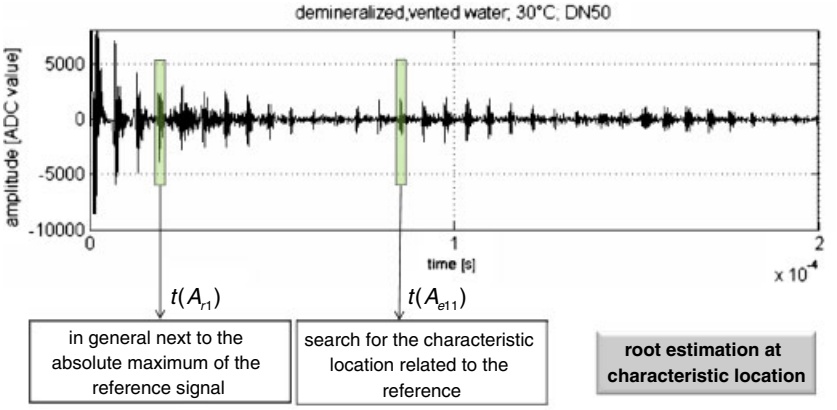


Figure 3. This figure shows a typical signal from the above-presented setup. Prediction of time-of-flight (TOF) is achieved by temporal comparison of the echo impulse t(Ae11) and the complementary buffer reflection t(Ar1).

resulting two vectors per signal are taken as input in rows combined with the corresponding TOF measured for each signal.

Since phase of a signal is sensitive to noise [17], the used bandwidth for analysis has to be adapted. For extracting the frequency domain information, the signal is divided into equal segments with 90% overlapping, and each segment was scaled by Blackman window function ($w(n)$, see Eq. 2) to ensure less spectral leakage. Later, the windowed segment was transformed to frequency domain using Fast Fourier Transform (FFT) method, at which the output frequency domain representation (i.e. power spectrum versus frequency) is the average of all the single segment transformations.

$$w(n_f) = A_0 - A_1 \cdot \cos\left(\frac{2 \cdot \pi \cdot n_f}{N-1}\right) + A_2 \cdot \cos\left(\frac{4 \cdot \pi \cdot n_f}{N-1}\right) \quad (2)$$

where N is the length of each window (in this contribution 1/10 of signal length), A_{0-2} are constants with values of 0.42, 0.5 and 0.08, respectively. The region of highest variation in the signals was found in a bandwidth of 0.5–3.5 MHz applying the method mentioned above (Fig. 4). This bandwidth was taken for (unfiltered) magnitude as the model input. The same bandwidth was used calculating the spectral phase representation of single signals. For statistical investigations, the data used require pre-processing. The following abstract explains the organization and calculations on matrices for the regression analysis. The target values (apparent extract (%), measured offline with the Anton Paar®) are stored in a column vector. The corresponding signal properties (magnitude, phase and TOF at a certain temperature) are stored in rows of a predictor matrix **X**.

$$\mathrm{X} = [tof, \mathbf{P}(f), \varphi(f)] \quad (3)$$

Previous to the decomposition the matrices were autoscaled by centering each column to its mean value and scaling to unit variance dividing by its standard deviation.

2.1.3 Introduction to partial least squares

The calculations on data sets were programmed and carried out applying the most commonly used nonlinear iterative partial least-square (NIPALS) algorithm developed by Wold [20]. This algorithm is calculating the PC iteratively. This statistical method uses reduced amount of latent

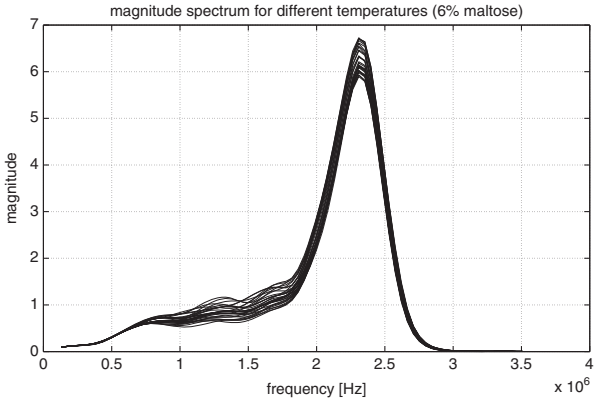


Figure 4. Magnitude spectra of smoothed magnitude presentation using Blackman window. The figure shows spectra of different signals for changing temperature at the same maltose concentration; Bandwidth area with high variance from ~0.5 to ~3 MHz.

variables compared with the descriptor variables found by highest covariance between in **Y** and **X**. The used algorithm for PLS calculating k components was carried out in home-built subroutines programmed using MATLAB®. A brief description of the algorithm can be found in the literature [19, 20, 52–54] (http://folk.uio.no/henninri/pca_module/pca_nipals.pdf).

2.1.4 Calculation of regression coefficients

After iteration is finished, the principal components are used to calculate the parameters in **B** of the regression model

$$\hat{\mathbf{Y}} = \mathbf{X} \cdot \mathbf{B} + \mathbf{1} \cdot \mathbf{b}_0 \quad (4)$$

Vector **b** is estimated as follows:

$$\mathbf{b}_0 = \mathbf{y}_m - \mathbf{x}_m \cdot \mathbf{B} \quad (5)$$

where the vectors $\mathbf{y_m}$ and $\mathbf{x_m}$ contain the mean values of the corresponding columns of **X** and **Y**. Matrix **B** is estimated as

follows:

$$\mathbf{B} = \mathbf{S}_X^{-1} \cdot [\mathbf{W}(\mathbf{P}^T \cdot \mathbf{W})^{-1} \cdot \mathbf{Q}^T] \cdot \mathbf{S}_Y \quad (6)$$

where the diagonal matrices **S** include the standard deviation of the corresponding columns of **X** and **Y** [52, 55].

The codes for calibration were programmed and carried out using MATLAB® (Version 7 Release 14, The MathWorks, USA).

2.1.5 Estimation of model size and accuracy

Calculation of regression coefficients (matrix **B**) is carried out by the following Eq. (6), respectively. Choosing the most reliable model order (number of PC taken for estimation of the parameter matrix **B**) causes the most problems in terms of accuracy and stability of the calculated regression model. One possible criterion is choosing the model size by the minimum predictive error (for example, root mean-square error (RMSE); the formal description of RMSE is shown in Eq. 8). In this contribution, the RMSE of cross-validation is used for model order prediction (RMSECV). Calculating the error was absolved leaving every line of each data set out once using it for validation after PLSR. The method is known as LOOCV.

$$\mathrm{RMSE} = \sqrt{\sum_{i=1}^{n} (\mathbf{y}_i - \hat{\mathbf{y}}_i)^2 / n} \quad (7)$$

The variables used in Eq. (7) are n for the number of samples, $\mathbf{y}$ for the known target value and $\hat{\mathbf{y}}$ for the corresponding predicted value.

2.2 Mechanistic growth model

2.2.1 Materials and methods

Validation of the kinetic model was accomplished collecting experimental data from fermentations of *S. pastorianus var. carlsbergensis* W43/70 (common bottom-fermenting yeast). To ensure comparability between each of the fermentations, the following settings were kept constant. The inoculum used was pre-cultivated in a sterile wort reaching a final volume of 2 L. The fermentations at three different temperatures were carried out using a B Braun® System (B Braun® Biostat UD-30) filled with 20 L brewing wort with 12 g/100 g original gravity. Stirrer speed was adjusted to 200 U/min, aeration rate at 5 L/min. A volume of around 5 mL antifoam solution was added to the medium to prevent foam. Additionally, the system was kept with an overpressure of 250 mbar. The fermentation time differed between 13 and 42 h according to the number of yeast cells per milliliter at the end of fermentation.

The simulations for the aerobic yeast growth were calculated using Berkeley Madonna® (v8.3.18, Berkeley Madonna)

2.2.2 Theory of aerobic yeast growth

The lack of suitable sensor systems as well as insufficient process knowledge inspires the development of mathematical approximations to get deeper insight into the organisms' behavior. Those approximations are used among others to predict and monitor the progress of cultivation. Such systems are useful to predict the relevant process parameters, which are of particular interest. One of the possibilities is the establishment of a reliable growth model.

The typical aerobic growth behavior of *Saccharomyces* yeast strains is described by Monod-type equations. The substrate uptake (Eq. 8) follows Monod kinetics, inhibited in the presence of ethanol as shown by Hoppe and Hansford [56].

$$q_s = q_{s,\max} \cdot \frac{C_s}{C_s + K_s} \cdot \frac{K_{ie}}{K_{ie} + C_E} \cdot L_t \cdot f_{temp} \quad (8)$$

For description of lag phase, when microorganisms are inoculated to the growth medium, a more detailed look on biochemical pathways has to be made. Basic Michaelis–Menten or Monod-type equations describe only the transition phase of depletion in substrate of interest but not the intrinsic enzymatic adaption in the beginning of the process. Therefore, an unstructured model using basic equations would not describe the complete process sufficient enough. Thus, a sigmoidal function simulating the lag phase is used to overcome this disadvantage, represented by the factor L_t (Eq. 9, Fig. 5).

$$L_t = \frac{1}{1 + e^{-(t - t_{lag})}} \quad (9)$$

The second factor f_{temp} is introduced to simulate the temperature-dependent glucose uptake. Gathering first insight, the model output of the preliminary investigations in this contribution is compared with the square root model (Eq. 10).

$$f_{temp} = (b \cdot (T - T_{min}) \cdot (1 - e^{c \cdot (T - T_{max})}))^2 \quad (10)$$

In contrast to the model approaches presented by Kurz, Sonnleitner and Kaeppeli or Barford [36, 39, 57], the division of substrate into different pathways was considered as

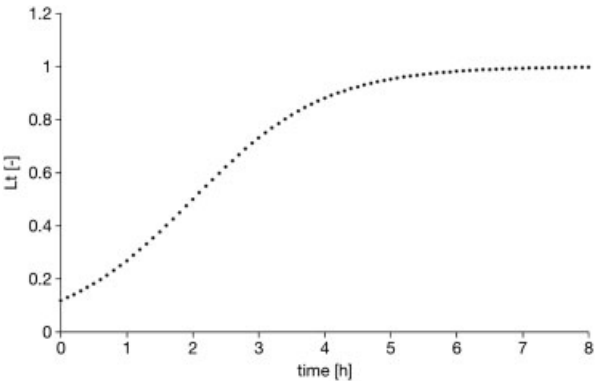


Figure 5. Course of the exponential function for simulation of the lag-time in the beginning of cultivation. With $t_{lag} \sim 2$ h, the factor L_t asymptotic convergence to one is achieved at around 1.

proposed by Irvine et al. [58]:

$$\alpha = \frac{1}{1+C_S/K} \tag{11}$$

This assumption is suitable for "blackbox" modeling and valid, as long as oxygen saturation in the medium is ensured.

The growth of yeast cells is considered as autocatalytic reaction (Eq. 12), whereas the total specific growth rate μ is described as the sum of the two substrate pathways partitioned by the value α, similar to Sonnleitner and Kaeppeli [39].

$$r_x = \mu \cdot C_X \tag{12}$$

The final system of differential equations only considers the mass balances of biomass, sugar and ethanol in the liquid phase. Growth on ethanol is not taken into account, since substrate concentration of the respective propagation process is always higher than the critical respiration limit (>1 g/100 g, [48]). Therefore, ethanol uptake can be neglected.

2.2.3 *Offline analysis and processing*

The experimental values used for the validation of the simulation had to be preprocessed. The fermentable portion of carbohydrates in wort is estimated as 70% of the total extract (total solids contained in a liquid) measured with the Anton Paar®. Furthermore, simulations were carried out in unit mmol/L. The conversion from unit g/100 g was carried out using glucose as the sugar equivalent. The biomass concentration was determined by microscopic counting via "Thoma-Kammer".

$$C_X = \frac{m_s \cdot n_{cells}}{M_{biomass}} \tag{13}$$

Conversion to unit mmol/L was absolved using Eq. (13) where m_s resembles the specific dry weight of a cell (2.5×10^{-11} g/cell [48]), n_{cells} the yeast cell count (cells/mL) and M the molar weight of 1 mol biomass. The mean biomass composition ($CH_{1.79}N_{0.15}O_{0.5}$) is taken from the literature [39].

2.3 Fuzzy logic theory

The supply of vital yeast in sufficient quantities and quality at the right time point is a crucial factor of the process and production planning. In Birle et al. [49], an expert system based on the fuzzy logic control is proposed, which allows a flexible, dynamic and demand-oriented process control of brewer's yeast propagation instead of static and inflexible, step sequence-based process control systems that are hitherto existing.

The set up of the fuzzy controller and the rule bases was accomplished with the software Virtual Expert® (Gim*bio* GmbH). The operator adjusts the desired point of yeast harvest via a user interface. The final cell concentration should be at least 80×10^6 cells/mL in achieving this fixed time point. This concentration is needed to guarantee a fast start of the subsequent process of alcoholic beer fermentation. The decline of apparent extract concentration is measured by a laser refractometer (ACM®). For extract trajectory, an objective function was defined. The deviation between the current trajectory and the objective one was taken as the input for the fuzzy temperature controller. As a second input variable, the deviation of current and objective slope of cell growth was applied. After assessing the deviations via a simple rule base in the manner of "if⋯then" rules, the fuzzy controller delivers an incremental increase or decrease in the temperature in the propagation tank. The strategy of ventilation is dedicated to an optimal supply with oxygen. To cope with this requirement, the ventilation controller uses the yield coefficient (cell growth/% extract) and the absolute cell concentration as input

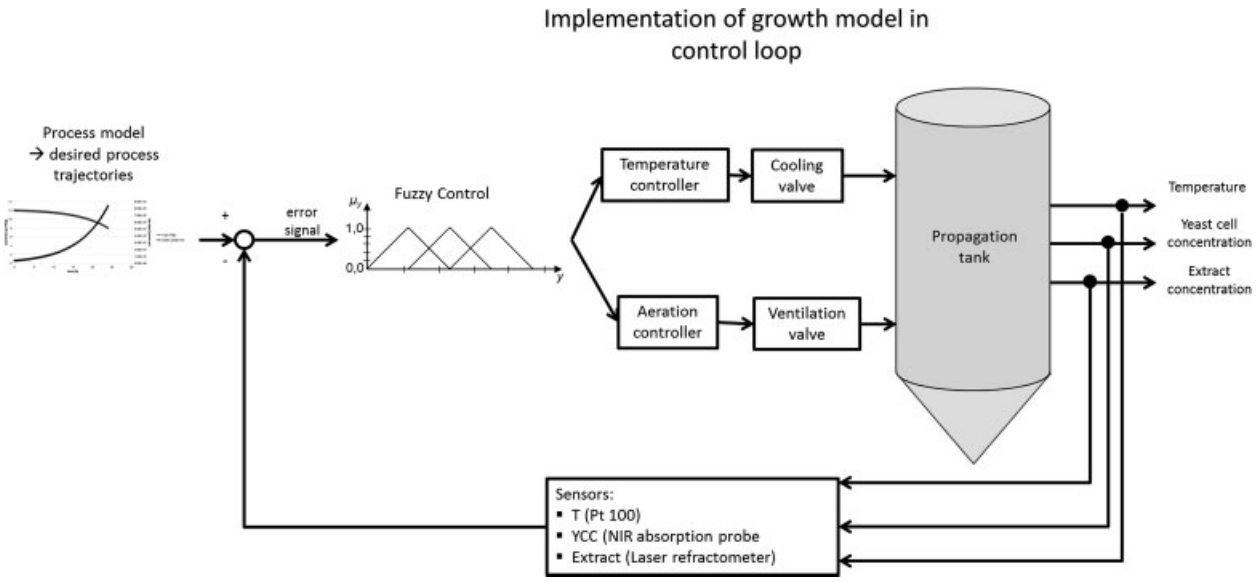


Figure 6. Applied control loop of the fuzzy propagation system; current values of involved sensors are compared to respective desired values out of the model – deviations entering the fuzzy controller; fuzzy outputs are used for tuning the corresponding actuators (cooling and ventilation valve).

variables to control the ventilation cycles and aeration volume. The applied fuzzy inference system follows the implication principle of Mamdani linking the membership functions in the precondition part of a distinct rule via MIN-operator (logical AND). The basic idea of this alternative of implication is that the conclusion's content of truth should not exceed the one of the precondition parts [59]. The fuzzy sets themselves were kept quite simple using standard forms of triangular and trapezoid sets. Although the rule base was kept very little, comprising only six rules for temperature regulation and nine rules for aeration control, the presented controller (see Fig. 6 for schematic explanation of the control loop) works quite reliable.

3 Results and discussion

3.1 Ultrasonic measuring device

In the above-mentioned paragraphs, it was shown that the influences of temperature as well as different solutes in aqueous solutions are visible changing from time domain to frequency domain. For finding the best solution in the determination of unknown concentrations, the use of statistical tools such as Chemometrics is proposed. Before calculating the regression model, several processing steps had to be carried out as explained in the sections before.

First, the signals were analyzed to detect the correct frequency band for all signals used for further calculations. The processing steps and the calculation reaching to the final regression model are collected to provide better understanding in a final scheme shown in Fig. 7. The final regression model was calculated using PLS regression, model size and accuracy were determined by cross-validation.

The prediction errors (RMSECV) of the 33 different models for each temperature point (6–22, 0.5 K steps) achieved by using nine PLS components for each model are shown in Table 1. The maximum error is <0.5 g/100 g. The two models showing higher a prediction error (9 and 15.5°C) contain

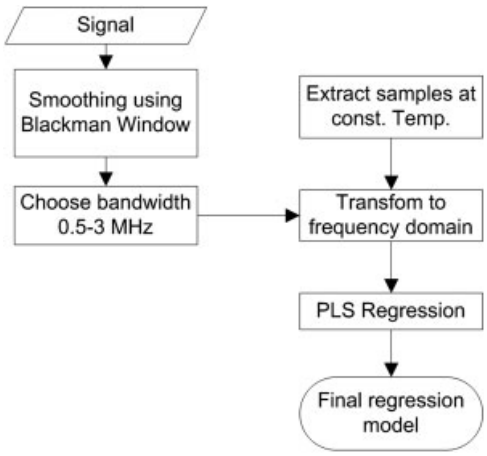


Figure 7. Schematic sum of the algorithm used for regression predicting the apparent extract in liquid medium.

Table 1. RMSECV for each model in the temperature range

T in (°C)	RMSECV (g/100 g)	T (°C)	RMSECV (g/100 g)
6	0.4	14.5	0.24
6.5	0.38	15	0.24
7	0.31	**15.5**	**0.87**
7.5	0.27	16	0.18
8	0.28	16.5	0.37
8.5	0.44	17	0.38
9	**0.61**	17.5	0.48
9.5	0.23	18	0.2
10	0.26	18.5	0.27
10.5	0.31	19	0.32
11	0.37	19.5	0.3
11.5	0.33	20	0.32
12	0.25	20.5	0.25
12.5	0.31	21	0.29
13	0.29	21.5	0.35
13.5	0.37	22	0.26
14	0.29		

samples predicted as outliers calculating the leverage of each sample [19, 60] (result not shown) and were thus not taken into account. The method of outlier detection still needs further investigation.

The parity plot in Fig. 8A shows a good relation between the measured and predicted values for respective solutions. Thus, the variations of the wave shape due to medium as well as temperature influences are visible considering the frequency transform of the total signal (transmitted pulse by the transducer and pulse traveled through the medium). The statistical method gathers this information by summing up the significant variance in few PLS components.

The final regression model structure is given as

$$\hat{\mathbf{y}} = \mathbf{1} \cdot \mathbf{b}_0 + \mathbf{X} \cdot \mathbf{B} \tag{14}$$

where vector $\mathbf{B}$ is structured as follows:

$$\mathbf{B} = [b_{tof}, b_{P(f_1)}, b_{P(f_2)}, \ldots, b_{P(f_n)}, b_{\varphi(f_1)}, b_{\varphi(f_2)}, \ldots, b_{\varphi(f_n)}]^{\mathrm{T}} \tag{15}$$

Validation of the models was achieved using samples out of whole data set, which were not included in the calibration. To calculate the concentration at the right temperature, the result was extracted by interpolation:

$$\hat{\mathbf{y}}_1(T_1 \geq T_0 - 0.5) = 1 \cdot \mathbf{b}_0(T_1 \geq T_0 - 0.5) + \mathbf{X} \cdot \mathbf{B}(T_1 \geq T_0 - 0.5)$$

$$\hat{\mathbf{y}}_2(T_2 \leq T_0 + 0.5) = 1 \cdot \mathbf{b}_0(T_2 \leq T_0 + 0.5) + \mathbf{X} \cdot \mathbf{B}(T_2 \leq T_0 + 0.5)$$

$$\hat{\mathbf{y}}_0(T_0) = \hat{\mathbf{y}}_1 + \frac{\hat{\mathbf{y}}_2 - \hat{\mathbf{y}}_1}{T_2 - T_1} \cdot (T_0 - T_1)$$

Using linear interpolation as first approximation causes a prediction error of 0.5 g/100 g for apparent extract with a marginal error of 45 ns for the TOF as well as 0.5 K temperature variation. This estimation is shown in Fig. 8C. The model combination over temperature still needs further investigation.

Nevertheless, high variations in the predicted results of used regression model may come from qualitatively corrupt ultrasonic

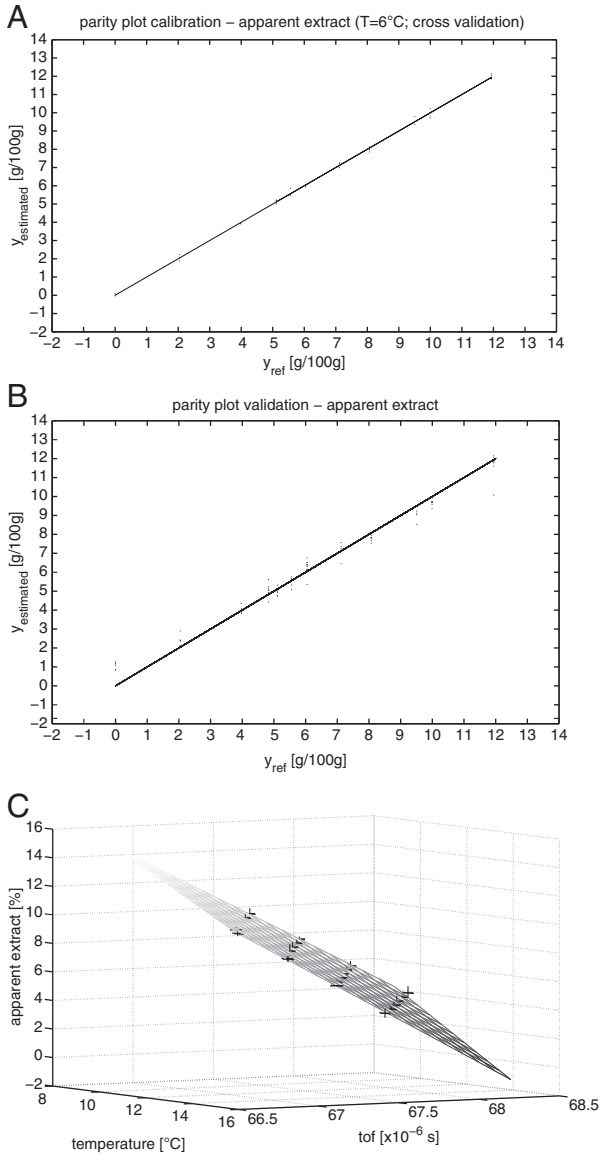


Figure 8. (A) Parity plot for apparent extract (g/100 g); calibration with samples at 6°C (±0.03 K); maximum predictive error over all temperature models (RMSECV) < 0.5 g/100 g; number of component: 9. (B) Parity plot for maltose; validation with samples not taken into calibration; maximum absolute error between 4 and 10 g/100 g maltose is less than 0.8 g/100 g. (C) Estimation of marginal error; the plane shows the behavior for maltose solutions. Error bars plotted were estimated by comparing the values for the models with corresponding temperature points (°C) with the output by linear interpolation. The TOF (10^{-6} s) was taken to resemble the ultrasonic property.

signals caused by the used setup (see parity plot in Fig. 8B). This is on the one hand due to the material used as buffer and on the other hand the data evaluation by the used electronics. Future steps include effort in investigating different buffer materials as well as revised electronic equipment. Furthermore, detailed investigation of the frequency domain using the sensitivity analysis as well as variable selection methods to investigate the most informative parts of the spectra in correlation to the targets of interest by means of regression will be carried out. This includes as well subdivision of signals into start and echo with independent analysis to investigate their relations.

The possibility to apply this sensor system online in a process later on will be one of the biggest benefits compared with the existing measuring systems in the industry. So far, online solutions are typically in direct contact with the medium and thus, service and maintenance as well as construction are of high effort. Final estimation of the substrate concentration using the presented approach would need simple vector multiplication. Therefore, online usage due to low processing time is possible. Further research will be applied on the one hand to reach a higher accuracy and on the other hand to implement the system online in a process.

3.2 Mechanistic growth model

Following the trend of temperature-dependent yeast growth with respect to substrate, ethanol and biomass concentration, a preliminary approach using simplified equations from the literature was tested. Therefore, three laboratory fermentations were absolved validating the principle accuracy as well as the temperature dependency. The parameters used for simulation are shown in Table 2.

The results of the simulations compared to validation runs are shown in Fig. 9A. The absolute errors compared with the offline measured data are summarized in Table 3. It is shown that the accuracy of this simple model is already quite acceptable. The parameter f_{temp} was adapted according to the respective temperature.

The value of biomass yield over the fermentative pathway (Y_{XSf}) had to be adopted compared with the literature [36, 39] to narrow the simulation output to the experimental values. This can be explained by several reasons. Owing to preprocessing of values for yeast cell count using specific yeast cell dry weight, which ranges between 2.5×10^{-11} [48] and 4×10^{-11} g/cell [36], values for biomass in mmol/L thus vary in a big span. Therefore, it has to be investigated, how this value behaves under reported conditions between different strains of *Saccharomyces* yeasts. Furthermore, the method of microscopic cell counting is an error-prone analytical method, although it is broadly accepted [61]. The final deviation between simulation results and

Table 2. Parameters of the Black-Box Model applied for validation

Parameter	Value	Unit	References
$q_{s,max}$	0.486	mmol/(mmol · h)	[1]
K_s	2.8	mmol/L	[1]
K_{ie}	500	mmol/L	[2]
K	5.5	mmol	*
Y_{ES}	1.748	mol/mol	Stoichiometry
Y_{XSox}	3.527	mol/mol	[1]
Y_{XSf}	0.98	mol/mol	*

Parameters with units, applied values and references are presented. Parameters marked with * are determined via the parameter estimation procedure.

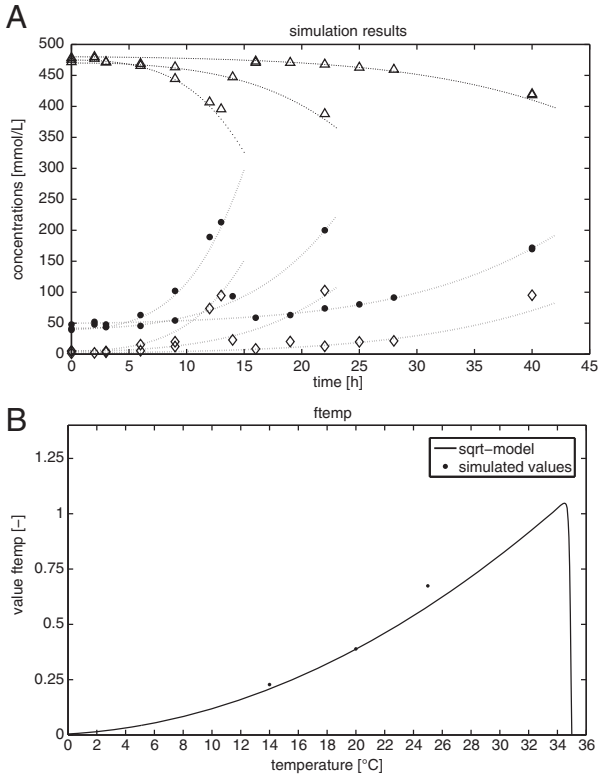


Figure 9. (A) Results of three fermentations compared to model simulations; it could be shown that using the proposed literature model with simplifications it is possible to simulate such processes. (Triangles and dark dotted lines resemble substrate trend, filled dots and light dotted lines resemble ethanol and diamond with dashed lines biomass trends; all concentrations are given in unit mmol/L). (B) Comparison between the square root model [36] for f_{temp} (hard lined) and simulation results from own runs (dots). Each dot resembles a single fermentation.

Table 3. Absolute errors between model simulations and experimental data form absolved fermentations

	Substrate (g/100 g)	Biomass (g/100 g)	Ethanol (g/100 g)
$T = 14°C$	0.18	0.062	0.011
$T = 20°C$	0.16	0.021	0.04
$T = 25°C$	0.3	0.041	0.091

Table 4. Values for parameters f_{temp} and t_{lag} achieved through parameter estimation for different temperatures

	f_{temp}	t_{lag}
$T = 14°C$	0.228	1.71
$T = 20°C$	0.39	3.3
$T = 25°C$	0.674	1.91

experimental data is shown in Table 3. Those variations of around 0.3, 0.06 and 0.09 g/100 g for substrate, biomass and ethanol, respectively, are believed to be due to abovementioned reasons. Nevertheless, further investigations are needed to get a deeper insight in those differences.

Furthermore, parameter t_{lag} was adapted in each run, since lag-time in the beginning of fermentation is dependent on the temperature as well as on the vitality/viability of yeast. This organism-specific properties could not be guaranteed to be always the same (for values see Table 4). It also noteworthy that variety in the wort composition taken from different batches could be a reason for small differences. The parameter f_{temp} is shown in Fig. 9B compared with the square root model proposed by Kurz [36].

Those preliminary results show that the assumptions made for simplifying the model equations are not far from the validated literature data. It has to be mentioned that the three points in Fig. 9B have to be validated by further test runs at several temperature points as well as varying temperature profiles. In addition, the influence of the Monod kinetics on the substrate decline presented in this work is rather small. Nevertheless, using the existing fundamental knowledge is believed to be more reliable in modeling growth of microorganisms. Therefore, future work will be accomplished increasing the amount of data to proof the simulation results as well as the mathematical background in detail. The underlying aim of raising the accuracy of the fuzzy controller by adapting the objective function for trend estimation using the apparent extract via simple approximation equations is possible. Vice versa, the controller could be also adopted to use other input variables like biomass as well as virtual concentration of sugar or ethanol.

3.3 Fuzzy logic control

First trials in the field of brewer's yeast propagation control via fuzzy logic [49] show very promising results. The applied fuzzy controller receives online field data from a sensor array comprising the measurements of dissolved oxygen (Clark electrode, Mettler Toledo®), turbidity (yeast cell count; Optek®), apparent extract concentration (laser refractometer; ACM®), temperature and pressure. The sensor array was implemented in the circulation pipe of the system. The controlled parameters are the temperature (glycol cooling system), ventilation intervals and the volume flow of ventilation. For this research, objective functions for extract decline as well as growth of cell count concentration were defined. The functions (slopes) are variable in time and therefore can be adjusted by setting the desired point of yeast harvest. As there was no experience about the metabolic and anabolic behavior at permanently changing temperatures, the objective functions were kept linear at first. The applied fuzzy system consists of two controllers, one for temperature control and another one adjusting aeration parameter such as ventilation periods (length of ventilation intervals) and volume flow of ventilation. The differences of real and objective rates are used as input variables for the fuzzy temperature controller that continuously adjusts temperature in the propagation vessel. Figure 10 shows schematically the mode of operation of the fuzzy temperature controller. The deviations of the first derivatives of extract decline and yeast cell growth compared with their model

functions are used as the input variables:

$$e_{\text{Extract}} = \frac{\mathrm{d}Extract_{\text{Model}}}{\mathrm{d}t} - \frac{\mathrm{d}Extract_{\text{Output}}}{\mathrm{d}t} \quad (16)$$

$$e_{\text{YCC}} = \frac{\mathrm{d}YCC_{\text{Model}}}{\mathrm{d}t} - \frac{\mathrm{d}YCC_{\text{Output}}}{\mathrm{d}t} \quad (17)$$

The two plots show the real trajectories (solid lines), model (dashed lines) and the deviation as control difference (dotted lines) already indicating potential for optimization. The input fuzzy sets used in this work are defined by trapezoidal membership functions and the applied rule base (Table 5) comprises six rules allocating the linguistic terms via simple control algorithms in the form of "if-then" rules. Change in the temperature is achieved via incremental temperature variation

$$T_{\text{out}} = T(t-1) + \frac{T_{\text{fuzzy}}}{n} \quad (18)$$

where T_{out} is the new calculated set point of temperature and $T(t-1)$ the temperature one time step before. T_{fuzzy}/n denotes the fuzzy output value referred to one hour divided by the number of control cycles per hour, each control cycle being 30 s. The fluctuations in those trajectories may come from inadequate objective functions used in the presented approach. Those model functions do not consider an initial lag phase and

Table 5. Rule base of temperature controller

Temperature	
IF	THEN
Delta_Es_Rate = slower	d_Temperature = pos.
Delta_Es_Rate = matched	d_Temperature = null
Delta_Es_Rate = faster	d_Temperature = neg.
Delta_HZZ_Rate = slower	d_Temperature = pos.
Delta_HZZ_Rate = matched	d_Temperature = null
Delta_HZZ_Rate = faster	d_Temperature = null

Delta_ES_Rate: Deviation of the current uptake rate of extract from the corresponding objective function; Delta_HZZ_Rate: Deviation of the current growth rate of biomass from the corresponding objective function; d_Temperature: incremental increase or decrease in temperature.

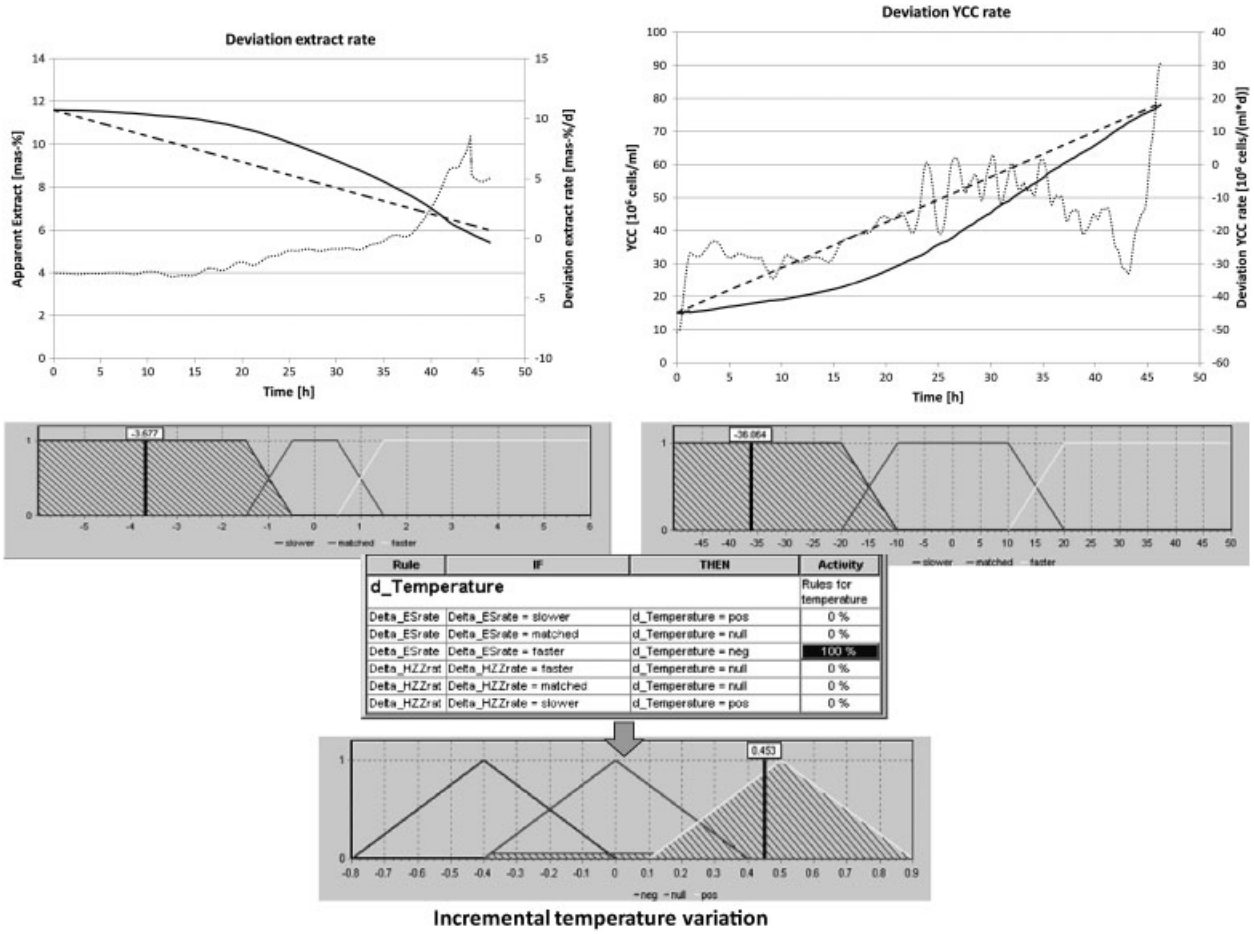


Figure 10. Schematic outline of the applied fuzzy temperature controller. Shown are the deviations of the first derivatives between modeled and real extract decline (g/(100 g · d)) and yeast growth rate (mio. cells/(mL+d)). Fuzzy sets are trapeziodal for both inputs. The rule base includes six rules. Fuzzy output variable is a temperature increment referred to as increase or decrease of temperature per hour.

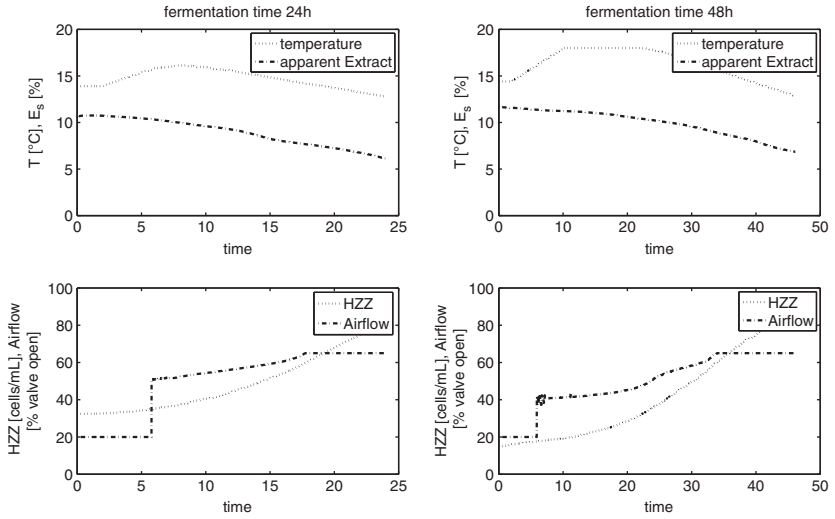


Figure 11. Plots show the results for fuzzy controlled propagations finished after 24 and 48 h, respectively. It is visible that the controller adapts the temperature (⋯line; unit °C) as well as the airflow (-.-line; resembled by percent opening of the aeration valve); the fermentation aim (∼6% apparent extract, -.-line; 80–100 Mio cells/mL, ⋯line) could be reached in both cases; the big jump in the trend of the airflow is due to the inflexible valves.

transition phase to exponential growth of the culture at process start. Thus, the deviations in the beginning are strongly negative. As a consequence, the temperature controller tends to increase the temperature in the initial process phase. Furthermore, bigger fluctuations in growth rate are visible and supposed to be due to effects of gassing out and gas bubbles influencing the turbidity measurement. Those indications show some of the major research points for future work already in progress. The second fuzzy controller used for aeration control based on the yeast cell count in combination with the yield coefficient was also established. The corresponding rules are not shown. The fermentation results achieved with the established controller are shown in Fig. 11.

The rule base decisions of the fuzzy controller for the aeration volume and the aeration cycles are based on the input variables yeast cell count concentration and yield (relation of cell growth and extract decline). One of the drawbacks observed while propagating was that due to the extreme formation of foam in the vessel, a continuous aeration strategy could not be pursued. Thus, pulsed ventilation was applied. These aeration intervals caused a non-negligible noise of the turbidity signal (result not shown). It is assumed that during the paused time period the hydrostatic pressure decreases and leads to gassing out. The gas bubbles are then detected by the sensor and interpreted as yeast cells by mistake. This effect needs further investigation, since optical sensors as well as acoustic ones are affected by bubbles. Another point is that the culture partly tends to switch to the anaerobic metabolism what can be seen from the trajectories of yield and the difference of the substrate uptake rate to the model function.

4 Concluding remarks and outlook

The paper is thought as a perspective in combining new ultrasonic measurement devices for qualification of fermentation progress together with a fuzzy logic control scheme, which is thought to be enhanced by trend estimation using mathematical modeling. Furthermore, the fuzzy controller will be enhanced embedding additional sensor information, which will make the process control more independent to sensor failures. Therefore, a new plant design is established containing a more comprehensive sensor array to enlighten the process of yeast propagation. The disadvantages of inflexible valve control mentioned for the reported trials will be diminished by implementing a flexible mass flow controller. Furthermore, analysis of the exhaust air will be introduced to obtain deeper insight into the metabolic activity of the yeast. The major drawback while propagating was the formation of foam as stated in Section 3. Therefore, pulsed ventilation was chosen causing a non-negligible noise in measuring the turbidity, which is assumed to be caused by degassing. This phenomenon as well as the aeration strategy itself has to be further investigated, since optical sensors as well as acoustic ones are both sensitive to gas bubbles in the medium. Another disadvantage will be diminished by replacing the inaccurate objective functions used so far by more ideal trends. Those simplified functions showed a good compromise in the preliminary investigation (variations of around 0.3, 0.06 and 0.09 g/100 g for substrate, biomass and ethanol, respectively). The discussed influence of the Monod part in the presented substrate decline will be proved in detail by future investigations, since it is believed that using the existing fundamental knowledge is more reliable when explaining growth behavior of microorganisms. The presented model equations need further validation, proof of assumptions as well as investigations on limitations like nitrogen source or trace elements. Furthermore, taking balances for oxygen and carbon dioxide in both liquid and gaseous phase into account would enhance model output with respect to the growth behavior. Research has to be made on the implementation of such equations into a process control scheme. With respect to an additionally enhanced control strategy, the procedure of temperature control has to be reconsidered. The system reacts relatively slow to temperature changes. Therefore, the controller might increase or decrease the temperature to its upper or lower limit until a significant change of cell growth, respectively extract consumption occurs. Thus, the controller is likely prone to

oscillation. To avoid this phenomenon, it is intended to implement a more predictive controller based on numerical state estimation, taking into consideration the temperature-dependent growth kinetics of *Saccharomyces sp.*

The third part of this contribution showed the processing and calibration of ultrasonic measuring equipment. Detection of relevant process parameters like sugar and ethanol concentrations during cultivation progress online is one of the major challenges in biotechnical applications. The independent detection of both solutes in aqueous solutions using one measuring device based on single physical background is not possible without assumptions so far. In brewing processes, the detection of extract content using sucrose or maltose as sugar equivalent with the aid of density is typically established. Since ethanol is influencing the density of respective solutions inversely to the sugar amount, the estimation of an "apparent extract" was used. This value resembles the mixture of major solutes, total sugar and ethanol content together.

It could be shown that regression on the target apparent extract is possible using the presented algorithms analyzing the ultrasonic signals in frequency domain. The calibration procedure was absolved using PLS regression reaching an error of <0.5 g/100 g by LOOCV. Although the variation in predicting the concentrations is still high, the advantages of this spectral analysis in combination with the presented sensor setup in service and maintenance due to completely contactless investigations of the fluid of interest is quite noteworthy. Furthermore, it is shown that using only ultrasonic characterization of such mixtures is enough to get insight into fluid properties without knowledge about progress of fermentation. However, this analysis can only be regarded as perspective, since influences of real fermentations like dissolved CO_2 are still not considered. Furthermore, the use of frequency spectra has to be regarded carefully, since it is, among others, influenced by noise caused by bubbles. This attenuation effect has to be studied by detailed bubble size analysis. Additionally, investigations in choosing the correct frequency band using sensitivity analysis as well as applying suitable variable selection methods for multivariate analysis have to be studied. Further, combining ultrasonic properties from time domain (TOF) with frequency domain ($P(f)$, Phase(f)) by means of multivariate statistics have to be studied in detail. Future aim will be the online prediction of concentrations correct to one decimal, which is known to be accurate enough in monitoring processes of brewing industry like cooking or fermentation. However, the stability and robustness of this approach in combination with setup optimization has to be the objective of further research.

One of the most critical points considering online process application is the influence of temperature. The study showed that the direct calibration of ultrasonic properties to offline measured apparent extract values at single temperatures is possible in a range of 6–22°C. Therefore, the connections between those independent models have to be evaluated.

Those presented results show a promising path to better process performance and understanding. The perspective of implementing the ultrasonic sensor online in such a system including an accurate correlation to the process-related parameters will increase the stability and robustness together with the flexible and adopted fuzzy controller to reach a high-quality end productct.

Nomenclature

A_i	[–]	constants of Blackman window function
B		vector of Regression parameters
b, c, T_{min}, T_{max}	[–], [–], [K], [K]	constants for f_{temp} – square root function
b_0	[g/100g]	first regression parameter
COA		center of area
C_X, C_S, C_E	[mmol/L]	concentration of biomass, substrate, ethanol
f_{temp}	[–]	temperature coefficient
K	[g/100 g]	constant for simulating Crabtree-effect
K_{ie}	[mmol/L]	Inhibition constant (ethanol)
K_S	[mmol/L]	half saturation constant (Substrate)
L_t		lag time function
M		maintenance
MAX		maximum operator
$M_{biomass}$	[g/mol]	molar weight of biomass
MIN		minimum operator
m_s	[g/cell]	specific dry weight of yeast cells
N	[–]	number of windows
N	[–]	number of samples
n_f	[–]	number of frequency pin
n_{Cells}, YCC	[mio. cells/mL]	yeast cell count
$P(f)$		magnitude spectrum
p, P		loading vector/matrix of X
q, Q		loading vector/matrix of Y
q_s, $q_{s,max}$		(maximum) specific substrate uptake
$\mathbf{S_x}$, $\mathbf{S_y}$		diagonal matrix with standard deviation
T	[°C, K]	temperature
$t(A_i)$		reflection of ultrasonic pulse
t, T		score vector/matrix of X
t_{lag}	[s]	lag time
tof	[s]	time of flight
u, U		score vector/matrix of Y
$W(n)$		window function
w, W		weighted loading vector/Matrix
X		predictor matrix
Y		target matrix
Y	[mol/mol]	yield coefficient
$\varphi(f)$		phase spectrum

Subscripts, Exponents

	"Predicted"
e11	Echo impulse
E	Ethanol
f	Fermentative
i	Counter
m	Mean
ox	Oxidative
r1	Buffer reflection
S	Substrate
X	Biomass

This work was partially funded by the "Bundesministerium für Wirtschaft und Technologie" (via AiF) through the research project KF-2039605-MD-9 and AIF-14790.

The authors have declared no conflict of interest.

5 References

[1] Becker, T., Krause, D., Softsensorsysteme–Mathematik als Bindeglied zum Prozessgeschehen. *Chem. Ing. Tech. Themenheft Prozessanalytik* 2010, *82*, 429–440.

[2] Food and Drug Administration, Guidance for industry: PAT – a framework for innovative pharmaceutical development, manufacturing and quality assurance 2004.

[3] O'Leary, R., Method of analysis for correcting dissolved CO2 content for specific gravity and alcohol variations in beer available (from: http://www.iul-instruments.de/pdf/vital sensors_2.pdf [cited 2009.19.03].

[4] Neue Hefereinzuchtanlage bei den Kölner Verbund Brauereien (KVB). *Brauwelt* 2008, 41–42, 1161.

[5] Henning, B., Rautenberg, J., Process monitoring using ultrasonic sensor systems. *Ultrasonics* 2006, *44*, e1395–e1399.

[6] Hauptmann, P., Hoppe, N., Püttmer, A., Application of ultrasonic sensors in the process industry. *Meas. Sci. Technol.* 2002, *13*, R73–R83.

[7] Bamberger, J. A., Greenwood, M. S., Measuring fluid and slurry density and solids concentration non-invasively. *Ultrasonics* 2004, *42*, 563–567.

[8] Bjørndal, E., *Acoustic Measurement of Liquid Density with Applications for Mass Measurement of Oil*, University of Bergen, Norway 2007.

[9] McClements, D. J., Fairly, P., Ultrasonic pulse echo reflectometer. *Ultrasonics* 1991, *29*, 58–62.

[10] Püttmer, A., Hauptmann, P., Henning, B., Ultrasonic density sensor for liquids. *IEEE 2Trans. Ultrason. Ferroelec. Freq. Contr.* 2000, *47*, 85–92.

[11] Contreras, N. I., Fairley, P., McClements, D. J., Povey, M. J. W., Analysis of the sugar content of fruit juices and drinks using ultrasonic velocity measurements. *Int. J. Food Sci. Technol.* 1992, *27*, 515–529.

[12] Kuo, F.-J., Sheng, C.-T., Ting, C.-H., Evaluation of ultrasonic propagation to measure sugar content and viscosity of reconstituted orange juice. *J. Food Eng.* 2008, *86*, 84–90.

[13] Resa, P., Elvira, L., Montero de Espinosa, F., González, R. et al., On-line ultrasonic velocity monitoring of alcoholic fermentation kinetics. *Bioprocess. Biosyst. Eng.* 2009, *32*, 321–331.

[14] Resa, P., Elvira, L., Montero de Espinosa, F., Gómez-Ullate, Y., Ultrasonic velocity in water-ethanol-sucrose mixtures during alcoholic fermentation. *Ultrasonics* 2005, *43*, 247–252.

[15] Cha, Y.-L., Hitzmann, B., Ultrasonic measurements and its evaluation for the monitoring of Saccharomyces cerevisiae cultivation. *Bioautomation* 2004, *1*, 16–29.

[16] Schäfer, R., *A contribution to extend the capabilities of ultrasonic process instrumentation*, in *Fakultät für Elektronik und Informationstechnik*, U. Theory, Editor. 2008, Otto-von-Guericke-Universität Magdeburg.

[17] Schäfer, R., Carlson, J. E., Hauptmann, P., Ultrasonic concentration measurement of aqueous solutions using PLS regression. *Ultrasonics* 2006, *44*, e947–e950.

[18] Henrion, R., Henrion, G., *Multivariate Datenanalyse*, Springer, Berlin 1995.

[19] Kessler, W., *Multivariate Datenanalyse*, Wiley, Weinheim 2007.

[20] Wold, H., Causal flows with latent variables. Partings of the ways in the light of NIPALS modelling. *Eur. Econ. Rev.* 1974, *5*, 67–86.

[21] Aynsley, M., Hofland, A., Morris, A., Montague, G. A., Di Masrimo, C., Artificial intelligence and the supervision of bioprocesses (real-time knowledge-based systems and neural networks). *Adv. Biochem. Eng. Biotechnol.* 1993, *48*, 1–27.

[22] James, S., Legge, R., Budman, H., Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. *J. Process Control* 2002, *12*, 113–121.

[23] Haake, C., Landgrebe, D., Scheper, T., Rhiel, M., Online-Infrarotspektroskopie in der Bioprozessanalytik. *Chem. Ing. Tech.* 2009, *81*, 1385–1396.

[24] Schuegerl, K., Progress in monitoring, modeling and control of bioprocesses during the last 20 years. *J. Biotechnol.* 2001, *85*, 149–173.

[25] Ulber, R., Frerichs, J.-G., Beutel, S., Optical sensor systems for bioprocess monitoring. *Anal. Bioanal. Chem.* 2003, *376*, 342.

[26] Zhang, H., Software sensors and their applications in bioprocess, in: Nicolletti, M. D. C., Jain, L. C. (Eds), *Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control*, Springer, Berlin, Heidelberg 2009, 25–56.

[27] Becker, T., Hitzmann, B., Muffler, K., Pörtner, R. et al., Future aspects of bioprocess monitoring. *Adv. Biochem. Eng. Biotechnol*. 2007, *105*, 249–293.

[28] Chen, L., Nguang, S., Li, X., Chen, X., Soft sensors for on-line biomass measurements. *Bioprocess. Biosyst. Eng.* 2004, *26*, 191–195.

[29] Junker, B. H., Wang, H. Y., Bioprocess monitoring and computer control: Key roots of the current PAT initiative. *Biotechnol. Bioeng.* 2006, *95*, 226–261.

[30] Nicoletti, M. C., Jain, L. C., Giordano, R. C., Computational intelligence techniques as tools for bioprocess modelling, optimization, supervision and control, in: Nicolletti, M. D. C., Jain, L. C. (Eds), *Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control*, Springer, Berlin Heidelberg 2009, 1–24.

[31] Simutis, R., Oliveira, R., Manikowski, M., Azevedo, S. F. d. et al., How to increase the performance of models for process optimization and control. *J. Biotechnol.* 1997, *59*, 73–89.

[32] Jones, K. D., Kompala, D. S., Cybernetic model of the growth dynamics of Saccharomyces cerevisiae in batch and continuous cultures. *J. Biotechnol.* 1999, *71*, 105–131.

[33] Lodolo, E. J., Kock, J. L., Axcell, B. C., Brooks, M., The yeast Saccharomyces cerevisiae–the main character in beer brewing. *FEMS Yeast Res.* 2008, *8*, 1018–1036.

[34] Papagianni, M., Boonpooh, Y., Mattey, M., Kristiansen, B., Substrate inhibition kinetics of Saccharomyces cerevisiae in fed-batch cultures operated at constant glucose and maltose concentration levels. *J. Ind. Microbiol. Biotechnol.* 2007, *34*, 301–309.

[35] Van Hoek, P., Van Dijken, J. P., Pronk, J. T., Effect of specific growth rate on fermentative capacity of Baker's yeast. *Appl. Environ. Microbiol.* 1998, *64*, 4226–4233.

[36] Kurz, T., Mathematically based management of Saccharomyces sp. batch propagations and fermentations, in: *Lehrstuhl für Fluidmechanik und Prozessautomation*. Muenchen, TU Muenchen 2002.

[37] Barford, J. P., Hall, R. J., An examination of the crabtree effect in *Saccharomyces cerevisiae*: the role of respiratory adaptation. *J. Gen. Microbiol.* 1979, *114*, 267–275.

[38] Postma, E., Verduyn, C., Scheffers, W. A., Van Dijken, J. P., Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of Saccharomyces cerevisiae. *Appl. Environ. Microbiol.* 1989, *55*, 468–477.

[39] Sonnleitner, B., Kaeppeli, O., Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. *Biotechnol. Bioeng.* 1986, *28*, 927–937.

[40] Zadeh, L. A., Fuzzy sets. *Information Control* 1965, *8*, 338.

[41] Herrera, E., Exact fuzzy observer for a baker's yeast fermentation process. *Int. IFAC Symp. Computer Appl. Biotechnol.* 2007, *1*, 309–314.

[42] Nyttle, V. G., Chidambaram, M., Fuzzy logic control of a fed-batch fermentor. *Bioprocess. Biosyst. Eng.* 1993, *9*, 115–118.

[43] Venkateswarlu, C., Gangiah, K., Fuzzy modeling and control of batch beer fermentation. *Chem. Eng. Commun.* 1995, *138*, 89–111.

[44] Venkateswarlu, C., Naidu, K. V. S., Dynamic fuzzy model based predictive controller for a biochemical reactor. *Bioprocess. Biosyst. Eng.* 2000, *23*, 113–120.

[45] Besli, N., Türker, M., Gul, E., Design and simulation of a fuzzy controller for fed-batch yeast fermentation. *Bioprocess. Biosyst. Eng.* 1995, *13*, 141–148.

[46] Mahjoub, M., Mosrati, R., Lamotte, M., Fonteix, C. et al., Fuzzy control of baker's yeast fed-batch bioprocess: A robustness study. *Food Res. Int.* 1994, *27*, 145–153.

[47] Miœkiewicz, T., Kasperski, A., A fuzzy logic controller to control nutrient dosage in a fed-batch baker's yeast process. *Biotechnol. Lett.* 2000, *22*, 1685–1691.

[48] Annemüller, G., Manger, H.-J., Lietz, P., *Die Hefe in der Brauerei Hefemanagement – Kulturhefe/Hefereinzucht – Hefepropagation im Brauprozess*, VLB, Berlin 2008.

[49] Birle, S., Fellner, M., Lehmann, J., Wening, H. et al., Yeast Propagation Manager (YPM). *Brauwelt Int.* 2010, *28*, 26–29.

[50] Birle, S., Entwicklung, Einführung und Optimierung eines fuzzy-gesteuerten Expertensystems zur vollautomatischen Regelung einer untergärigen Hefepropagation im großtechnischen Maßstab, in: *Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt, Lehrstuhl für Brau- und Getränketechnologie*, Technische Universität Muenchen, Muenchen 2010.

[51] Hoche, S., Hussein, W. B., Hussein, M. A., Becker, T., Time of flight prediction for fermentation process in-line application. *Eng. Life Sci.* 2011, doi: 10.1002/elsc.201000177.

[52] Mitzscherling, M., Prozeßanalyse des Maischens mittels statistischer Modellierung, in: *Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt*, Technische Universität, München 2004.

[53] Risvik, H. *Principal Component Analysis (PCA)* & *NIPALS algorithm*. 2007, available from: http://folk.uio.no/henninri/pca_module/pca_nipals.pdf [cited November 20, 2008].

[54] Krause, D., Schöck, T., Hussein, M. A., Becker, T., Ultrasonic characterization of aqueous solutions with varying sugar and ethanol content using multivariate regression methods. *J. Chemom.* 2011, doi: 10.1002/cem.1384.

[55] Martens, H., Næs, T., *Multivariate Calibration*, Wiley, New York 1991.

[56] Hoppe, G. K., Hansford, G. S., Ethanol inhibition of continuous anaerobic yeast growth. *Biotechnol. Lett.* 1982, *4*, 39–44.

[57] Barford, J. P., Hall, R. J., A mathematical model for the aerobic growth of Saccharomyces cerevisiae with a saturated respiratory capacity. *Biotechnol. Bioeng.* 1981, *23*, 1735–1762.

[58] Dunn, I. J., Heinzle, E., Ingham, J., Prenosil, J. E., *Biological Reaction Engineering – Dynamic Modelling Fundamentals with Simulation Examples*, Wiley, Weinheim–New York–Basel–Cambridge 2003.

[59] Mamdani, E., Advances in the linguistic synthesis of fuzzy controllers. *Int. J. Man-Machine Studies* 1976, *8*, 669–678.

[60] Botella, C., Ferré, J., Boqué, R., Outlier detection and ambiguity detection for microarray data in probabilistic discriminant partial least squares regression. *J. Chemom.* 2010, *24*, 434–443.

[61] MEBAK, *Mitteleuropäische Brautechnische Analysenkommission MEBAK e.V., Band III*, 1996.

2.2.2 Ultrasonic sensor for predicting sugar concentration using multivariate calibration

Ultrasonic sensor for predicting sugar concentration using multivariate calibration

D. Krause, W.B. Hussein, M.A. Hussein *, T. Becker

Center of Life and Food Sciences Weihenstephan, Group of Bio-Process Analysis, TU Muenchen, Weihenstephaner Steig 20, 85354 Freising, Germany

ARTICLE INFO

Article history:
Received 5 August 2013
Received in revised form 10 February 2014
Accepted 19 February 2014
Available online 11 March 2014

Keywords:
Ultrasound
Sugar concentration
Feature extraction
Multivariate data analysis
Partial least squares (PLS)

ABSTRACT

This paper presents a multivariate regression method for the prediction of maltose concentration in aqueous solutions. For this purpose, time and frequency domain of ultrasonic signals are analyzed. It is shown, that the prediction of concentration at different temperatures is possible by using several multivariate regression models for individual temperature points. Combining these models by a linear approximation of each coefficient over temperature results in a unified solution, which takes temperature effects into account. The benefit of the proposed method is the low processing time required for analyzing online signals as well as the non-invasive sensor setup which can be used in pipelines. Also the ultrasonic signal sections used in the presented investigation were extracted out of buffer reflections which remain primarily unaffected by bubble and particle interferences.

Model calibration was performed in order to investigate the feasibility of online monitoring in fermentation processes. The temperature range investigated was from 10 °C to 21 °C. This range fits to fermentation processes used in the brewing industry. This paper describes the processing of ultrasonic signals for regression, the model evaluation as well as the input variable selection. The statistical approach used for creating the final prediction solution was partial least squares (PLS) regression validated by cross validation. The overall minimum root mean squared error achieved was 0.64 g/100 g.

1. Introduction

In many biotechnological processes it is of great interest to detect the concentration of ingredients in the working medium. Further, monitoring and control of such industrial processes need reliable online measuring devices. Those sensor systems should have the ability to be robust, easy to use and non-invasive, even when it comes to food related applications. Typically, the sensor equipment used for analyzing fluid properties is invasive to the medium of interest [1–4]. Over the last few decades the importance of ultrasonic sensors for such applications have become more and more standard [5,6]. Several groups have studied the possibility of ultrasonic devices for measuring sugar concentration in various different ways [7,8]. Further, frequency and time domain representation of ultrasonic features is reported to contain information about the changes in density of respective fluids. In literature there are several possibilities presented which use indirect prediction of acoustic impedance via reflection coefficient combined with ultrasonic velocity estimated from time of flight measurement to finally calculate the density of the fluid [9–14]. However, relevant information in relation to changes in sugar concentration and therefore density changes should be visible in acoustic impedance. Extracting this feature based on the known physical relations with the presented setup is quite difficult. Nevertheless, one possibility to extract the feature impedance is based on the decay of temporal echo amplitudes in the buffer material [10]. Other possibilities presented in literature are based on the frequency domain. Further, influences like superposition as well as signal resolution do have a high impact on those approaches. Despite all that, literature reports difficulties in predicting impedance accurately under the given circumstances [15,16]. In this work, the information was gathered using several acoustic features calculated on time and frequency domain representation. These were extracted on buffer reflections of ultrasonic signals. Therefore, influences caused by i.e. gas bubbles can be avoided by only analyzing the signal reflection. Although those signal parts do not penetrate the medium, they still carry medium information in the reflection coefficient. Generally, the chosen features individually capture the information included in each signal. This results in a new feature based multivariate representation covering signal attenuation as well.

* Corresponding author. Tel.: +49 (0) 8161 71 3277; fax: +49 (0) 8161 71 3883.
E-mail address: hussein@wzw.tum.de (M.A. Hussein).

http://dx.doi.org/10.1016/j.ultras.2014.02.019
0041-624X/

Nomenclature

a	regression parameters for temperature dependence (−)
$\boldsymbol{B}$, $\boldsymbol{b}$	matrix/vector of regression parameters (−)
b_0	first regression parameter (g/100g)
BW	bandwidth (−)
cen	centroid (−)
cf	crest factor (−)
EDFT	extended discrete Fourier transform
eng	energy (−)
ent	entropy (−)
f_s	sampling rate (Hz)
h	leverage of sample (−)
k	number of iterations extracting the latent vectors
kur	kurtosis (−)
m	number of variables (−)
mag	magnitude (−)
MLR	multiple linear regression
n	number of samples (−)
N	number of data points in ultrasonic signal sequence (−)
NIPALS	nonlinear iterative partial least squares
$\boldsymbol{p}$, $\boldsymbol{P}$	loading vector/matrix of X
PMMA	poly(methyl methacrylate)
PVDF	polyvinylidene flouride
PC	principal component
PCA	principal component analysis
PLS(R)	partial least squares (regression)
$\boldsymbol{q}$, $\boldsymbol{Q}$	loading vector/matrix of Y
RMSECV	root mean squared error of cross-validation (g/100g)
$\boldsymbol{S}$	diagonal matrix with standard deviation
s^2	sample variance
ske	skewness (−)
spr	spread (−)
T	temperature (°C K)
$t(A_i)$	reflection of ultrasonic pulse
$\boldsymbol{t}$, $\boldsymbol{T}$	score vector/matrix of $\boldsymbol{X}$
$\boldsymbol{u}$, $\boldsymbol{U}$	score vector/matrix of $\boldsymbol{Y}$
VIP	variable importance in the projection
$\boldsymbol{w}$, $\boldsymbol{W}$	weighted loading vector/matrix
$\boldsymbol{X}$	predictor matrix
$\boldsymbol{Y}$	target matrix

Subscripts, exponents

^	"predicted"
a,i,j	counter
s	scaled
s	spectral
t	temporal

Analyzing material properties as control variables for industrial applications based on physical modelling is not always possible due to the lack of knowledge. This leads to the use of multivariate statistics. These methods have been used for years in several fields when dealing with large volumes of data [17]. For evaluating data based on its statistical variance, multivariate regression models, such as principal component regression (PCR) or PLS, are used to handle the large amount of mostly collinear variables. These methods are used for correlation of target values ($\boldsymbol{Y}$) with direct measurable descriptor variables ($\boldsymbol{X}$). The background of such approaches is multiple linear regression (MLR). In this contribution partial least squares regression (PLSR) is used for modelling the variations in ultrasonic signals transmitted through aqueous solutions with varying maltose concentrations. This method uses a reduced number of latent variables compared to the descriptor variables found by cross correlation of variance in $\boldsymbol{Y}$ and $\boldsymbol{X}$.

It was shown earlier, that changing concentration of a dissolved substance causes changes in the fluid properties such as density and bulk modulus. This directly influences the properties of the ultrasonic waves travelling through the fluid [18]. It was reported, that features of ultrasonic signals like reflection coefficient is frequency dependent [18]. Developing a physical relationship between pulse distortion and the fluid properties is quite complicated due to the complexity of the system. Because of this the usage of multivariate data analysis is of benefit in cases of fast modelling of the phenomena of interest [18]. With the aid of multivariate analysis, it is possible to extract the most dominant information with respect to density. At the same time, the noise caused by arbitrary influences such as temperature inaccuracies, superposition, and bubble induced distortion will be discarded. This is the goal of the presented study. It was shown earlier in a similar approach using PLS, that it is possible to predict substance concentrations using ultrasonic signal features [18].

This study presents a system, which is fully non-invasive. Further, it is less dependent on influences caused by bubble interferences or particles suspended in the medium of interest. A signal with corresponding wavelength travelling through the fluid could lose energy caused by scattering or dissipation at these bubbles or particles [19]. Up till now, the calibration covers a defined temperature range and is capable of detecting sugar concentration. The method is based on feature extraction of ultrasonic pulses and multivariate regression based on PLS. Up to the author's knowledge, the combination of both parts is new in the field of ultrasonic measurements. Finally, the method is simple and easy to implement. It is also possible, to extend the detection on ternary mixtures containing ethanol as well. Benefit of this extension would be the possibility to use just on sensor device as well as uncoupling the detection from any relation to the process behavior like in Resa et al. or Cha and Hitzmann [20–22].

2. Materials and methods

2.1. Experimental

The experimental setup as well as the measuring principle is explained schematically in Fig. 1 including photography of the setup (middle). The shown container is indirectly heated over a tempering mantle which is supplied with tempering fluid by an external thermostat. To investigate maltose-water mixtures the container is first filled with the solution of interest. Additionally, the solution is heated/cooled continuously as well as permanently mixed by a magnetic stirrer to reach a temperature distribution as homogeneous as possible. The ultrasonic signals are recorded over a temperature range from 10 up to 21 °C at constant container pressure over a time frame of around 3 h. The temperature points used in this study for calibration were extracted in steps of 0.5 K out of this continuous spectrum.

The in-house produced piezo electric transducer (built using a piezo ceramic with a center frequency of 2 MHz) is used for creating an ultrasonic pulse by excitation with a rectangular electrical pulse (width of 200 ns, amplitude of 5 V). After passing container wall (wall 1) and fluid the pulse is reflected at the backside (wall 2) and caught by the transducer which works as a receiver in the

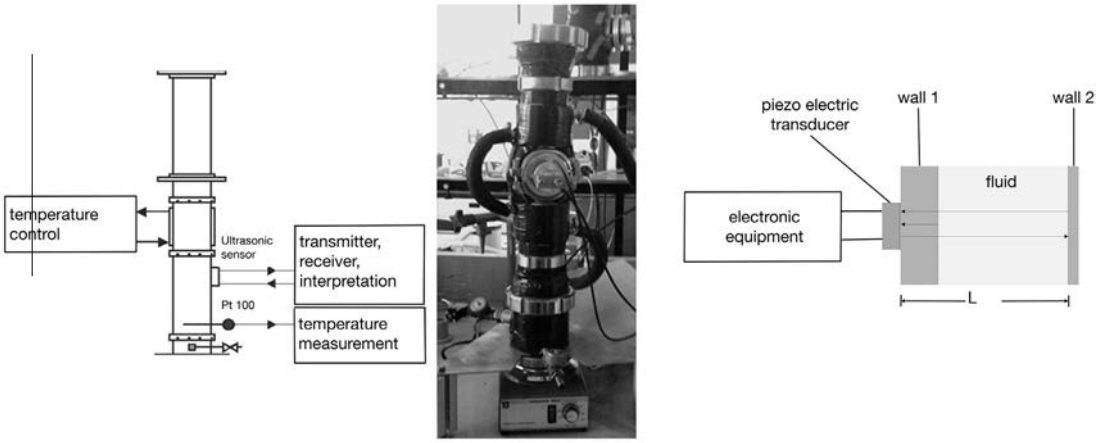


Fig. 1. On the left hand side a schematic drawing of the experimental setup is shown; it includes a tempering mantle provided with tempering fluid by an external thermostat, the piezo electric transducer (coupled to a microcontroller) and a Pt 100 resistance thermometer for monitoring the temperature inside the chamber; further, there is an inlet for filling the chamber with fluid; the same setup is shown in the photography (middle); the right hand side shows the schematic measuring principle (pulse echo method) with piezo electric transducer, excited via the electronic equipment; after passing the first wall (steel) the pulse travels through the sample and is reflected at the backside (wall 2); transducer works as receiver at the same time, signal is collected via the electronic equipment.

same time (pulse-echo method). The signal is recorded via microcontroller connected to the measuring device. The temperature of the probe fluid is measured by Pt 100 temperature sensors (maximum accuracy of ±0.1 °C). The experiments were carried out for each concentration separately. Several mixtures of maltose-water in a range of 2–12% (per weight) for measuring the ultrasonic pulse behavior were prepared. Homogenization was reached with a magnetic stir bar. For each investigation a sample volume of 3 L was prepared dissolving a known mass of crystalline maltose (D(+)-Maltose Monohydrate, Roth®) using a Sartorius Laboratory® weight (L 2200 S) in distilled water reaching a defined final weight. The reference concentration for control was measured using an Anton Paar® Density Meter (DMA 4500). Signals for analysis were extracted from the total set of continuous measured temperature range resulting always in data sets of around 70 objects with varying concentration at individual temperature points.

2.2. Temperature influence

In the presented work individual temperature independent multivariate relations were calculated. Therefore, signals for analysis were extracted from the total set of continuous measured temperature spectrum resulting always in data sets of around 70 objects with varying concentrations for each temperature point. Finally, each regression parameter of individual temperature model was plotted against temperature to gather the influence (see Section 3, Fig. 6C). The overall model was built by polynomial regression of each regression parameter. This relation was developed using first order polynomial regression due to lowest overall residual error (see Section 3, Fig. 6A and B).

2.3. Signal processing

Ultrasonic signals collected using the presented setup were influenced by noise and superposition phenomena. Therefore, data was preprocessed averaging around 10 signals of approximately equal conditions. Further, most relevant reflections created in the steel buffer of the presented sensor system were extracted.

The transformation of time domain representation into frequency domain was accomplished using extended discrete Fourier transform (EDFT) and the used signal parts were extracted from buffer reflections of the used setup. Since they are influenced by superposition phenomena, the region of interest was limited to a comparably small time frame. Due to limited sampling frequency, the data representation can be seen as incomplete resulting in numerical frames of ~200 data points. Literature reports the mentioned extended discrete Fourier transform algorithm (EDFT) for incomplete data. Additionally, a higher frequency resolution can be reached [23,24], which was necessary in extracting meaningful frequency spectra for spectral feature extraction. This algorithm is able to extrapolate input sequence to a defined length N. In contrast to DFT, an increase in the frequency resolution up to $1/(N * f_s)$ can be achieved (where f_s resembles the sampling rate) [23].

Since the frequency domain of a signal is sensitive to noise [18] the interval of interest for frequency analysis has to be adapted. To locate this interval, the signals were processed with window functions. One benefit of windowing is the reduction of frequency magnitude in regions of non-interest (just containing random noise) to a mean of zero in a statistical manner. Therefore, uninformative signal parts will be weighted down. To choose an adequate method according to the needs, a sinusoidal signal including several frequencies was analyzed using different windows. Those windows were rectangular, Hamming, Hanning, flat top, Blackman, Kaiser, Bartlett and Gaussian window function (Fig. 2).

Minimal spectral leakage is an important criteria in the window choice. Thus, rectangular, flat top and Kaiser window can be excluded comparing the results presented in Fig. 2. Further, Bartlett and Gaussian window were excluded taking the accuracy of magnitude calculation into account. The results of this investigation showed, that the differences between Hanning, Hamming and Blackman window function were neglectable. Therefore, Blackman window function was applied to investigate the representative bandwidth of the US-signals in this work. Each signal was analyzed window-wise to cover the main magnitude changes over the whole frequency band like reported in Krause et al. [25]. It is shown in Fig. 3, that the highest variations in the signals take place in a bandwidth of 0.5–3.5 MHz. This bandwidth was taken for feature extraction. The final feature analysis was carried out on the unwindowed signal. The frequency region of interest was set to the predefined bandwidth from window investigation.

2.4. Feature extraction

In this work 12 features are suggested for further analysis. Those features are summarized in Table A1 (Appendix) at the end of this contribution. They are taken as model input for multivariate regres-

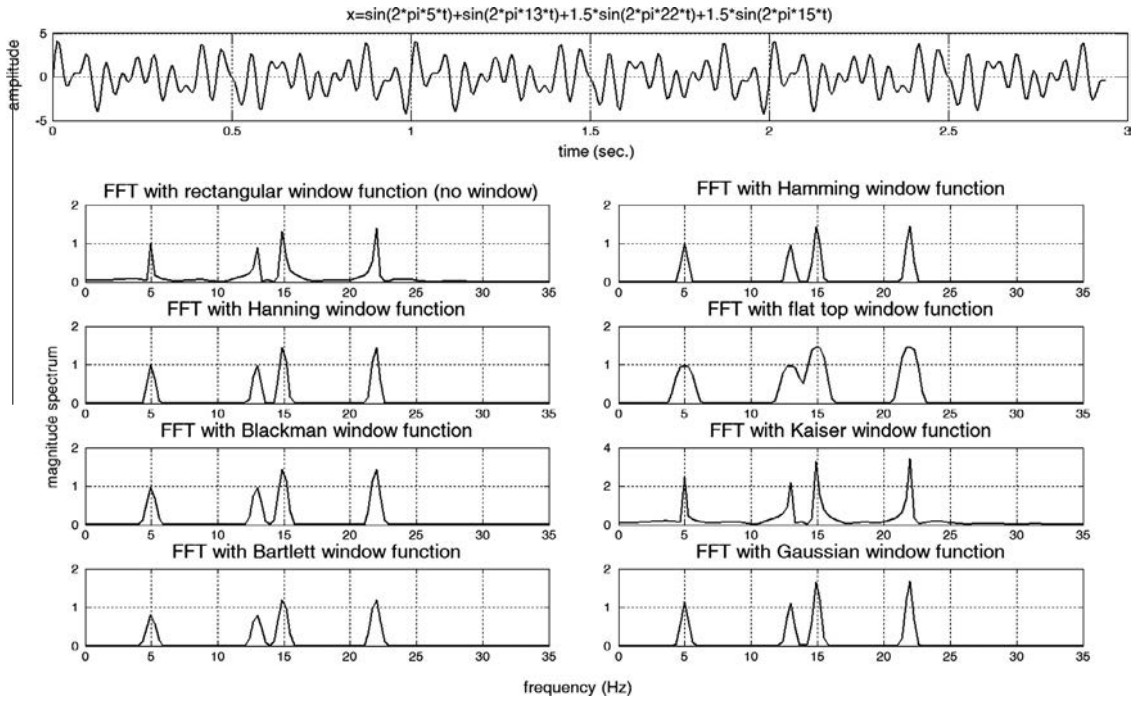


Fig. 2. Investigation of spectral leakage in different window functions applied to a sinusoidal signal including different frequencies.

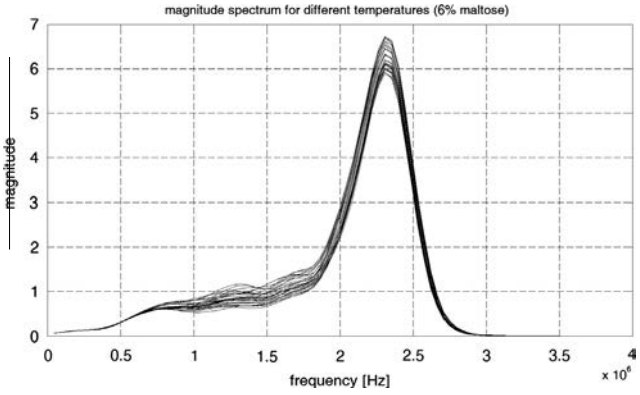


Fig. 3. Magnitude spectrum of raw ultrasonic signals analyzed with Blackman window method.

sion. The used features were chosen by their correlation index in a statistical manner. Furthermore, already published contributions investigating those mentioned acoustic features for signal analysis showed the general applicability [26–28]. Sensitivity analysis on features in several bioprocesses proved dependencies in various applications. As already mentioned above, the chosen features capture the information included in each signal in a multivariate sense individually.

2.5. Multivariate regression

Data used for statistical analysis requires statistical pre-processing. The next section explains the organization of matrices for regression analysis. The target values (concentration of maltose [g/100 g]) are stored in a column vector. The corresponding signal properties (mag, BW, cf_s etc.; all at a certain temperature) are stored in rows of a predictor matrix $\boldsymbol{X}$ (Eq. (1)).

$$\boldsymbol{X} = [\mathrm{mag}, \mathrm{BW}, \mathrm{kur}_s, \mathrm{ske}_s, \mathrm{eng}_t, \mathrm{eng}_s, \mathrm{ent}_t, \mathrm{ent}_s, \mathrm{cf}_s, \mathrm{cen}_s, \mathrm{spr}_s, \mathrm{cf}_t] \quad (1)$$

Previous to decomposition the matrices were autoscaled by centering each column to its mean value and scaling to unit variance dividing by its standard deviation. This is necessary to exclude influences due to absolute values of single columns.

2.6. Theory of partial least squares

The calculations on each data set (~70 signals) were programmed and carried out applying the most commonly used nonlinear iterative partial least squares (NIPALS) Algorithm developed by Wold [17,29–32]. This algorithm is calculating the PLS components iteratively. The used algorithm for PLS calculating k components can be formally described as

For $i = 1, 2, \ldots, k$

1. $\boldsymbol{w}_i = (\boldsymbol{X}_{i-1}^T \cdot \boldsymbol{u}_{i-1})/||(\boldsymbol{X}_{i-1}^T \cdot \boldsymbol{u}_{i-1})||$

whereas $\boldsymbol{u}_{i-1}$ in the first iteration will be equal to the column of $\boldsymbol{Y}$ with highest magnitude. The calculated vector $\boldsymbol{w}$ is called loading weight vector of $\boldsymbol{Y}$ and $\boldsymbol{X}$ obtained as the components of the cross-covariance between targets and predictors. The score vector $\boldsymbol{t}$ and the corresponding loading vector $\boldsymbol{q}$ are calculated as follows:

1. $\boldsymbol{t}_i = \boldsymbol{X}_{i-1} \cdot \boldsymbol{W}_i$
2. $\boldsymbol{q}_i = (\boldsymbol{Y}_{i-1}^T \cdot \boldsymbol{t}_i)/(\boldsymbol{t}_i^T \cdot \boldsymbol{t}_i)$

The vector $\boldsymbol{u}$ in the first iteration is chosen as mentioned above. In the following iterations it is calculated via

3. $\boldsymbol{u}_i = (\boldsymbol{Y}_{i-1} \cdot \boldsymbol{q}_i)/(\boldsymbol{q}_i^T \cdot \boldsymbol{q}_i)$

Further, the loading vector $\boldsymbol{p}$ is calculated as follows:

4. $\boldsymbol{p}_i = (\boldsymbol{X}_{i-1}^T \cdot \boldsymbol{t}_i)/(\boldsymbol{t}_i^T \cdot \boldsymbol{t}_i)$

Finally, the "working matrices" for the next iteration are calculated as

5. $\boldsymbol{X}_i = \boldsymbol{X}_{i-1} - \mathbf{t}_i \cdot \boldsymbol{p}_i^T$

6. $\boldsymbol{Y}_i = \boldsymbol{Y}_{i-1} - \boldsymbol{t}_i \cdot \boldsymbol{q}_i^T$

The vectors $\boldsymbol{p}_i$, $\boldsymbol{q}_i$, $\boldsymbol{u}_i$, and $\boldsymbol{t}_i$ are stored as columns in corresponding matrices shown in Eq. (2).

$$\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_k] \tag{2}$$

2.7. Calculation of regression coefficients

After iteration is finished, the principal components are used to calculate the parameters in $\boldsymbol{B}$ of the regression model (Eq. (3)).

$$\hat{\boldsymbol{Y}} = \boldsymbol{X} \cdot \boldsymbol{B} + \boldsymbol{1} \cdot \boldsymbol{b}_0 \tag{3}$$

Vector $\boldsymbol{b}_0$ is estimated as shown in Eq. (4).

$$\boldsymbol{b}_0 = \bar{\boldsymbol{y}} - \bar{\boldsymbol{x}} \cdot \boldsymbol{B} \tag{4}$$

whereas the vectors $\boldsymbol{y}$ and $\boldsymbol{x}$ contain the mean values of the corresponding columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Matrix $\boldsymbol{B}$ is estimated following Eq. (5).

$$\boldsymbol{B} = \boldsymbol{S}_X^{-1} \cdot [\boldsymbol{W}(\boldsymbol{P} \cdot \boldsymbol{W})^{-1} \cdot \boldsymbol{q}^T] \cdot \boldsymbol{S}_Y \tag{5}$$

whereas the diagonal matrices $\boldsymbol{S}$ include the standard deviation of the corresponding columns of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Further details can be found amongst others in Krause et al. [33].

2.8. Estimation of model size and accuracy

Calculation of regression coefficients (matrix $\boldsymbol{B}$) is carried out following Eqs. (4) and (5), respectively. Choosing the most reliable model order (number of PLS components taken for estimation of the parameter matrix $\boldsymbol{B}$) causes the most problems in terms of accuracy and stability of the calculated regression model. One possible criterion is to choose the model size by the minimum predictive error (for example root mean square error [RMSE]; the formal description of RMSE is shown in Eq. (6) determined over the most commonly used cross-validation. In this contribution the root mean square error of cross-validation is used for model order prediction (RMSECV).

$$\mathrm{RMSE} = \sqrt{\sum_{i-1}^{n} (y_i - \hat{y}_i)^2 \Big/ n} \tag{6}$$

The variables used in Eq. (6) are n for the number of samples, y for the reference value and $\hat{y}$ for the corresponding predicted value.

In this work, the combination of individual temperature independent regression models are combined by polynomial fit of regression parameters to temperature. Therefore, it is assumed that RMSECV tends to overfitting with respect to the final model structure. Thus, the measure used for decision on the number of components was the prediction error of the final model (including temperature dependence) on a validation dataset (RMSEV) covering the whole range of temperatures and concentrations (data not included in the calibration).

2.9. Variable importance/feature selection

In multivariate regression it is of great interest to extract most relevant information for robustness reasons. It is also necessary to decide, whether an input variable is important to the model or not. Moreover, it is beneficial to exclude non-informative variables from the input to prevent influences of those due to noisy or defective data. Extracting informative variables in multivariate regression using PLS can be done analyzing regression coefficients combined with variable importance in the projection (VIP) [34–36]. Therefore, the "autoscaled" regression coefficients (Eq. (7)) and the VIP (Eq. (8)) were calculated for each model and variable, respectively.

$$\boldsymbol{B}_s = \boldsymbol{W}(\boldsymbol{P} \cdot \boldsymbol{W})^{-1} \cdot \boldsymbol{Q}^T \tag{7}$$

$$VIP_j = \sqrt{n \sum_{a=1}^{A} (\boldsymbol{Q}_a^2 \boldsymbol{t}_a^T \boldsymbol{t}_a (w_{ja} / ||\boldsymbol{w}_a||)^2 \Big/ \sum_{a=1}^{A} \boldsymbol{Q}_a^2 \boldsymbol{t}_a^T \boldsymbol{t}_a} \tag{8}$$

Here, index a stands for the number of used latent vectors, j for the respective variable. All the presented methods reaching to the final regression model are collected in a final scheme for better understanding (Fig. 4).

3. Results and discussion

To find the best solution for detection of unknown concentrations, several iterations and processing steps had to be accomplished. First, the signals were averaged and analyzed to detect the correct frequency band from 0.5 to 3.5 MHz for all signals to allow further calculations. Second, buffer reflections were determined and transformed to frequency domain for extracting acoustic features. After autoscaling the dataset was statistically analyzed for outliers (presented later in the results section). Additionally, variable selection methods were applied to reduce the

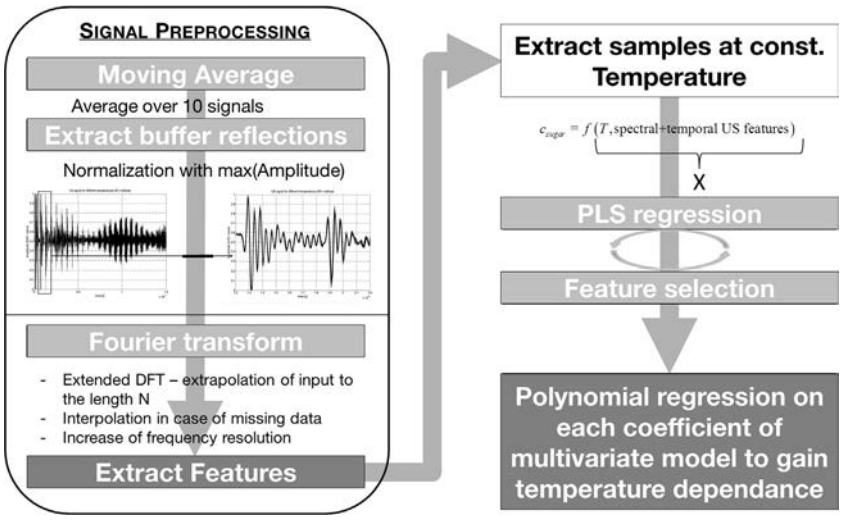


Fig. 4. Schematic summary of methods used for pre-processing of ultrasonic signals of the dataset for model investigation.

input matrix to the most relevant variables. Finally, individual isothermal regression models were calculated using PLS regression. The whole investigation presented covers a range from 2 to 12 g/100 g maltose diluted in distilled water and a temperature spectrum ranging from 10 to 21 °C, where the datasets of around 70 samples at each temperature point were analyzed using cross-validation. For each sample set models with increasing number of PLS components and corresponding errors are calculated. Furthermore, each individual model was analyzed for importance of input variables using variable importance in the projection (**VIP**) in combination with scaled regression coefficients ($\boldsymbol{B}_s$) to select the most relevant features for calibration. There for each variable selection vector was normalized individually and the resulting numerical values were used as gray scale color representation (Fig. 5). To increase the accuracy of the model, features were excluded according to their lack of importance indicated by the importance measures ($\boldsymbol{B}_s$ and **VIP** (Fig. 5)) and model evaluation was repeated. With the aid of the presented importance calculation, it was possible to exclude 4 out of the 12 presented features. Those remaining features are spectral centroid, energy, bandwidth, crest factor and magnitude as well as temporal crest factor, entropy and energy. The excluded features were spectral kurtosis, skewness, entropy and spread. These four features indicate low sensitivity in correlation to maltose concentration as well as to temperature (see Table A1). This highlights the strength of those importance measures used in this contribution for data sorting.

The final regression structure implying the temperature dependent coefficients was tested using polynomial regression of first and second order on the individual regression coefficients. The overall RMSEV was found to be lower using first order polynomial approaches (Fig. 6).

Finally, the overall regression model structure is given by Eq. (9):

$$\hat{\boldsymbol{Y}} = \boldsymbol{X} \cdot \boldsymbol{b} \tag{9}$$

Whereas regression vector $\boldsymbol{b}$ is structured as follows:

$$\boldsymbol{b} = [b_0, b_{mag}, b_{BW}, \ldots, b_{cf_t}]^T \tag{10}$$

The temperature influence of each regression coefficient in vector $\boldsymbol{b}$ is calculated as shown in Eq. (11).

$$\boldsymbol{b} = \boldsymbol{a}_{0,T} + \boldsymbol{a}_{1,T} \cdot T \tag{11}$$

with

$$\boldsymbol{a}_{0,T} = [a_{0,0}, a_{0,mag}, a_{0,BW}, \ldots, a_{0,cf_t}]^T$$

$$\boldsymbol{a}_{1,T} = [a_{1,0}, a_{1,mag}, a_{1,BW}, \ldots, a_{1,cf_t}]^T$$

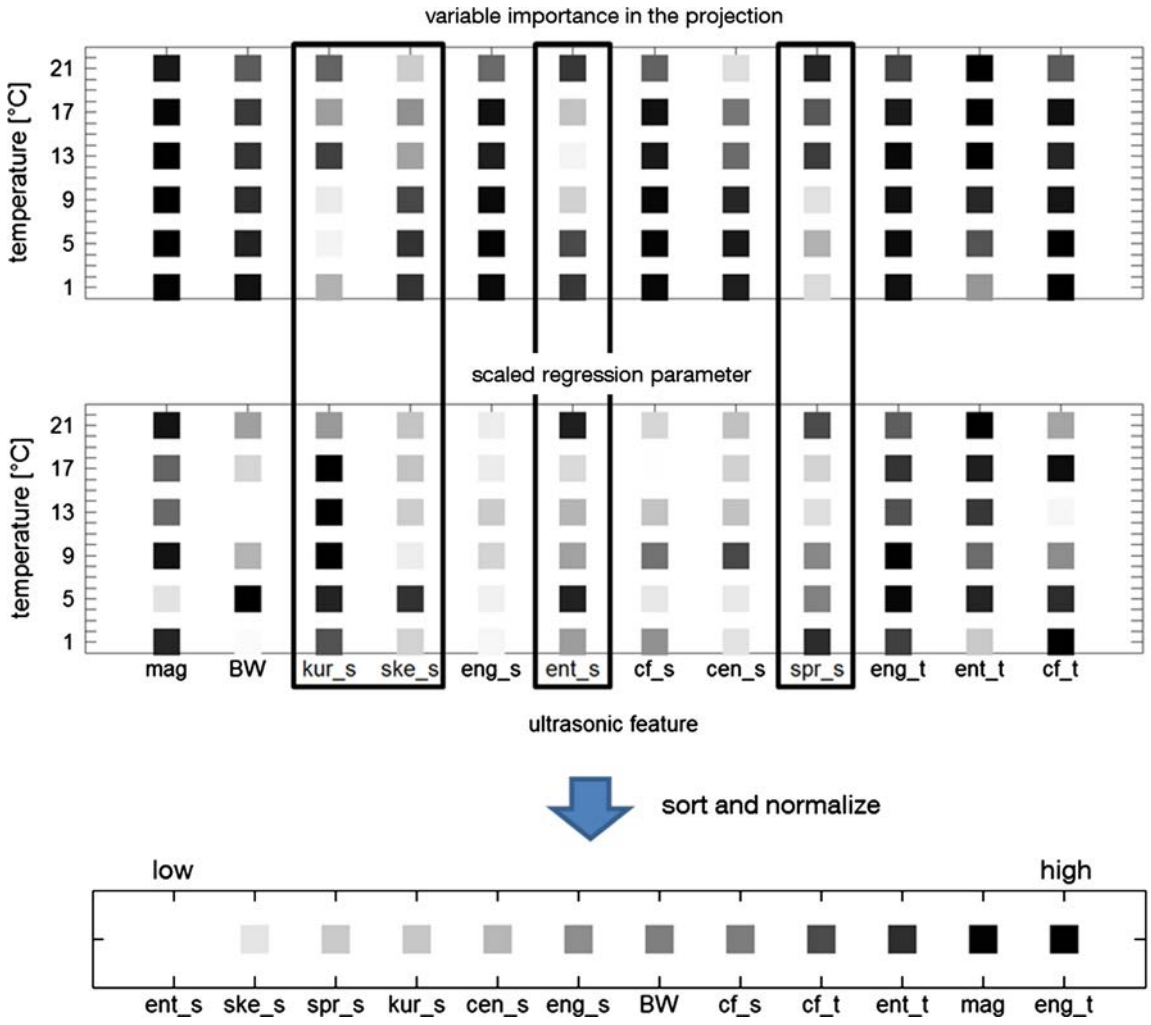


Fig. 5. Color map (gray scale) presenting the importance of each variable (feature) in individual multivariate temperature model; for clarity reasons only values each 4th temperature model are shown; the single line plot at the bottom represents the summed and normalized importance measure of both presented methods over the whole temperature profile investigated.

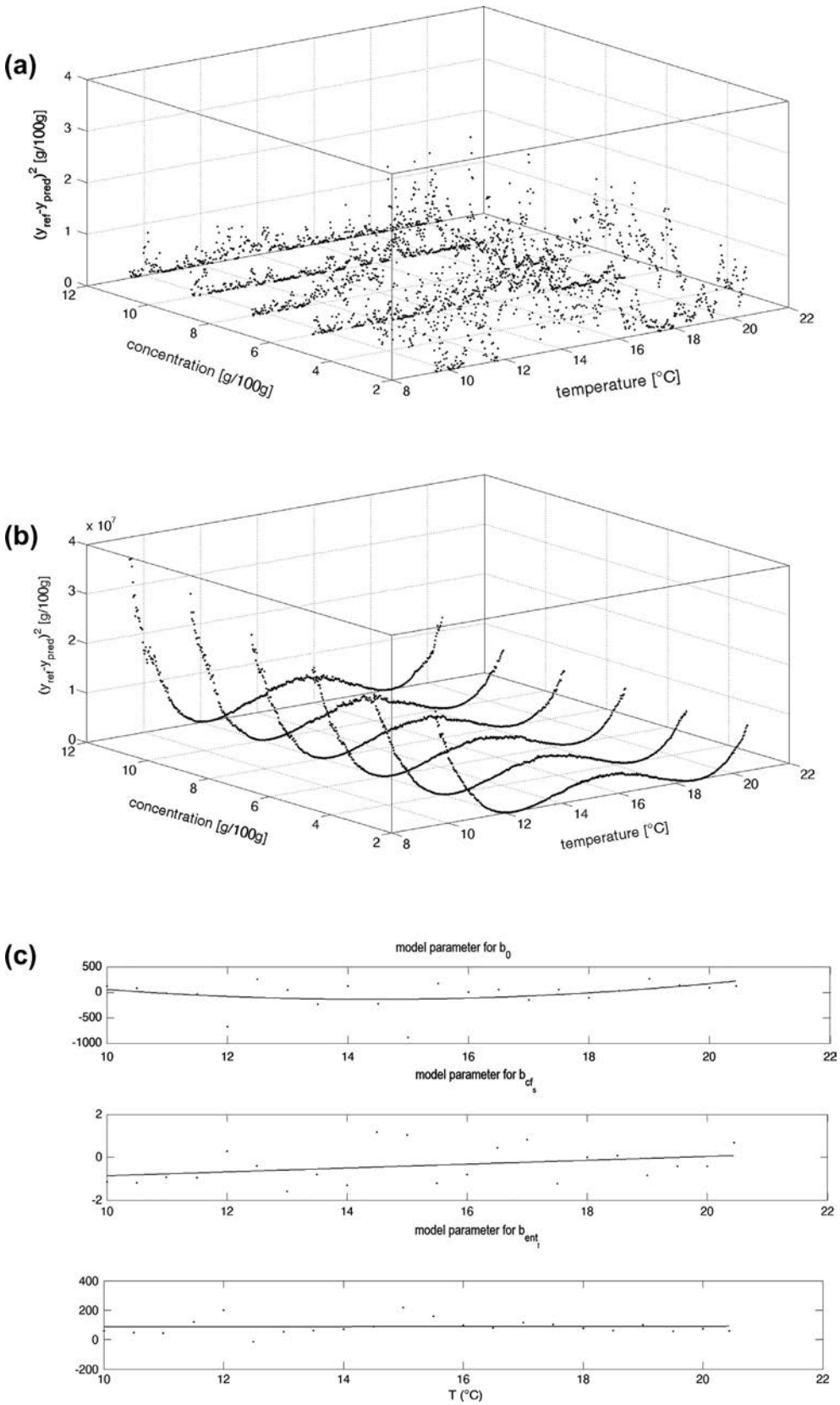


Fig. 6. Comparison between 1st (A) and 2nd (B) order polynomial fits used for temperature dependent regression coefficients of individual multivariate models. Although behavior of some parameters do indicate nonlinear temperature dependency (C), the overall RMSE achieved with only 1st order polynomial fits showed best result.

The variable T (1×1) resembles the temperature value of the respective signal.

Decisions on the correct number of components used in the final overall solution, including polynomial regression on temperature dependence, were made by averaging the prediction errors for an external dataset (overall RMSEV). The minimum prediction error of 0.64 g/100 g was achieved by using 5 PLS components and 8 features as inputs in each individual regression model. The decrease in the overall RMSE (calculated on ~2900 signals with varying temperature after combining the individual regression models) is presented in Fig. 7 and Table 1. Although the reported error is still quite high, the detection of different concentrations at varying temperature was possible (see parity plot Fig. 8).

This figure indicates that the established solution lacks accuracy in range of lower concentrations for two reasons: Firstly, there appears to be a higher sensitivity to information ratio for lower sugar concentrations, and secondly a possible non-linearity of the features used in the presented concentration range. The plot of the smoothed histogram indicates skewed distribution for prediction at 2 g/100 g, whereas the other histograms are almost normally distributed. This further supports the points mentioned above.

Nevertheless, predicting concentrations in g/100 g which are correct to one decimal place is known to be accurate enough for monitoring processes online in brewing industry such as cooking or fermentation. However, the calculated validation error of 0.64 g/100 g on laboratory data is still higher than a theoretically acceptable absolute deviation for online monitoring of +/−0.5 g/100 g.

Statistical outlier detection was used to find and reduce the number of signals containing noise, which is caused by several unknown influences. This strategy was used to enhance robustness of the solution as well as achieving an acceptable prediction error. In this work an approach describing the influences of each signal on the calibration is used to exclude outliers from the dataset. Therefore, the leverage (Eq. (12)) as well as the residual $\boldsymbol{Y}$-variance (Eq. (13)) of each sample was calculated [17,37].

$$H_i = \frac{1}{N}\sum_{a=1}^{A}(t_{ia}^2/\boldsymbol{t}_a^T\boldsymbol{t}_a) \tag{12}$$

$$S_{Ri}^2 = (y_i - \hat{y}_i)^2 \tag{13}$$

Samples with higher leverage than 3 times the mean of h (possibly representing samples with unusual x data [37]) and additionally higher residual variance than 3 times the mean of s_R^2 were excluded in each dataset. This caused a decrease of the inner model prediction error (RMSE of calibration, Table 2).

Table 1
Comparison between number of features and the number of PLS components chosen for the models.

PLS components Input variables	1	2	3	4	5	6
Full	1.02	0.83	0.77	0.74	0.69	0.66
Minus ent_s	0.98	0.83	0.71	0.66	0.65	0.66
& Minus ske_s	0.92	0.83	0.73	0.67	0.66	0.65
& Minus spr_s	0.99	0.83	0.71	0.67	0.65	0.68
& Minus kur_s	1.07	0.80	0.69	0.65	0.64	0.66
& Minus cen_s	1.02	0.76	0.66	0.65	0.68	0.67

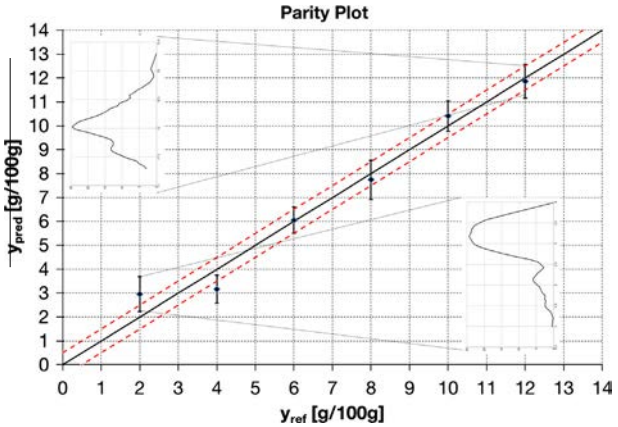


Fig. 8. Parity plot – estimated vs. reference concentration of ~2900 samples with varying temperature covering the whole temperature range investigated. The continuous line resembles $y_{pred} = y_{ref}$. The plotted error bars resemble a $2 * \sigma$ deviation (~95% confidence interval to predicted mean); the dashed line resembles a +/−0.5 g/100 g absolute deviation; little figure: smoothed histogram of predicted concentration at 2 g/100 g.

Nevertheless, overall RMSE including temperature dependence of each regression coefficient resulted in a lower value when using full dataset in individual regression models. Therefore, further investigations to detect possible outliers such as distorted signals using density based approaches, such as the local outlier factor (LOF) [38], will be part of future analysis.

The possibility of applying this non-invasive sensor system online to a process will be one of the biggest benefits as compared to existing measuring systems in industry. Until now most online solutions are based on invasive setup designs or bypass solutions

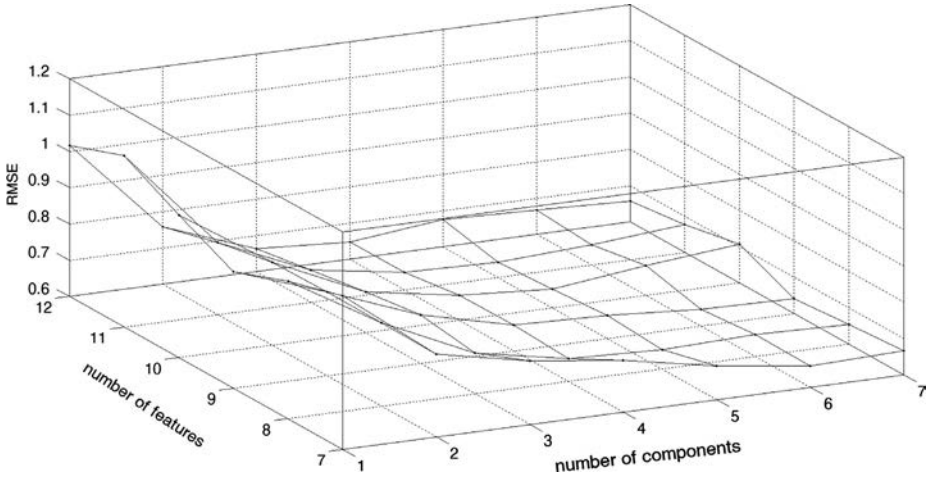


Fig. 7. Comparison between number of features (decreasing due to the variable importance order) and the number of PLS components chosen for the models; minimum prediction error (RMSE) of ~0.64 g/100 g achieved using 8 out of 12 features.

Table 2
Comparison taking all datasets (including 12 ultrasonic features) and those excluding the outliers indicated by the outlier detection algorithm.

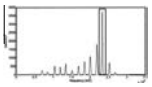
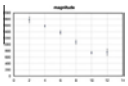
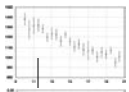
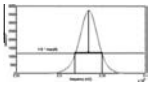
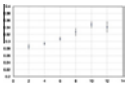
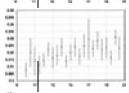
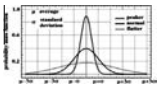
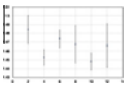
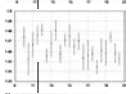
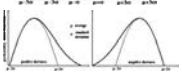
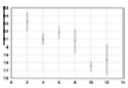
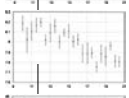
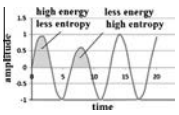
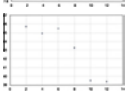
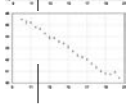
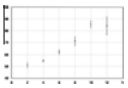
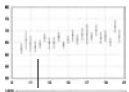
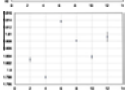
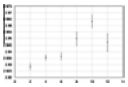
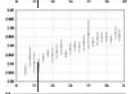
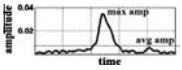
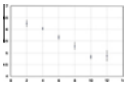
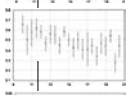
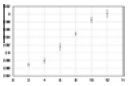
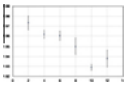
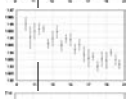
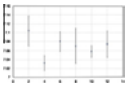
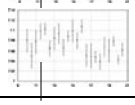
PLS components	All data points		w/o Outliers	
	Mean RMSE	Overall RMSE	Mean RMSE	Overall RMSE
3	0.63	0.78	0.59	0.79
4	0.58	0.74	0.53	0.75
5	0.54	0.7	0.47	0.74
6	0.5	0.66	0.42	0.67
7	0.48	0.67	0.39	0.68

to detect relevant concentrations, which are relatively complicated from a service, maintenance and economic point of view. Other possibilities are offline systems, which include a high effort in preparing samples for measurement. The used setup together with the presented multivariate data analysis presents a possibility to predict maltose concentration by using only reflected signal parts which did not penetrate the medium of interest. Therefore, influencing effects such as gas bubbles and particles can be neglected. Additionally, the presented multivariate method for feature selection showed its strength towards choosing the necessary input variables by their sensitivity towards the substrate concentration. Nevertheless, further research will be applied to reach an even higher accuracy of ultrasonic signals by investigating different setup materials, near field and superposition phenomena as well as electronic circuit adaptions. Additionally, the system should be implemented in a process to investigate dynamic process influences online. Finally, calibration needs to be extended to include influences due to fermentation products such as alcohol.

4. Conclusions

In this contribution it is shown, that prediction of maltose dissolved in water at different temperatures using features from time and frequency domain of ultrasonic pulses combined with statistical modelling as regression tool is possible. The method used for choosing model size (number of principle components) was cross validation. The overall RMSE value for prediction was 0.64 g/100 g using 5 components and 8 features. Those features used are: maximum spectral magnitude, bandwidth, temporal and spectral energy, temporal entropy, temporal and spectral crest factor and spectral centroid. This highlights the strength of the methods

Table A1
Explanation of used acoustic features; the shown correlations are examples extracted from ~10 °C and 6 g/100 g maltose, respectively – those plots vary between each measuring point; the presented bar resemble the standard deviation of each value.

Feature name	Abbr.	Formula	Graphical interpretation	Correlation to conc.	Correlation to T
Maximal magnitude	mag	$\max(\lvert x(n)\rvert)$			
Bandwidth	BW	$f_n(\lvert x(n)\rvert == \frac{1}{3}\max(\lvert x(n)\rvert) - f_n(\lvert x(n)\rvert == \frac{1}{3}\max(\lvert x(n)\rvert)$			
Spectral kurtosis	kur_s	$\frac{\sum_{n=1}^{N}(n-C_t)^4 * \text{pm}f(n)}{\text{ssp}^2}$ $\begin{cases} < 3 & \text{flatter distribution} \\ = 3 & \text{normal distribution} \\ > 3 & \text{peaker distribution} \end{cases}$			
Spectral skewness	ske_s	$\frac{\sum_{n=1}^{1024}(n-\text{cen}_t)^3 * \text{pm}f(n)}{\text{ssp}^{3/2}}$ $\begin{cases} < 0 & \text{more energy on the right} \\ = 0 & \text{symmetric distribution} \\ > 0 & \text{more energy on the} \end{cases}$			
Temporal energy	eng_t	$\sum_{k=1}^{N}(x(k))^2$			
Spectral energy	eng_s	$\sum_{n=1}^{N}\lvert X(n)\rvert^2$			
Temporal entropy	ent_t	$-\sum_{k=1}^{N}\left(\frac{\lvert x(k)\rvert}{\sum_{k=1}^{N}\lvert x(k)\rvert}\right)^2 \ln\left(\frac{\lvert x(k)\rvert}{\sum_{k=1}^{N}\lvert x(k)\rvert}\right)^2$			
Spectral entropy	ent_s	$-\sum_{n=1}^{1024}\text{pm}f(n) * \ln(\text{pm}f(n))$			
Spectral crest factor	cf_s	$\frac{\max(\lvert X(n)\rvert)}{\frac{1}{N}\sum_{n=1}^{N}\lvert X(n)\rvert}$			
Temporal crest factor	cf_s	$\frac{\max(\lvert X(n)\rvert)}{\frac{1}{N}\sum_{n=1}^{N}\lvert X(n)\rvert}$			
Spectral centroid	cen_s	$\text{cen}_s = \frac{\sum_{n=1}^{N} n.\lvert X(n)\rvert}{\sum_{n=1}^{N}\lvert X(n)\rvert}$			
Spectral spread	spr_s	$\text{ssp} = \sum_{n=1}^{N}(n - \text{cen}_s)^2 * \text{pm}f(n),\ \text{pm}f(n) = \frac{\lvert X(n)\rvert}{\sum_{n=1}^{N}\lvert X(n)\rvert}$			

used to detect less sensitive inputs in correlation to respective targets.

Although validation error in this approach is still higher than the one which is already presented in Krause et al. [25], the presented approach is more realistic, since simple linear interpolation is not feasible for the shown temperature behavior. Further, the presented error is still higher than a theoretically acceptable absolute deviation of +/−0.5 g/100 g in monitoring online processes in brewing industry. However, the lack in accuracy presented in lower regions of sugar concentration is believed to be due to signal sensitivity issues and non-linear behavior of presented features. Therefore, the stability and robustness of this approach in combination with setup optimization will be the aim of further research (setup material, setup design, near field and superposition phenomena, electronic circuit adaptions). This will include the enhancement of signal quality in the measurement setup. One of the biggest influences is coming from buffer material and design. This will be further optimized by changing to material with properties more suitable to process demands [e.g. poly(methyl methacrylate) (PMMA), polyvinylidene flouride (PVDF)] and by adapting the design of the buffer itself. The mentioned materials were chosen according to investigations on the impedance differences between fluid and buffer. In theory, PMMA is more sensitive with respect to density changes in the fluid. Those changes have an impact on the wavelet amplitude. The drawback is the loss of buffer reflections. PVDF could be used as reflector material, since it has high absorptivity [39,40].

Further, iterative schemes for enhancing variable selection as shown in literature [41] are thought to enhance the outcome of the presented approach and will be investigated. Additionally, appropriate outlier detection algorithms will be implemented. Future investigations should also focus on thermal gradient and other dynamic process influences. Nevertheless, this measuring system is highly competitive with current existing solutions in the industry from both an economical point of view as well as regarding service and maintenance concerns.

Acknowledgements

This work was partially funded by the "Bundesministerium fuer Bildung und Forschung" through the research project AZ – 994 – 11 in cooperation with "Bayerische Forschungsstiftung".

Appendix A.

See Table A1.

References

[1] R. O'Leary, Method of Analysis for correcting dissolved CO_2 content for specific gravity and alcohol variations in beer, <http://www.iul-instruments.de/pdf/vitalsensors_2.pdf> (cited 2009 19 03); 3.
[2] Webpage, Neue Hefereinzuchtanlage bei den Kölner Verbund Brauereien (KVB), Brauwelt, 41–42 2008 p. 1161.
[3] Webpage, <http://www.anton-paar.com/Dichtesensoren/59_Germany_de?productgroup_id=117> <http://www.anton-paar.com/Dichtesensoren/59_Germany_de?productgroup_id=117>. (2009 25.04.2009)
[4] Webpage, <http://www.sensotech.com/.2010.05.04.2010> <http://www.sensotech.com/>.
[5] P. Hauptmann, N. Hoppe, A. Püttmer, Application of ultrasonic sensors in the process industry, Meas. Sci. Technol. 13 (2002) R73–R83.
[6] B. Henning, J. Rautenberg, Process monitoring using ultrasonic sensor systems, Ultrasonics 44 (Supplement 1) (2006) e1395–e1399.
[7] N.I. Contreras et al., Analysis of the sugar content of fruit juices and drinks using ultrasonic velocity measurements, Int. J. Food Sci. Technol. 27 (5) (1992) 515–529.
[8] F.-J. Kuo, C.-T. Sheng, C.-H. Ting, Evaluation of ultrasonic propagation to measure sugar content and viscosity of reconstituted orange juice, J. Food Eng. 86 (1) (2008) 84–90.
[9] J.C. Adamowski, F. Buiochi, R.A. Sigelmann, Ultrasonic measurement of density of liquids flowing in tubes. in: Ultrasonics Symposium, Proceedings, IEEE, Seattle, WA, USA, 1995
[10] J.A. Bamberger, M.S. Greenwood, Measuring fluid and slurry density and solids concentration non-invasively, Ultrasonics 42 (2004) 563–567.
[11] E. Bjørndal, Acoustic Measurement of Liquid Density with Applications for Mass Measurement of Oil, University of Bergen, Norway, 2007.
[12] M.S. Greenwood et al., On-line ultrasonic density sensor for process control of liquids and slurries, Ultrasonics 37 (2) (1999) 159–171.
[13] J.M. Hale, Ultrasonic density measurement for process control, Ultrasonics 26 (6) (1988) 356–357.
[14] W. Sachse, Density Determination of a Fluid Inclusion in an Elastic Solid from Ultrasonic Spectroscopy Measurements. in: Ultrasonics Symposium, 1974.
[15] S. Hoche, M.A. Hussein, T. Becker, Ultrasound based density determination via buffer-rod-1 techniques: a review. J. Sens. Sens. Syst. (2013) (submitted).
[16] A. Püttmer, P. Hauptmann, Ultrasonic density sensor for liquids. in: Proceedings IEEE International Ultrasonics Symposium, Sendai (Japan), 1998.
[17] W. Kessler, Multivariate Datenanalyse, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2007.
[18] R. Schäfer, J.E. Carlson, P. Hauptmann, Ultrasonic concentration measurement of aqueous solutions using PLS regression, Ultrasonics 44 (1) (2006) e947–e950.
[19] H. Elfawakhry, M.A. Hussein, T. Becker, Investigations on the evaluation of rheological properties of cereal based viscoelastic fluids using ultrasound, J. Food Eng. 116 (2) (2013) 404–412.
[20] P. Resa et al., On-line ultrasonic velocity monitoring of alcoholic fermentation kinetics, Bioprocess Biosyst. Eng. 32 (2009) 321–331.
[21] P. Resa et al., Ultrasonic velocity in water–ethanol–sucrose mixtures during alcoholic fermentation, Ultrasonics 43 (4) (2005) 247–252.
[22] Y.-L. Cha, B. Hitzmann, Ultrasonic measurements and its evaluation for the monitoring of *saccharomyces cerevisiae* cultivation, Bioautomation 1 (2004) 16–29.
[23] V.Y. Liepin'sh, An algorithm for evaluation a discrete Fourier transform for incomplete data, Autom. Contr. Comput. Sci. 30 (3) (1996) 27–40.
[24] Q. Zhang et al., A precise and adaptive algorithm for interharmonics measurement based on iterative DFT, IEEE Trans. Power Del. 23 (4) (2008) 1728–1735.
[25] D. Krause et al., Bioprocess monitoring and control via adaptive sensor calibration, Eng. Life Sci. 11 (4) (2011) 402–416.
[26] W.B. Hussein, M.A. Hussein, T. Becker. Application of audio signal processing in the detection of the red palm weevil. in: Proceeding of European Signal Processing Conference EUSIPCO2009, Glasgow, Scottland, 2009.
[27] E. Wallhäußer et al., On the usage of acoustic properties combined with an artificial neural network – a new approach of determining presence of dairy fouling type A, J. Food Eng. 103 (4) (2011) 449–456.
[28] W.B. Hussein, M.A. Hussein, T. Becker, Detection of the red palm weevil using its bioacoustics features, J. Bioacoustics 19 (3) (2010) 177–194.
[29] H. Abdi, Partial least squares regression and projection on latent structure regression (PLS Regression), Wiley Interdiscipl. Rev.: Comput. State 2 (1) (2010) 97–106.
[30] R. Henrion, G. Henrion, Multivariate Datenanalyse, Springer Verlag, Berlin, 1995.
[31] M. Mitzscherling, Prozeßanalyse des Maischens mittels statistischer Modellierung, in: Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt, Technische Universität: München, 2004.
[32] H. Wold, Causal flows with latent variables. Partings of the ways in the light of NIPALS modelling, Eur. Econ. Rev. 5 (1) (1974) 67–86.
[33] D. Krause et al., Ultrasonic characterization of aqueous solutions with varying sugar and ethanol content using multivariate regression methods, J. Chemom. 25 (4) (2011) 216–223.
[34] L. Eriksson, et al., Multi- and megavariate data analysis – principles and applications, Umetrics Academy, 2006.
[35] Webpage, <http://www.crgraph.de/PLS.pdf.2011.05.08.2011> <http://www.crgraph.de>.
[36] D. Lee et al., A variable selection procedure for x-ray diffraction phase analysis, Appl. Spectrosc. 61 (12) (2007) 1398–1403.
[37] C. Botella, J. Ferré, R. Boqué, Outlier detection and ambiguity detection for microarray data in probabilistic discriminant partial least squares regression, J. Chemom. 24 (7–8) (2010) 434–443.
[38] M.M. Breunig, et al. LOF: identifying density-based local outliers. in: Proceedings of SIGMOD'00, Dallas, Texas, 2000.
[39] S. Hoche, M.A. Hussein, T. Becker, Ultrasound-based density determination via buffer rod techniques: a review, J. Sens. Sens. Syst. 2 (2) (2013) 103–125.
[40] A. Püttmer, P. Hauptmann, B. Henning, Ultrasonic density sensor for liquids, ultrasonics, ferroelectrics and frequency control, IEEE Trans. 47 (1) (2000) 85–92.
[41] R.F. Teófilo, J.P.A. Martins, M.M.C. Ferreira, Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression, J. Chemom. 23 (1) (2009) 32–48

2.2.3 NIR and PLS – Discriminant Analysis for predicting the processability of malt during lautering

This article is originally published in European Food Research and Technology, 2015. 240(4): p. 831-846 and reused with the permission of Springer.

ORIGINAL PAPER

NIR and PLS discriminant analysis for predicting the processability of malt during lautering

D. Krause · C. Holtz · M. Gastl · M. A. Hussein · T. Becker

Received: 29 July 2014 / Revised: 21 October 2014 / Accepted: 19 November 2014

Abstract This work is focused on a new strategy for quality analysis of brewing malt using near infrared (NIR) spectra taken from malt kernels in reflection as fingerprint to classify directly to processability of malt. One part of the study deals with calibrating a partial least squares discriminant analysis (PLS-DA) model with NIR spectra classifying malt into the three different classes resulting in a five-component model. Therefore, suitable pre-processing algorithms for spectra were tested. The target for calibration is given by an expert opinion on lautering runs (filtration step in brewing). The accuracy achieved using pilot plant data in relation to the expert classification "good", "normal" and "bad" was 90.6 and 92.7 % in validation and calibration, respectively. The second part of the study is presenting the transfer of these analytical tools to industrial scale. This was established via adjustment to corresponding system conditions. The accuracy achieved using similar algorithms as mentioned before was 93.6 and 76.6 % in calibration and validation, respectively. Independent from this, two numerical possibilities were established for automatic process evaluation classifying the different processes in three categories (good, normal, bad): the first is calculating the residual standard deviation of a process based on multivariate statistical process control and the second is discretizing each process individually based on its single online trends. Both methods were compared to the expert opinion coinciding with 84 and 85 %, respectively.

Keywords PLS discriminant analysis · Near infrared spectroscopy · Malt quality · Multivariate statistical process control

Abbreviations

μ	Centroid of class S
A	Number of components
ABOF	Angle-based outlier factor
B	Matrix of regression parameters
$\mathbf{b_0}$	Vector of intercepts
d()	Distance
$\mathbf{e_i}$	Vector of residuals
FAN	Free amino nitrogen
h_i	Leverage
MLR	Multilinear regression
MSC	Multiplicative scatter correction
MSPC	Multivariate statistical process control
n	Number of sampling points
NAS	Net analyte signals
NIR	Near infrared
p	Polynomial degree
p, P	Vector/matrix of X-loadings
PC	Principal components
PCA	Principal component analysis
PLS-DA	Partial least squares discriminant analysis
PLS(R)	Partial least squares (regression)
q, Q	Vector/matrix of Y-loadings
RMSE	Root mean squared error
RSD	Residual standard deviation
S	Classes
$\mathbf{s_P}$, **b**, **c**	Vectors representing data point in multidimensional space

D. Krause (✉) · M. A. Hussein · T. Becker
Research Group Bio-Process Analysis Technology, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany
e-mail: d.krause@tum.de

C. Holtz · M. Gastl
Research Group Raw Materials, Center of Life and Food Sciences Weihenstephan, Technische Universität München, Weihenstephaner Steig 20, 85354 Freising, Germany

Published online: 07 December 2014

SSE	Sum of squares
STA	Single trend analysis
sVAST	Supervised variable stability scaling
SVD	Singular value decomposition
$\mathbf{S_X}$, $\mathbf{S_Y}$	Diagonal matrix with standard deviation
t, T	Vector/matrix of X-scores
VAST	Variable stability scaling
VIP	Variable importance in the projection
V_t	Ratio of variance
w	Vector of weight factors
w, W	Weighted loading vector/matrix
x	Vector of input values/NIR spectrum
X	Matrix of input data
$\mathbf{X_T}$	Matrix of restructured PLS scores
Y	Matrix of targets
σ	Variance
τ	Vector of scaled time frames

Subscripts

^	Predicted
¯	Mean
a, i, j, k, n, p	Counter
s	Scaled
turb	Turbidity

Introduction

Malt is a natural product which is exposed to vintage-induced variations. Consequently, occurring harvest quality variations on this raw product causes adaption in processing. Nevertheless, high and constant quality is an essential requirement concerning economic production of valuable beer.

Particularly, lautering is influenced by variations in malt properties. Lautering is known as one of the limiting processes in several breweries. Therefore, short lautering time accompanied by corresponding wort quality represents a central demand. Particularly, the variations in malt quality are critical with respect to compliance of this necessity. Thus, breweries demand for malt fulfilling narrow specifications to prevent lautering problems prior to processing. However, it is shown quite often that standard malt analytics are not enough to predict malt with bad lautering abilities. This statement is confirmed by investigations of the workgroup VLB Berlin [1]. Factors like homogeneity or β-glucan content are known influencing factors on lautering time, indeed. Nevertheless, isolated factors are not suitable for a reliable performance criterion (e.g. Nischwitz et al. [2]). In contrast to filtration theory, investigations revealed that viscosity has limited impact on lautering velocity. Instead, it is obviously more influenced by particle size distribution of used grain. Particle size distribution is lately reasoned by malt texture and composition. Malt solution (detectable as friability via friabilimeter as a fast method) [3], which directly influences groating results, seems to play an important role on particle size distribution. Composition and change of protein fraction during mashing appear to have further impact. Protein starts to precipitate during mashing [4] and generates protein–carbohydrate complex, which proved to be lautering inhibiting (e.g. Moll et al. [5] or Sjoholm et al. [6]). The common malt analytics do not cover these areas of complex formation and interactions.

Further, studies investigate laboratory systems, which should predict lautering time in industrial scale. Such systems are usually time intensive and minor effective, since laboratory mash is filtered, but transfer of the results to industrial scale is only partially possible.

Each brewery aims at predicting the reasons for bottlenecks in lautering as early as possible to react with suitable retaliatory actions. This means that a fast and reliable method has to be available to reject single malt batches already at delivery. This would also help in preparation for specific technological as well as control actions whilst brewing to minimize lautering problems.

NIR analytics provide the possibilities to close the gaps in malt analytics as well as in fast methods to predict lautering ability. The content of information is even higher once the whole spectral region is considered. This background is used in the current study. There are more or less distinct differences in the absorption bands which could be used for analysis. In the field of microbiology, bacteria and yeasts are distinguished on strain level after defined cultivation using a broader IR region as a fingerprint (e.g. Ellis et al. [7]). In addition, the potential of spectral evaluation was shown in earlier studies by extracting inaccessible process parameters [8]. The potential of NIR spectroscopy in predicting raw material quality as well as analysing the composition of food stuff (e.g. Nicolaï et al. [9]) was shown in diverse areas, too. In cereal technology, a wide range of applications based on NIR is presented (e.g. by using artificial neural networks for calibration, review by Goyal [10]). Amongst others, applications in brewing industry, measuring protein and water, β-glucan content or solution progress of green malt are known (e.g. Meurens and Yan [11]). Appropriate calibrated devices are therefore standard in incoming goods inspection of malt houses. NIR spectroscopy was also applied to analyse germination properties of barley [12] or calibration on quality index of malt [13].

Aim of study

The aim of the present study is the establishment of a prediction method for lautering ability of different malt batches using NIR spectroscopy. Therefore, next to the existing quality evaluation of malt, a fingerprint adjusted

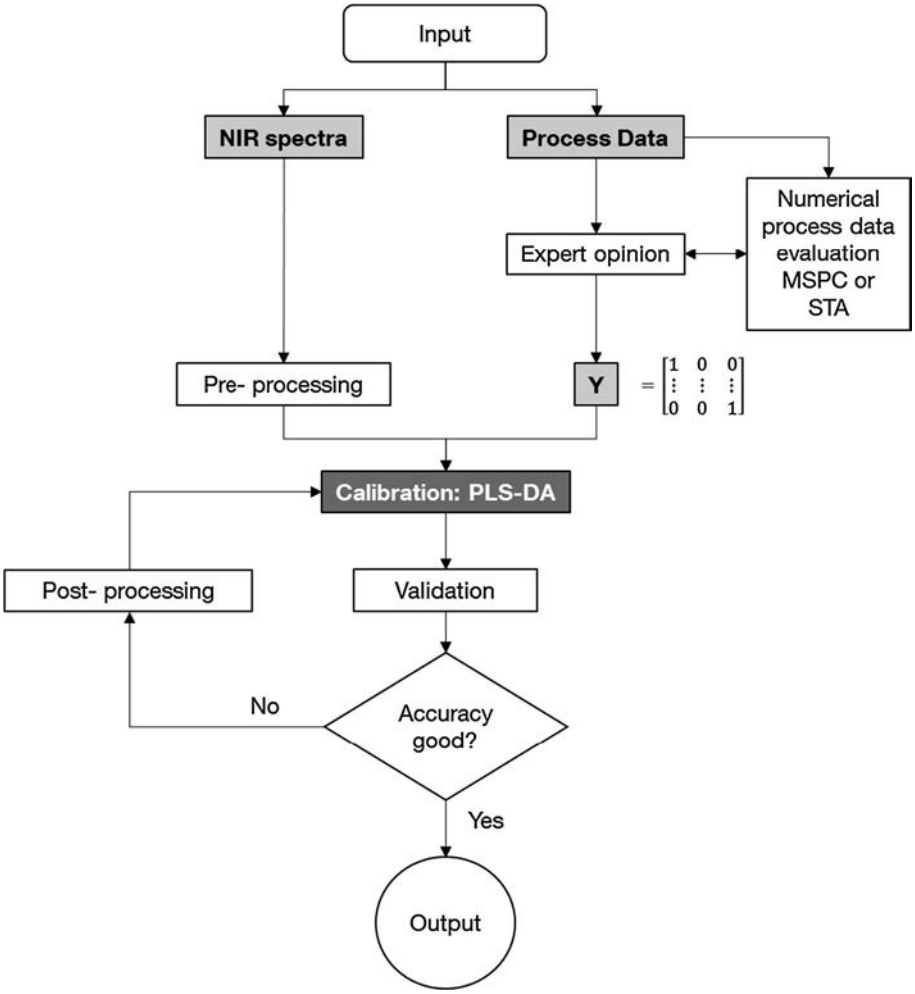


Fig. 1 A flow chart of the whole calibration process [the two different input sources (pilot and industrial scale) are investigated independently]; NIR data serve as predictor, the expert opinion as target in the artificial matrix Y; as a side process, two numerical possibilities were tested on the process data and compared to the expert opinion

to the corresponding processing system is established. This fingerprint should be capable in classifying problematic batches. To achieve this goal, the process step before (mashing) is kept constant. This should minimize its influence on the correlation between NIR spectra and lautering parameters. Finally, the NIR spectrum of a malt batch should be directly used as a rapid evaluation method for quality inspection. The NIR spectra were used for calibrating a PLS-DA model classifying the malt into the three different classes. Therefore, a "dummy" matrix of three columns with entry one for class membership or zero for no class membership is needed as target for the PLS algorithm. This dummy matrix is created from the expert opinion on the lautering runs. Independent from this, two numerical possibilities were established for automatic process evaluation and compared to expert opinion.

The following steps had to be developed and are explained in the next sections:

(1) Evaluation and classification of lautering processes (either by expert knowledge or by numerics)
(2) Processing of NIR spectra
(3) Establishing a fingerprint for prediction of quality by discriminant analysis.

The whole approach including all necessary and investigated steps is summarized in the flow chart shown in Fig. 1.

Materials and methods

Data pool

Malt

Samples for adjustment of NIR measuring system were harvested between 2010 and 2012. The pilot plant trials were realized on two different barley types (Marthe and Grace) with differing malt qualities (variation of malting to reach different cytolytic and proteolytic solubility properties). Altogether, 11 pilsner malt single batches for lautering were employed. The reference analysis for malt

Table 1 Specification ranges of used malt samples (bold marked values are out of norm)

Analytics	MEBAK method	Unit	Norm values	Minimum	Maximum
Moisture	3.1.4.1	%	3–5	4.0	5.6
Extract	3.1.4.2.2	% d.m.	79–82	81.3	**85.2**
Viscosity (8.6 %)	3.1.4.4	mPas	1.450–1.600	**1.426**	**1.607**
Friability value	3.1.3.6.1	%	>80	83.3	**99.0**
Steely kernels		%	<2.5	0.0	**2.1**
Appearance	3.1.4.2.6	Clear or opal		Clear	**Opa**l
pH value			5.6–6.0	5.7	6.1
Raw protein	3.1.4.5.1.1	% d.m.	8.0–13.5	9.1	11.2
Soluble nitrogen	3.1.4.5.2.1	mg (100 g)$^{-1}$ d.m.	550–750	**538**	**793**
Kolbach index	3.1.4.5.3	%	35–45	35.8	**49.6**
FAN	3.1.4.5.5.1	mg (100 g)$^{-1}$ d.m.	120–160	123	202
β-Glucan	3.1.4.9.1.2	mg L^{-1}		15	**370**

specifications was performed according to the standard methods of "Mitteleuropäische Brau- und Analysenkommission" (Mebak) [14]. The ranges of malt specifications used are shown in Table 1. Malt samples with fat marked values were used to cover diverse quality.

NIR spectra

NIR-spectra were recorded using the Multi Purpose FT-NIR spectrometer (MPA, Bruker Optik GmbH) without any sample preparation prior to analysis. The spectra were measured in reflection on full-corn samples in the range of 800–2,500 nm with resolution of 8 cm^{-1} and 64 scans and transformed to absorption spectra by Fourier transform. Each spectrum was recorded in three different sample beakers to support the robustness of the calibration model. Investigating on sample shape (full corn, milled), the resolution and number of scans as well as the choice of different malt specification and the reference measurement can be found in Holtz et al. [15].

Process data lautering trials

The lautering process is used for separating the mash into insoluble parts of the grain and the liquid wort needed as base media for brewing. The process consists of three main steps: mashing out, recirculation and sparging. The trials for this study were performed in a pilot brew house scaled to 60 L of cast wort. This plant is equipped with a programmable logic control (PLC) system (PCS-7, Siemens). Several sensors to measure differential pressure, turbidity, original gravity, flow rate, temperature and total volume are mounted to the lauter tun.

On each malt type, three to nine repetitions were carried out resulting in a total of 51 lautering runs. Each run was carried out using 10 kg malt, ground using the institute's two roller mill (Künzel Maschinenbau GmbH, Kulmbach, Germany) with a consistent milling gap of 0.8 mm. The mashing program with temperature rests of 62, 72, and 78 °C/10 min, and a heating rate of 1 °C/min was fixed for all trials. Mashing was carried out with a ratio of 1 kg grist to 4 L water in each run. These two processing steps prior to lautering were kept as constant as possible to have most probable direct relation between malt properties and lautering performance.

The lautering itself was accomplished with a standard procedure described in the following. This recipe included a lauter rest of 10 min followed by 5 min turbid wort pumping with consistent valve and pump settings. Further, the lautering valve was set to a fixed value over the whole duration of lauter wort run (no control), since the mass flow was used as indicator for malt quality. Sparging was carried out with 20 L of water (78 °C) after a volume of 20 L as well as 40 L of run-off wort. After finishing mashing out, the raking machine was left out of the spent grain and first lowered to 50 mm at the same time the first sparging started in each process. The raking machine was lowered earlier just in case of no flow. The raking system was rotating with 0.75 m/min until the end of lautering. End of the process was manually set at a total lauter wort volume of 60 L. Finally, the lautering plots were sorted and interpreted to categories "good", "normal" and "bad" using expert knowledge. These criteria are explained in the following paragraphs.

Data from industrial processes were collected over 1 year from a German brewery. In processes investigated, only 100 % pilsner malt was used. The lautering program as well as the processing steps before was absolved by a typical step control, and conditions were kept constant according to the internal standards.

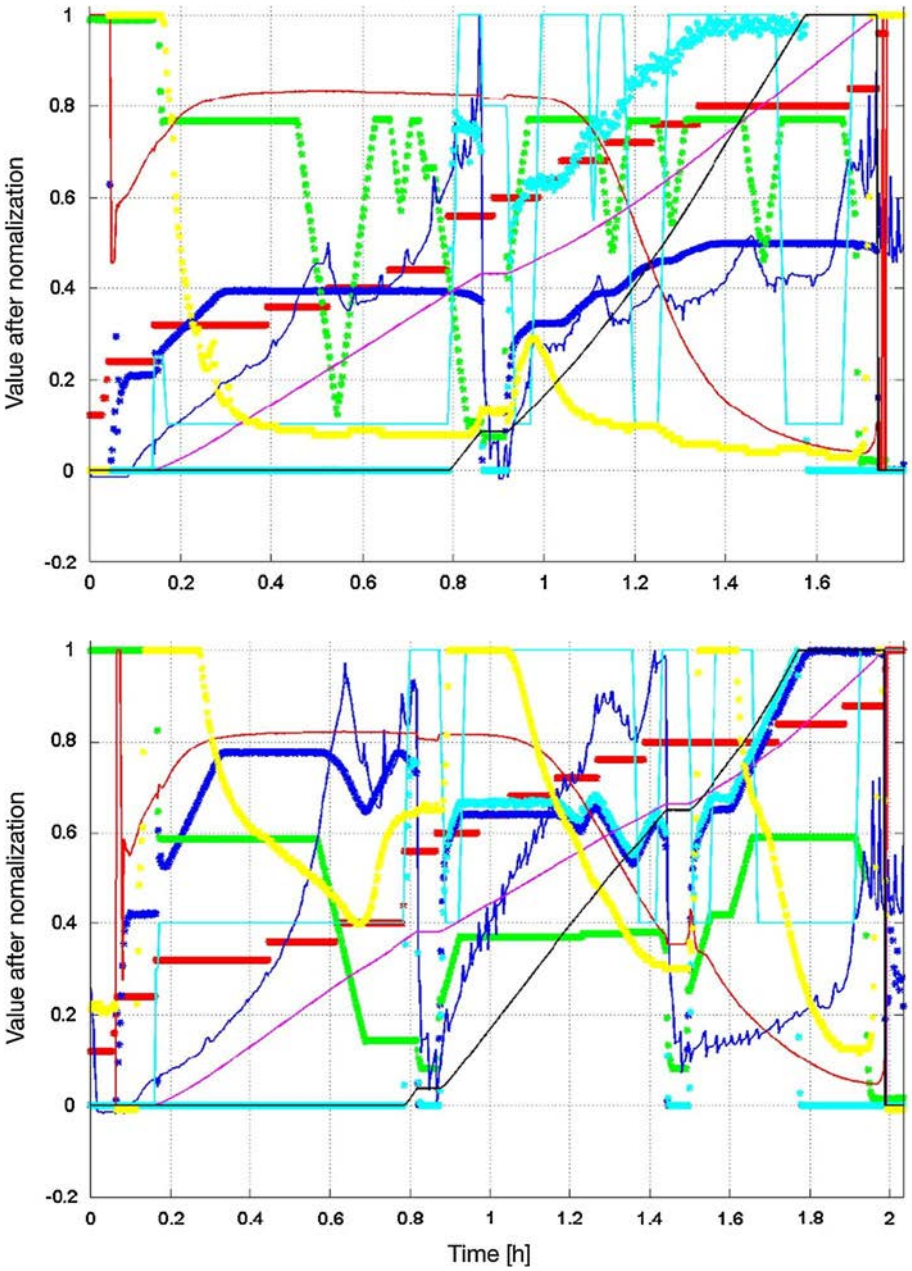


Fig. 2 Good (*top*) and bad (*bottom*) lautering progress of industrial brewery; *red dotted* step number, *blue dotted* lautering wort, *green dotted* rake system, *blue line* pressure difference, *light blue dotted* mass flow sparging water, *red line* original gravity, *light blue line* speed of rake system, *magenta line* integral of mass flow, *yellow dotted* turbidity, *black line* sparging water volume; both diagrams are showing two different qualities of process runs. This is clearly visible when just taking the yellow dotted trend (turbidity), the green dotted trend (height of rake system) as well as the duration of each process (time axis). Two of the main criteria are turbidity trend (below or above 40 EBC) as well as the total process time—in these two examples, these criteria are clearly fulfilled for both classifications (good and bad)

Data processing

Interpretation and analysis of process data

There are three different approaches for the classification in this study; the first one is based on expert knowledge. Experienced brewers were asked to evaluate the typical process chart (as seen in Fig. 2). This figure shows the complexity of this task. Even though the main indicators are the trends for turbidity, rake system height, the pressure difference or the fill height (depending on the used equipment) as well as the total process time, there are a lot of more influencing factors. Therefore, it is quite hard to give a detailed explanation here.

The second approach is based on a multivariate method taken from the toolbox of multivariate statistical process control (MSPC). This method uses the full-process data as input for a PLS regression (here the volume as target) and a subsequent PCA on the resulting T-scores. The resulting residual matrix from this investigation is taken for calculating the "residual standard deviation"—the more similar a process is to the calibration data, the lower this value is. The whole approach was not influenced by the expert classification except for the used calibration processes.

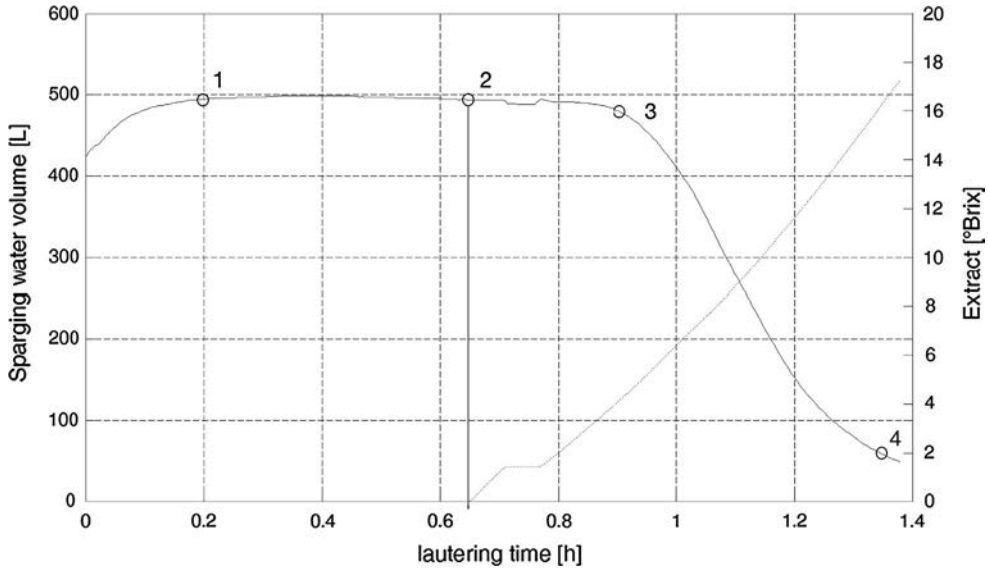


Fig. 3 Schematic extract trend line whilst lautering; the markers indicate characteristic time points: *1* defined start point, *2* start of sparging, *3* inflection point and *4* defined end point

The third numerical possibility is named single trend analysis (STA). Here, each of the mentioned trends is analysed individually to achieve a numerical representation of each process individually. The result is an $[n \times 1]$ vector for each process, where n depends on the number of taken trends for representation. This method was followed by an automatic k-means cluster analysis. The only connection to the expert opinion is the matching error used whilst iterating over the numerical values from individual trends taken into the n-dimensional vector for each process.

The used categories or classes were “good”, “normal” and “bad”. The timeline of the extract in combination with start of sparging water is used to divide the process into relevant investigation areas. Figure 3 shows the characteristic points of the extract trend. Point one indicates the start of first wort run. Second point indicates start of sparging water which has to be considered as a disturbance of the system. Each procedure induced by manual as well as automatic control implicates an indirect influence on malt properties or conceals the direct influence of properties, respectively. With respect to the controlled industrial processes, this fragmentation provides the possibility to give different weights to each of the different areas. The inflection point in the extract curve implies the end of first wort run (point 3, the whole first wort is displaced from spent grains). Finally, point 4 represents the end of lautering run, the break-even point of last water.

Evaluation of processes by experts is time-consuming and contains the possibility for errors. In this study, it was tried to establish mathematical procedures for process evaluation as well. The next paragraphs will explain the possibility of either single trend analysis (STA) of significant process trends or methods of multivariate statistical process control (MSPC).

Measurements in original magnitude are not presented in this work, since each lautering tun is having its individual settings, sensors and sensor readings, respectively.

Single trend analysis (STA) First, different approaches for discretization of single sensor trends were investigated. Therefore, single processes were divided into temporal sections according to the extract trend. The different relevant time points such as first recast are indicated in Fig. 3. Further, the trends of turbidity, volume flow, volume, position of rake system, fill level and pressure difference were discretized in different approaches and taken into account. These are explained in the following paragraphs.

Turbidity One of the main quality criteria is the maximal turbidity of a lautering process. Here, the trend was divided into several horizontal areas for numerical expression. Afterwards, the time the trend line is present in individual areas is summed up. Finally, these values are related to the total time and multiplied with a vector **w** of weight factors to one single value (Eqs. 1, 2).

$$\boldsymbol{\tau}_j = \frac{\sum_{i=1}^{n} \left(t_{out,i} - t_{in,i}\right)}{t_{total}} \tag{1}$$

$$\boldsymbol{x_{turb,p}} = \boldsymbol{\tau}_p^T * \boldsymbol{w} \tag{2}$$

The vector $\boldsymbol{\tau}$ contains the scaled values for the corresponding areas. Here, j resembles the respective area (d; 16 for industrial, 21 for pilot plant trials) and n how often the trend entered this area.

The vector is multiplied with a vector of weight factors resulting in a representative numerical value. The weights are distributed linearly from 1/d (first area, 0–5 EBC) until

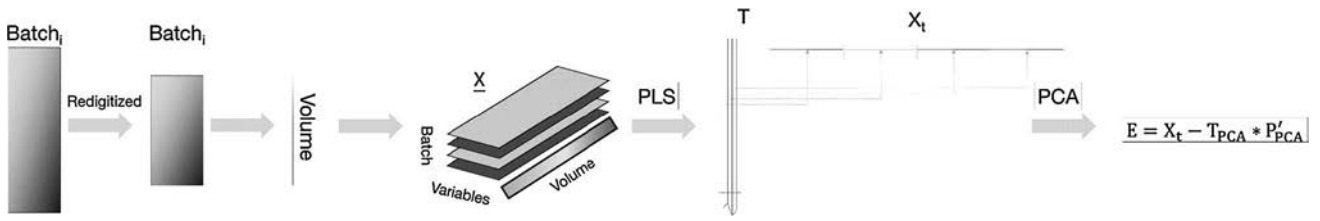


Fig. 4 Schematic explanation of MSPC method by S. Wold et al.; "re-digitizing" of processes to equal data length (orientation on maximal volume), PLS calibration on volume, restructuring of PLS score vectors, PCA analysis of this matrix, calculation of residual matrix E after extraction of a defined number of components

1 (last area). This results in higher final values for processes with higher turbidity. This is caused by bigger temporal values in high turbidity areas which are additionally weighted stronger.

Rake system position and volume flow The numerical values for rake system position are calculated as sum of time under a specific threshold. This threshold is plant specific and represents a deep cut. These cuts are necessary as soon as the cake density is too high. The higher the density, the lower the volume flow rate falls. Therefore, adjusts the process to a more comparable status despite taking deep cuts into account. A similar value was calculated for the volume flow.

Pressure difference The pressure difference reflects the progress of filter cake resistance. The average slope of the difference until the first recast as well as the ratio between start and end difference were taken as representative numerical values.

All these extracted values were compared in different combinations as input to a cluster analysis using K-means. Afterwards, the result was compared to the choice of the experts.

K-means K-means classifies *n* data points into *k* different classes S_j, whereas μ_j resembles the centroid of each class. The data point *n* is represented as a vector **x** containing the different combinations of the above-mentioned extracted values. To reach optimal class division, the sum-of-squares criterion in Eq. 3 is minimized.

$$SSE = \sum_{j=1}^{k} \sum_{x_n \in S_j} |\boldsymbol{x_n} - \mu_j|^2 \tag{3}$$

New data points are classified by the minimal distance to each of the cluster centre (Eq. 4).

$$class = \min\left(d\left(S_j, \boldsymbol{x_i}\right)\right) \tag{4}$$

Multivariate statistical process control (MSPC) MSPC is used to create statistically supported process trajectories for the control of processes. To extract these trajectories, a number of optimal processes and their sensor trends are used. The approach used in the present study is based on a method developed by Wold et al. [16, 17]. Compared to other methods of MSPC, this method is advantageous in its adaption to the existing issue. One of the reasons for this is the usage of "unfold-PLS" (usage of PLS on unfolded three-dimensional data matrix). Here, the interpretation of results is somewhat easier compared to other methods, such as multiway-PCA, multiway-PLS or truly tri-linear decomposition [16–18]. Further, the method is divided into three different levels of process control. This helps to modulate different aspects of analysis such as individual observations (Level 1), the evolution of processes (Level 2) and the processes as a whole (Level 3).

The quality comparison of individual processes is aimed in the present study. The aspects of process control of the presented method are therefore renounced. The principle of extracting relevant information out of the existing data pool is shown in Fig. 4. Therefore, the process data is adapted to equal length. Here, a defined wort volume served as criteria for the new "sampling frequency". Afterwards the re-digitized data are calibrated to the volume as target vector using PLS. The calculated scores are restructured for a subsequent PCA analysis to investigate the variance of these scores over the temporal process trends.

Finally, the residual matrix is calculated using a defined number of components (criteria: explained variance of restructured matrix $\mathbf{X_T}$). In the end, the residual standard deviation (RSD) is calculated for each process (Eq. 5):

$$RSD_i = \sqrt{\frac{\sum_1^n \mathbf{e_i^2}}{(k-d)}} \tag{5}$$

where $\mathbf{e_i}$ resembles the residual vector of a single process, n the number of used PLS components for generation of matrix $\mathbf{X_T}$, k the number of used variables (online-sensor data) and d the number of used PCA components (**T**-scores). The used processes for calibrating this procedure are qualified as "good" based upon expert knowledge. Extracted PLS und PCA components are further used to

decompose new processes calculating their RSD. This value describes how good a certain process fits to the model based on good processes; hence, it gives gradual deviation of single processes.

Pre-processing of NIR spectra

Extracting relevant information often proves to be difficult due to a variety of influences. This underlines the necessity for appropriate pre-processing. The data matrix **X** consist of NIR spectra oriented row wise (objects). Each column represents the absorption values calculated form the measured reflection spectra at a specific wavelength (variables). One can differentiate between row- and column-wise pre-processing. Row-wise pre-processing is used to reduce physical influences such as light scattering in spectra. This can be caused by differing particle material. Such effects can be multiplicative and thereby explain a general slope difference compared to the reference signal. In this study, the algorithms named multiplicative scatter correction (MSC) and standard normal variate (SNV) were applied. Further, algorithms to calculate the first or second derivative used to correct a possible baseline shift were used. The named methods are most common in pre-processing of light spectra in multivariate data analysis [19].

Column-wise pre-processing is aiming at comparability in absolute values of different variables (auto-scaling), easier interpretation of results (mean centring) and reduction in signal noise (variable stability scaling, VAST). In case of discriminant analysis, group-specific differences can be implemented by supervised variable stability scaling (sVAST) [19, 20].

All mentioned transformations (spectral and variable pre-processing) were used in the present study. Those methods are quite common in the field of multivariate data analysis and described well in the literature. A good summary of scaling and pre-processing algorithms can be found amongst other methods in Axelson [19].

Spectral smoothing and derivatives using Savitzky–Golay filter One of the most used algorithms in pre-processing light spectra is the Savitzky–Golay function. This algorithm can be used as a numerical smoothing algorithm. Here, the spectral values are smoothed by fitting a polynomial using a predefined number of sampling points.

Another possibility to increase the useful data content is the baseline correction via numerical derivatives. This can be done using the Savitzky–Golay algorithm as well by differentiating the polynomials over the chosen sampling points. The parameters “sampling points” and “polynomial degree” have to be adapted according to the present data. In this study, the optimal choice is investigated by iterating between 11 and 25 sampling points (n) and a polynomial degree between 3 and 7 (p) (results not shown). The best prediction accuracy of 85 % is achieved using $n = 13$ and $p = 3$ for data from pilot trials.

Multivariate model development

The data evaluation, projecting spectra on quality criteria received from lautering processes, is based on multivariate data analysis. These methods are used to build a relation between dependent and independent variables [21]. Over the last decades, those multivariate data analysis algorithms were established in the field of spectral NIR analysis (see Kessler [20]). Particularly, the use of partial least squares regression (PLSR) showed to be beneficial compared to other methods. In contrast to multilinear regression (MLR) or principal component regression (PCR), PLSR is taking the relation between dependent and independent variables into account. This is achieved by the correlation between dependent and independent variables prior calculating the components. The success of PLSR in interpretation of NIR spectra can be seen in several fields (e.g. Nicolaï et al. [9]).

The background of PLSR is PCA. The data matrix is transformed into latent variables. Based on the algorithm, the decomposition results amongst others in scores (**T**) and loadings (**P**). The final regression model is calculated between these components and matrix **Y** (targets) rather than based on the original variables. The underlying mathematical driving force for decomposition is given by the statistical variance in the data. A detailed description of the background can be found in Kessler [20]. The algorithms used for decomposition were either “nonlinear iterative partial least squares” (NIPALS) (see Wold [22]) or kernel algorithm (see Lindgren et al. [23]). Description of NIPALS algorithm can be found in various literatures (e.g. Kessler [20] or Krause et al. [24]). In addition, it was shown earlier that calculating principal components for PLS regression, using the kernel algorithm, requires less computational effort than NIPALS algorithm. This method is based on eigenvectors of the kernel matrix [17, 23], which is calculated by $\mathbf{X}^{\mathrm{T}}\mathbf{Y}\mathbf{Y}^{\mathrm{T}}\mathbf{X}$. In contrast to the well-known singular value decomposition (SVD) [25], the components are calculated iteratively according to the magnitude of eigenvalues. Prior calculating the new next component, the data matrix is deflated [23]. The used code in this study was adapted for more objects than variables by De Jong [26], based on Lindgren.

After calculation of latent variables, calculation of regression coefficients **B** between **X** and **Y** was carried out using a relation proposed by Martens and Naes [27]. The final model calculating the group membership is given by Eq. 6.

$$\hat{\boldsymbol{Y}} = \boldsymbol{X} \cdot \boldsymbol{B} + 1 \cdot \boldsymbol{b}_0 \tag{6}$$

where $\hat{\mathbf{Y}}$ resembles the predicted target matrix, $\mathbf{X}$ the matrix of predictors, $\mathbf{B}$ the matrix of regression coefficients, $\mathbf{1}$ a column vector of ones and $\mathbf{b}_0$ a vector of intercepts.

Choice of appropriate number of components The selection of the suitable number of principal components (PCs) for modelling the desired correlation is often critical. In the present study, it is dependent on how good the relationship between lautering data and NIR spectra is reflected. In general, the choice can be based on different criteria [27]. Using the minimal predicting error, as a condition for determining the number of components, is one of the most popular methods. Here, the root mean square error (RMSE) is calculated. In this study, the RMSE of cross-validation as well as the RMSE of validation is used. The background of calculation can be found in the literature (e.g. Kessler [20]). Finally, the choice of the number of components is indicated by the lowest error or highest accuracy. In parallel, this method was supported by calculating the explained variance $\sigma_{\text{explained}}$ of $\mathbf{X}$ data (Eq. 7) [28].

$$\frac{\sigma_{\text{explained}}}{\sigma_{\text{total}}} = \frac{\sum_{a=1}^{A} \boldsymbol{t_a^T t_a p_a^T p_a}}{tr\boldsymbol{X^T X}} \tag{7}$$

where $\mathbf{t}_a$ represents the score, $\mathbf{p}_a$ the loading vector of the ath component and $\mathbf{X}$ the data matrix of predictors. Both methods were used for modelling the prediction of lautering ability.

Partial least squares: discriminant analysis Partial least squares (PLS) as standard method for regression of light spectra was originally not established for statistical discrimination. But it was already used for this purpose quite often [29]. However, plenty of other possibilities are existing in the area of cluster analysis. Examples like K-means or support vector machines (e.g. Serrano et al. [30]) are used as algorithms either on scores of a PCA or PLS analysis or directly on the original variables or properties of the present problem [30, 31]. Further, an extension of the PLS regression was proposed (see Barker and Rayens [29] or Nocairï et al. [29, 32]). In this method, the target values (groups or classes) are replaced by a dummy matrix of orthogonal unit vectors (Eq. 8).

$$Y_{ki} \underset{=}{\text{def}} \begin{cases} 1, & y_i = l_k \\ 0, & y_i \neq l_k \end{cases} \tag{8}$$

The group membership to either of the k groups is indicated by l_k, i resembles the corresponding object (1…n spectra). Afterwards, the matrix $\mathbf{X}$ of spectra is analysed in combination with dummy matrix $\mathbf{Y}$ using PLS. The extracted loadings of this decomposition are used to establish a regression model. The calculated regression coefficients can be used for class membership prediction. The predicted decimal values have to be transformed to binary values. The thresholds for this transformation were detected using iteration on each class individual (e.g. 0.45 for class one, results not shown). The threshold values with the maximal correct classified objects for each class (training data) were taken for further analysis.

Finally, using PLS-DA needs adaption of prediction accuracy calculation. The calculation is shown in Eq. 9. In this study, only predictions with clear classification were accepted. Double classifications such as [1 1 0] were treated as wrong. Exceptions were made in case of unclear classification even within the expert's choice.

$$Q\,\% = \frac{\sum \text{correct predicted}}{\sum \text{all objects}} \cdot 100 \tag{9}$$

Post-processing

Variable selection The full NIR spectrum of a malt sample contains a variety of information. There are regions, which are not really contributing with importance to predict certain features. Particularly in the case of high signal-to-noise ratio, in specific signal regions, it is important to exclude non-informative or noisy areas, which strongly contributes to model stability.

The possibilities of variable selection are manifold [33]. Focus in this study was given on "net analyte signal" (NAS, Eq. 10), the scaled regression coefficients $\mathbf{B}_s$ and the variable importance in the projection ((VIP), Eq. 11).

$$\boldsymbol{x_i^{nas}} = \frac{\hat{y}_i}{\boldsymbol{b^T b}} \cdot \boldsymbol{b} \tag{10}$$

For calculating the NAS, regression vector $\mathbf{b}$ and the predicted value y are necessary. Using Eq. 10, a signal $\mathbf{x}$ is calculated which just contains values of high magnitude at important areas of the signal. Calculating and averaging all signals on samples of calibration make it possible to isolate important signal regions.

The calculation of scaled regression coefficients is achieved automatically whilst constructing the regression model. They also contain information due to importance of single variables (see Teófilo et al. [33]).

The third method (VIP, Eq. 11) is based on calculating the projection part of single variable j in relation to y. This is done with increasing number of components (A, a) using scores $\mathbf{t}$ and the corresponding loadings ($\mathbf{w}$ and c). If these single numeric values provide a significant contribution, higher contributions of this variable to the target are expected [34].

$$VIP_j = \sqrt{\frac{n \cdot \sum_{a=1}^{A} w_{ja}^2 c_a^2 \boldsymbol{t_a^T t_a}}{\sum_{a=1}^{A} c_a^2 \boldsymbol{t_a^T t_a}}} \tag{11}$$

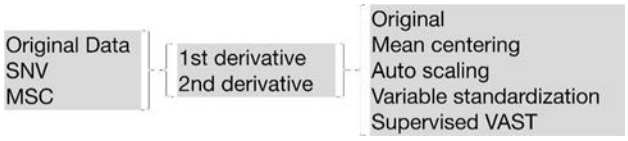


Fig. 5 Possible combinations of single pre-processing algorithms for NIR spectra; the pre-processing due to light scattering (*left*) can be combined with derivatives (*middle*) and with different variable pre-processing (*right*)

This is done with increasing number of components (A, a) using scores **t** and the corresponding loadings (**w** and c). If these single numeric values provide a significant contribution, higher contributions of this variable to the target are expected [34].

The effectiveness of these methods is dependent, amongst others, on the present data [33]. Further it was reported that reducing the input variables by less informative areas should be supported by several methods.

Outlier detection (*Angle-based outlier factor*) The calculation of an ABOF is shown in Eq. 12. Here, each point in multidimensional space is regarded as outlier if the variance of its angle to all the other points in the cluster is low. If the variance is big, the point is more likely located in the centre of the cluster.

$$ABOF(\boldsymbol{s_p}) = Var_{\boldsymbol{b},\boldsymbol{c},\varepsilon M/\boldsymbol{b_p}}\left(\frac{\langle \overline{\boldsymbol{bs_p}}, \overline{\boldsymbol{cs_p}} \rangle}{\left\|\overline{\boldsymbol{bs_p}}\right\|^2 \cdot \left\|\overline{\boldsymbol{cs_p}}\right\|^2}\right) \tag{12}$$

The calculated angle between $\mathbf{s_p}$, **b** and **c** (numerator) is relativized by its spatial distance (denominator). The principle is explained in Kriegel et al. [35]. The data extracted from STA analysis and their combinations were taken as basis for this investigation. The ABOF algorithm was used to analyse results from STA investigation to check the **Y** data for outliers.

Statistical outlier detection Various methods calculating outliers in a multivariate data pool are presented in the literature (e.g. Martens and Naes [27]). In the present study, a method based on the leverage of an object combined with the ratio of residual variance after decomposition to the maximum variance [20, 36] is used. Calculation of sample leverage (Eq. 13) was done using the extracted scores from decomposition procedure. This is reflecting the influence of one sample on the final regression model. The residual variance (Eq. 14) is huge for objects, which contain irrelevant information (**X** data) or are badly described by the model (**Y** data), respectively.

$$h_i = \frac{1}{n}\sum_{a=1}^{A}\left(\frac{t_{ia}^2}{\boldsymbol{t_a^T t_a}}\right) \tag{13}$$

$$V_t = \frac{residualsamplevariance}{totalvariance} \tag{14}$$

where n is the number of objects (samples), t_{ia} the score value of sample i and $\mathbf{t}_a$ the score vector of the ath component.

Results and discussion

Combination of data pre-processing algorithms

Based on the present issue, the decision on pre-processing algorithms as well as the areas of importance in the spectra a priori is not possible. Such a decision could lead to unwanted loss of information, since even wet chemical analysis did not lead to clear relations between single chemical compounds and their influence on filtration performance yet. Therefore, the aim of this study is a data-driven "fingerprint". As a consequence, data-driven choice of algorithms based on whole spectral information was

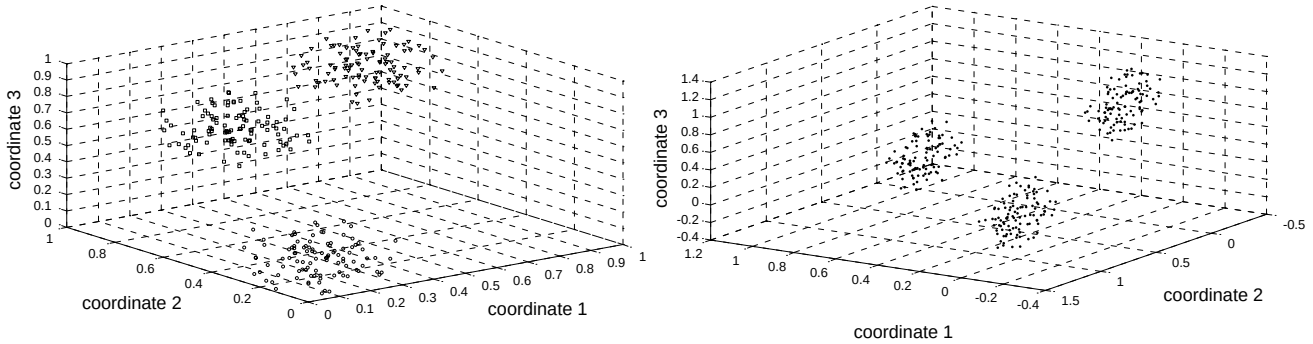


Fig. 6 Graphical proof of generated (artificial) unmasked three-dimensional three-class problem; *left* three original classes (*triangles*, *squares* and *circles*); *right* prediction of classes via PLS-DA–classes are clearly predicted

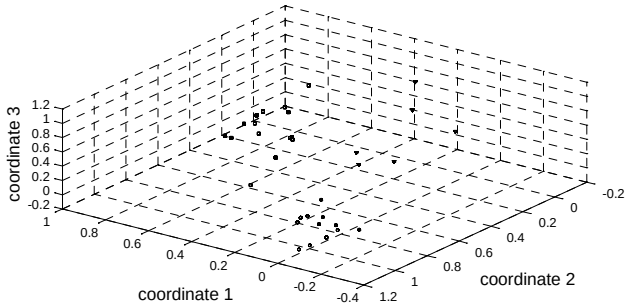


Fig. 7 Three-class problem—predicting malt quality; proof for no masking effect occurring; *circles* predicted as group one, *triangles* predicted as group three, *squares* predicted as group two, *crosses* samples completely predicted wrong

established. Iterations were performed to decide, which combination of pre-processing algorithms fits best, whereas the prediction error serves as decision criterion. All possible combinations are displayed in Fig. 5. The best prediction accuracy was achieved combining either SNV or MSC together with the first derivative and VAST.

Creating the prediction model using discriminant analysis

The aim of this study was the prediction of malt quality in relation to its lautering performance. Therefore, a rough classification should be established. These classes are provided by experts grading the process trends. The subsequent prediction should be based on the correlation between NIR spectra on malt kernels collected in laboratory and expert classification. Such kind of regressions falls within the scope of discriminant analysis. One of the main problems reported in the literature is the so-called masking problem. This means that predicting one out of three classes could not be achieved, since it is masked by the other groups [31, 37]. If these classes are aligned in the three-dimensional space, the predictions using the coordinates in the **X**-matrix lead to one class taking values around 0.5 in each **Y**-coordinate. One solution would be extending the original **X**-matrix prior to regression by a nonlinear extension. Extending the **X**-matrix by a polynomial of second order to crisper classification, each class coordinate (**Y**-matrix) gets values which supports the class membership. Furthermore, Fig. 6 shows another trial of a generated (artificial) three-dimensional three-class problem. The classes presented here are unmasked. The predictions also present a crisp classification of individual classes. This graphical representation proofs that masking is just occurring in special cases, where classes are somewhat lined up with each other.

For this reason, the presented classification problem for processability of malt was also graphically investigated, if masking is occurring. The result is shown in Fig. 7

Table 2 Final configuration of prediction model; size of the used data set, pre-processing of data matrix, number of used components for regression and prediction accuracy

Data set	52 lautering trials; each 3 NIR spectra (malt kernel)	
Pre-processing methods	First derivative via Savitzky–Golay filter, $n = 13$ and $p = 3$	
	Multiplicative scatter correction or standard normal variate (MSC or SNV)	
	Variable stability scaling (VAST)	
Number of components	5 PLS components	
	Calibration	Validation
Number of objects	124	32
Prediction accuracy	92.7 %	90.6 %
Number of wrong predicted processes	3	1

indicating that each class has clearly different coordinates than the others. It can be concluded that masking effect is not existing. Since this multidimensional problem is more difficult to proof graphically, further studies on proofing a masking effect are necessary in future.

The final results of the classification model achieved on the data collected via pilot plant trials using the presented algorithms as well as the corresponding calibration and validation sets are summed up in Table 2.

Variable selection

The presented methods (NAS, $\mathbf{B}_s$ and VIP) were compared to reduce the amount of input variables without deterioration in the prediction accuracy. The achieved results using VIP, which proved best performance, are shown in Fig. 8. The representation on top shows the absolute values of VIP (summed for all three groups). The figure at the bottom represents a typical NIR spectrum with areas of characteristic influence on model predictions, detected by VIP (red marked). A reduction in input variables (wavelengths) of 51.6 % could be achieved using this kind of variable selection whilst prediction error not influenced.

Verification using industrial data

The lautering data of an industrial brewery were analysed by experts to fit into the categories “good”, “normal” and “bad”. These process data are influenced by the control, which is necessary to achieve most optimal results. In this evaluation, the height of the rake system has to be taken into account. Therefore, it has to be analysed that whether an increase in turbidity is caused by the influence of rake system control, lautering valve or the malt quality

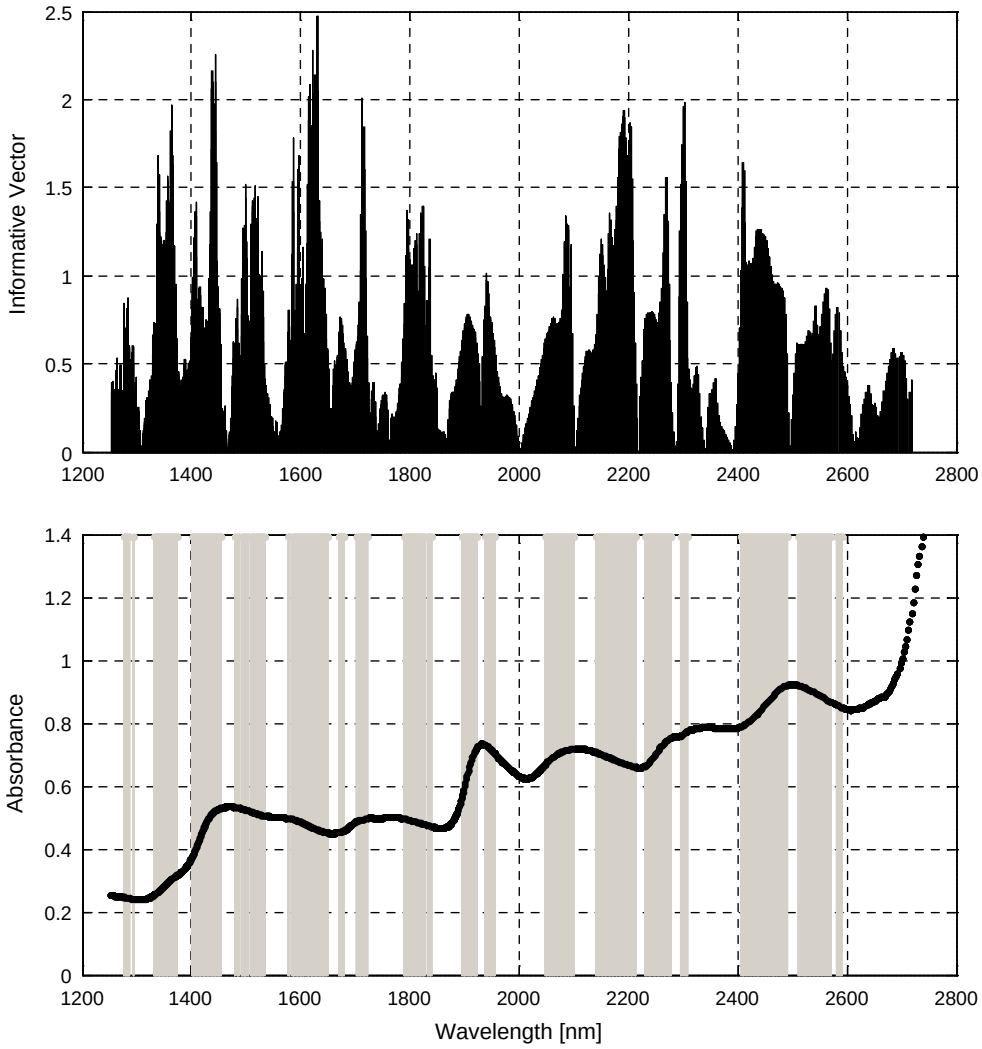


Fig. 8 *Top* absolute values of "variable importance in projection" (VIP) for variable selection (calculated and summed up for all three groups); bottom: typical NIR spectrum for malt kernels; *grey marked* areas were detected as distinctive by VIP reducing the input matrix by 51 % of total variables (wavelength)

itself. To characterize lautering performance according to the malt quality, these technically induced effects were not taken into account by the experts. Additionally, the same criteria for classification into these three groups were taken to ensure comparability to pilot trials. A good lautering progress is shown in Fig. 2, top. The turbidity decreases in the beginning and stays below 40 EBC. A bad lautering progress is shown in Fig. 2, bottom. The turbidity proceeds almost the whole lautering time above 40 EBC. The rake system is run deep to keep the volume flow upright.

Single trend analysis

A comparison between expert classification and K-means analysis on the basis of rake system, turbidity and lautering time is shown in Fig. 9. The colours resemble classification by K-means on STA values, and labels indicate classification by experts. This comparison coincide with 85 % (14 processes of 94 were not matched).

Multivariate statistical process control (MSPC)

Iteration on threshold to achieve a classification in three classes according to the quality using aforementioned RSD had to be realized. This iteration was executed in comparison with the expert's classification. The calculated RSD of all investigated processes is shown in Fig. 10. The colours resemble the membership to the individual group, which are defined by the iterative thresholds. The comparison between experts and RSD coincide with 84 % (15 processes out of 94 not matched).

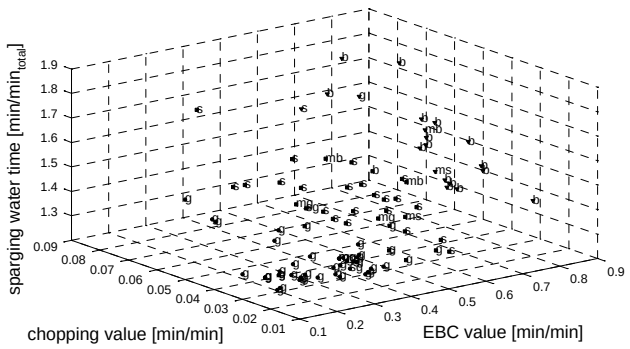


Fig. 9 Graphical representation of K-means results on the basis of rake system, turbidity value and the end point of lautering in comparison with the expert classification; triangles indicate bad, squares the normal and circles the good processes classified by K-means; *g* resembles good, *s* normal and *b* bad processes classified by experts. In case of data points labelled with two letters, a clear classification by experts was not possible; the maximum matching achieved between experts, and K-means classification on single trend analysis (STA) was 85 %

Discriminant analysis of industrial data

The coherences achieved with pilot trials should finally be validated using industrial data. The used data sets, the final processing and the best prediction error are summarized in Table 3. The result of this analysis is shown in Fig. 11. Green circles indicate correct predictions, and red crosses indicate incorrect predictions. The bases of this visualization are values extracted via STA. The created calibration model for industrial data is based on 21 PLS components. This comparably high number could, amongst others, be related to the lower resolution of industrial NIR spectra (16 wave numbers). Additionally, several processes were just provided with two malt spectra. Finally, it may be assumed that the different number of objects per group influences the result as well [32].

In addition, pattern not included in calibration could not be correctly predicted either. All these mentioned points will be basis for further research. Another aspect is the analysis of outliers. Two possible ways which were applied to the present issue are described in the following paragraphs.

Outlier detection

Outlier detection was performed due to the lower accuracy in industrial data. According to the literature, a variety of methods based on different background are existing. One of the presented methods is the angle-based outlier factor (ABOF), and the other method is used in multivariate statistics [20, 36].

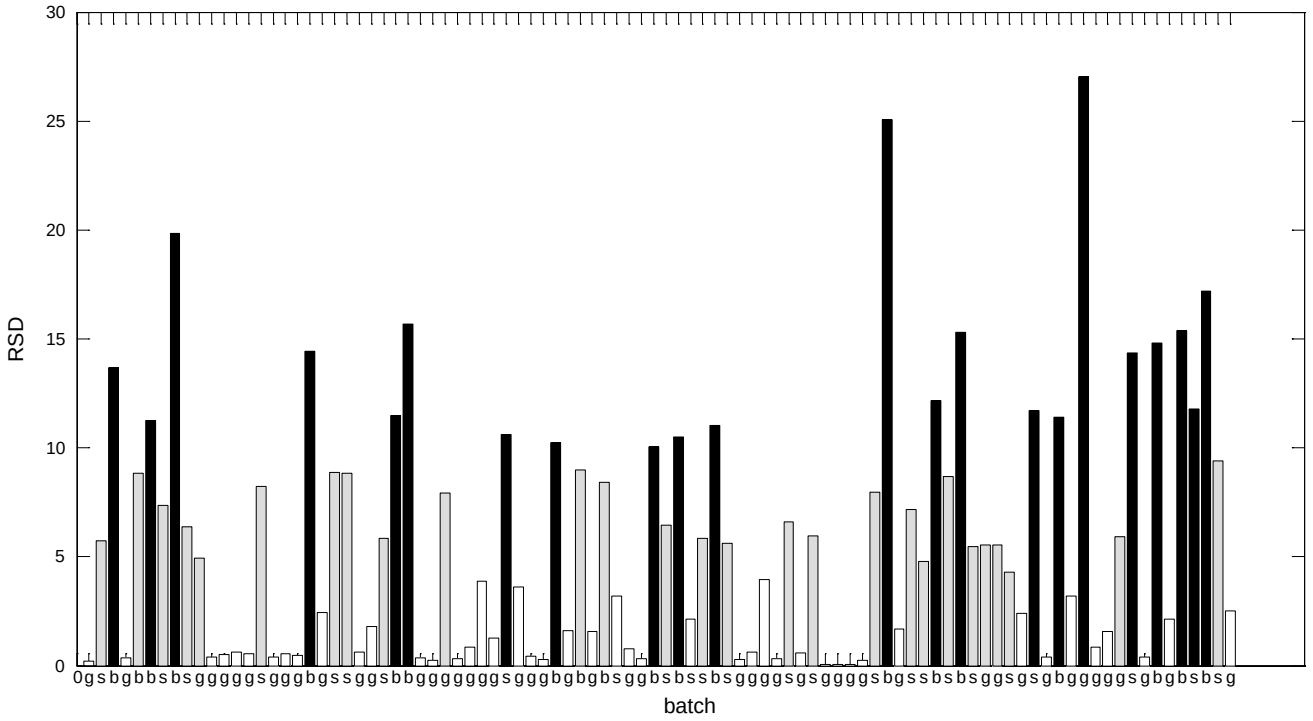


Fig. 10 Residual standard deviation (RSD) of all 94 industrial processes; these were available for model development; red indicate bad, blue the normal and green the good processes; the maximum matching achieved between experts, and MSPC-based classification was 84 %. Bad (b) batches are coloured *black*, normal (s) as *grey* and good (g) as *white*

Table 3 Final configuration of prediction model for industrial processes; size of the used data set, pre-processing of data matrix, number of used components for regression and prediction accuracy

Data set	94 Lautering trials; each min 2 NIR spectra (malt kernel)	
Pre-processing methods	First derivative via Savitzky–Golay filter, $n = 25$ and $p = 3$	
	Multiplicative scatter correction (MSC)	
	Variable stability scaling (VAST)	
Number of components	21 PLS components	
	Calibration	Validation
Number of objects	188	46
Prediction accuracy	93.7 %	76.6 %
Number of wrong predicted processes	11	9

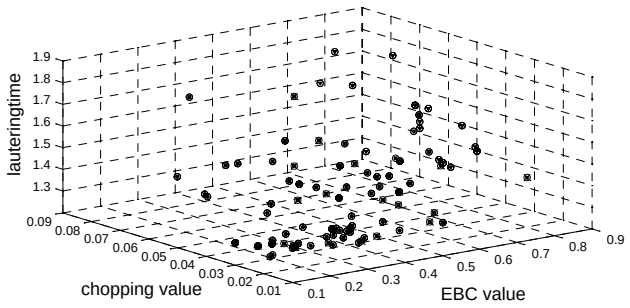


Fig. 11 Result of calibration and validation of industrial data; circles indicate correct, crosses false predictions; in total, 20 out of 234 data points were predicted incorrect (11 out of 188 in calibration, 9 out of 46 in validation); graphical presentation of results based on STA analysis

For ABOF analysis, the data extracted from STA were used. Iterations on these different parameters in different combinations are performed aiming to find thresholds to define the four misclassified processes as outliers. Using this method, the y data (processes) were analysed for outlying data points. In the optimal combination of parameters and thresholds, these four processes could be defined as outliers. Nevertheless, 20 other processes were classified as outliers as well. Consequently, the presented approach is not suitable under the given circumstances and assumptions. The result of the comparison between leverage and residual variance of industrial x data is shown in Fig. 12. Although just a few points could be regarded as outliers, since they are in the area up right resembling high influence on the model and still being poorly described, no coherence between these points and the misclassified processes was visible. In conclusion, it can be said that further investigations are necessary for achieving a meaningful outlier analysis.

Conclusion and outlook

In this study, it was shown that NIR fingerprints of malt kernels can be used to predict lautering performance. This was achieved by calibrating NIR spectra to expert classification on lautering runs. Three classes based on expert knowledge were defined, namely “good”, “normal” and “bad”. The study was based on investigating different data pre-processing, processing and post-processing algorithms.

The development of best possible algorithm combination for data processing and model generation was performed based on pilot plant trials. Afterwards, it was validated using industrial data. The best prediction accuracy was reached using either standard normal variate (SNV) or multiplicative scatter correction (MSC) combined with the first derivative as spectral and variable stability scaling (VAST) as variable pre-processing. The calibration model was based on discriminant analysis and PLS (PLS-DA). The prediction accuracy for validation in relation to the expert classification “good”, “normal” and “bad” was 90.6 % with five components and 76.6 % with 21 components for pilot and industrial data, respectively. Model development was performed on full kernel NIR spectra with a resolution of 8 wave numbers and 64 scans. The lower accuracy as well as the high number of components for industrial data is amongst others caused by lower resolution of spectra. Further influences are the limited number of patterns for calibration. It might be reasonable to assume that non-existing patterns in calibration result in faulty predictions. Both could be solved by either change of the data recording or continuous extension of data pool. Additionally, different number of objects per group for calibration can influence the prediction accuracy [32]. Another aspect is the loss of information whilst pre-processing of data. The used algorithms reduced noise caused by physical effects such as light scattering. It is conceivable that this information could provide a contribution to a better prediction using adapted treatment. To what extend these points in the present investigation can be influenced and solved will be the aim of further research. In addition, the method for classification should be compared to other approaches presented in the literature (such as combination of support vector machines (SVM) and particle swarm optimization (PSO), Melgani and Bazi [38] or tabu search, PSO and SVM, Chuang et al. [39]). Furthermore, spectroscopic data treatment could be enhanced by multiple regression systems (see Benoudjit et al. [40]). It has to be mentioned that the studies on variable selection and outlier detection could also contribute to stability and improvement of models. The investigations performed so far showed first positive results. Nevertheless, these points as well as the relation between the presented results and the biochemical background should be further treated. The latter could be achieved for example by empirical or synthetic calibration

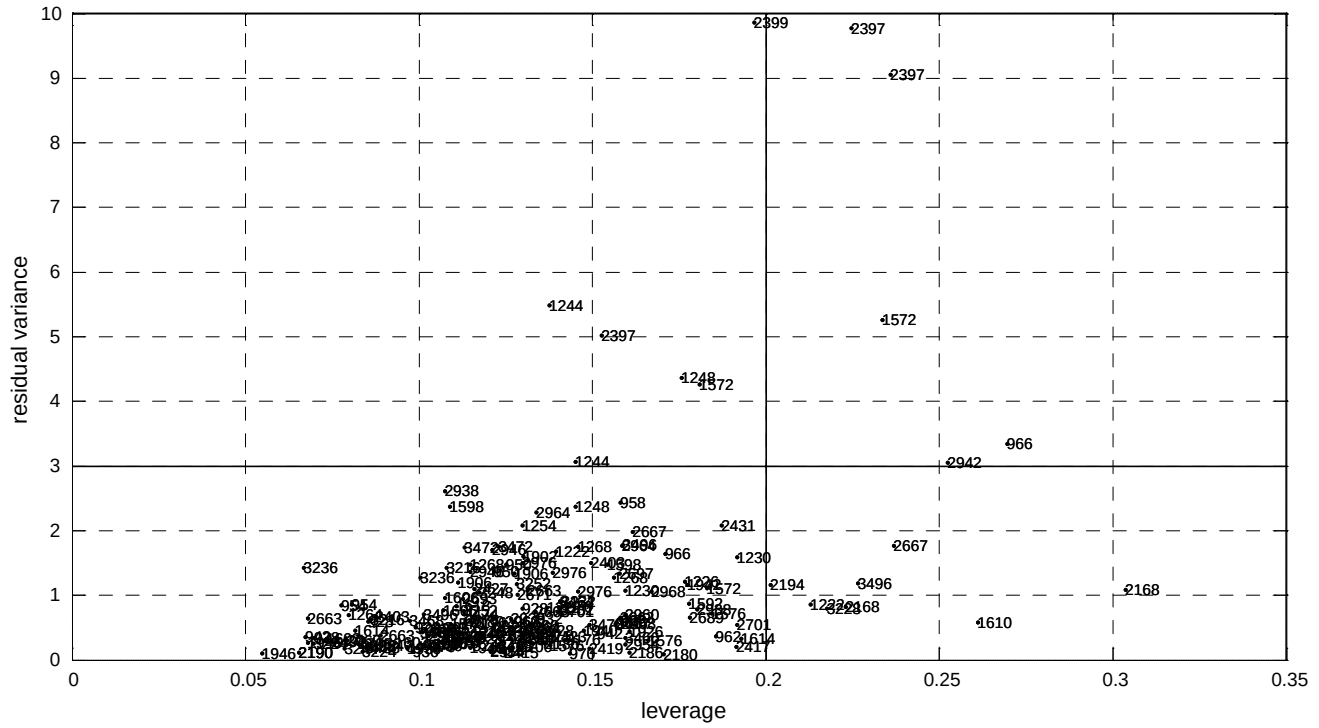


Fig. 12 Graphical presentation of sample influence on calibration model using leverage on *x* axis and unexplained part of variance on *y* axis; the processes, which would be classified as outlier did not match with the misclassified processes by PLS-DA

(see Shi et al. [41]) as well as more intense investigations on the relation to the reference methods presented.

In the present study, the use of uniform malt samples was intentionally to prevent inhomogeneity in the raw material. Further investigations should also take malt blends and special malts into account. Aim should be to validate the functioning of the established method. A further aim should be the possibility to directly influence the lautering performance by specific manipulation of malt grist through blending of differently qualified raw material. Here, investigations on variation of process steps grinding and mashing should be performed. These can have a big impact on lautering, as well. To what extend these investigations could be used for recommendations according to adjustments of these two process steps remains to be analysed.

The possibilities of automatic qualification of lautering processes showed quite promising results. This leads to reduction in expert efforts, respectively. A match of 85 % between single trend analysis and expert qualification as well as 84 % between multivariate statistical process control (MSPC) and expert qualification was achieved. Since almost each lautering tun is having its individual settings, sensors and sensor readings, respectively, usage of any of the presented approaches has to be adapted to each problem of interest individual, which is counting for most of the existing, data-driven models. It has to be investigated that how far these methods can be adapted to be easily transferred to other brewing plants and recipes. Therefore, inputs and methods for processing should be analysed in greater depth.

Finally, it remains to be said that the presented investigations were limited due to harvesting years. This does not cover seasonal variations of raw material quality. It is most reasonable to expect an influence on the presented method by differences in harvesting year, since changes in the composition of raw material is most likely. Those changes already influence more simple correlations like NIR to protein content. For this purpose, a method for automatically adapting the models by means of a moving average filter should be investigated.

Acknowledgments This work was partially funded by the “Bundesministerium fuer Wirtschaft und Technologie” (via AiF) through the research project AIF 16529 N/1 in cooperation with “Wissenschaftsförderung der Deutschen Brauwirtschaft e.V”.

Conflict of interest None.

Compliance with Ethics Requirements This article does not contain any studies with human or animal subjects.

References

1. Rath F (2006)Möglichkeiten der Vorhersage von Verarbeitungsrisiken im industriellen Läuterprozess auf der Basis von Laboranalysen des Malzes. In: 35. Internationales Braugerstenseminar

2. Nischwitz R, Cole NW, MacLeod L (1999) Malting for brewhouse performance. J Inst Brew 105(4):219–227
3. Schwill-Miedaner A, Flocke R, Sommer K (1997) Zusammenhänge zwischen Malzauflösung und Partikelgrößenverteilung des Schrotes. Brauwelt 12:412–416
4. Ferenczy L, Bendek G (1991) Untersuchung der Veränderung der Würzeproteinzusammensetzung während des Maischprozesses mit der SDS-PAGE-Gelelektrophorese und deren Einfluß auf das Abläutern. Monatsschrift fuer Brauwissenschaft 44(5):191–200
5. Moll M, Lenoel M, Flayeux R, Laperche S, Leclerc D, Baluais G (1989) The new Tepral method for malt extract determination. ASBC J 47(1):14–17
6. Sjoholm K, Pietila K, Home S, Aarts R, Jaakkola N (1994) Evaluation of Lautering Performance of Malt. Monatsschrift fuer Brauwissenschaft 47(5):165–171
7. Ellis DI, Broadhurst D, Kell DB, Rowland JJ, Goodacre R (2002) Rapid and quantitative detection of the microbial spoilage of meat by fourier transform infrared spectroscopy and machine learning. Appl Environ Microbiol 68:2822–8282
8. Stippl V, Delgado A, Becker T (2002) Optical method for the in situ measurement of pH-value during high pressure treatment of foods. High Pressure Res 22:757–761
9. Nicolaï BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, Lammertyn J (2007) Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. Postharvest Biol Technol 46 (2):99–118. doi:10.1016/j.postharvbio.2007.06.024
10. Goyal S (2013) Predicting properties of cereals using artificial neural networks: A review. Sci J Crop Sci 2(7):95–115
11. Meurens M, Yan SH (2002) Applications of vibrational spectroscopy in brewing. Handbook of Vibrational Spectroscopy
12. Møller B (2004) Near infrared transmission spectra of barley of malting grade represent a physical-chemical fingerprint of the sample that is able to predict germinative vigour in a multivariate data evaluation model. J Inst Brew 110(1):18–33
13. Nielsen JP, Bro R, Larsen J, Munck L (2002) Application of Fuzzy Logic and Near Infrared Spectroscopy for Malt Quality Evaluation. J Inst Brew 108(4):444–451. doi:10.1002/j.2050-0416.2002.tb00574.x
14. Anger H-M (2006) Brautechnische Analysenmethoden-Rohstoffe. Selbstverlag der Mitteleuropäische Brautechnische Analysenkommission, Freising
15. Holtz C, Krause D, Hussein MA, Gastl M, Becker T (2014) Lautering performance prediction from malt by combining whole near-infrared spectral information with lautering process evaluation as reference values. J Am Soc Brew Chem (in press)
16. Wold S, Kettaneh N, Fridén H, Holmberg A (1998) Modelling and diagnostics of batch processes and analogous kinetic experiments. Chemometr Intell Lab Syst 44 (1–2):331–340. doi:10.1016/S0169-7439(98)00162-2
17. Whitehead IJ (2012) Soft sensing: using multivariate analysis for yeast propagation monitoring. Diplomarbeit, TU München
18. Smilde A, Bro R, Geladi P (2004) Multi-way analysis. Applications in the chemical sciences. Wiley, Chichester
19. Axelson D (2010) Data Preprocessing for Chemometric and Metabonomic Analysis. Kingston, Ontario
20. Kessler W (2007) Multivariate Datenanalyse. Wiley-VCH Verlag GmbH & Co, KGaA, Weinheim
21. Geladi P, Kowalski BR (1986) Partial least squares regression: a tutorial. Anal Chim Acta 185:1–17
22. Wold H (1974) Causal flows with latent variables. Partings of the ways in the light of NIPALS modelling. Eur Econ Rev 5(1):67–86
23. Lindgren F, Geladi P, Wold S (1993) The kernel algorithm for PLS. J Chemom 7(1):45–59. doi:10.1002/cem.1180070104
24. Krause D, Schöck T, Hussein MA, Becker T (2011) Ultrasonic characterization of aqueous solutions with varying sugar and ethanol content using multivariate regression methods. J Chemometr 25(4):216–223. doi:10.1002/cem.1384
25. Golub GH, Reinsch C (1970) Singular value decomposition and least squares solutions. Numer Math 14(5):403–420. doi:10.1007/BF02163027
26. De Jong S, Ter Braak CJF (1994) Comments on the PLS kernel algorithm. J Chemom 8(2):169–174. doi:10.1002/cem.1180080208
27. Martens H, Næs T (1991) Multivariate calibration. Wiley
28. Jørgensen B, Goegebeur Y (2012) Module 8: partial least squares regression II. http://statmaster.sdu.dk/courses/ST02/index.html. Accessed 03.05. 2012
29. Barker MR, Rayens W (2003) Partial least squares for discrimination. J Chemom 17(3):166–173
30. Serrano CC, Gutierrez NB (2011) Partial least square discriminant analysis (PLS-DA) for bankruptcy prediction
31. Indahl UG, Martens H, Næs T (2007) From dummy regression to prior probabilities in PLS-DA. J Chemometrics 21:529–536. doi:10.1002/cem.1061
32. Hicham Nocairi EMQ, Evelyne Vigneau, Dominique Bertrand (2005) Discrimination on latent components with respect to patterns. Application to multicollinear data original. Comput Stat Data Anal 1 (1):139–147
33. Teófilo RF, Martins JPA, Ferreira MMC (2009) Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression. J Chemom 23(1):32–48. doi:10.1002/cem.1192
34. Sorol N, Arancibia E, Bortolato SA, Olivieri AC (2010) Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice A test field for variable selection methods. Chemometr Intell Lab Syst 102(2):100–109. doi:10.1016/j.chemolab.2010.04.009
35. Kriegel H-P, Hubert MS, Zimek A (2008) Angle-based outlier detection in high-dimensional data. Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pp 444–452. doi:10.1145/1401890.1401946
36. Botella C, Ferré J, Boqué R (2010) Outlier detection and ambiguity detection for microarray data in probabilistic discriminant partial least squares regression. J Chemom 24(7–8):434–443. doi:10.1002/cem.1304
37. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning, vol 1. Springer, New York
38. Melgani F, Bazi Y (2008) Classification of electrocardiogram signals with support vector machines and particle swarm optimization. IEEE Trans Inform Technol Biomed 12(5):667–677
39. Chuang L-Y, Yang C-H, Yang C-H (2009) Tabu search and binary particle swarm optimization for feature selection using microarray data. J Comput Biol 16(12):1689–1703
40. Benoudjit N, Melgani F, Bouzgou H (2009) Multiple regression systems for spectrophotometric data analysis. Chemometr Intell Lab Syst 95(2):144–149. doi:10.1016/j.chemolab.2008.10.001
41. Shi Z, Cogdill RP, Martens H, Anderson CA (2010) Optical coefficient-based multivariate calibration on near-infrared spectroscopy. J Chemom 24(5):288–299. doi:10.1002/cem.1301

2.2.4 Online monitoring of bioprocesses via multivariate sensor prediction within Swarm Intelligence Decision Making

Online monitoring of bioprocesses via multivariate sensor prediction within swarm intelligence decision making

D. Krause *, M.A. Hussein *, T. Becker

Research Group Bio-Process Analysis Technology, Technische Universität München, Center of Life and Food Sciences Weihenstephan Weihenstephaner Steig 20, 85354 Freising, Germany

ARTICLE INFO

Article history:
Received 31 December 2014
Received in revised form 1 April 2015
Accepted 12 April 2015
Available online 23 April 2015

Keywords:
Swarm intelligence
Particle swarm optimization
Swarm sensors
Fermentation
Online monitoring
Multivariate statistical process control

ABSTRACT

In this work, a methodology combining process knowledge with computational efforts is presented. The aim is to create a self-organizing sensor network capable of getting over sensor failures. First, multivariate linear and non-linear models are utilized creating a search space based on the multi-sensor data pool. Simple correlations between the raw data retrieved from several sensors are used for extracting multivariate statistical process control trajectories. Those different models are scored by swarm intelligence (particle swarm optimization) leading to the optimal sensor/model combination at certain time step aiming at determination of the fermentation trajectory in combination with sensor output validation.

The results on online data indicate the possibility of more robust online monitoring using the swarm sensing idea for biotechnological processes to insure optimal and timely effective processing as well as sensor failure detection. Adjustments of the basic algorithms, cost function, accuracy of output as well as the dynamic behaviour are addressed. This methodology is not restricted to the number of sensor inputs as well as the use of specific sensor readings, which makes it beneficial over other approaches.

1. Introduction

Increase of transparency of individual processing steps achieving accurate production is one of the major goals in bioprocess monitoring. This includes online monitoring of system variables, where immediate online measurement of leading process variables like substrate as well as biomass concentration is generally not possible.

Furthermore, comprehensive quality control of running processes has to be addressed. Additionally, stable and effective monitoring of industrial biotechnological applications requires a variety of sensor systems. Those are generally exposed to influencing side effects like temperature dependency associated with changes in by-products, gas bubbles and complex process behaviour resulting in corrupted sensor information. Also, existing sensors make up a big part in costs and safety for the whole plant concerning service, maintenance and usage (overview in [1]). Moreover, sensor data are used mostly independently from each other for monitoring. The full potential of such a data pool, if all information is used for control, is rarely tapped. The connectivity of sensors in most biotechnological applications is not among each other but embedded in a central hierarchical structure (field, process and enterprise level). The measured signals merge for the first time on the process level and can be used by the operator for process guidance. A direct communication with each other is not existing and feedback to field level is rare. Literature reports studies, which deal with interacting sensors [2–5] as well as processing of their data [6,7]. Thereby, mobile sensor networks as an early warning system of forest fires [5] or the maintenance for historical buildings [2] were established. The single components act independently from each other but react to signals of other sensors taking their position and environmental conditions into account. The application, which will be controlled by the corresponding network will only take decisions based on several independent signals. Further investigations on swarm intelligent based sensor network communication can be found in Hase [8]. The principle of an "intelligent swarm" of sensors is established in other studies [4,5], too. The usage of swarm algorithms for optimization increased through the past decades [9–13]. One of the most famous algorithms is "particle swarm optimization" (PSO), which is based on the behaviour of birds or fish swarm [11]. This algorithm is known to solve highly non-linear and dynamic problems such as bioprocesses [14]. Furthermore, this algorithm has the ability to perform independent from integrated process knowledge using a suitable cost function. Reports from different fields like process engineering or mechanical engineering [9,14–20] as well as the choice of suitable wave length for multivariate prediction models show the successful application of swarm intelligence. Wolf et al. (2009) reports on the optimization and control of a biogas plant comparing PSO and a genetic algorithm (GA). Here, PSO needed less computational time with respect to convergence on a simulation result [14]. Another comparison between GA, ant colony optimization (ACO), PSO and simulated annealing (SA) by Abraham et al. (2006) resulted in better solutions from PSO and ACO (mostly better standard deviations compared to SA and GA) [9]. Finally it has to be mentioned, that the usage of swarm

* Corresponding authors. Tel.: +49 8161 71 3277; fax: +49 8161 71 3883.
E-mail address: d.krause@wzw.tum.de (D. Krause).

http://dx.doi.org/10.1016/j.chemolab.2015.04.012
0169-7439/

intelligence (SI) is favourable with respect to simplicity, flexibility and fast convergence [9,16,21] as well as robust and almost optimal solutions [11]. Additional benefits and advantages are their memory and learning capability (e. g. ACO, Blum and Roli [22]). Comparing PSO to other metaheuristic algorithms results usually in more parameters to tune and/or more equations and thus more calculation steps (e. g. GA, ACO). Nevertheless, it is generally possible to tune any algorithm and its parameters to find tracks on which they work better than other ones. One difference and advantage of PSO over many other metaheuristic methods is the small amount of tuneable parameters and the simplicity of only having two equations. The challenge is always the adaption of algorithm parameters, the choice of the cost function and the convergence criterion as well as the boundary conditions.

The implementation in biotechnological applications with the mentioned properties is mostly realized on isolated applications of single methods or devices acting on their own. Those are rarely applied in industry. In contrast to the mentioned applications, the presented study reports a methodology which will find the optimal solution out of a pool of process solutions (multivariate trajectories). Therefore, computational intelligence in combination with multivariate statistical process control is used to establish a sensor network being capable of online measurements, self-control and detection of corrupted data/sensors, respectively. Such data driven models are one possibility to deal with a multivariate data pool (see [23]). The aim is to detect faulty sensors and corrupted signals by simple multivariate linear and non-linear statistical process control charts based on sensor data. Swarm intelligence (particle swarm optimization) is then used to find the optimal sensor/model combination out of those different models at a certain time step. Thus, trajectory of fermentation will be determined and a sensor output validation will be performed. The approach combines process knowledge with corresponding measuring systems aiming at cross-linked sensors for their self-control.

2. Material & methods

2.1. Experimental

Modelling and validation background of the presented approach were experimental data from aerobic fermentations of *Saccharomyces pastorianus* var. *carlsbergensis* W43/70 (common bottom-fermenting yeast). Those were realized in a pilot scale industrial fermentation tank (Co. Esau & Hueber, the R&I scheme is shown in Fig. 1). This reactor has a working volume of around 70 L. The medium taken was beer wort (~12 g/100 g original gravity, 100% barley malt, based on a commercial malt extract, Weyermann Bavarian Pilsner). The inoculum used was pre-cultivated yeast for ~5 Mio·cells/mL (counted under the microscope via Thoma Chamber). The aeration rate and pulse/pause frequency as well as the temperature were adjusted and controlled according to the predefined total process time using fuzzy control. The total process time is defined by the time the bacterial population reaches its end point of 100–120 Mio·cells/mL as well as an extract decline of maximum 4 g/100 g. The pump speed was fixed to 30% pump power (approx. 1100 L/h). To prevent foam, another fuzzy controller was used to adjust the overpressure to maximum 1 bar. The whole controller principle can be found elsewhere [24,25]. The processes taken for the investigations presented were in the range from ~10.5 to ~20.5 °C resulting in batch length from 25 to 51 h. The online data were collected in sampling frequency of 10 s and reduced to a frequency of approximately 50 s for computational issues.

2.2. Data pre-processing

The batch data used for the presented approach contained sensor data for pH, dissolved oxygen, temperature, turbidity, density, ultrasound velocity and pressure. Each data set was inspected and cut to

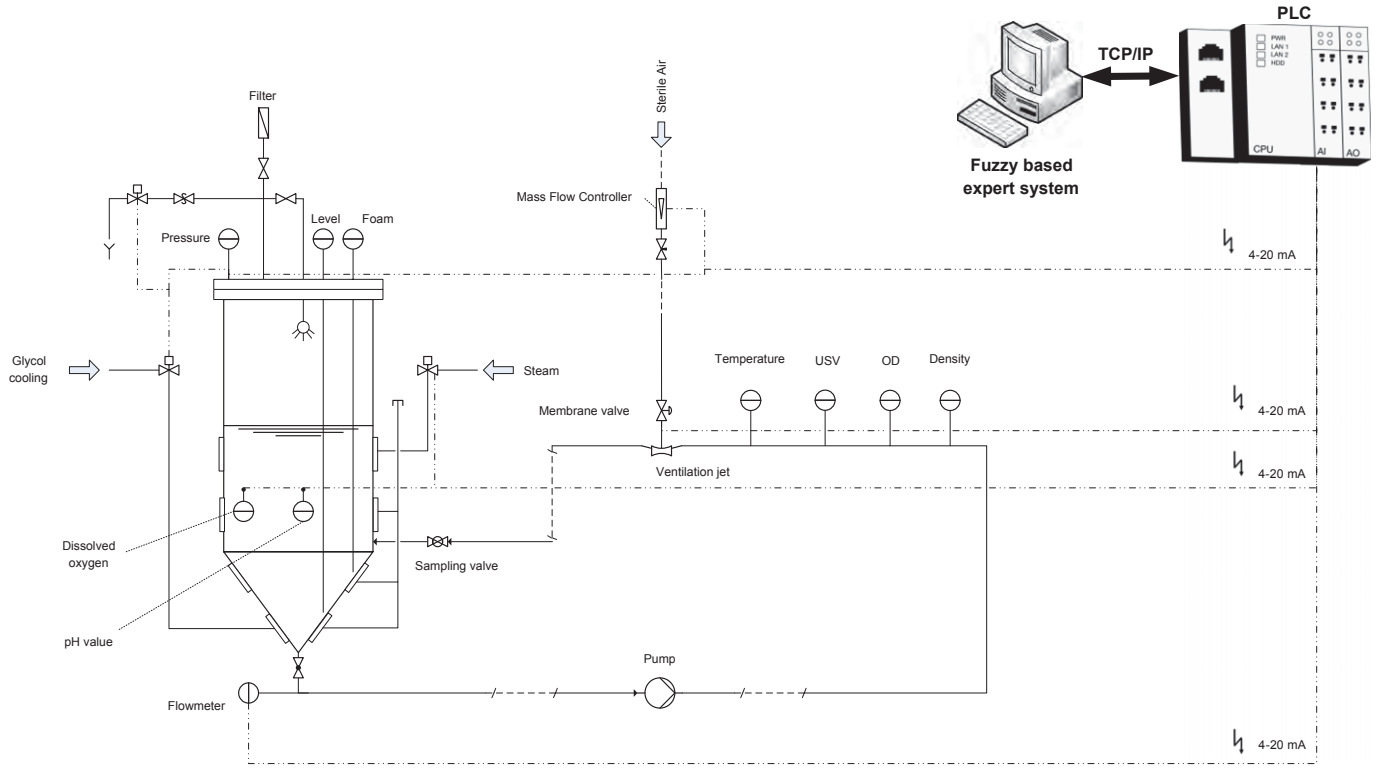


Fig. 1. P&I scheme and periphery of pilot propagation plant; the reactor vessel of 120 L total volume is equipped with level, foam and pressure sensors, dissolved oxygen and pH probes as well as temperature sensors. Tempering of the vessel can be done over cooling and heating jackets. Homogenization is reached using a pump in a piping circle including aeration via a jet (TurboAir, Co. Esau & Hueber) and ultrasound, density as well as turbidity measurement units. All instruments are connected electronically to a programmable logic controller (PLC, Co. Beckhoff), digital signals are pre-processed via Software TwinCat (Co. Beckhoff), and online data processing as well as Fuzzy Control is accomplished via Software VirtualExpert (Co. Gimbio).

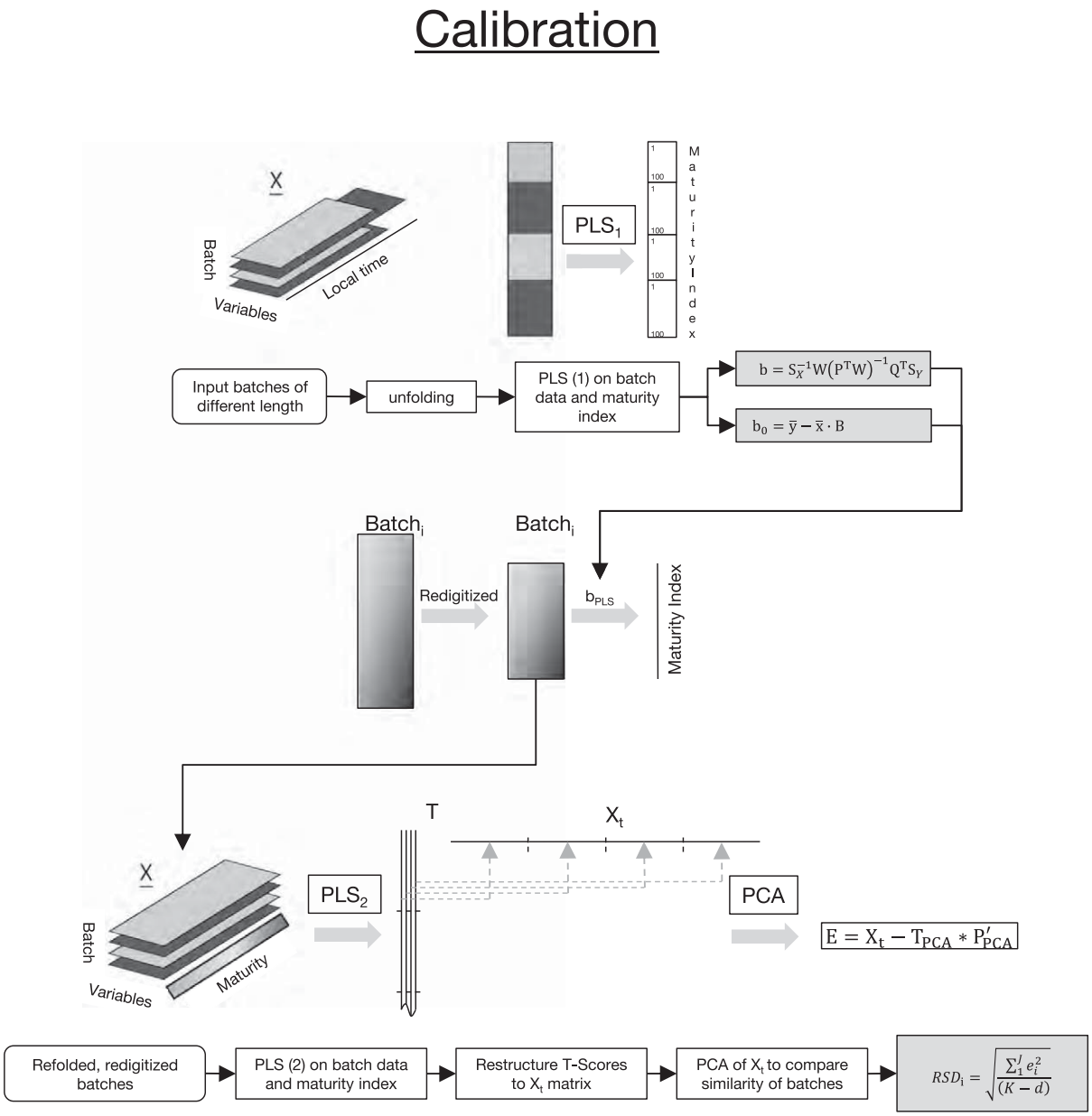


Fig. 2. Schematic explanation of MSPC method by S. Wold et al. [32] for calibration: (1) "unfolding" of batches an PLS regression on maturity of batches ➔ regression parameters for online prediction of maturity; (2) "re-digitalizing" of processes to equal data length (orientation on percent batch finished), subsequent PLS regression on re-digitized batches and maturity; and (3) reorganization of PLS-scores resulting in matrix $\boldsymbol{X}_t$, PCA on this matrix and using a certain number of components reaching residual matrix $\boldsymbol{E}$ ➔ calculation of RSD for individual batches as value for batch similarity; other variables are diagonal matrices $\boldsymbol{S}_x$ and $\boldsymbol{S}_y$ containing the respective standard deviations, matrix $\boldsymbol{W}$ with weighted loadings, matrix $\boldsymbol{Q}$ with loadings of targets $\boldsymbol{Y}$, $\boldsymbol{b}_0$ as intercept, vector $\boldsymbol{e}_i$ as residual for individual batch of certain score, J as counter for number of second PLS components, K the number of used variables (online-sensor data) and d the number of used PCA components.

the length from start- (inoculation with yeast) to endpoint (90–100 Mio·cells/mL). Furthermore, the following conversions of single sensor data were accomplished.

2.2.1. Turbidity

The turbidity sensor (Co. Optek, measuring in CU units) was calibrated to zero using water. Therefore, the turbidity value of wort (index wo) was subtracted from the online data (index on) in each batch (Eq. (1)).

$$OD_s = OD_{on} - OD_{wo}. \tag{1}$$

2.2.2. Density

The sensor (vibrating U-tube, Co. Centec) measures density and temperature. Each online value is transformed by Eq. (2), where $\rho(T_t)$ resembles the density at the temperature T measured at time t (index wa for water and m for medium, respectively).

$$\rho_s(T_t) = \frac{\rho_m(T_t)}{\rho_{wa}(T_t)}. \tag{2}$$

2.2.3. Ultrasound velocity

The sensor (Co. Centec) measures speed of sound and temperature. Each online value is transformed by Eq. (3), where $v(T_t)$ resembles the speed of sound at the temperature T measured at time t (index wa for water and m for medium, respectively).

$$v_s(T_t) = \frac{v_m(T_t)}{v_{wa}(T_t)}. \tag{3}$$

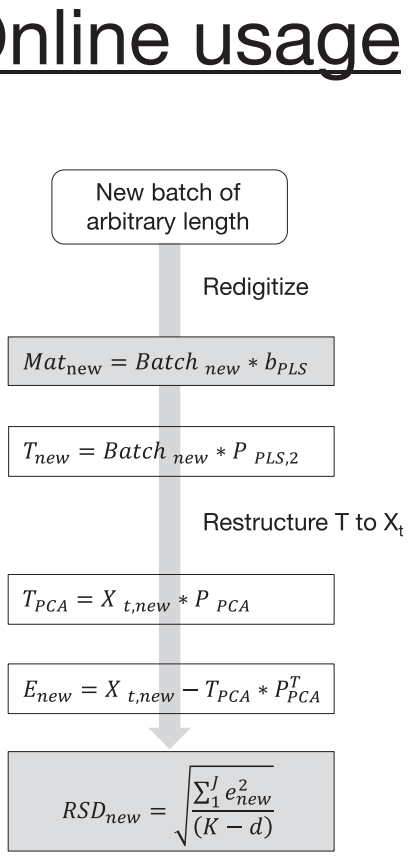


Fig. 3. Online usage of MSPC: the new batch will be re-digitized, maturity calculated using regression parameters $\mathbf{b}_{PLS}$ from first PLS, scores $\mathbf{T}_{new}$ calculated using loadings $\mathbf{P}$ from second PLS, resorting and calculation of Scores $\mathbf{T}_{PCA}$ using loadings $\mathbf{P}_{PCA}$ — finally, calculating residual matrix and the RSD value.

2.2.4. Dissolved oxygen

The oxygen sensor (Clark probe, Co. Mettler Toledo) was calibrated against air to measure the concentration of dissolved oxygen in mg/L. As concentration changes were influenced by system pressure and temperature next to microbial consumption, values were scaled from 0 to 100% using Eq. (4).

$$c_{O_2,\%}(T_t, p_t) = \frac{c_{O_2,mg/L}(T_t, p_t)}{c_{O_2 max,mg/L}(T_t, p_t)} * 100 \tag{4}$$

whereas $C_{O_2\,max,\mathrm{mg/L}}$ is calculated according to Henry's law

$$c_{O_2 max,\mathrm{mg/L}}(T_t, p_t) = H_{c,p}(T_t) * M_{O_2} * p_t * p_{O_2} * f \tag{5}$$

where M_{O_2} is the molar mass of oxygen (g/mol), p the total system pressure measured online (atm), p_{O_2} the partial pressure of oxygen (0.2095) and f a correction factor for the extract concentration (~dissolved solids, mainly sugar) of the medium (adopted from Annemüller et al. [26]). The extract concentration resembles the amount of dissolved material (mainly sugar). The constant $H_{c,p}$ is temperature dependent and calculated following Eq. (6).

$$H_{c,p}(T_t) = H_{c,p,s} * \exp\left(C * \left(T^{-1} - T_s^{-1}\right)\right) \tag{6}$$

where $H_{c,p,s}$ ($1.2296 * 10^{-3}$ mol/(L * atm)) resembles Henry's constant for standard conditions and C a constant (1895.8 K), both adapted from literature [27]. T_s resembles the temperature for standard conditions (298.15 K).

The values for pH (Co. Mettler Toledo), pressure and temperature (built-in temperature sensor of ultrasound setup) were left original.

Besides, any column of following matrices used for multivariate data analysis were pre-processed using auto-scaling (z-transform) [28,29].

2.3. Multivariate statistical process control (MSPC)

This method is used to create statistically supported process trajectories for process control [30,31]. There are a number of different possibilities to extract those trajectories out of a number of optimal processes and their sensor trends. Here, the approach is based on a method developed by Wold et al. [32,33], which is advantageous in case of adaption to the existing issue compared to other methods. The method is based on "unfold-PLS" (usage of PLS on unfolded three dimensional data matrices), where the interpretation of results is somewhat easier compared to other multi-way decomposition approaches used in multivariate data analysis [32–34].

Three different levels of process control can be extracted using this method. Those are individual observations (Level 1), the batch trace level (Level 2) and the batch level (Level 3) [32]. For the presented approach in this study only two levels were used. Level 1 (referred to as "maturity prediction", MP) and Level 3 (referred to as "residual standard deviation", RSD).

The calculations and algorithms can be found in Wold et al. [32]. The procedure as well as the most relevant equations is shown in Figs. 2 and 3. The used algorithms for PLS regression and model building were programmed and carried out with home-built subroutines programmed in MATLAB (Version 7 Release 14, The MathWorks, Inc., USA). A brief description of the algorithms can be found in literature [29,32,35–40].

2.4. Data post-processing

In the presented work seven sensors were used as inputs. Additionally, the inputs were extended by a full polynomial extension of second order including mixed terms. Therefore, a total of 35 inputs is available. Modelling each possible combination of those inputs results in 2^{35} possibilities. Since each single input is not of same importance to the modelled target of interest, a post processing algorithm called "variable importance in the projection" (VIP) is applied. The description of calculation can be found in various literature (e.g. Eriksson et al. Lee et al. or Chong and Jun [41–43], dependent on the method of decomposition). In the presented approach, a model for maturity prediction using al 35 inputs was established to investigate the importance of each variable. Since the mean squared VIP equals one, the generally applied cut-off criterion is "greater than one" [43].

2.5. Particle swarm optimization

One of the most famous algorithms in swarm intelligence is "particle swarm optimization" (PSO). This algorithm is based on the behaviour of birds or fish swarm [11]. The virtual members in such a swarm (particle or individual) consist of a binary code resembling a certain combination of sensors for one process trajectory. In each iteration, the individual model solution will be calculated. Afterwards, the cost function will be evaluated (explained in the next paragraph). This cost value will be compared with the local and the global best solution at this step to calculate the new velocity (v_i) for changing the position (p_i) of the particle, if necessary [see Eqs. (7) and (8)].

$$v_{i,p+1} = r_1 * w * v_{i,p} + c_1 * r_2 * \left(p_{i,best,local} - p_i\right) + c_2 * r_3 * \left(p_{best,global} - p_i\right). \tag{7}$$

$$p_{i,p+1} = p_{i,p} + v_{i,p+1}. \tag{8}$$

With inertia weight w (a fixed number typically around 0.9 or dynamic [19]), c_1 and c_2 as constants known as "self-confidence" and "swarm-confidence" [11] ($c_1 + c_2 = 4$ [19]) as well as r_1, r_2, and r_3 as uniformly distributed random numbers between zero and one. The random number r_1 multiplied to the inertia weight w was implemented to introduce random-dynamic change. The choice of those constants in this work was accomplished by iterations (see Results & discussion section). Furthermore, $p_{i,best,local}$ represents the local best solution or position of each individual particle and $p_{best,global}$ the global best solution of the swarm.

Those equations will be iteratively calculated on each particle until a breakup criterion is reached. The breakup criterion in this investigation was a certain number of iterations.

2.5.1. Constraints

Typically, the velocity of each particle in a swarm searching on a continuous n-dimensional search space is limited between $-v_{max}$ and v_{max}. Since the presented problem resembles a discrete search space, following constraints are applied after each other instead.

$$p_{i,k+1}(j) = \begin{cases} p_{i,k}(j) & \text{if } |v_{i,k+1}(j)| < 0.25 \\ p_{i,k}(j) + \dfrac{v_{i,k+1}(j)}{|v_{i,k+1}(j)|} & \text{otherwise} \end{cases}. \tag{9}$$

$$p_{i,k+1}(j) = \begin{cases} 0 & \text{if } p_{i,k+1}(j) \le 0 \\ 1 & \text{otherwise} \end{cases}. \tag{10}$$

Here, index k stands for the number of iteration, j the input number and i for the particle.

Further, the random number r_1 is inserted to create more flexibility and dynamical behaviour of the swarm. Preliminary results (data not shown) on the presented discrete problem showed an early convergence at local minima in case of fixed inertia weights.

2.5.2. Cost function

The cost function in this work consists of three parts. The first part ($cost_{i,1}$) resembles the full process evaluation in the corresponding data window compared to the historic process trajectory on the chosen number of inputs. Here, i resembles the particle, j the batch number, n the total number of batches and c stands for calibration.

$$cost_{i,1} = \frac{\mathrm{RSD}_{on}}{\sum_{j=1}^{n} \mathrm{RSD}_{c,j}/n}. \tag{11}$$

$$cost_{i,1} = \begin{cases} 1 & \text{if } cost_{i,1} < 1 \\ cost_{i,1} & \text{otherwise} \end{cases}. \tag{12}$$

The second part includes the single sensor evaluation. To evaluate the sensor input for the cost in each of the possible input combinations, each input data window was evaluated compared to the historic background. Therefore, the inputs of the five batches were z-transformed. The mean z-transformed trend as well as the 6σ band was used as valid data frame for each individual sensor. If a certain input violates the corresponding alarm limit more than a defined number of times (95% confidence interval, r_{95}), the cost of this input rises. This approach is adapted to the "Alarm limit violation similarity factor" presented in Johannesmeyer et al. [44].

Following the approach presented by Johannesmeyer et al. (2002), the inverse of the cumulative binomial distribution at a probability

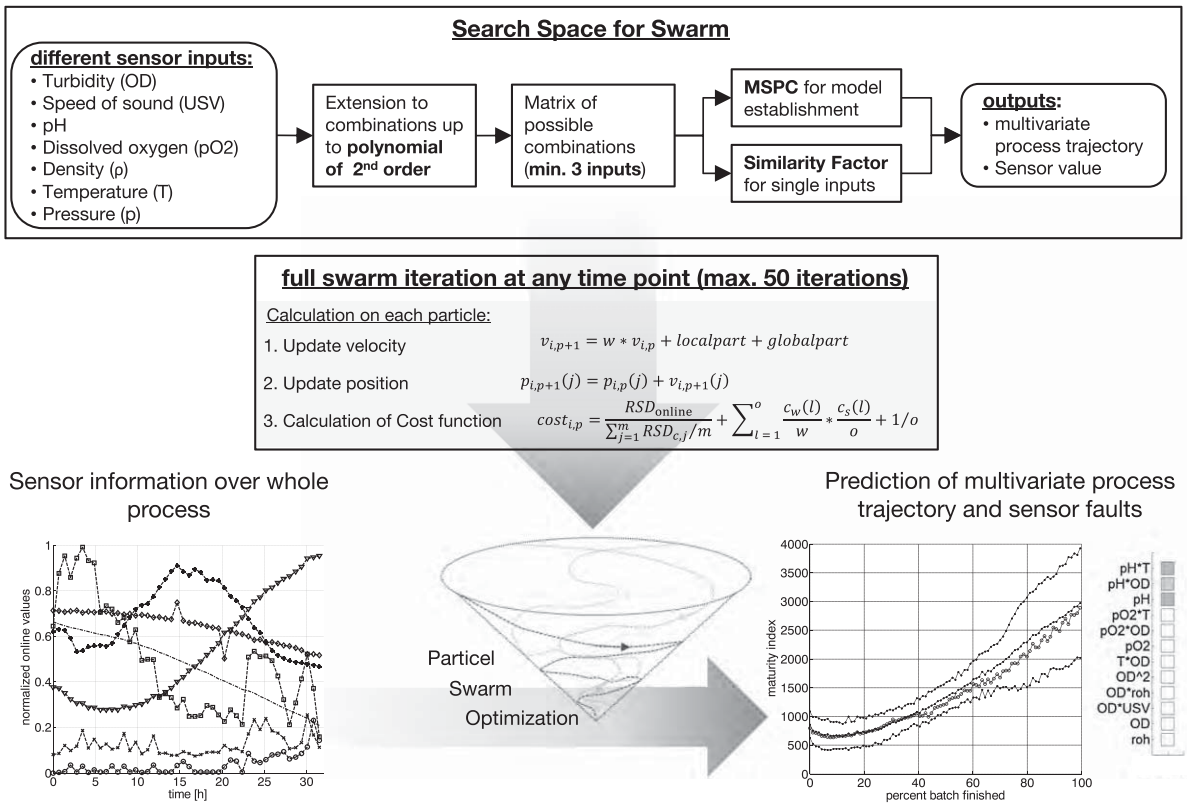


Fig. 4. Process monitoring platform and interaction with swarm sensing system to deliver a process corridor out of the most suitable sensor input combination together with sensor output validation; $\mathbf{v}_i$ = velocity of particle i (index $p+1$ = current iteration step, p = step before), w = inertia weight, p_i = position of particle i (for $j = 1$ to n number of inputs), RSD = residual standard deviation to historical batches, m = number of batches for calibration, index l = 1 to o number of inputs used in the current proposed model.

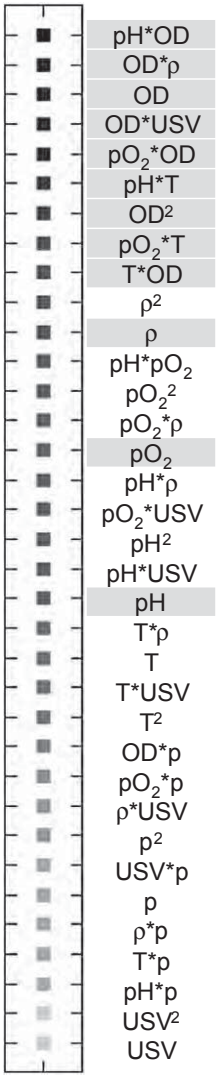


Fig. 5. VIP values in grayscale of the first PLS regression on maturity index using 35 inputs; the first nine terms are found by the "greater than one" rule; density, dissolved oxygen and pH were taken into the selection manually to include them in the evaluation of the swarm intelligent based system; *OD* — optical density/turbidity, ρ — density, *USV* — speed of sound, pO_2 — dissolved oxygen, T — temperature, p — pressure and pH — pH-value.

value of 0.95 is calculated first to define the "critical number of violations" [allowed hits of the boundaries Eq. (13)] [44].

$$r_{95} = \max l, \quad \text{such that} \quad \sum_{i=0}^{l} \binom{n}{i} * p^i * (1-p)^{n-i} < 0.95 \tag{13}$$

where l is the number critical number of violations, n the number of trials (length of the data frame) and p the probability value (0.025 for a two sided distribution).

Moreover, the number of violations of each sensor in each data frame is counted and divided by the maximum number of allowed violations (***counter***$_s$). Afterwards, a data frame counter for each sensor as well as the cost for the chosen sensor combination is calculated (Eqs. (14), (15) and (16)).

$$1.\quad counter_w(j) = \begin{cases} counter_{w-1}(j)+1 & \text{if } counter_s(j) > 1 \\ counter_{w-1}(j)-1 & \text{otherwise} \end{cases}. \tag{14}$$

$$2.\quad counter_w(j) = 0 \text{ if } counter_w(j) < 0. \tag{15}$$

$$3.\quad cost_{i,2} = \frac{1}{n} * \sum_{j=1}^{n} \frac{counter_w(j)}{w} * counter_s(j). \tag{16}$$

Here, the index/variable w resembles the data frame number, s for "sensor", i the particle, n the total number of sensors/inputs used and j the individual input.

The third part is introduced (Eq. (17)), since the chosen model should include as much valid inputs as possible for evaluating the sensor network. Otherwise, the swarm tends to stop at local optima with lower input number but similar low cost.

$$cost_{i,3} = 1/n. \tag{17}$$

Here, n stands for the number of inputs chosen.

Finally, the evaluation of each particle is calculated by summing all three parts (Eq. (18)).

$$cost_i = \sum_{l=1}^{3} cost_{i,l}. \tag{18}$$

Table 1
Ranges for parameter investigation via iteration.

	Min	Step size	Max
c_1	0.6	0.2	3.4
c_2	$=4-c_1$		
w	0.7	0.1	3.2
No. particles	10	2	32

3. Results & discussion

3.1. Swarm sensing idea

The aim in the present study was to create a swarm intelligence based system to evaluate sensor readings as well as process performance online. Additionally, the network should be capable to replace sensor readings in case of failure (drift, wrong calibration, total failure). The search space for this swarm was based on sensor readings extended to a polynomial of second order including mixed terms. Each possible combination of those generated search space dimensions was used to create multivariate statistical process control charts. Furthermore, single sensor evaluations with historical trends based on statistical areas were used. Both of the latter were used to prove the similarity of online data to historic background data. Those different models are scored by swarm intelligence (particle swarm optimization) leading to the optimal sensor/model combination at a certain time step. This leads to enlightening of the trajectory of fermentation in combination with sensor output validation. Those points include the challenge of vastly varying batch length as well as dealing with corrupted sensor inputs. The whole methodology is summarized in Fig. 4.

The results achieved in the present study are based on online data from five batches of aerobic yeast fermentation under brewing relevant conditions for calibration and two batches for validation. This number of background data is principally not enough in a statistical sense. Nevertheless, it is enough for showing the ability of the proposed method.

In the following paragraphs, choosing most relevant data for MSPC, adjustments of the settings for basic swarm algorithms as well as calibration and validation of the swarm for process and sensor evaluation are addressed.

3.1.1. Search space

The background of the search space is a multivariate linear and non-linear combination of sensor inputs. Starting with seven sensors (see Fig. 4), a full polynomial of second order including mixed terms

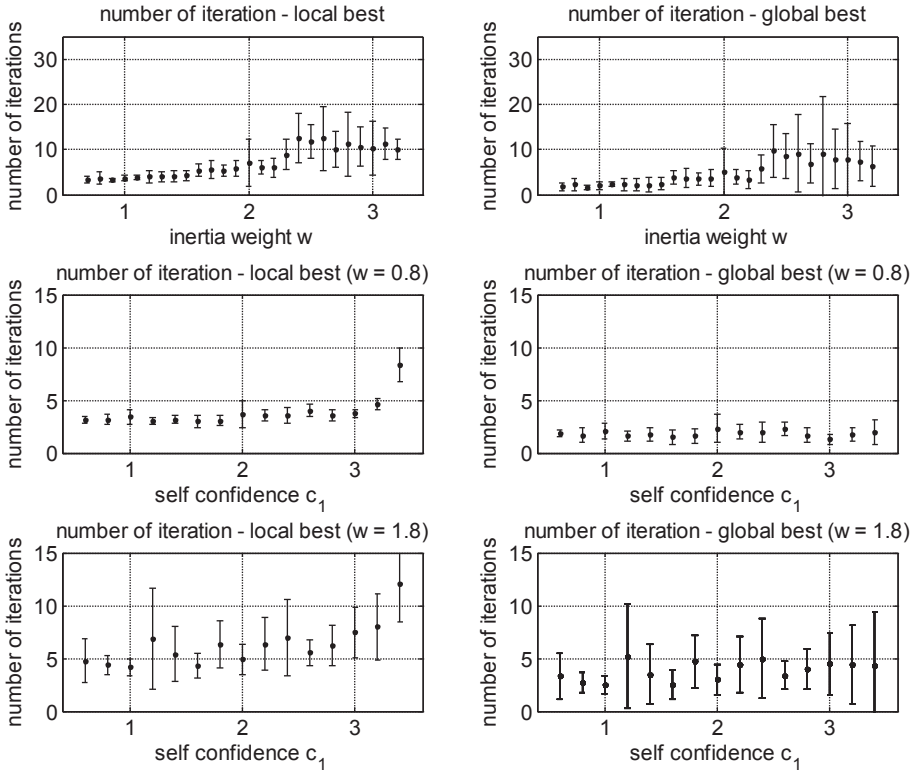


Fig. 6. Comparison between number of iterations reaching convergence to an individual local as well as a proposed global best solution with different values for parameters c_1, c_2, w and 20 particles; the data background were 10 data frames (50 data points, 50% overlap) out of a calibration batch; according to the number of iterations and the standard deviations, the upper two diagrams lead to the assumption, that neither $w = 0.8$ ($w \approx 0.4$) nor $w = 1.8$ ($w \approx 0.9$) but $w = 2.2$–3 due to higher flexibility (higher standard deviations) should be chosen (assuming, that higher number of iterations provides more solutions invested by swarm); the other 4 diagrams support this assumption to bigger values of w. Further, settings for c_1 seem to be reasonable between 1 and 1.5 as well as 2.2 and 3.2 (c_2 between 3 and 2.5 as well as 1.8 and 0.8, respectively).

gives 35 terms/inputs, leading to 2^{35} possible combinations of those inputs.

On the one hand, this causes a quite high computational effort. On the other hand, the importance or impact of each individual input is unclear. Therefore, preliminary investigations were conducted to find the most suitable input terms for further investigations. Here, the first step of the mentioned MSPC methodology was performed using the same five batches as calibration background.

The resulting matrices $\boldsymbol{P}$ (x-loading), $\boldsymbol{W}$ (weighted x-loading), $\boldsymbol{q}$ (y-loading) and $\boldsymbol{T}$ (x-scores) are used for an investigation based on variable importance in the projection (VIP) (e.g. [43]). This method is often used to prove the influence of a certain input variable on the problem of interest. In general, the mean of all VIP values is equal to one. Therefore, the rule of "bigger than one" is mostly applied. Under the conditions presented, this resulted in 20 important inputs and therefore 2^{20} possibilities. To further reduce the computational effort of the approach for online application, the cut-off value was set to 1.1 resulting in 12 inputs and ~4000 possible combinations. The result of this investigation is shown in Fig. 5. The values pH, turbidity and density or their combinations inside the selection do reflect the most informative values due to their high online stability and correlation to the progress of the fermentation. Besides, dissolved oxygen is quite important due to the faster growth of yeast under oxidative conditions. Pressure has no big impact from the biological process point of view, as long as it stays in the presented boundaries. Speed of sound is very sensitive to bubble interferences and therefore noisy. Additionally, influence of temperature on speed of sound is higher than concentration differences caused by the organisms' metabolism. Therefore, both values show a lower numerical impact. The highest influence on the duration of the biological process is given by temperature, which is included in mixed terms T^*OD, pH^*T and $pO_2{}^*T$. The only impacting term ($pO_2{}^*USV$) including speed of sound might be reasonable, since one influence on the speed of sound is the aeration resulting in gas bubbles. This fluctuation occurs in both online trends and might be diminished by the combination of them. Those results show the power of this data driven modelling approach in combination with process knowledge. Nevertheless, further investigations are necessary to fund those statements. Such investigations should include sensitivity analysis of inputs and the effect provided by polynomial extension with regard to the cost function to investigate their additional information. Those 12 inputs are taken for further investigations. The search space for the swarm is established out of all possible combinations of those inputs. This results in 2^{12} (4096) possibilities reduced by the combinations with less than three inputs involved. For each possibility,

Table 2
Settings for swarm after applying condition one on ten averaged results from ten data frames out of the calibration data; the minimum cost value was reached with zero standard deviation in all cases; P — Particles, No. P_c — Number of particles converged to global best solution, No. $I_{L/G}$ — Number of iterations for converged local/global best solution.

	No. P	c_1	c_2	w	No. P_c	No. I_L	Std.	No. I_G	Std.
1.	24	2.6	1.4	2.2	23.7	9.6	2.8	6.6	2.9
2.	26	2.4	1.6	2.4	25.4	14	9	10.3	10.2
3.	26	2.6	1.4	2.6	24.6	15.3	5.5	10.2	6.9
4.	28	2.2	1.8	2.2	28	10.2	6	7.9	6
5.	28	3	1	2.3	27.2	14.9	8.2	11.1	9.4
6.	28	1.8	2.2	2.4	27.9	9	3.3	5.8	3
7.	28	2.4	1.6	2.6	26.5	13.2	5.1	8.9	6.1
8.	30	2.6	1.4	2.5	29.3	10.3	2.9	6.7	3.3
9.	30	2.8	1.2	2.6	27.7	10.9	3.4	5.8	2.8
10.	30	2.4	1.6	2.8	27.4	11.4	2.8	6.5	3.2
11.	32	3	1	2.3	31.7	10.9	5.9	6.9	6.8
12.	32	3	1	2.6	29.7	13.1	3.4	6.7	3.2
14.	32	1	3	2.7	30.8	8.5	3.2	5	2.4

Table 3
Investigation on the effect of increasing number of particles for probability in finding best solution in each modelled case over the whole process; for calibration batch the model with n = 12 inputs is expected; increasing number of particles are followed by increased processing time on each data frame — the mean time on data frames is shown as Ø t; settings for the swarm: $c_1 = 2.4$, $c_2 = 1.6$ and w = 2.2.

No. of particles / No. of inputs	20	25	30	35	40	45	50	55	60	65	70	75	80
$n = 12$ [%]	58.9	62.8	70.7	76.4	79.6	81.9	83.5	85	86.4	87.7	88.6	89.4	90.2
$n = 11$ [%]	31.1	29.4	24.1	19.7	17.3	15.6	14.3	13.1	11.9	10.8	10	9.3	8.6
$n = 10$ [%]	7.8	6.7	4.4	3.3	2.7	2.2	1.9	1.7	1.5	1.3	1.2	1.1	1
$n = 9$ [%]	2.2	1.1	0.7	0.6	0.4	0.4	0.3	0.3	0.2	0.2	0.2	0.2	0.2
Ø t [s]	6.6	8.4	10.3	12	13.6	15.5	17.2	19	20.8	22.6	24.3	26	27.7

level 2 (maturity prediction for graphical interpretation) and 3 (residual standard deviation for cost function) are calculated. The number of PLS components (PLS regression) as well as principal components (PCA) were chosen on a minimum of 85% explained X-variance.

3.2. Parameter investigation

Values for the parameters of PSO algorithms can be taken from literature. Nevertheless, correct setting should be investigated using iterations on the current problem of interest. Therefore, parameters c_1, c_2, w and the numbers of particles itself were examined (Table 1).

The search space was used like mentioned above, results were achieved using 10 data frames (50 data points, 50% overlap) out of one batch from the calibration set.

First, iterative results on parameter settings were analysed compared to values of literature. A good start for parameters c_1 and c_2 at 2 was mentioned [19]. Values for inertia weight w were reported at 0.9 with reduction to 0.4 later on [19]. Assuming a high efficiency for reaching global minimum would be a comparably big number of iterations until convergence. Fig. 6 is showing results for constant settings of c ($c_1 = c_2 = 2$) and $w \approx 0.4$ and $w \approx 0.9$. Since the random number r_1 is inserted for dynamic issues, the values for w were approximated, assuming for a sufficiently high number of repetitions ($\gg 10$) to reach statistically ½ of the value multiplied with the random number. Therefore, $w = 0.8$ and $w = 1.8$ are the values of interest for comparison. The figure leads to the assumption, that due to higher flexibility the inertia weight should be chosen between 2.2 and 3. Further, settings for c_1 seem to be reasonable between 1 and 1.5 as well as 2.2 and 3.2 (c_2 between 3 and 2.5 as well as 1.8 and 0.8, respectively).

The results presented in Fig. 6 do not include the condition to find global minimum. Additionally, number of particles was arbitrarily chosen. Thus, the results were further analysed with the following conditions:

1. The minimum cost should be 1.0833 (see Cost function section; data from calibration have to be correct, all inputs should be used).
2. At least 90% of all particles have to converge to global best.
3. Minimum number of particles.

The results after applying condition one and two are shown in Table 2. The results in Table 2 emphasizes, that the treated task in this investigation need a higher self-confidence of the swarm. The average value for c_1 is 2.4 and for c_2 1.6. Moreover, the values for inertia weight tend towards 2.5. However, for final settings, condition three as well as the minimum numbers of iterations including the minimal standard deviation for each, local and global best solution are applied, the latter leads to minimum computational effort. Therefore, only setting one fulfils all definitions. Besides, those settings are consistent with the assumptions from Fig. 6. Therefore, they were taken for all further investigations.

3.3. Calibration of swarm

Calibration of swarm was established using the settings investigated. Further, the search space based on historic trajectories was prepared using five calibration batches. The data were analysed in frames of 50 data points and 50% overlap. After applying those settings on the calibration run, in more than 50% of the modelled cases the model with maximum number of inputs ($n = 12$) and more than 20% with one input lower ($n = 11$) over the whole process time was found. No sensor was indicated as false. Even though, none of the inputs is outside its statistical boundaries, the swarm does not reach the known global best solution in all cases. Thus, the expectations for an ideal case are

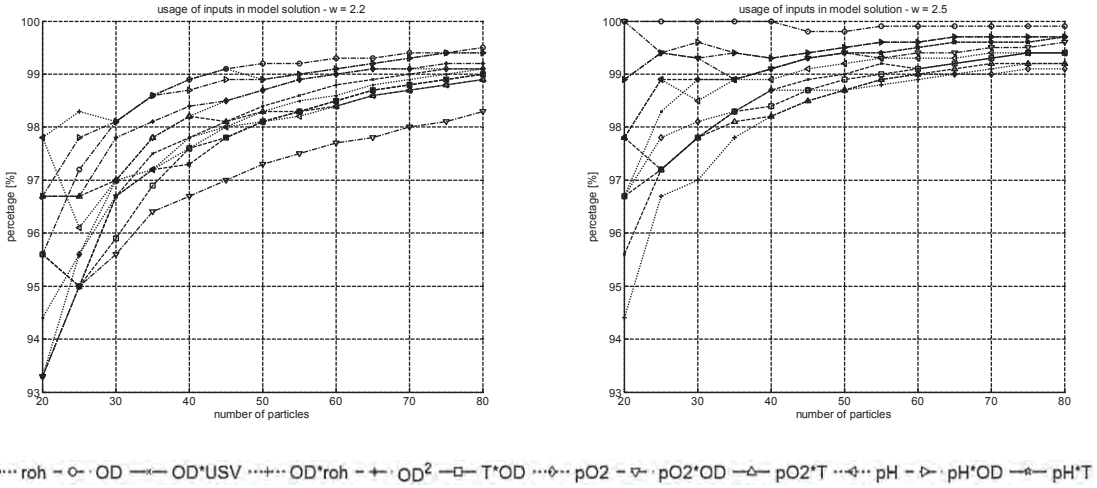


Fig. 7. Investigation on increasing number of particles on the inputs' usage for inertia weight $w = 2.2$ (left) and 2.5 (right) in the model solutions per iteration over the whole process; *roh* = density, *OD* = turbidity, *USV* = speed of sound, T = temperature, pO_2 = dissolved oxygen, pH = pH value as well as some mixed or squared terms.

Table 4
Investigation on the effect of increasing number of particles for probability in finding best solution in each modelled case over the whole process by increasing inertia weight w to 2.5; probabilities are given in %, the averaged processing time on each time point iteration in seconds.

No. of particles / No. of inputs	20	25	30	35	40	45	50	55	60	65	70	75	80
$n = 12$ [%]	73.3	81.7	84.1	86.1	87.8	89.4	90.8	91.7	92.3	93	93.5	94	94.4
$n = 11$ [%]	24.4	17.2	14.8	13.1	11.6	10	8.7	7.9	7.3	6.7	6.2	5.7	5.4
$n = 10$ [%]	2.2	1.1	1.1	0.8	0.7	0.6	0.5	0.4	0.4	0.3	0.3	0.3	0.3
Ø t [s]	6.6	8.4	10.1	11.7	13.5	15.3	17	18.6	20.4	22.2	24	25.6	27.3

not fulfilled sufficiently. In case of a good batch, a higher percentage of the known best model with maximum number of inputs should be reached. To reach this goal, one particle could be initialized with the best model case. This choice would lead to almost 100% usage but is in contrast to the sense of particle swarm optimization. Another solution could be the introduction of a part time memory, which helps the swarm to partially forget the global best solution and introduces a higher dynamic behaviour. This issue is partially integrated already by the inertia weight w. Additionally, this option could result in no convergence of the swarm at all, which is against its nature. The third proposed option is simply increasing the number of particles which increases the probabilities converging at global best solution at least until a certain limit of particles. The disadvantage is the increased processing time. The results of this investigation are shown in Table 3.

Not surprisingly those results show, that with higher number of particles the optimal solution will be found more often in all cases over the whole process time. They further show that despite the more

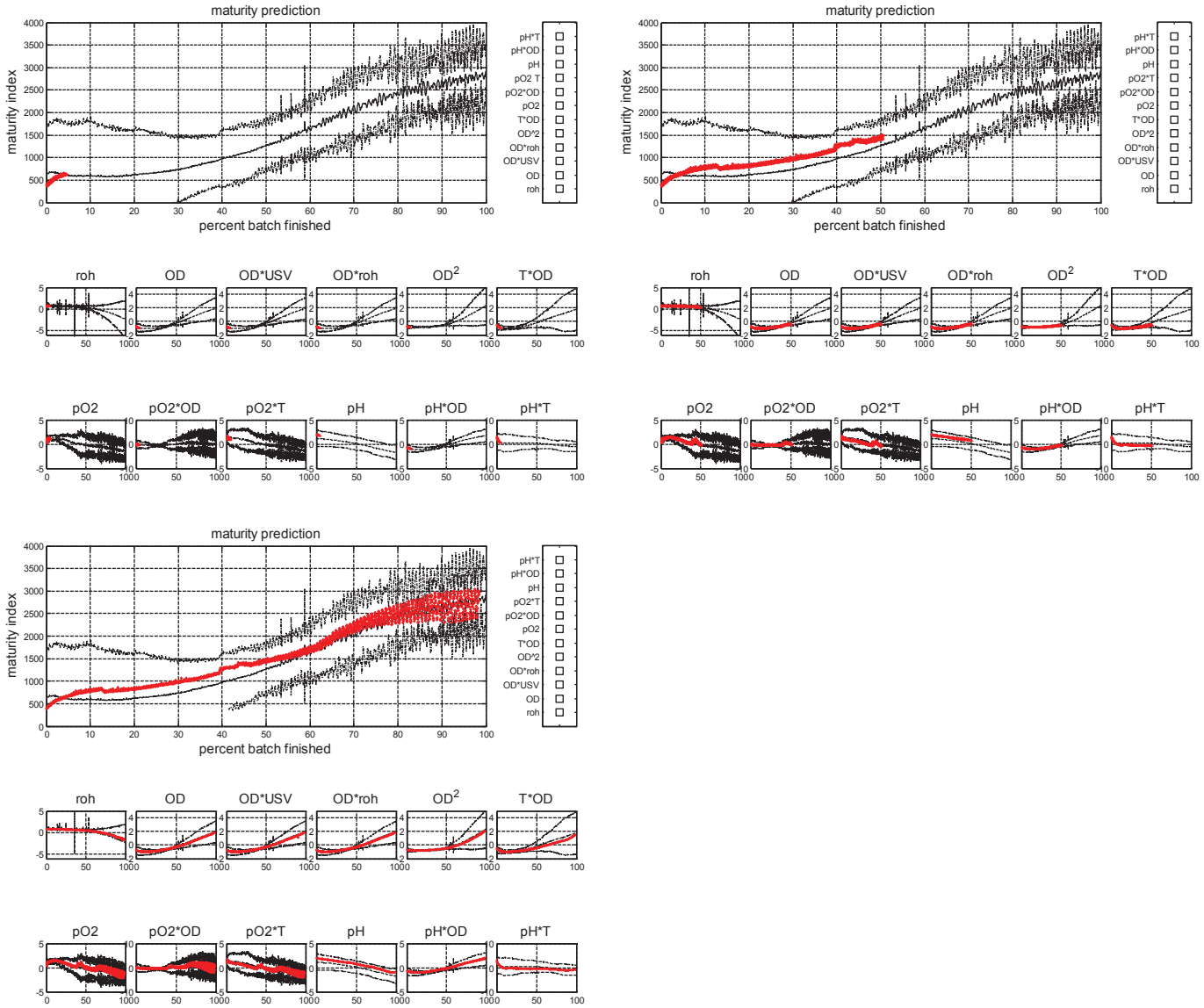


Fig. 8. Three time sectors for one calibration batch are shown even though the end of this batch (lower left) is fluctuation, no input value is indicated as corrupted (colour scale on the right of the trajectory). Moreover, each individual scaled input trend (small figures at the bottom) shows good similarity with historical data. It is also visible, that the process trajectory is in each frame identical, which indicates proper working of the swarm finding the model with maximum number of inputs. In 90.8% of the modelled cases 12 inputs were taken, in the other 8.7% with one input lower.

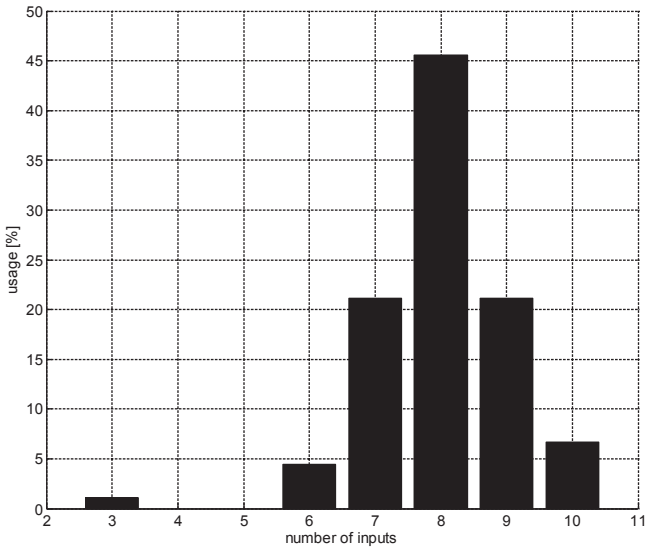


Fig. 9. Usage of inputs in swarms decision on model trajectory; most cases were modelled with eight inputs, second most are almost equally distributed on nine and seven inputs.

than three times bigger time needed per iteration for 80 in comparison to 25 particles, it is still in an acceptable range. The sampling frequency is around 50 s, the windows analysed with 50 data points and 50% overlap. Therefore, around every 20th minute, one output of the swarm would be necessary in online usage. Further, Fig. 7 shows the increasing usage of each input over the whole process time with increasing number of particles. This underlines the assumption, that higher number of particles increases the probability of finding the best model with maximum number of inputs.

The last trial for reducing the number of particles with respect to computational effort was increasing the value of inertia weight from 2.2 to the mean of 2.5 (see results Table 2). The results in Table 4 show a reduction to 50 particles with similar accuracy and ~1.5 times less processing time.

In addition, usages of individual inputs (Fig. 7, right) underline the efficiency of the swarm. Each input was used more than 98% over the whole process time. Therefore, those settings are taken for further analysis. The result of calibration for three time sectors is shown in Fig. 8. It is visible, that no sensor is indicated as corrupted (colour scale on the right of the trajectory). Besides, each individual scaled input trend (small figures at the bottom) shows good similarity with historical data. It is also visible, that the background process trajectory in each frame is identical. This indicates proper working of the swarm in finding the model with maximum number of inputs. Even though, a model with less number of inputs might be more feasible due to less computational effort as well as the reduced necessity of sensors in general, the aim of the proposed swarm sensing solution was twofold. Next to the more stable monitoring with a trustable online trajectory, the whole sensor network should be evaluated, regardless of the number of individual sensors at this point. In 90.8% of the modelled cases, 12 inputs were taken, in 8.7% one input lower. The trajectory always keeps the $3^*\sigma$ boundaries even in the noisy area at the end of the process. Moreover, each input was used more than 98% of the cases. This indicates the random "failure" of the swarm in choosing models with less inputs and therefore underlines the efficiency of the swarm.

All in all, those settings were used to investigate a batch with sensor faults in the following section.

3.4. Validation of swarm

Validation of swarm was established using the same settings as well as the search space like mentioned before. The data were similarly analysed in frames of 50 data points and 50% overlap. As a preliminary sensor failure testing, a batch, where the pH sensor failed was chosen. This failure was due to a mechanical damage and leakage of internal buffer solution. Therefore, the expected models found by the swarm should have maximally nine inputs. The decisions of the swarm applying the settings from before on the validation run resulted in around 20% of the modelled cases by using the maximum valid number of inputs ($n = 9$), around 45% with one input lower ($n = 11$) and around 20% with two inputs lower (results shown in Fig. 9) over the whole process time. Table 5 shows the usage of the 12 available inputs in the model choice of swarm decision over the whole process time. Furthermore, the amount of inputs' individual percentage outside the historical statistics is displayed.

The results presented in Fig. 10 underline the power of the presented method. In more than 20% of the modelled cases the decision of the swarm tends towards the model with maximum accurate inputs, each input related to pH is indicated as corrupted in the upper right colour scale. In around 65% of the cases eight and seven inputs are used. Similar results were achieved using 80 particles (no significant changes in the overall model choices, results not shown), which supports the chosen settings above.

Thus, the usage lower than 95% of some inputs can be explained by inputs related to dissolved oxygen are outside their boundaries in some cases (see also Fig. 10, small figures at the bottom). In addition, turbidity values leave their optimal region in minor circumstances. Furthermore, the number of inputs in the swarms' choice lower than nine is also valid. The differences in the usage percentage might be a higher correlation of individual inputs as well as slightly different RSD values in certain time frames.

4. Conclusion & outlook

It could be shown, that a discrete swarm with suitable parameter settings on a search space based on multivariate statistical process control charts is able to overcome sensor failures. It would be further possible to replace corrupted sensor readings by historical data based on swarm decision using the same approach but calibrating to the sensor value as target instead. Moreover, investigations regarding integration and stability are necessary. Furthermore, behaviour in the case of multiple failures are necessary. In addition, limits and numerical stability for bigger number of particles used on the respective task have to be investigated. Even though processing times presented are quite low, the investigations were performed with respect to an integration of the proposed approached on a microcontroller. Those devices have usually a comparably low processing power. This issue can therefore still not be neglected.

Nevertheless, the results indicate the possibility of more robust online monitoring using the swarm sensing idea for biotechnological processes to insure optimal and timely effective processing as well as sensor failure detection. The methodology was successful in predicting false input information. Furthermore, it was possible to still predict the progress of fermentation in a multivariate statistical process control

Table 5
Percentage of inputs used in multivariate trajectories over the whole process time. All of the valid inputs were used more than 95%.

	ρ	OD	OD^*USV	$OD^*\rho$	OD^2	T^*OD	pO_2	$pO_2{}^*OD$	$pO_2{}^*T$	pH	pH^*OD	pH^*T
Usage [%]	95,6	98,9	95,6	94,4	98,9	98,9	92,2	47,8	68,9	0	6,7	1,1
Outside statistics [%]	0	1,1	2,2	1,1	0	2,2	4,4	10	12,2	100	93,3	98,9

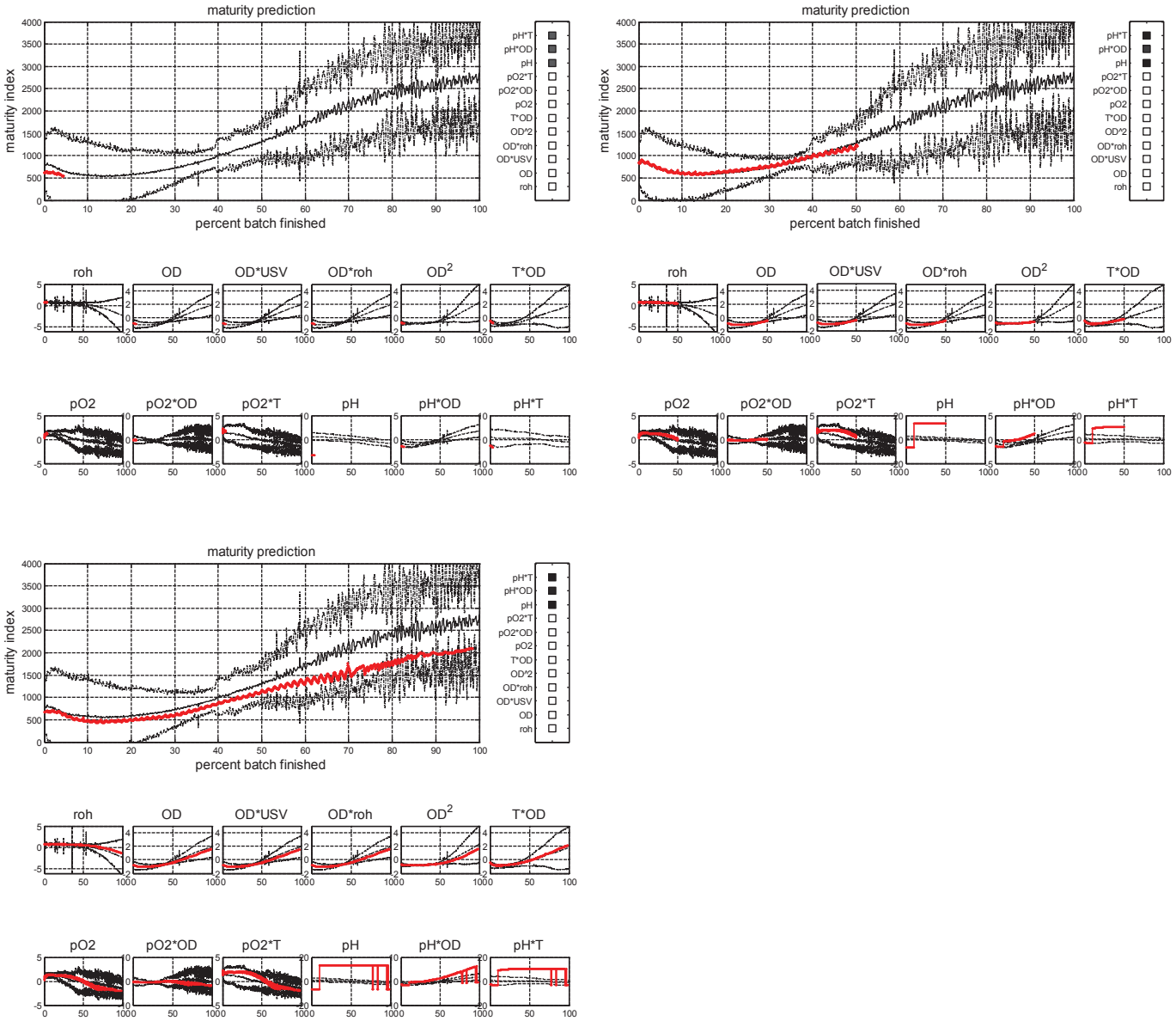


Fig. 10. Three time sectors for one validation batch are shown; three input values are indicated as corrupted (colour scale on the right of the trajectory, each of them combined with pH). The individual scaled input trends (three small figures at the lower right bottom) show the non-conformity with historical data. It is also visible, that the process trajectory changes between the frames. The swarm found the model with maximum valid number of inputs in more than 20%, more than 65% one or two inputs lower over the whole process time.

sense. The swarm supported over the whole process the control charts in validation run, in 85% the decision was towards the models with almost maximum inputs. In combination with the historical similarity of sensors it was possible to find false inputs in 100%. The presented methodology is not restricted to the number of sensor inputs as well as the use of specific sensor readings, which makes it beneficial over simple MSPC or other approaches. In the presented work, adjustments of the basic algorithms, cost function, accuracy of output as well as the dynamic behaviour of the swarm are addressed. Nevertheless, chosen cost functional need to be further investigated (e.g. weighting of individual parts). Furthermore, the shown results on the settings of the swarm indicate additional effort. Also, constraints and the chosen parameters inside (e.g. c_1 or c_2 for inter-particle communication to control the spread of relevant information) as well as the influence of window size and pre-processing of inputs have to be investigated. In case of the applied constraints, a comparison to other binary PSO algorithms such as presented by Chuang et al. (2009) has to be performed [45]. Besides, comparison to other multi-model fitting or model fusion methods as well as deterministic optimization methods or other evolutionary algorithms (e.g. GA or Bayesian Optimization Algorithm) as well as their multi-objective versions should be accomplished to see the functionality as well as the benefits according to computational effort and accuracy of the modified PSO version presented. Finally, comparable methods to the presented online monitoring approach to investigate the power like the price theory model need to be established [46].

Conflict of interest

There is no conflict of interest.

References

[1] T. Becker, B. Hitzmann, K. Muffler, R. Poertner, K.F. Reardon, F. Stahl, R. Ulber, Future aspects of bioprocess monitoring, (Book section), Adv. Biochem. Eng. Biotechnol. 105 (2006) 249–293.
[2] Bluetooth Sensornetzwerke, Hochschule_Fulda, 2008.
[3] K. Daniel, S. Rohde, N. Goddemeier, C. Wietfeld, Cognitive Agent Mobility for Aerial Sensor Networks, IEEE Sensors Journal, 2011. 1.

[4] H. Reichl, eGrain - Elektronischer Staub, Fraunhofer Magazin, 4 (2001) 22–23.
[5] C. Buschmann, S. Fischer, J. Koberstein, N. Luttenberger, F. Reuter, SWARMS — Software Architecture for Radio-Based Mobile Self-Organizing Systems, 2008.
[6] M.A. Sharaf, J. Beaver, A. Labrinidis, P.K. Chrysanthis, TiNA: a scheme for temporal coherency-aware in-network aggregation, Proceedings of the 3rd ACM International Workshop on Data Engineering for Wireless and Mobile Access 2003, pp. 69–76 (San Diego, CA, USA).
[7] S.R. Madden, M.J. Franklin, J.M. Hellerstein, W. Hong, TinyDB: an acquisitional query processing system for sensor networks, J. ACM Trans. Database Syst. 30 (2005) 122–173.
[8] C. Hase, Schwarmbasiertes Multipath-Routing in Sensornetzen, 2006.
[9] A. Abraham, H. Guo, H. Liu, Swarm intelligence: foundations, perspectives and applications, Studies in Computational Intelligence, Springer Verlag, Berlin/Heidelberg 2006, pp. 3–25.
[10] E. Bonabeau, M. Dorigo, G. Theraulaz, Swarm Intelligence From Natural to Artificial Systems, Oxford Univ. Press, New York, 1999.
[11] S. Das, A. Abraham, A. Konar, Swarm intelligence algorithms in bioinformatics, in: A. Kelemen, A. Abraham, Y. Chen (Eds.), Computational Intelligence in Bioinformatics, Springer Berlin Heidelberg 2008, pp. 113–147.
[12] M. Nicoletti, L. Jain, R. Giordano, Computational intelligence techniques as tools for bioprocess modelling, optimization, supervision and control, Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control, Springer Berlin, Heidelberg 2009, pp. 1–23.
[13] H. Zhang, Software sensors and their applications in bioprocess, in: M.d.C. Nicolletti, L.C. Jain (Eds.), Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control, Springer-Verlag, Berlin Heidelberg 2009, pp. 25–56.
[14] C. Wolf, S. McLoone, M. Bongards, Biogas plant control and optimization using computational intelligence methods, at-Automatisierungstechnik, 57 2009, pp. 638–649.
[15] Q. He, L. Wang, An effective co-evolutionary particle swarm optimization for constrained engineering design problems, Eng. Appl. Artif. Intell. 20 (2007) 89–99.
[16] F. Moussouni, S. Brisset, P. Brochet, Comparison of two multi-agent algorithms: ACO and PSO for the optimization of a brushless DC wheel motor, Intelligent Computer Techniques in Applied Electromagnetics, Springer Berlin, Heidelberg 2008, pp. 3–10.
[17] K.E. Parsopoulos, M.N. Vrahatis, Unified particle swarm optimization for solving constrained engineering optimization problems, Lect. Notes Comput. Sci 3612 (2005) 582–591.
[18] K. Sedlaczek, P. Eberhard, Using augmented Lagrangian particle swarm optimization for constrained problems in engineering, Struct. Multidiscip. Optim. 32 (2006) 277–286.
[19] Y. del Valle, G.K. Venayagamoorthy, S. Mohagheghi, J.-C. Hernandez, R.G. Harley, Particle swarm optimization: basic concepts, variants and applications in power system, IEEE Trans. Evol. Comput. 12 (2008) 171–195.
[20] J. Xiao, Z.-k. Zhou, G.-x. Zhang, Ant colony system algorithm for the optimization of beer fermentation control, J. Zhejiang Univ. Sci. 5 (2004) 1597–1603.
[21] R. Hassan, B. Cohanim, O. De Weck, G. Venter, A comparison of particle swarm optimization and the genetic algorithm, Proceedings of the 46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 2005.
[22] C. Blum, A. Roli, Metaheuristics in combinatorial optimization: overview and conceptual comparison, ACM Comput. Surv. (CSUR) 35 (2003) 268–308.
[23] T. Becker, D. Krause, Softsensorsysteme — Mathematik als Bindeglied zum Prozessgeschehen, Chem. Ing. Tech. 82 (2010) 429–440.
[24] S. Birle, M. Fellner, J. Lehmann, H. Wening, H. Kühnl, E. Pötzl, T. Becker, Yeast propagation manager (YPM), Brauwelt Int. 28 (2010) 26–29.
[25] S. Birle, M.A. Hussein, T. Becker, On-line yeast propagation process monitoring and control using an intelligent automatic control system, Eng. Life Sci. 15 (2015) 83–95.
[26] G. Annemüller, H.-J. Manger, P. Lietz, Die Hefe in der Brauerei Hefemanagement - Kulturhefe/Hefereinzucht - Hefepropagation im Brauprozess, 2 ed. VLB Berlin, 2008.
[27] R. Sander, Compilation of Henry's Law Constants for Inorganic and Organic Species of Potential Importance in Environmental Chemistry, Max-Planck Institute of Chemistry, Air Chemistry Department, 1999.
[28] D. Krause, W. Hussein, M. Hussein, T. Becker, Ultrasonic sensor for predicting sugar concentration using multivariate calibration, Ultrasonics 54 (2014) 1703–1712.
[29] D. Krause, C. Holtz, M. Gastl, M. Hussein, T. Becker, NIR and PLS discriminant analysis for predicting the processability of malt during lautering, Eur Food Res Technol (2014) 1–16.
[30] B. Lennox, G.A. Montague, H.G. Hiden, G. Kornfeld, P.R. Goulding, Process monitoring of an industrial fed-batch fermentation, Biotechnol. Bioeng. 74 (2001) 125–135.
[31] J.F. MacGregor, T. Kourti, Statistical process control of multivariate processes, Control. Eng. Pract. 3 (1995) 403–414.
[32] S. Wold, N. Kettaneh, H. Fridén, A. Holmberg, Modelling and diagnostics of batch processes and analogous kinetic experiments, Chemom. Intell. Lab. Syst. 44 (1998) 331–340.
[33] I.J. Whitehead, Soft Sensing — Using Multivariate Analysis for Yeast Propagation Monitoring, TU München, 2012.
[34] A. Smilde, R. Bro, P. Geladi, Multi-way Analysis. Applications in the Chemical Sciences, Wiley, Chichester, UK, 2004.
[35] W. Kessler, Multivariate Datenanalyse, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2007.
[36] H. Martens, T. Naes, Multivariate Calibration, Wiley, Chichester, UK, 1991.
[37] M. Mitzscherling, Prozeßanalyse des Maischens mittels statistischer Modellierung, Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt, Technische Universität, München, 2004.
[38] H. Risvik, Principal Component Analysis (PCA) & NIPALS algorithm, 2007.
[39] H. Wold, Causal flows with latent variables, Eur. Econ. Rev. 5 (1974) 67–86.
[40] D. Krause, T. Schöck, M.A. Hussein, T. Becker, Ultrasonic characterization of aqueous solutions with varying sugar and ethanol content using multivariate regression methods, J. Chemom. 25 (2011) 216–223.
[41] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, Multi- and Megavariate Data Analysis — Principles and Applications, Umetrics Academy, 2006.
[42] D. Lee, H. Lee, C.-H. Jun, C.H. Chang, A variable selection procedure for X-ray diffraction phase analysis, Appl. Spectrosc. 61 (2007) 1398–1403.
[43] I.-G. Chong, C.-H. Jun, Performance of some variable selection methods when multicollinearity is present, Chemom. Intell. Lab. Syst. 78 (2005) 103–112.
[44] M.C. Johannesmeyer, A. Singhal, D.E. Seborg, Pattern matching in historical data, AIChE J. 48 (2002) 2022–2038.
[45] L.-Y. Chuang, C.-H. Yang, C.-H. Yang, Tabu search and binary particle swarm optimization for feature selection using microarray data, J. Comput. Biol. 16 (2009) 1689–1703.
[46] P. Chavali, A. Nehorai, Managing multi-modal sensor networks using price theory, IEEE Trans. Signal Process. 60 (2012) 4874–4887.

3. Discussion

Data driven analysis of sensor and process data has tremendously increased over the last decade. This statistical driven area is very helpful in case of (partially) unknown physical or chemical background as well as multi-influenced conditions. Those drawbacks are typically present with bioprocess data.

The major goal of process analysis is to find and understand a causal relation between a measurement and a response. By using data driven approaches, causality is not always obvious. Therefore, it is always necessary to proceed corresponding data storage and processing together with knowledge inclusion [7]. Further, development of first principle solution is one aim in most applications. Nevertheless, those are not always reachable [7]. Another target can be to establish soft-sensors predicting not directly measurable data or the combination of online data with predictive mechanistic models. However, those topics do not often find the way to industrial application and thus have a lack in broad acceptance.

The present work is showing applications of newest multivariate data analytics used in sensor calibration and process data analysis including sensor network inspection of bioprocesses. Those fields are coping multidimensional temporal and spectral data including aspects of outlier analysis, variable importance, and model robustness.

The mentioned applications show, that it is hard to accomplish a data driven model solution with standard algorithms and without knowing the tuning parameters and surrounding conditions. Thus, single software solutions will not be capable in all cases to handle those multiple aspects. This implies knowledge of data and objectives as well as understanding of algorithms and should always be absolved with care and deliberation. This aspect is somehow also emphasized by Kourti, 2005 with respect to multivariate process control [64]. Kourti reports that “process knowledge is a must”, since it affects choice of weights and transformations of variables, for example [64].

The success of data driven modelling, including the mentioned feature analysis, is shown in the publications not included in this work as well [116, 119] and underlined by the additional results shown in this thesis. Those results, even though not yet accurate, show the general applicability and power of both, feature analysis for fluid inspection as well as data driven modelling with multivariate data analysis.

The used variable selection method can of course be enhanced. More investigations would be necessary to give a fundamental conclusion on the power of individual methods, but the tendency to general solutions and holistic approaches rather than solutions fitted to a single challenge and hard to transfer is shown by the numerous publications in literature and the possibilities shown in this work.

Another aspect, especially in supervised classification learning, could be the question if a feature is relevant for a learning algorithm rather than for the classification task [79]. This aspect implies the size of training set compared to number of features [79], also in combination with model robustness. This was not considered to full extend in this work and should be taken into account in future research. Additionally, pre-modelling variable selection has to be considered, and application such as model population analysis (MPA) seem powerful and should be investigated. Even though, each individual problem will have its own aspects, but generalization and wide applicability as well as more or less easy interpretation capabilities to solve many different challenges is always preferable.

General results

In the first part of this thesis, results with respect to fuzzy control and mechanistic growth modelling are presented. Even though not the core of the presented work, it is worth mentioning, that the aim of replacing simple linear objective function without the knowledge of the process might be not always acceptable, even though a fuzzy controller is able to deal with such. Not surprisingly, there were big deviations to the linear trajectories visible. Under the given circumstances, these might have even a positive impact on the quality of the product (yeast cells), since the controller forces fast growth by overshooting the ideal temperature in the beginning of the process and

undershoot at the end, resulting in cooling of the fermentation broth if the aimed number of cells is reached. Nevertheless, aim of further studies was to replace those functions by a more predictive controller based on numerical state estimation including the temperature dependent growth kinetics of *Saccharomyces sp.* (e. g. Birle et al., 2015 [118]). Further, the discussed minor influence of the Monod based growth velocity on the presented substrate decline were investigated based on oxygen induced growth limitations e. g. by McHardy, 2013 [117]). It was shown, that the used process systems are limited in oxygen distribution, but neglecting the aerobic part of glycolysis reaching an eight to 16 times higher ATP yield is not possible.

Empirical and data driven calibration of ultrasonic sensors

The validation error of maltose, ethanol or apparent extract concentrations published on the presented ultrasonic setup (first and second thesis publication) prove to be high and usually higher than a theoretically acceptable absolute deviation of +/- 0.5 g/100 g for online monitoring in brewing industry. Further, calibration background (binary mixture with only maltose or temperature spectrum as input) are not feasible neither for the more complicated product matrix (side effects of bubbles, CO_2 or ethanol concentration) nor online application of the respective device. The published validation errors and used calibration background predicting maltose or apparent extract are summarized in Table 3.1.

Table 3.1: summary of published validation errors on sugar concentration predicted by different multivariate calibration models on ultrasonic signal properties

	Target	Model background	Inputs	boundaries	Error (RMSEV)	Limitations	publication
1.	Sucrose (Ethanol)	PLS, PCR	Temperature spectrum of USV	0-12 (0-6) g/100g; 2-30 °C	0.5 g/100g (0.18 g/100g)	Not online	[116]
2.	Apparent extract	PLS	Phase and magnitude spectrum (frequency domain)	maltose 0-12 g/100g ethanol 0-3 g/100g 6-22 °C temp 6-22 °C	< 0.5 g/100g	simple linear interpolation between temperatures	First thesis publication
3	Maltose	PLS + polynomial regression	Temporal and spectral Features	maltose 2-12 g/100g temp 10-21 °C	0.64 g/100g	Only binary calibration samples	Second thesis publication

The influences reported causing these deviations are signal sensitivity issues, possibly causing lack of accuracy in regions of lower sugar concentration, setup material and setup design, near field and superposition phenomena due temperature effects as well as electronic circuit adaptions. These more physical and setup specific aspects are investigated and discussed in more detail by Hoche *et al.* [13, 120, 121], aiming at the known physical relationship between density and speed of sound to predict concentrations of maltose and ethanol. Even though successful and definitely proving the concept in specific aspects, the used materials (PVDF, PMMA, PEEK), for example, are not fully accepted in food related applications (temperature limits or chemical resistance compared to steel; possibility to contain bisphenol-A (PMMA)). The results presented in Krause *et al.*, 2011 [116] are not included in this thesis, since they are not very advantageous with respect to online application. Nevertheless, it is worth mentioning, that it could be shown to predict both concentrations (substrate and product) independently by using just one measuring device.

The other drawbacks mentioned in this thesis such as binary mixtures, model stability and robustness as well as non-linearity of the presented approach are preliminarily accomplished in further studies, described in the following paragraphs.

a. Considerations of Non-linearity

The second thesis publication presents the prediction of sugar, respectively maltose concentration in binary mixtures with varying temperature. The first obvious non-linearity is visible in temperature dependence. The effect of additional dissolved ethanol (ternary mixture) introduces another probable non-linear effect. A preliminary investigation on three temporal US features and support vector machines showed that it is possible to

cluster binary from ternary mixtures with respect to a certain concentration. It was further visible, that support vectors were similar and deviating in a non-linear form with temperature (Figure 3.1).

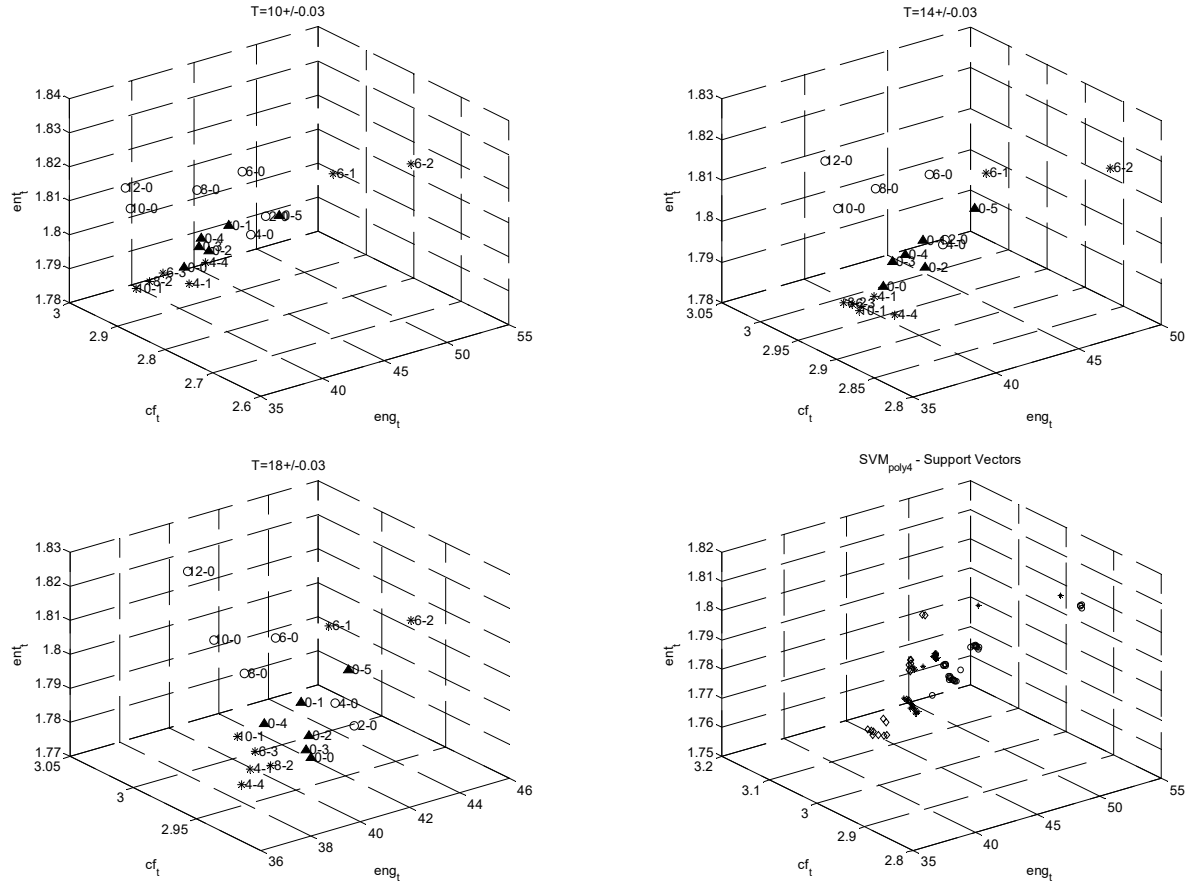


Figure 3.1: Preliminary trial of support vector clustering on three temporal US features to discriminate binary maltose from binary ethanol and ternary mixtures; upper left and right as well as lower left figure display the mixtures at different temperatures, circles for binary maltose, star for ternary and triangle for binary ethanol mixture; lower right figure displays the support vectors for three different temperatures, circle for 10 °C, star for 14 °C and diamond for 18 °C

Therefore, it is most promising to use support vector regression to establish a continuous calibration model predicting maltose or ethanol concentration. These algorithms are used to enhance the presented output from the second thesis publication. Therefore, the extended dataset of ternary mixtures over the temperature range of 10 to 20 °C from the experimental setup presented in the chapter "material and methods" of the first and second thesis publication was used as basis.

b. Support vector Regression

For the first investigation, a sample set of concentrations between zero and 14 g/100g maltose and zero to 5 g/100g ethanol at an almost constant temperature T = 10 °C (+/- 0.15 °K) was extracted. The predictor matrix **X** consisted of three temporal features (presented in the

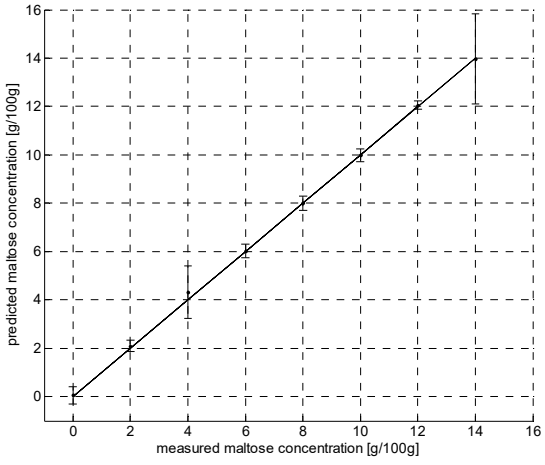


*Figure 3.2: parity plot for measured vs. predicted maltose concentration using LS-SVR with RBF kernel ($RBF_{\gamma = 143.53,\ \sigma 2 = 0.19}$), T = 10 °C; error bars presented resemble the 2*σ of each concentration level; except for 14 and 4 g/100g, the prediction accuracy is very good compared to the investigation using standard PLS regression*

preliminary investigation above), namely energy, entropy and crest factor. Three of the most popular kernels were tested (Table 3.2). The most promising accuracy is given by using the RBF kernel, where a validation error of 0.33 g/100g was reached. The parity plot is shown in Figure 3.2, presenting the prediction of maltose concentration.

This data set was analyzed with a toolbox (LS-SVM v1.8, Suykens, Leuven, Belgium) in Matlab R2010a (The MathWorks, Inc., Natick, USA).

Table 3.2: prediction accuracy presented by RMSEC and RMSEV for calibration with LS-SVM using different kernel functions; the dataset consists of three temporal features (entropy, energy and crest factor) from US signals on samples with concentrations between zero and 14 g/100g maltose (2 g/100g steps) and zero to 5 g/100g ethanol (1 g/100g steps) at an almost constant temperature T = 10 °C (+/- 0.15 °K); the linear kernel just has one parameter (γ), the polynomial kernel additionally the intercept (b = 2.652) and the polynomial degree (p = 6), the RBF kernel the variance (squared standard deviation, $\sigma^2 = 0.19$); the calibration data set contained $n_c = 116$, the validation set $n_v = 115$ samples

	γ	RMSEC	RMSEV
linear	5.72	2.97	3.14
poly	0.37	0.38	0.48
RBF	143.53	0.26	0.33

The parity plot for the model using LS-SVR with RBF kernel ($RBF_{\gamma = 143.53, \sigma2 = 0.19}$) for the dataset at T = 10 °C with a prediction accuracy shown by the error bars is superior in comparison to the investigation using standard PLS regression, except for 14 and 4 g/100g. Taking out these two levels of concentration, the results change as shown in Table 3.3. The most promising accuracy is given again by using the RBF kernel, where a validation error of 0.13 g/100g was reached. The increase in the prediction performance is quite noteworthy and justifies the reduction of the inputs by these two measuring points.

Table 3.3: prediction accuracy presented by RMSEC and RMSEV for calibration with LS-SVM using different kernel functions; the dataset consists of three temporal features (entropy, energy and crest factor) from US signals on samples with concentrations between zero and 12 g/100g maltose (2 g/100g steps, without 4 g/100g) and zero to 5 g/100g ethanol (1 g/100g steps) at an almost constant temperature T = 10 °C (+/- 0.15 °K); the linear kernel just has one parameter (γ), the polynomial kernel additionally the intercept (b = 8.6581) and the polynomial degree (p = 5), the RBF kernel the variance (squared standard deviation, $\sigma^2 = 0.17$); the calibration data set contained $n_c = 92$, the validation set $n_v = 92$ samples

	γ	RMSEC	RMSEV
linear	5.64	1.57	1.56
poly	30.44	0.16	0.31
RBF	2547	0.04	0.13

Additionally, the parity plot is shown in Figure 3.3. Furthermore, the whole approach was tested for other temperatures and for ethanol concentration and has proven similar good results.

Even though, the results of SVR model show very promising accuracy, a deeper investigation on the problem of interest and the interpretability of interim results of algorithm used is not as simple as described for typical PLS results discussed in the sections above. Thus, the approach discussed in the following section was realized.

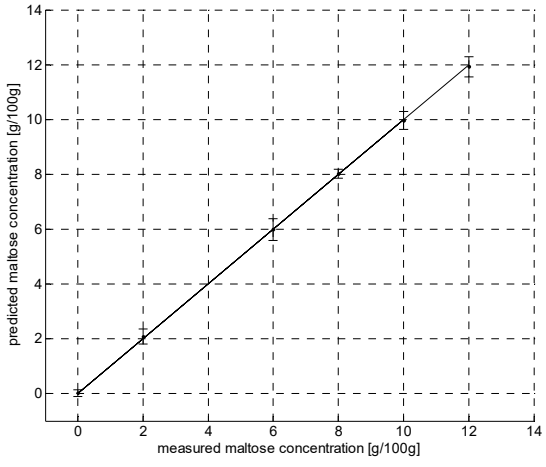


*Figure 3.3: parity plot for measured vs. predicted maltose concentration using LS-SVR with RBF kernel ($RBF_{\gamma = 2546.96, \sigma2 = 0.17}$), T = 10 °C; error bars presented resemble the 2*σ of each concentration level; the prediction accuracy is increased compared to the investigation before*

c. Kernel PLS

Another possibility in modelling non-linearity is given by kernel pre-processing as described in section 1.2.1.6 and PLS as mentioned in

section 1.2.2.2 “Multivariate data analysis for regression”. The results using these algorithms and the same dataset as reported for SVR one paragraph before are summarized in Table 3.4. Even though lower than for the LS-SVR model, the most promising accuracy is again given by using the RBF kernel, where a validation error of 0.23 g/100g was reached. Consequently, the parity plot is given in Figure 3.4, presenting the prediction of maltose concentration. The number of components were chosen on the Akaide Information Criterion (AIC, shown in the right figure of Figure 3.4), which is more suitable than simple RMSE since it takes model complexity by number of components or latent vectors (LV) into account. The error bars in the parity plot for the model using kernel-PLS with RBF kernel ($RBF_{\sigma^2=0.17}$), for the dataset at T = 10 °C, show very good prediction accuracy. The result is still quite good compared to the investigation using standard PLS regression. This approach also works for other temperatures and for ethanol concentration in a similar good manner.

Table 3.4: prediction accuracy presented by RMSEC and RMSEV for calibration with Kernel-PLS using different kernel functions; the dataset consists of three temporal features (entropy, energy and crest factor) from US signals on samples with concentrations between zero and 14 g/100g maltose and zero to 5 g/100g ethanol at an almost constant temperature T = 10 °C (+/- 0.15 °K); the linear kernel just has one parameter (γ), the polynomial kernel additionally the intercept (b = 8.66) and the polynomial degree (p = 5), the RBF kernel the variance (squared standard deviation; $\sigma^2 = 0.17$); the calibration data set contained nc = 92, the validation set nv = 92 samples

	LV	RMSEC	RMSEV
linear	3	2.97	3.14
poly	11	0.64	0.63
RBF	28	0.15	0.23

Both kernel methods show the possibility for predicting concentrations out of binary mixtures at single temperatures. Even though, the used number of LV is high, the less complex kernel-PLS shows a comparably acceptable error taking into account that only three temporal features were taken as inputs. This raises the following issues:

- Is it possible to include temperature influences?, and
- What features are relevant?

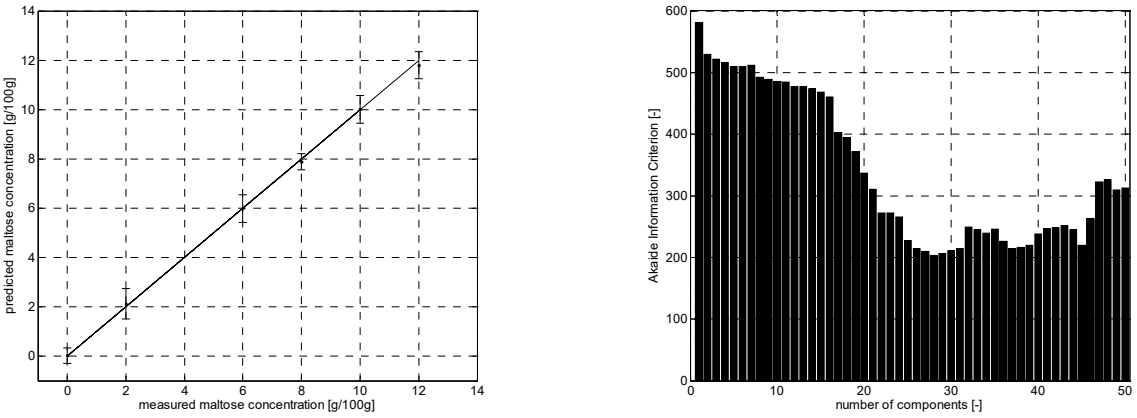


*Figure 3.4: (left) parity plot for measured vs. predicted maltose concentration using kernel-PLS with RBF kernel ($RBF_{\sigma^2=0.17}$), T = 10 °C; error bars presented resemble the 2*σ of each concentration level; except for 14, 4 and 0 g/100g, even though less accurate than the LS-SVM solution, the prediction accuracy is still acceptable compared to the result presented in the second thesis publication (where only binary mixtures were analyzed); (right) presentation of decreasing AIC by increasing number of components*

The second point can be investigated by variable selection (section 1.2.3.1” Variable selection/inspection”) or by simply iterating on the inputs and their combination for the best model output. The latter is more beneficial, if the inputs are altered, which is the case when using kernels.

For the first issue, literature presents a methodology called external parameter orthogonalisation (EPO, see Roger *et al.*, 2003 and 2004 [47, 122]). This potential method tries to divide the input space in two subspaces, one with useful information with respect to the target of interest and one with the information altered by external parameters. Beneficially, the approach tries to find subspaces without relation to the respective target Y and does

not need the response to be kept constant. In the basic EPO algorithm, it is not necessary to use accurate values of the respective external parameter as responses [47]. Roger *et al.* defines, amongst others, two main correction strategies – availability of external parameter (e. g. measured) or not [47].

d. External Parameter Orthogonalisation (EPO)

The presented approach was used on the data set of 6000 samples with concentrations between two and 12 g/100g maltose, one and 5 g/100g ethanol and temperatures between 10 and 20 °C. Firstly, the same inputs as presented for LS-SVM und kernel-PLS were investigated. EPO was performed using PCA, the final model building using PLS. The EPO reduction of the input matrix resulted in a better prediction error, but still in an unacceptable magnitude. Thus, the whole feature space of 12 features was investigated iteratively to find the best input combination. Furthermore, model building on the EPO reduced input matrix was accomplished by kernel-PLS. The iteration was performed on:

- steps of temperature for EPO from 0.5 to 4 °K in 0.5 °K step size
- number of chosen EPO components (maximum number limited to 99.99% explained X-variance)
- number of chosen PLS LV based on the AIC

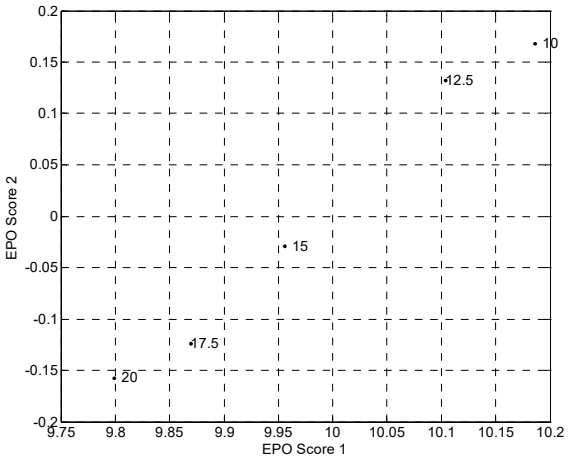


Figure 3.5: Score plot of EPO investigation on the dataset of step size 2.5 °K and all samples with concentrations between zero and 14 g/100g maltose and zero to 5 g/100g ethanol; it is visible, that the first component is enough to clearly distinguish between the different

This investigation resulted in the combination of features highest magnitude in the frequency representation, spectral kurtosis, skewness, entropy, centroid as well as temporal energy and entropy. Furthermore, the best temperature step for EPO was 2.5 °K with a final model validation prediction error of 0.73 g/100g, three EPO components and seven latent PLS vectors. The concentrations in the data set were randomly distributed - half of the samples were taken for calibration, the others for validation. For EPO investigation, a calibration set of temperatures with step size 2.5 °K was extracted from the full data set. For the proof of EPO concept (inputs were chosen according to the output of iteration result) the presented algorithm of section 1.2.1.7 was followed; the result is shown as score plot in Figure 3.5. Thus, it is visible, that one EPO component seems to be sufficient in discriminating between the different temperature levels.

e. Final model building including robust calibration

Additionally to EPO pre-processing, the iteration output from the analysis before was investigated by kernel-PLS using RBF kernel. These algorithms are combined with the model robustness scheme in section 1.2.3.3 and the over fitting issues presented in section 1.2.2.3 to support the choice of final model size. Next to the question of the number of EPO components, the issue of model size by number of LV has to be investigated. Thus, the mean prediction error for validation (RMSE) and the corresponding standard deviation together with the number of EPO components for reducing the effect of temperature as external parameter in the input data are compared (Figure 3.6). The figure indicates, that two EPO components are obviously too much for the presented case (error as well as standard deviation are mostly higher). Although the differences between original and reduced data set by one EPO are quite low, there are better accuracies achieved around eight to 10 and 14/15 LV by the latter.

Furthermore, nine LV should be ideal considering the standard deviation, since the deviation rises for both data set afterwards, even though the error decreases after 12 LV again. The choice of a model with lower prediction accuracy but better standard deviation of prediction error for robustness reasons is favorable.

Investigating the proposed measures for validity and precision explained in section 1.2.2.3 “Model Validation” also supports those choices (Figure 3.9 at the end of this chapter). The ratio MS_{rc}/MS_{rv} (validity calibration/validity validation) decrease with rising number of components due to overfitting, ratio MS_{av}/MS_{rv} (precision/validity) indicate by convergence no necessity to include more components or latent vectors. Further, $MS_{av}(p-1)/MS_{av}(p)$ (precision) and $MS_{rv}(p-1)/MS_{rv}(p)$ (validity) converge or fluctuate around one starting at a certain model size, which again supports the assumption from before. In the separate plots of MS_{av} (precision) and MS_{rv} (validity) a (local) minimum would indicate the preferable number of latent vectors. Thus, in each plot a number of latent vectors between eight and 10 would be enough.

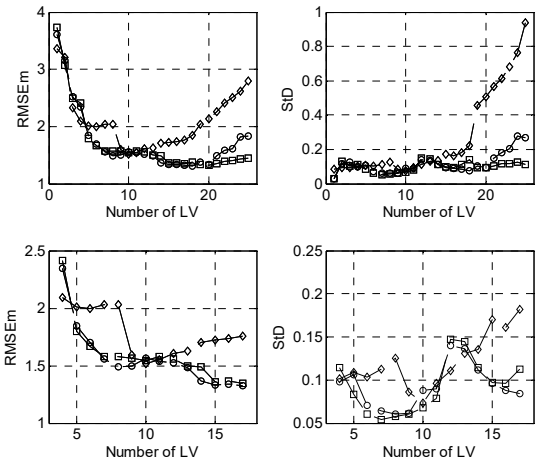


Figure 3.6: comparison between mean prediction error (left side) and their standard deviation (right side) for original data set models (squares), data set reduced by one EPO component (circles) and reduced by two EPO components (diamonds); the top two figures show plot from one to 25, the figures at the bottom from three to 18 LV; two EPO components are obviously too much for the presented case (error as well as standard deviation are mostly higher); although the differences between original and reduced data set by one EPO are quite low, there are better accuracies achieved around eight to 10 and 14/15 LV by the latter. Even though the error decreases after 12 LV again, nine LV should be ideal considering the standard deviation, since the deviation rises for both data set afterwards.

Finally, using one EPO component on the input data results in the lowest mean sum of squares for precision indicating one EPO component as sufficient. This choice is underlined by Figure 3.7, where each model is compared in the pre-chosen area between five and 16 LV. Inspection of parity plots additionally also supports the assumptions from before, since precision with only one EPO component in higher concentration levels are slightly better whereas the validity difference between one and two EPO components are much less (Figure 3.10 at the end of this chapter).

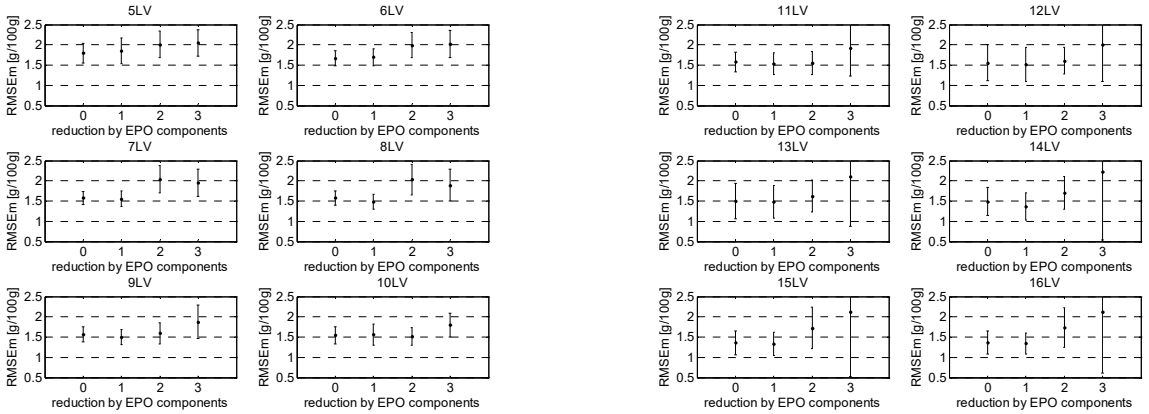


*Figure 3.7: to underline the choice from above, each model is compared individually in the pre-chosen area plotting residual error against number of used EPO components; error bars = three times standard deviation (3*σ); due to robustness, the choice would be the same as for the figure before (slightly higher error but lower standard deviation); thus nine LV and one EPO component seem to be suitable*

To support the choice of latent vectors once more, visual inspection of error statistics are shown in Figure 3.8. Even though their might be slight evidence of better validity as well as better precision in some levels and the figure indicates a slight better shape of residual distribution for 15 LV, the normal probability plot for nine LV is not deviation as much as the one for 15 LV in the boundary areas of the data. This might indicate overfitting in case of 15 latent vector model. Single level statistics as well as individual standard deviations of levels/samples

are shown in appendix A.5 for nine and 15 LV, visually proving the assumption of nine LV as sufficient (some individual distributions with 15 LV deviate stronger). Nevertheless, those could be used for further analysis with respect to error-prone or outlying samples.

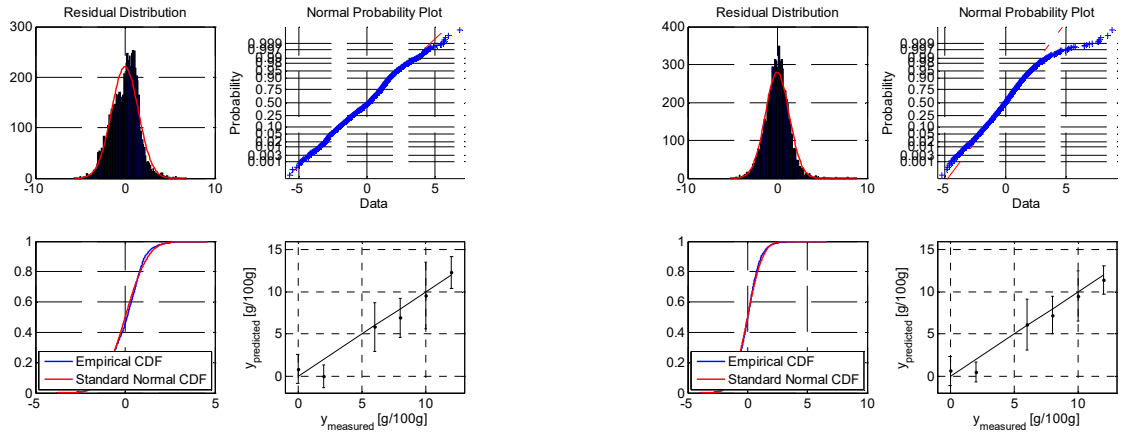


Figure 3.8: comparison of four graphical validation methods proving normal distribution of residual errors; residual distribution (top left), linearized normal probability plot (top right), cumulative distribution function (down left) and parity plot with two times standard deviation as error bars (down right); even though residual distribution using 15 LV (top left subfigure on the right side) fits better and their might be slight evidence of better validity (level 2 g/100g less deviating from expected values) as well as better precision (deviation of level 10 g/100g slightly lower), the normal probability plot for nine LV is not deviation too much. This might indicate overfitting in case of 15 LV model

Finally, several different model algorithms as well as data set configurations were tested. Presenting all the results of this feasibility and possibility investigation shown in the last paragraphs would be beyond the scope. Therefore, they are just summarized in Table 3.5. All reported prediction errors are far away from being acceptable, but those results show the power of data driven modelling combined with modern algorithms for data processing and knowledge about the problem of interest.

Table 3.5: different model algorithms including robust approach in PLS based models, used on seven US- features as input and RBF kernel as non-linear pre-processing; boundaries are samples between zero and 12 g/100g maltose and 10 to 20 °C (binary mixtures) and additional zero to 5 g/100g ethanol (binary and ternary mixtures)

Model type	Data set	γ	σ^2 (RBF)	EPO	LV	RMSEC	Mean RMSEV	σ (RMSEV)
SVR	6600 samples, Temperature as 8th input variable, ternary	232.12	5.63	-	-	0.39	1.22	0.08
Kernel-PLS	6600 samples, polynomial extension (2nd degree) of input data, ternary	-	65.32	1	15	1.22	1.34	0.06
Kernel-PLS	2450 samples, Only binary maltose	-	169.43	2	8	0.65	0.895	0.11
Kernel-PLS	6600 samples, ternary	-	11.34	1	9	1.47	1.5	0.06

Additionally, it is worth to mention, that the error achieved using the Kernel-PLS approach on binary maltose samples (see Table 3.5, third line) resulted in a comparable error then reported in the second thesis publication. The improvement here is a single and robust model solution. The figures for those models including the same visual inspection possibilities as discussed above are shown in the appendix A.4 to A.7. The comparably big errors presented in this thesis have multiple sources. Amongst others, the mentioned non-linearity in the second thesis publication is partially investigated in the section before. Further, robustness as well as possible outliers inside the data set, also indicated by the skewed error distribution in the second thesis publication are treated in the mentioned section. Even though preliminary, those investigations are similar to the approach MPA presented by Li *et al.* [70]. Nevertheless, one of the biggest influences is coming from buffer material and design.

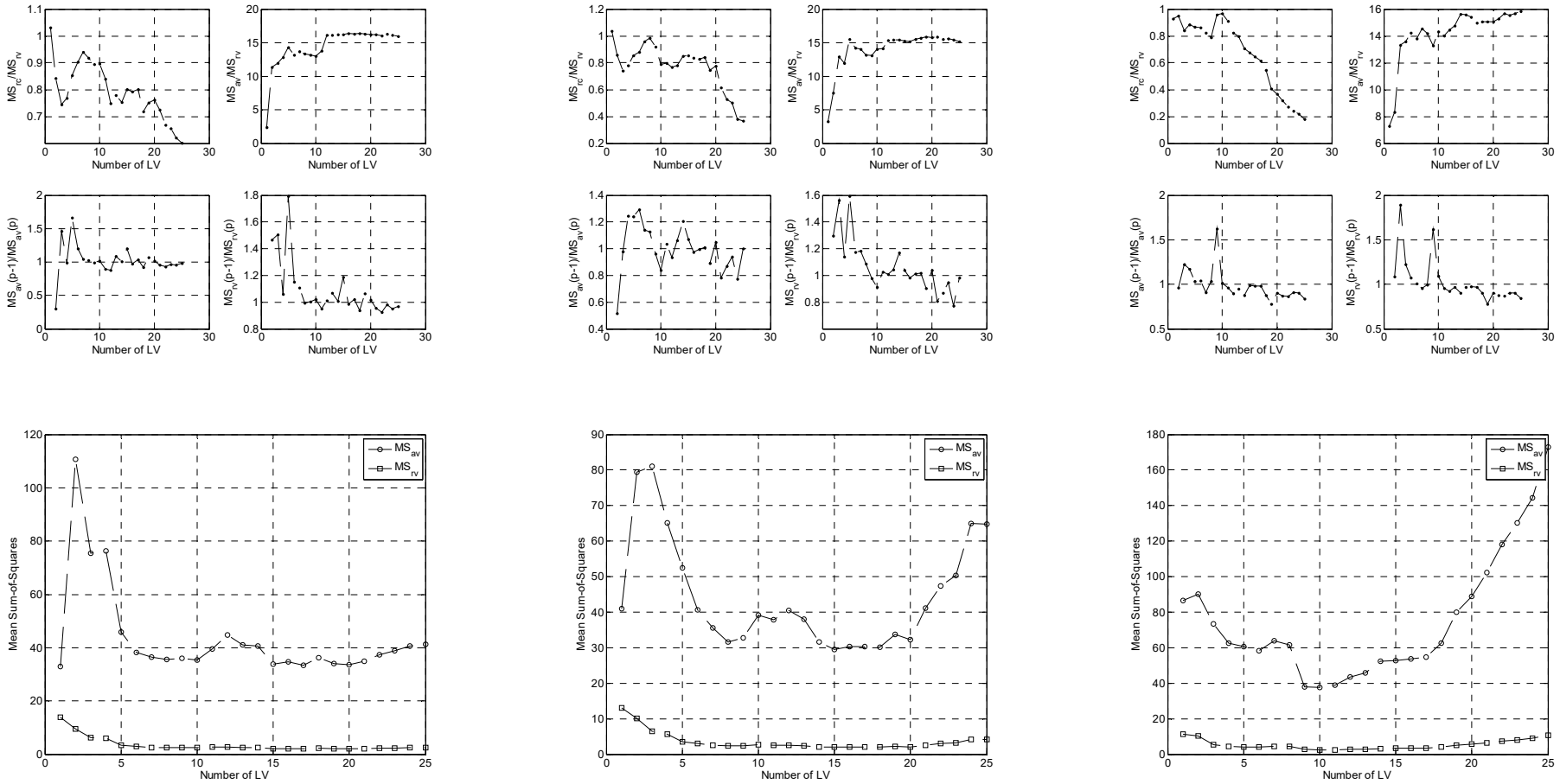


Figure 3.9: plots of validity and precision as well as ratios of mean sum of squares in different variants (explanation section 1.2.2.3) as one step in robust model generation using original data set (left), reduced by one (middle) and two EPO components (right). Figures on the top present ratios MS_{rc}/MS_{rv} (validity calibration/validity validation, top left, decreasing with rising number of components due to overfitting), MS_{av}/MS_{rv} (precision/validity, top right, convergence supports the assumption of no necessity to include more components or latent vectors), $MS_{av}(p\text{-}1)/MS_{av}(p)$ (precision, bottom left) and $MS_{rv}(p\text{-}1)/MS_{rv}(p)$ (validity, bottom right) – the latter both converge around one, which supports again the assumption of no necessity to include more components or latent vectors; figures at the bottom present precision and validity mean sum of squares as separate plot - dashed line with circles – MS_{av} (precision), dashed line with squares - MS_{rv} (validity) – a (local) minimum in both is preferable – thus, in each plot a number of latent vectors between eight and 10 would be enough; using one EPO component on the input data results in the lowest mean sum of squares for precision indicating one EPO component as sufficient

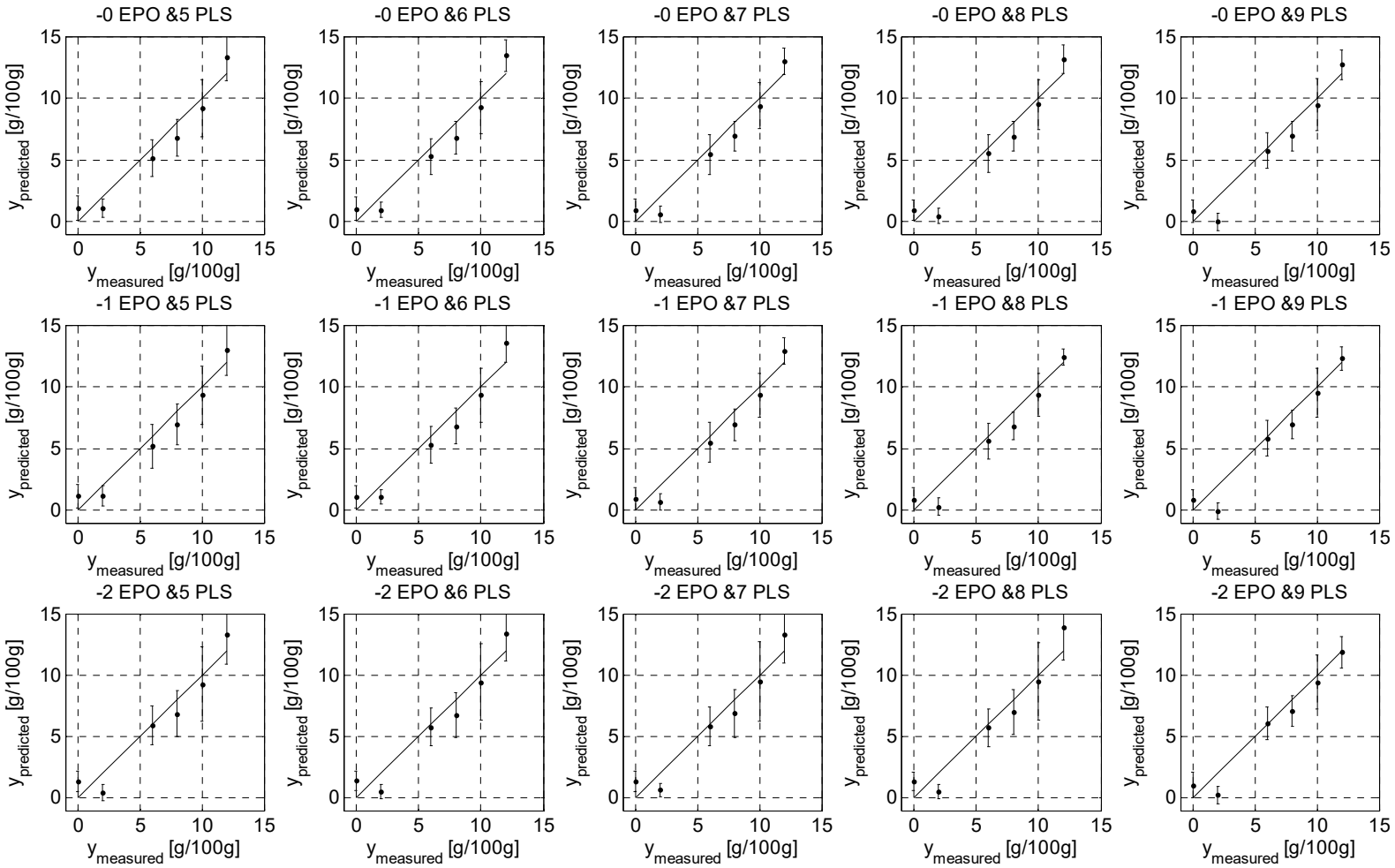


Figure 3.10: visual inspection of parity plots (error bars of one times standard deviation) with rising number of EPO components (vertical) and rising number of latent vectors (horizontal); the differences here are only minor, but precision with only one EPO component in higher concentration levels (e. g. seven, eight and nine LV) are slightly better whereas the validity difference between one and two EPO components are much less (e. g. level 2 g/100g at eight or nine LV)

The setup used was non-optimized with respect to the targets of interest. Those were first of all speed of sound and acoustical impedance, which results in density. Density and speed of sound open the possibility to detect concentration changes in ethanol and sugar in the same time. For measuring acoustical impedance in the necessary accuracy, the materials used should have comparable impedance not as with the used steel buffer. Further, design and electronics as well had to be optimized. Thus, material changes with properties more suitable to process demands [e.g. polymethylmethacrylate (PMMA), polyvinylidene fluoride (PVDF)] and by adapting the design of the buffer and the setup itself, were investigated (amongst others). These materials were taken according to investigations on the differences between sample and buffer impedance. All those aspects are discussed in detail in Hoche *et al. [13, 120, 121, 123, 124]*. Nevertheless, optimization in the direction of those aspects might also not reach to the goal when focusing on feature analysis (e.g. loss of buffer reflections).

Furthermore, the use of frequency spectra has to be regarded carefully. The frequency domain is highly sensitive to noise caused by bubbles, for example. In future investigations, this attenuation effect has to be studied by detailed bubble size analysis. In addition, combining ultrasonic properties from time domain such as time of flight (TOF) with frequency domain (P(f), Phase(f)) by means of multivariate statistics have to be studied in detail. This includes investigations on block wise (pre-)processing (Skov *et al*., 2008 [125]) and combining physical knowledge with data driven approaches (TOF/USV and PLS on frequency spectra or multiple specific features).

One possibility to exclude distorted signals in case of interfering gas bubbles or particles for example, such as in the mentioned propagation process (pulsed aeration, see first thesis publication) is given by outlier detection prior model analysis. This aspect is investigated with Angel Based Outlier Analysis (ABOF) on US signals. The setup used to collect those signals was a buffer-reflector design with ~50mm path length measuring in pulse echo mode. The fermentation was aerated in pulse-pause, therefore temporal occurrence of gas bubbles has to be expected. Features were extracted on the temporal region where ultrasonic echoes are expected (50 to 200µs). This results in a three-dimensional vector $\mathbf{S}_P$, with was further used for ABOF investigation. At the start of fermentation, a buffer of 80 signals is filled (~ 15 min) which are not distorted by any circumstance. The minimal variance of this buffer resembles the threshold for the ongoing investigation. Afterwards, the algorithm moves in a moving window sense comparing each new measured signal to the buffer. Just signals above the variance threshold are taken into the buffer and outliers are left out. The success of this possibility is shown in Figure 3.11, which proves the potential on the extracted time of flight clearly influenced by gas bubbles (causing visible shifts in trend lines). Nevertheless, it is also visible, that the cross correlation method adapted by Hoche *et al.,* 2011 [120] is also capable of detecting the time of flight, even if the signals are distorted (see results around 1.5 h). Thus, the whole approach should be investigated in more detail.

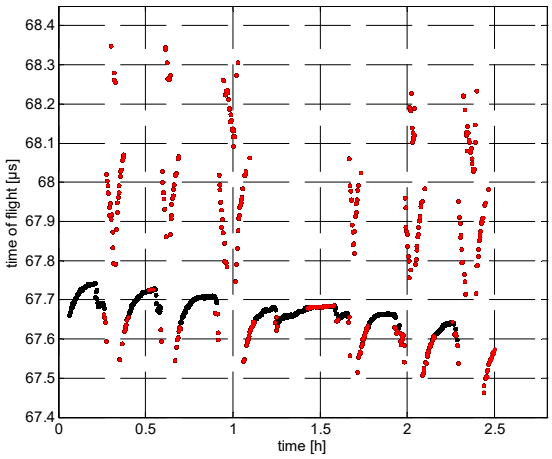


Figure 3.11: outlier detection of ultrasonic signals out of a yeast propagation process represented by US time of flight detection over time; outlier detection based on ABOF and three dimensional temporal feature space; outliers (red dots) are signals distorted by gas bubbles induced by aeration; algorithm used for time of flight detection was not successful to detect correct pulses, since gas inclusions deviate the US-signal too much

Assuming to receive ideal signals, the stability and robustness of the multivariate prediction model has to be verified. First, the discussed non-linearity was additionally investigated in this thesis. Although the presented errors are way too high for online as well as laboratory conditions, it could be shown, that the tendency of

predictions, even under changing temperature as well as on ternary mixtures (water-sugar-ethanol) with only US features of buffer reflections of the mentioned setup (steal buffer mounted in a pipe) is following the expected trend. Not surprisingly, the errors for support vector regression using RBF-kernel as pre-processing showed comparably best performance. Nevertheless, interpretability and overfitting issues lead to further trials using kernel-PLS with RBF-kernel as pre-processing. This trial is showing a holistic analysis starting with extensive feature selection using a combination of VIP and regression parameters. Further, studies of robustness are integrated by model building via robust calibration using prediction error statistics. Additionally, overfitting issues are handled by adapted, ANOVA related model qualification measures. All those steps lead to a more stable and trustable choice of latent vectors for the PLS regression model. Although the presented final models do not reach necessary accuracy, the approach reported combines several aspects of model qualification following the reports of Teófilo *et al.,* 2009 and Mehmood *et al.,* 2012 (variable selection) [2, 77], Axelson, 2012 and Faber, 1999 (data sets) [39, 100], Bruns, *et al.,* 2006 (regression ANOVA) [72] as well as Li *et al.*, 2012 and Chen *et al.*, 2014 (MPA and robustness) [70, 76]. Thus, it is showing, although still simple, the complexity of multivariate data analysis. If possible, it is highly recommended to take those aspects into account in future investigations. Especially statistical analysis of model output has a big advantage over standard solutions with respect to robustness, overfitting, outliers or variable selection (e. g. Li *et al.,* 2012 [70]). Further, using non-linear regression background, such as kernel pre-processing, highlights the potential of data driven modelling with respect to the presented example US feature investigation.

In conclusion, all presented calibration solutions are different – but compared to approaches in the literature based on ultrasound, the presented methods are independent of any assumption in process behaviour. Although the variation in all presented predictions is still high, the advantages of the final spectral analysis in combination with the presented sensor setup due to completely contactless investigations of the fluid of interest is quite noteworthy. So far, online solutions are typically in direct contact with the medium and thus design is of high effort. Furthermore, most monitoring solutions using ultrasound are based on at least two measuring principles to detect two different solutes. Nevertheless, there are still many points to investigate, such as iterative variable selection, sensitivity, robustness and outlier analysis.

Quality inspection using NIRS

Applying multivariate data analysis on quality inspection is another aspect handled in this work. It is shown in the third thesis publication, that NIRS can be used as data-driven fingerprints for estimating the processability with respect to lautering performance. Due to the complexity of NIR spectra the direct use for chemical analysis is not recommended (Sileoni, 2011 [105]). Nevertheless, it is possible to use them as a first approach for qualitative analysis. NIRS is widely used to quickly and directly identify starting products in pharmaceutical industry, for example [105]. As already mentioned before, data pre-processing is one of the major challenges. This study showed several state-of-the-art methods and its combinations reaching most promising accuracies by data driven choices (SNV or MSC in combination with first derivative for spectral treatment, VAST for variable treatment). It is worth to mention, that those trials were accomplished on full kernel spectra and validated in relation to the expert classification “good”, “normal” and “bad” on pilot plant (90.6 %, five LV) as well as industrial scale data (76.6 %, 21 LV). Reasons for lower accuracy as well as comparably high number of latent vectors for industrial data could be either the lower resolution of spectra or a limited number of patterns used for calibration procedure as well as diverging number of objects per class or group [126]. In the second case it is reasonable to assume that non-existing patterns for model calibration results in faulty predictions. The latter both could be solved by continuous extension of data pool. For the first reason, a general recommendation is to always collect as much data as possible (higher resolution) – reduction will be always possible afterwards. This is shown by variable selection applied on the presented data. The results of 51 % less wavelength as input to the pilot plant model lead to a more robust model solution and further indicate wave band regions of higher importance. This results underlines the necessity of variable selection once more. Nevertheless, the relation of results to the

biochemical background should be further treated. This can be achieved by for example more intense investigations on the relation to the laboratory reference methods or by empirical or synthetic calibration (see Shi *et al.* [127]). In future research on this topic it is recommended to follow the report of Sileoni *et al.* using knowledge from wave band characteristics to understand the results more with respect to the chemistry and/or physics. Sileoni further recommend creating knowledge by correlations between chemistry and spectroscopic features using relevant experiments. Those could support in finding qualitative as well as quantitative relations with respect to external parameters or waveband regions [105].

An additional influence on accuracy of any model is the loss of information whilst pre-processing of data. Kourti reports, that in case of pre-processing the multivariate nature of the respective data needs to be preserved [64]. It is possible by univariate data compression to introduce interfering correlations [64]. Thus, algorithms for such issues should be carefully chosen. Those used for the presented case reduce noise caused by physical effects such as light scattering. It can be assumed that such information may reach a better prediction by using adjusted treatment. Therefore, this topic should be integrated in future research. Even though, Sileoni reports, that NIRS is not utilisable for structure clarification [105], others recommend to use both, absorbance for chemical and scatter for physical (including also structural) information to utilize more of the existent content inside spectra [7]. Huang *et al.* also reports the possible usage for structure determination [43].

The core of the presented classification models is PLS-DA. Although simple, this algorithm was suitable to handle the presented challenge. One major drawback in classification is the masking effect. It was possible to show, that this effect is not visible under the presented conditions. Even though, the used algorithm showed good performance and it is widely applied in literature, the outcome should be compared to other reported methodologies such as support vector machines (SVM) based solutions [128, 129].

Altogether, there are more aspects to investigate - instead of:

- using uniform malt samples (reason: prevent inhomogeneity in the raw material), malt blends and special malts have to be analysed to validate the established method.
- keeping all processing steps as constant as possible, varying the influence on lautering performance by specific manipulation of grinding, mashing and blending of malt to invest their impact.
- using samples limited in harvesting year, seasonal variations of raw material composition have to be taken into account by e.g. automatically adapting the models by means of a moving average filter.

Those investigations might be used for recommendations according to adjustments of any process steps prior to lautering, but it remains to be analysed.

However, the number of 21 PLS components used for the industrial model is quite high. This could be reasoned by the lower resolution of spectra, the different number of objects per group or patterns not included in calibration. One of the major challenges for multivariate calibration models in general always remaining is the transferability, even though the presented results show applicability and transferability in both, pilot and industrial scale. Usually, a calibration model is dependent on the data and its origin (e.g. environments or instruments). Therefore, one of the future task has to be research in transfer approaches and thus reaching universality [111]. This aspect is supported by Nicolai *et al.*, 2007, who point to more explorative than only empirical model building in multivariate data analysis [113]. They mention light transport simulations or Monte Carlo method as possibilities to support this issue, most likely reaching better separation of physical and chemical information [113]. These factors would lead, amongst others, to improve the technology and to increase the ability for highly demanded on-line analysis for food industry, for example [111]. Nevertheless, the brought application of such technologies will only succeed, if on the one hand, the cost/benefit ratio of measurement systems will improve by either decreasing investment cost or increasing quality awareness of customers (see Nicolai *et al.*, 2007 [113]).

However, one highlight of these investigations so far is, that even if physical, chemical or biological relations are (partly) known, hidden information might be found by data driven approaches. In the report of Procopio *et al.*, 2013 [130] the authors used PLS to receive a holistic view on importance of different amino acids as substrate and the final aroma composition of the aimed fermentation product including possible mixed effects. The most valuable outcome of the investigation was the match with already existing results reported in the literature. Thus, it was shown, to establish a fingerprint between the concentrations of amino acid and the different aromatic components using multivariate analysis. In conclusion, those results point out the usage of multivariate data analysis to uncover synergies between inputs and outputs in a statistical manner.

Multivariate Statistical Process Control (MSPC)

One additional and very powerful usage of multivariate methods is the application on process control. In summary, several processes are used to define a statistical background of virtual trajectories to check the quality of the progress of future processes at each temporal window. Kourti, 2005 reports, that using univariate data compression methods in such cases often destroy the multivariate nature or delete relevant information of the present data [64]. Therefore, basic multivariate data compression such as PCA or PLS is recommended. With respect to the dimensionality of process data, such algorithms need adaption of the data matrices prior to compression. The usage of such background was shown successfully (amongst others) by Mitzscherling, 2004 [44] or Whitehead, 2012 [67] and by the fourth thesis publication. Nevertheless, such methods are used in general to qualify process batches. This can be accomplished also for one complete batch, as shown in the third thesis publication. Here, the presented method was used to automatically qualify lautering processes to reduce future efforts of experts. Validation on three classes defined by expert knowledge, namely “good”, “normal” and “bad” resulted in a match of 84 % between MSPC and expert qualification. However, since each processing unit (e.g. lautering tun) is having its distinct settings and sensors, implying adjustments of the presented approach to each problem of interest individually. This is not surprisingly counting for most of the existing, data-driven models. As already mentioned, the transferability is one of the major challenges, and thus has to be investigated for all systems and solutions in the area of multivariate data analysis.

In addition, the applied method for MSPC in this thesis is still two-dimensional, since matrices are unfolded and batch lengths adopted by maturity indexing. The really trilinear methods like PARAFAC show also to be quite powerful and thus a rising interest in a variety of fields [44, 66]. With the increasing amount of data from growing number of available sensors including spectral solutions, these tools become of major importance in data analysis, since such data pool are not only three-dimensional anymore. Nevertheless, this is a very relevant path of future exploration and might be highly relevant [67].

However, it has to be mentioned once more, that knowledge, especially process knowledge is a vital necessity when using those methods. Most of the decisions on the choice in model parameters, the way of data treatment, the correct choice of data used for model generation as well as corresponding variable weights is dependent on this essential knowledge [64].

Swarm Sensing

One major challenge in online monitoring of bioprocesses is addressed in the fourth thesis publication. This is the validity of sensor readings with respect to progress of running process. As shown by several applications of MSPC for example, the immediate online measurement of system and leading process variables like substrate as well as biomass concentration is not necessary. However, comprehensive quality control of running processes is, amongst others, limited to failure and drift-free sensors and sensor readings. Additionally, the rising amount of sensor systems for the variety of measurement goals (rising typically due to higher quality demands) are all together exposed to influencing side effects. Thus, any multivariate model based on such variables will fail, if one or more sensor does not fulfil its original purpose. The presented swarm intelligence based approach was established to work even under those conditions. It was possible to show, that simple MSPC based trajectories of diverse origin

can be used as search space for a discrete swarm to find in any time step of running process a suitable solution for statistical process trajectory. Further, each sensor used as input for this search space can be evaluated on its validity. Amongst the successful monitoring by control charts of the validation process supported by the swarm in combination with the historical similarity of sensors, it was possible to find and neglect false inputs in 100% by still giving positive monitoring feedback. The biggest benefit of this approach is therefore being not restricted to the number of sensor inputs or the necessity of specific sensor readings compared to simple MSPC or other predictive models.

The whole investigations were performed aiming at an integration of the proposed method on a microcontroller later on. Since such devices are limited in processing power the issue of processing time cannot be neglected. Although it is feasible to assume, that models with minimum number of inputs might work much better (less computational effort, reduced necessity of sensors), than the solutions presented, the swarm is aiming at the model with the maximum number of individual inputs by purpose. In addition to more stable monitoring with a stressable online trajectory, an evaluation of the entire sensor network is absolved (with the feature, that the network could contain an arbitrary amount of members). As a result, on a process with barely false sensor readings, each input was used more than 98 % of the cases. Accepting a random “failure” of the swarm (choice of models with less inputs), this result underlines the efficiency of the swarm.

Another topic, which is partially investigated in the fourth thesis publication, is the aspect of missing data. Kourti reports the possibility to use latent variable methods together with the knowledge of highly correlated process variables as well as redundant information [64]. The method presented here is based on similar background, but using the choice of a particle swarm on different multivariate model solutions instead of loading vectors of a single latent variable method. This is obviously more advantageous in case of noisy data, always present in respective processes. Additionally, more than one sensor reading can be error prone at one specific time point investigated. Further, it is possible to aim at replacing a corrupted sensor reading by historical data based on a swarm decision. Therefore, calibration would be performed to the corresponding sensor as target instead following the approach presented. Nevertheless, in case of changing biological conditions, such as changing metabolism of respective organism, the presented particle swarm based solution might also fail.

Even though the reported results are based upon an initial approach, the possibility to predict the progress of fermentation in all cases in a multivariate statistical process control sense is shown. The presented results point towards more robust online monitoring for biotechnological processes insuring temporal effective processing as well as sensor failure detection. In conclusion, the swarm sensing approach presented resembles a combined multiple sensor investigation for holistic process and integrated sensor evaluation and control.

Even though it is generally possible to tune any algorithm and its parameters to find challenges on which they work better than another one (which does not mean, that one algorithm is always better), comparison to data-driven or other metaheuristic optimization algorithms including multi-objective versions should be examined. This includes a comparison based upon computational effort as well as accuracy between such methods and the modified PSO version presented. Amongst others, it is also necessary to determine, how this system behaves if several sensor failures occur. Furthermore, chosen cost functional with respect to weights of individual parts as well as parameters of the swarm indicate additional effort.

For all presented applications, a different model background results not surprisingly in different accuracies. It is one of the biggest difficulties to choose an optimal model background, understand and adapt it to the demands, choose the correct inputs, find outliers decently, validate and robustify the model correctly, create the correct design space and model the correct physicochemical and/or biological background. This is underlined by the fact, that the lack of causality in any relation is not always obvious, since tools of multivariate data analysis are able to establish relations between multiple variables and almost any target (e. g. quality data), even if there is no general relation existing [7]). Further, the usage of any of the presented approaches has to be adapted to each problem of

interest individually, which is counting for most of the existing, data-driven models, exemplarily shown in the third thesis publication on pilot an industrial scale process data.

Altogether it should be mentioned, that nothing is final and everything could be better. We are just approximating. With the rise of computational power, more possibilities are given. It is recommendable to ignore time and computational cost to get best outputs in calibration. With the issue of powerful computers, it is possible to go into much higher operations. Therefore, even simple methods like data-driven approaches can go for complex problem solutions.

The most conclusive and valuable outcome of this thesis would be a procedure for picking the right mathematical tool for:

- Data Pre-processing
- Model Generation
- Post-Processing

In contrary to expectations, no one can predict the right tool, and this comes from the heuristic nature of the problem. As shown in the third thesis publication, a choice of correct algorithm (in the respective example on pre-processing algorithms or areas of importance in spectra) suitable for the problem of interest *a priori* is not always possible. A fully inappropriate choice most often leads to complete failure and very poor accuracies. A partially improper choice might lead to loss of information. Nevertheless, in several cases even wet chemical analysis does not lead to clear statements of a component or a quality attribute (e.g. third thesis publication). Moreover, the physical background is often not completely known (e.g. first and second thesis publication). In any of those cases, it might be necessary to try different ways and generate knowledge to find an acceptable outcome of the respective investigation. Thus, rather only clever suggestions would be plausible based upon pattern recognition, linkage to quality assurance, regressions of new sensing systems or combined evaluation in a holistic process/sensor interaction perspective – all considering the sub steps and the major steps with regard to the target state estimator (or function). Finally, expanding the statement of Brosnan and Sun, 2004 on the potential of computer vision “a higher implementation and uptake (…) to meet the ever expanding requirements of the food industry” [103] is counting for new technologies in the whole PAT toolbox, such as sensors, data analysis, process monitoring and control implying the demand on novel developments.

4. References

1. Bakeev, K.A., *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries*. 2010: John Wiley & Sons.
2. Mehmood, T., et al., *A review of variable selection methods in partial least squares regression.* Chemometrics and Intelligent Laboratory Systems, 2012. **118**: p. 62-69.
3. Miller, C.E., *Chemometrics in process analytical technology (PAT)*, in *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries.* , K.A. Bakeev, Editor. 2010, John Wiley & Sons. p. 353-438.
4. Gendrin, C., Y. Roggo, and C. Collet, *Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review.* Journal of pharmaceutical and biomedical analysis, 2008. **48**(3): p. 533-553.
5. Nicoletti, M., L. Jain, and R. Giordano, *Computational intelligence techniques as tools for bioprocess modelling, optimization, supervision and control*. Computational Intelligence Techniques for Bioprocess Modelling, Supervision and Control. 2009: Springer. 1-23.
6. Kessler, W., *Multivariate Datenanalyse*. 2007, Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA.
7. Kessler, R.W., *Perspectives in process analysis.* Journal of Chemometrics, 2013. **27**(11): p. 369-378.
8. Munck, L., et al., *Chemometrics in food science—a demonstration of the feasibility of a highly exploratory, inductive evaluation strategy of fundamental scientific significance.* Chemometrics and Intelligent Laboratory Systems, 1998. **44**(1): p. 31-60.
9. Meurens, M. and S.H. Yan, *Applications of vibrational spectroscopy in brewing.* Handbook of Vibrational Spectroscopy, 2002.
10. Egly, D., *Einsatz der IR-Spektroskopie bei der Prozessverfolgung von Hefe-Fermentationen sowie bei der Qualitätskontrolle von Getränken, Obst und Hefen*. 2012, Universitätsbibliothek.
11. Lourenço, N., et al., *Bioreactor monitoring with spectroscopy and chemometrics: a review.* Analytical and bioanalytical chemistry, 2012. **404**(4): p. 1211-1237.
12. Sileoni, V., et al., *Near-Infrared Spectroscopy for Proficient Quality Evaluation of the Malt and Maize Used for Beer Production.* Journal of the Institute of Brewing, 2010. **116**(2): p. 134-139.
13. Hoche, S., M.A. Hussein, and T. Becker, *Ultrasound-based density determination via buffer rod techniques: a review.* J. Sens. Sens. Syst., 2013. **2**(2): p. 103-125.
14. Halstensen, M. and K.H. Esbensen, *Acoustic chemometric monitoring of industrial production processes*, in *Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries.*, K.A. Bakeev, Editor. 2010, John Wiley & Sons. p. 353-438.
15. Hauptmann, P., N. Hoppe, and A. Püttmer, *Application of ultrasonic sensors in the process industry.* Measurement Science and Technology, 2002. **13**(8): p. R73-R83.
16. *Kurzbericht: Ultraschall und Dichtemessung am Beispiel der Inversion von Saccharose*. 2009, Martin-Luther-Universität Halle-Wittenberg, Zentrum für Ingenieurwissenschaften, Institut für Verfahrenstechnik / Thermische Verfahrenstechnik: Halle, Germany.
17. Hauptmann, P., et al., *Ultrasonic sensors for process monitoring and chemical analysis: state-of-the-art and trends.* Sensors and Actuators A: Physical, 1998. **67**(1-3): p. 32-48.
18. Kessler, R.W., *Strategic Position of Spectroscopy in a PAT/QbD Environment*, in *PAT & QbD Forum*. 2014, Satorius AG: Göttingen, Germany.
19. David, J. and N. Cheeke, *Chapter 7: Reflection and Transmission of Ultrasonic Waves at Interfaces*, in *Fundamentals and applications of ultrasonic waves*. 2002, CRC Press: Physics Department Concordia University Montreal, Qc, Canada.
20. Padmavathi, G., D. Shanmugapriya, and M. Kalaivani. *Acoustic signal based feature extraction for vehicular classification*. in *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on*. 2010. IEEE.
21. Peeters, G. *A Large Set of Audio Features for Sound Description*. 2004; Available from: http://recherche.ircam.fr/anasyn/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf.
22. Peeters, G., et al., *The timbre toolbox: Extracting audio descriptors from musical signals.* The Journal of the Acoustical Society of America, 2011. **130**(5): p. 2902-2916.
23. Ben-Harush, O., H. Guterman, and I. Lapidot. *Frame level entropy based overlapped speech detection as a pre-processing stage for speaker diarization*. in *Machine Learning for Signal Processing, 2009. MLSP 2009. IEEE International Workshop on*. 2009. IEEE.
24. Hussein, W.B., *Digital Processing Based Solutions for Life Science Engineering Recognition Problems*. 2013, TU München: München.
25. Krause, D., et al., *Ultrasonic sensor for predicting sugar concentration using multivariate calibration.* Ultrasonics, 2014. **54**(6): p. 1703-1712.
26. Park, B., et al., *Ultrasonic spectral analysis for beef sensory attributes.* Journal of food science, 1994. **59**(4): p. 697-701.
27. Park, B., et al., *Measuring intramuscular fat in beef with ultrasonic frequency analysis.* Journal of animal science, 1994. **72**(1): p. 117-125.

28. Mörlein, D., et al., *Non-destructive estimation of the intramuscular fat content of the< i> longissimus</i> muscle of pigs by means of spectral analysis of ultrasound echo signals.* Meat Science, 2005. **69**(2): p. 187-199.
29. Tao, R., et al. *Music genre classification using temporal information and support vector machine*. in *Proc. of the 16th Advanced School for Computing and Imaging Conf.(ASCI 2010)*. 2010.
30. Tzanetakis, G. and P. Cook, *Musical genre classification of audio signals.* Speech and Audio Processing, IEEE transactions on, 2002. **10**(5): p. 293-302.
31. Rathore, A., R. Bhambure, and V. Ghare, *Process analytical technology (PAT) for biopharmaceutical products.* Analytical and bioanalytical chemistry, 2010. **398**(1): p. 137-154.
32. Pomerantsev, A.L. and O.Y. Rodionova, *Process analytical technology: a critical view of the chemometricians.* Journal of Chemometrics, 2012. **26**(6): p. 299-310.
33. Halstensen, M. and K. Esbensen, *New developments in acoustic chemometric prediction of particle size distribution—'the problem is the solution'.* Journal of chemometrics, 2000. **14**(5-6): p. 463-481.
34. Schäfer, R., J.E. Carlson, and P. Hauptmann, *Ultrasonic concentration measurement of aqueous solutions using PLS regression.* Ultrasonics, 2006. **44**(1): p. e947-e950.
35. Doble, M., *Avoid the Pitfalls of bioprocess development.* CEP, 2006: p. 34-41.
36. Szita, N., et al., *Development of a multiplexed microbioreactor system for high-throughput bioprocessing.* Lab on a Chip, 2005. **5**(8): p. 819-826.
37. Sonnleitner, B. and O. Kaeppeli, *Growth of Saccharomyces cerevisiae is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis.* Biotechnology and Bioengineering, 1986. **28**(6): p. 927-937.
38. Wold, S., et al., *Modelling and diagnostics of batch processes and analogous kinetic experiments.* Chemometrics and Intelligent Laboratory Systems, 1998. **44**(1–2): p. 331-340.
39. Axelson, D.E., *Data Preprocessing for Chemometric and Metabonomic Analysis*. 2nd ed. 2012, Kingston, Ontario: MRi_Consulting.
40. Hector C. Keun, T.M.D.E., Henrik Antti, Mary E. Bollard, Olaf Beckonert, Elaine Holmes, John C. Lindon, Jeremy K. Nicholson, *Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling.* Analytica Chimica Acta, 2003. **490**(1-2): p. 265-276.
41. Barnes, R.J., M.S. Dhanoa, and S.J. Lister, *Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra.* Appl. Spectrosc., 1989. **43**(5): p. 772-777.
42. Jørgensen, A., *Clustering excipient near infrared spectra using different chemometric methods*. 2000, University of Helsinki,.
43. Huang, H., et al., *Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review.* Journal of Food Engineering, 2008. **87**(3): p. 303-313.
44. Mitzscherling, M., *Prozeßanalyse des Maischens mittels statistischer Modellierung*, in *Lehrstuhl für Fluidmechanik und Prozeßautomation*. 2004, Technische Universität München.
45. Nicolaï, B.M., K.I. Theron, and J. Lammertyn, *Kernel PLS regression on wavelet transformed NIR spectra for prediction of sugar content of apple.* Chemometrics and Intelligent Laboratory Systems, 2007. **85**(2): p. 243-252.
46. Bennett, K. and M. Embrechts, *An optimization perspective on kernel partial least squares regression*, in *Advances in Learning Theory: Methods, Models and Applications, NATO Science Series: III. Computer and Systems Sciences*, J.A.K. Suykens, et al., Editors. 2003, IOS Press: Amsterdam. p. 227-250.
47. Roger, J.-M., F. Chauchard, and V. Bellon-Maurel, *EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits.* Chemometrics and Intelligent Laboratory Systems, 2003. **66**(2): p. 191-204.
48. Becker, T. and D. Krause, *Softsensorsysteme – Mathematik als Bindeglied zum Prozessgeschehen.* Chemie Ingenieur Technik – Themenheft Prozessanalytik, 2010. **82**(4): p. 429-440.
49. Lindgren, F., P. Geladi, and S. Wold, *The kernel algorithm for PLS.* Journal of Chemometrics, 1993. **7**(1): p. 45-59.
50. Whitehead, I.J., *Soft sensing – using multivariate analysis for yeast propagation monitoring.* 2012, TU München.
51. Burnham, A.J., R. Viveros, and J.F. MacGregor, *Frameworks for latent variable multivariate regression.* Journal of chemometrics, 1996. **10**(1): p. 31-45.
52. Lavine, B. and J. Workman, *Chemometrics.* Analytical chemistry, 2010. **82**(12): p. 4699-4711.
53. Hastie, T., et al., *The elements of statistical learning*. Vol. 2. 2009: Springer.
54. De Luca, M., et al., *Derivative FTIR spectroscopy for cluster analysis and classification of morocco olive oils.* Food chemistry, 2011. **124**(3): p. 1113-1118.
55. Brereton, R.G., *Chemometrics for pattern recognition*. 2009: Wiley.
56. Barker, M.R., W. , *Partial least squares for discrimination.* Journal of Chemometrics, 2003. **17**(3): p. 166–173.

57. Chauchard, F., et al., *Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes.* Chemometrics and Intelligent Laboratory Systems, 2004. **71**(2): p. 141-150.
58. Smola, A.J. and B. Schölkopf, *A tutorial on support vector regression.* Statistics and computing, 2004. **14**(3): p. 199-222.
59. Vapnik, V., S.E. Golowich, and A. Smola, *Support vector method for function approximation, regression estimation, and signal processing.* Advances in neural information processing systems, 1997: p. 281-287.
60. Burges, C.J.C., *A Tutorial on Support Vector Machines for Pattern Recognition.* Data Mining and Knowledge Discovery, 1998. **2**(2): p. 121-167.
61. Pelckmans, K., et al., *LS-SVMlab: a matlab/c toolbox for least squares support vector machines.* Tutorial. KULeuven-ESAT. Leuven, Belgium, 2002.
62. Goyal, S., *Predicting properties of cereals using artificial neural networks: A review.* Scientific Journal of Crop Science, 2013. **2**(7): p. 95-115.
63. Smilde, A., R. Bro, and P. Geladi, *Multi-way analysis: applications in the chemical sciences*. 2005: John Wiley & Sons.
64. Kourti, T., *Application of latent variable methods to process control and multivariate statistical process control in industry.* International Journal of Adaptive Control and Signal Processing, 2005. **19**(4): p. 213-246.
65. Krämer, N. and M.L. Braun. *Kernelizing PLS, degrees of freedom, and efficient model selection*. in *Proceedings of the 24th international conference on Machine learning*. 2007. ACM.
66. Bro, R., *Multi-way analysis in the food industry: models, algorithms, and applications*. 1998, Københavns Universitet
67. Whitehead, I.J., *Soft sensing – using multivariate analysis for yeast propagation monitoring*, in *Lehrstuhl für Brau- und Getränketechnologie*. 2012, TU München.
68. Grassi, S., et al., *Beer fermentation: Monitoring of process parameters by FT-NIR and multivariate data analysis.* Food chemistry, 2014. **155**: p. 279-286.
69. Administration, F.a.D., *Guidance for Industry: PAT - A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance.* 2004.
70. Li, H.-D., Y.-Z. Liang, and Q.-S. Xu, *Model Population Analysis for Statistical Model Comparison*, in *Chemometrics in Practical Applications*, K. Varmuza, Editor. 2012, InTech.
71. Faber, N.M. and R. Bro, *Standard error of prediction for multiway PLS: 1. Background and a simulation study.* Chemometrics and Intelligent Laboratory Systems, 2002. **61**(1–2): p. 133-149.
72. Bruns, R.E., I.S. Scarminio, and B. de Barros Neto, *Statistical design-chemometrics*. Vol. 25. 2006: Elsevier.
73. Esbensen, K.H. and P. Geladi, *Principles of Proper Validation: use and abuse of re-sampling for validation.* Journal of Chemometrics, 2010. **24**(3-4): p. 168-187.
74. Brereton, R.G., *Applied chemometrics for scientists*. 2007: John Wiley & Sons.
75. Brereton, R.G., *Chemometrics: data analysis for the laboratory and chemical plant*. 2003: John Wiley & Sons.
76. Chen, H.-Z., et al., *An optimization strategy for waveband selection in FT-NIR quantitative analysis of corn protein.* Journal of Cereal Science, 2014. **60**(3): p. 595-601.
77. Teófilo, R.F., J.P.A. Martins, and M.M.C. Ferreira, *Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression.* Journal of Chemometrics, 2009. **23**(1): p. 32-48.
78. Nadler, B. and R.R. Coifman, *The prediction error in CLS and PLS: the importance of feature selection prior to multivariate calibration.* Journal of Chemometrics, 2005. **19**(2): p. 107-118.
79. Kohavi, R. and G.H. John, *Wrappers for feature subset selection.* Artificial intelligence, 1997. **97**(1): p. 273-324.
80. Guyon, I. and A. Elisseeff, *An introduction to variable and feature selection.* The Journal of Machine Learning Research, 2003. **3**: p. 1157-1182.
81. Böhlmann, S., *Skript zur Lehrveranstaltung Prozessanalyse und Versuchsplanung.* Technische Universität Dresden, Fakultät Maschinenwesen, Institut für Verfahrensautomatisierung, 2012.
82. Matteo Cassotti and F. Grisoni. *Tutorial 6: Variable selection methods: an introduction*. [cited 2014; Available from: http://www.moleculardescriptors.eu/tutorials/T6_moleculardescriptors_variable_selection.pdf.
83. Chong, I.-G. and C.-H. Jun, *Performance of some variable selection methods when multicollinearity is present.* Chemometrics and Intelligent Laboratory Systems, 2005. **78**(1–2): p. 103-112.
84. Sorol, N., et al., *Visible/near infrared-partial least-squares analysis of Brix in sugar cane juice A test field for variable selection methods.* Chemometrics and Intelligent Laboratory Systems, 2010. **102**(2): p. 100-109.
85. Martens, H. and T. Næs, *Multivariate Calibration*. 1991: John Wiley & Sons.

86. Ferreira, A.P., T.P. Alves, and J.C. Menezes, *Monitoring complex media fermentations with near-infrared spectroscopy: Comparison of different variable selection methods.* Biotechnology and bioengineering, 2005. **91**(4): p. 474-481.
87. Meinshausen, N., *Hierarchical testing of variable importance.* Biometrika, 2008. **95**(2): p. 265-278.
88. Reinikainen, S.P. and A. Höskuldsson, *COVPROC method: strategy in modeling dynamic systems.* Journal of chemometrics, 2003. **17**(2): p. 130-139.
89. Garcia, H. and P. Filzmoser, *Multivariate Statistical Analysis using the R package chemometrics.* Vienna: Austria, 2011.
90. Cai, W., Y. Li, and X. Shao, *A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra.* Chemometrics and intelligent laboratory systems, 2008. **90**(2): p. 188-194.
91. Li, H.D., et al., *Model population analysis for variable selection.* Journal of chemometrics, 2010. **24**(7-8): p. 418-423.
92. Li, H.-D., et al., *Recipe for revealing informative metabolites based on model population analysis.* Metabolomics, 2010. **6**(3): p. 353-361.
93. Saeys, Y., I. Inza, and P. Larrañaga, *A review of feature selection techniques in bioinformatics.* bioinformatics, 2007. **23**(19): p. 2507-2517.
94. Mehmood, T., et al., *A Partial Least Squares based algorithm for parsimonious variable selection.* Algorithms for Molecular Biology, 2011. **6**(1): p. 27.
95. Westad, F. *Et tilbakeblik på MSPC med betragtninger om afvigergrænser, modelopdatering og batch-modellering og -monitorering*. Available from: https://mit.ida.dk/IDAforum/U0602c/Documents/20100128%20dsk.2010/06%20Frank%20Westad.pdf.
96. Häggblom, K.-E. *Basics of Multivariate Modelling and Data Analysis*. Available from: http://www.users.abo.fi/khaggblo/MMDA/MMDA6.pdf.
97. Botella, C., J. Ferré, and R. Boqué, *Outlier detection and ambiguity detection for microarray data in probabilistic discriminant partial least squares regression.* J. Chemom., 2010. **24**(7-8): p. 434-443.
98. Cao, D.S., et al., *A new strategy of outlier detection for QSAR/QSPR.* Journal of computational chemistry, 2010. **31**(3): p. 592-602.
99. Kriegel, H.-P., M.S. Hubert, and A. Zimek, *Angle-based outlier detection in high-dimensional data.* Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, 2008: p. 444-452.
100. Faber, N.M., *Estimating the uncertainty in estimates of root mean square error of prediction: application to determining the size of an adequate test set in multivariate calibration.* Chemometrics and Intelligent Laboratory Systems, 1999. **49**(1): p. 79-89.
101. Liebmann, B., P. Filzmoser, and K. Varmuza, *Robust and classical PLS regression compared.* Journal of Chemometrics, 2010. **24**(3-4): p. 111-120.
102. Plutowska, B. and W. Wardencki, *Application of gas chromatography–olfactometry (GC–O) in analysis and quality assessment of alcoholic beverages–A review.* Food Chemistry, 2008. **107**(1): p. 449-463.
103. Brosnan, T. and D.-W. Sun, *Improving quality inspection of food products by computer vision—a review.* Journal of Food Engineering, 2004. **61**(1): p. 3-16.
104. Wu, D. and D.-W. Sun, *Colour measurements by computer vision for food quality control–A review.* Trends in Food Science & Technology, 2013. **29**(1): p. 5-20.
105. Sileoni, V., C. Cavani, and G. Perretti, *Study of innovative methods of control in the cereal productive chain for the production of beer and spirits.* 2011.
106. Ratcliffe, M. and J. Panozzo, *The application of near infrared spectroscopy to evaluate malting quality.* Journal of the Institute of Brewing, 1999. **105**(2): p. 85-88.
107. Giovenzana, V., R. Beghi, and R. Guidetti, *Rapid evaluation of craft beer quality during fermentation process by vis/NIR spectroscopy.* Journal of Food Engineering, 2014. **142**: p. 80-86.
108. Gianinetti, A., et al., *Improving discrimination for malting quality in barley breeding programmes.* Field crops research, 2005. **94**(2): p. 189-200.
109. Munck, L. and B. Møller, *A New Germinative Classification Model of Barley for Prediction of Malt Quality Amplified by a Near Infrared Transmission Spectroscopy Calibration for Vigour "On Line" Both Implemented by Multivariate Data Analysis.* Journal of the Institute of Brewing, 2004. **110**(1): p. 3-17.
110. Lachenmeier, D.W., *Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra.* Food Chemistry, 2007. **101**(2): p. 825-832.
111. Cen, H. and Y. He, *Theory and application of near infrared reflectance spectroscopy in determination of food quality.* Trends in Food Science & Technology, 2007. **18**(2): p. 72-83.
112. Georgieva, M., et al., *Application of NIR spectroscopy and chemometrics in quality control of wild berry fruit extracts during storage.* Hrvatski časopis za prehrambenu tehnologiju, biotehnologiju i nutricionizam, 2014. **8**(3-4): p. 67-73.

113. Nicolaï, B.M., et al., *Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review.* Postharvest Biology and Technology, 2007. **46**(2): p. 99-118.
114. Gowen, A., et al., *Hyperspectral imaging–an emerging process analytical tool for food quality and safety control.* Trends in Food Science & Technology, 2007. **18**(12): p. 590-598.
115. Perrot, N., et al., *Fuzzy concepts applied to food product quality control: A review.* Fuzzy Sets and Systems, 2006. **157**(9): p. 1145-1154.
116. Krause, D., et al., *Ultrasonic Characterization of Aqueous Solutions with varying Sugar and Ethanol Content using Multivariate Regression Methods.* Journal of Chemometrics, 2011. **25**(4): p. 216-223.
117. McHardy, C., *Model-based Analysis of Growth Limitations during Batch Propagation of Brewer's Yeast*, in *Lehrstuhl für Brau- und Getränketechnologie*. 2013, TU München.
118. Birle, S., M.A. Hussein, and T. Becker, *On-line yeast propagation process monitoring and control using an intelligent automatic control system.* Engineering in Life Sciences, 2015. **15**(1): p. 83-95.
119. Procopio, S., et al., *Significant amino acids in aroma compound profiling during yeast fermentation analyzed by PLS regression.* Food Science and Technology 2012.
120. Hoche, S., et al., *Time-of-flight prediction for fermentation process monitoring.* Engineering in Life Sciences, 2011. **11**(4): p. 417-428.
121. Hoche, S., M.A. Hussein, and T. Becker, *Critical process parameter of alcoholic yeast fermentation: speed of sound and density in the temperature range 5–30° C.* International Journal of Food Science & Technology, 2014. **49**(11): p. 2441-2448.
122. Roger, J.-M., F. Chauchard, and P. Williams, *Removing the block effects in calibration by means of dynamic orthogonal projection. Application to the year effect correction for wheat protein prediction.* J Near Infrared Spectrosc, 2008. **16**(3): p. 311-315.
123. Hoche, S., M. Hussein, and T. Becker, *L3-Nicht-invasive, online Dichtebestimmung mittels ultraschallbasierender Mehrfach-Reflektions-Methode.* Tagungsband, 2013: p. 418-420.
124. Hoche, S., M. Hussein, and T. Becker, *Density, ultrasound velocity, acoustic impedance, reflection and absorption coefficient determination of liquids via multiple reflection method.* Ultrasonics, 2014.
125. Skov, T., D. Ballabio, and R. Bro, *Multiblock variance partitioning: A new approach for comparing variation in multiple data blocks.* Analytica chimica acta, 2008. **615**(1): p. 18-29.
126. Hicham Nocairi, E.M.Q., Evelyne Vigneau, Dominique Bertrand, *Discrimination on latent components with respect to patterns. Application to multicollinear data Original.* Computational Statistics & Data Analysi, 2005. **1**(1): p. 139-147.
127. Shi, Z., et al., *Optical coefficient-based multivariate calibration on near-infrared spectroscopy.* Journal of Chemometrics, 2010. **24**(5): p. 288-299.
128. Melgani, F. and Y. Bazi, *Classification of electrocardiogram signals with support vector machines and particle swarm optimization.* Information Technology in Biomedicine, IEEE Transactions on, 2008. **12**(5): p. 667-677.
129. Chuang, L.-Y., C.-H. Yang, and C.-H. Yang, *Tabu search and binary particle swarm optimization for feature selection using microarray data.* Journal of Computational Biology, 2009. **16**(12): p. 1689-1703.
130. Procopio, S., et al., *Significant amino acids in aroma compound profiling during yeast fermentation analyzed by PLS regression.* Food Science and Technology, 2013. **51**(2): p. 423-432.

A. APPENDIX

A.1 Model robustness – NIR REIP investigation

Dataset of 721 objects divided into seven subsets; iteration on windows and on datasets; best prediction error with N = 21 and p = 5 (SG filter), spectral smoothing (SG0) and variable stability scaling (VAST) resulting in prediction errors presented in Table A.1. Nevertheless, deeper investigations are needed (e. g. degrees of freedom, variable selection and number of latent vectors)

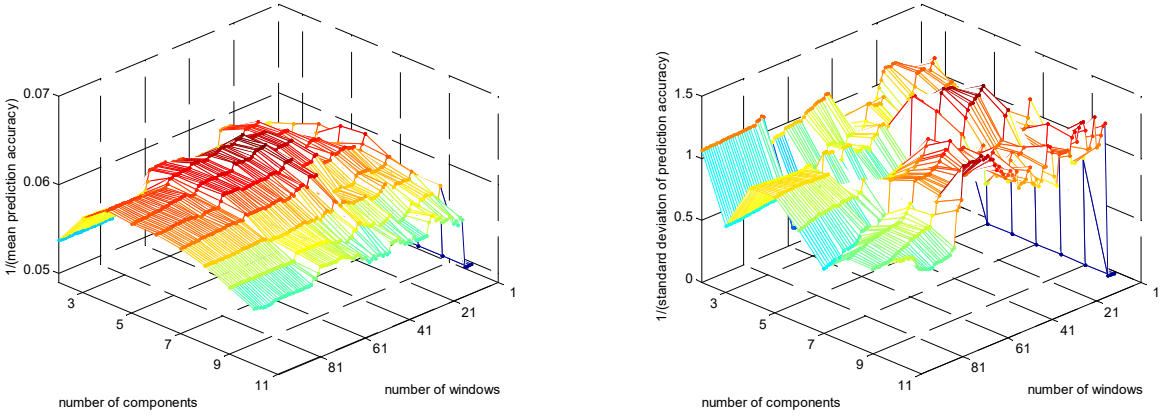


Figure A.1: reciprocal error and standard deviation of error for NIR-REIP investigation – iteration over windows of selected variables and number of component; the optimal choice should be in the region of the maximum in both plots.

Table A.1: choice of model size according to the figure above and the mentioned criteria (subtitle of figure)

	Latent Vectors	Variable Windows	Mean RMSEV	Standard Deviation
Best error	6	39	16.2	1.3
Best std	8	27	17.3	0.7
One option	6	27	16.4	0.84

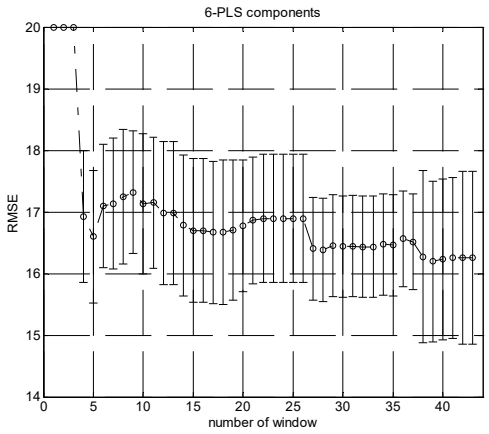


Figure A.2: plot of prediction error using six latent vectors; lowest error is visible at window 39, but standard deviation of error is higher than windows before

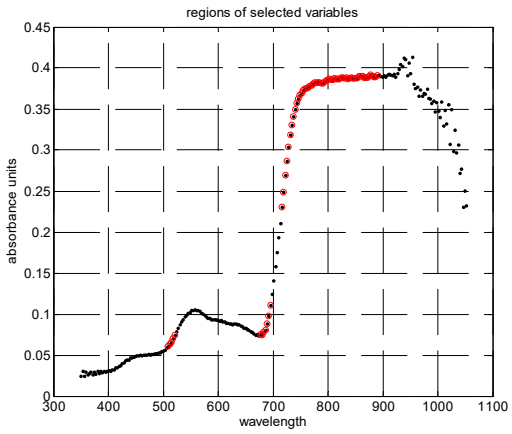


Figure A.3: chosen window contains the red marked variables (wave lengths)

A.2 Linearization – Kernel Matrix

Several different ways for linearization of potential non-linear data are reported in literature, such as non-linear extension of data matrix by e. g. **polynomial extensions** [44] or OSC correction [39]. Another common way is the construction of a **kernel matrix**. A certain sample set of input data **X** [n x m] is restructured using the direct dependence of two samples to each other. This means, the samples of **X** are transformed into a new feature space using nonlinear mapping [45]. This results in a new input matrix **K** [n x n] (Equation 1).

$$K = \begin{bmatrix} k_{1,1} & \cdots & k_{1,n} \\ \vdots & \ddots & \vdots \\ k_{n,1} & \dots & k_{n,n} \end{bmatrix} \tag{1}$$

Where $k_{i,j}$ can take the form of several different functions. The most famous are the linear or covariance, the polynomial and the radial basis function kernel (RBF, Equation 2). The description of those functions can be found in literature, for instance Nicolai *et al.* 2007 [45].

$$k_{i,j} = e^{\frac{-\left\|x_i^T - x_j^T\right\|^2}{2*\sigma^2}} \tag{2}$$

The Kernel width parameter σ is linked to the reliability or the SNR of the data. If this parameter is higher, the solution of the model becomes more linear. Over all, the value of $k_{i,j}$ becomes one in case of samples that are more similar and zero in case of less similar ones.

Further, it is recommended to always perform a centering of the Kernel matrix prior to analysis. Therefore, Bennett and Embrechts propose the following two equations [46]:

$$\boldsymbol{H} = \left(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_{\boldsymbol{n}}\mathbf{1}_{\boldsymbol{n}}^{\boldsymbol{T}}\right) \tag{3}$$

$$\boldsymbol{K}_{\boldsymbol{v,center}} = \left(\boldsymbol{K}_{\boldsymbol{v}} - \frac{1}{n}\mathbf{1}_{\boldsymbol{n_v}}\mathbf{1}_{\boldsymbol{n}}^{\boldsymbol{T}}\boldsymbol{K}\right)\boldsymbol{H} \tag{4}$$

$$\boldsymbol{K}_{\boldsymbol{center}} = \boldsymbol{HKH} \tag{5}$$

Where **I** is a [n x n] identity matrix, $\mathbf{1}_n$ and $\mathbf{1}_{nv}$ vectors of ones with [n] and [n_v] dimensionality.

A.3 External Parameter Orthogonalisation

The methodology used can be found in detail in Roger *et al.*, 2003 [47]. They present a data matrix as follows:

$$\boldsymbol{X} = \boldsymbol{XP} + \boldsymbol{XQ} + \boldsymbol{R} \tag{1}$$

where **P** contain the loadings of the projection onto the relevant target information (relevant subspace $\vec{\mathbf{C}}$) and **Q** onto the external parameter influence (subspace $\vec{\mathbf{G}}$, containing influence of external parameter); **R** resembles a residual matrix. In case of being able to calculate or guess an infectious subspace $\widehat{\mathbf{G}}$, an estimation of **Q** can be estimated by following expression:

$$\widehat{\boldsymbol{Q}} = \widehat{\boldsymbol{G}}\left(\widehat{\boldsymbol{G}}^{T}\widehat{\boldsymbol{G}}\right)^{-1}\widehat{\boldsymbol{G}}^{\boldsymbol{T}} \tag{2}$$

The pre-processing by EPO will subsequently transform **X** into the relevant part **X*** by:

$$\boldsymbol{X}^{*} = \boldsymbol{X}\left(\boldsymbol{I} - \widehat{\boldsymbol{Q}}\right) \tag{3}$$

Assuming one to k matrices $\mathbf{X}^{i}$ with n samples and m variables having k different values for an external parameter, for instance temperature T. Thus, it is possible to create a matrix **M** [k x p] with k averaged spectra on each $\mathbf{X}^{i}$, where each row is calculated as:

$$\boldsymbol{m}_{i} = \frac{1}{n}\sum_{j=1}^{j=n}\boldsymbol{x}_{j}^{i} \tag{4}$$

Further, matrix **D** [k x p] is defined by $\mathbf{d}_i = \mathbf{m}_i - \mathbf{m}_1$. This definition together with Equation 1 gives:

$$\boldsymbol{d}_{i} = (\boldsymbol{m}_{i} - \boldsymbol{m}_{1})\boldsymbol{P} + (\boldsymbol{m}_{i} - \boldsymbol{m}_{1})\boldsymbol{Q} + \frac{1}{n}\sum_{j=1}^{j=n}\left(\boldsymbol{r}_{j}^{i} - \boldsymbol{r}_{1}^{i}\right) \tag{5}$$

Taking into account, that all $\mathbf{m}_i$ are by definition the mean spectra of the same targets, $(\mathbf{m}_i - \mathbf{m}_1)\mathbf{P} = 0$. Thus, the final matrix form is given by:

$$\boldsymbol{D} = \boldsymbol{AQ} + \boldsymbol{R}' \tag{6}$$

Roger *et al.*, 2003 [47] propose to retrieve subspace $\widehat{\mathbf{G}}$ by a PCA of **D** resulting in:

$$\boldsymbol{D} = \boldsymbol{T}\widehat{\boldsymbol{G}}^{\boldsymbol{T}} + \boldsymbol{R}'' \tag{7}$$

Since the columns of $\widehat{\mathbf{G}}$ are therefore orthogonal and of unitary length, $\widehat{\boldsymbol{G}}^{\boldsymbol{T}}\widehat{\boldsymbol{G}} = \boldsymbol{I}$. Equation 27 will reduce to $\widehat{\boldsymbol{Q}} = \widehat{\boldsymbol{G}}\widehat{\boldsymbol{G}}^{\boldsymbol{T}}$ and

$$\boldsymbol{X}^{\boldsymbol{0}*} = \boldsymbol{X}^{\boldsymbol{0}}\left(\boldsymbol{I} - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{G}}^{\boldsymbol{T}}\right) \tag{8}$$

Finally, a calibration can be calculated between $\mathbf{X}^{0*}$ and $\mathbf{Y}^{0}$, any new sample is preprocessed by:

$$\boldsymbol{x}_{\boldsymbol{new}}^{*} = \boldsymbol{x}^{*}\left(\boldsymbol{I} - \widehat{\boldsymbol{G}}\widehat{\boldsymbol{G}}^{\boldsymbol{T}}\right) \tag{9}$$

The only open question is, how much components should be used to reduce the effect of the external parameter. Two of several possibilities are presented in Roger *et al.*, 2003 [47]: a k-fold cross validation on the different $\mathbf{X}^{i}$ resulting in an error as a function of EPO component and PLS latent variables number (which requires knowledge about corresponding responses to samples in **X**). The second approach is based on an analysis of variance measured by Wilk's ratio between the inner group and the total variance [47]. This can be explained in a geometrical way: without extracting the external parameter influence, two different samples measured at the same external parameter settings can match more than two equal samples measured at two different external parameter settings. Therefore, the n clusters of k equal samples do not separate. If they do not separate at all, the ration of variances equal zero. In case of perfect separation, the ratio equals one.

The approach used in the presented work is slightly different. The choice is done on the final error (similar to the approach one above). But the error is calculated on a validation data set. This way was chosen to combine the approach with kernel PLS and the method of robust calibration.

A.4 Additional results for SVR-model on full data set

The figures are additional to the results presented in the main document and present the distribution of samples used in the calibration and validation data based on concentrations and temperatures (Figure A.4), final model error statistic (Figure A.5 + A.6) and the single concentration level error statistics (Figure A.7).

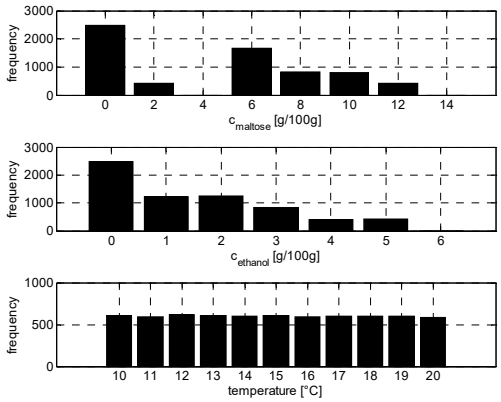


Figure A.4: frequency of respective concentrations and temperatures in samples taken for all investigations in calibration and validation

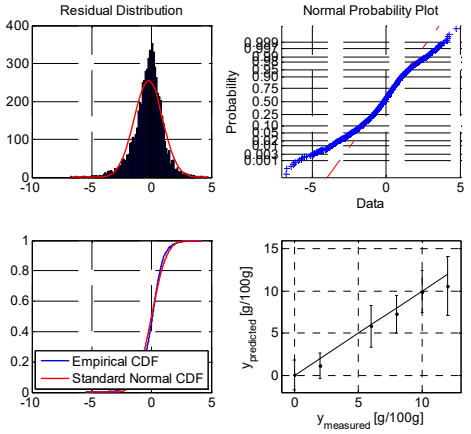


Figure A.5: comparison four graphical validation methods proving normal distribution of residual errors of SVR model solution; residual distribution (top left), linearized normal probability plot (top right), cumulative distribution function (down left) and parity plot with two times standard deviation as error bars (down right)

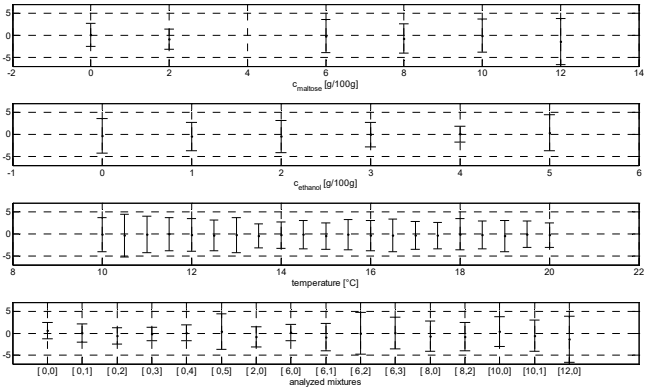


Figure A.6: resulting error and its standard deviation separated according to concentration level, temperature and the concentration combination in the samples; there are differences visible, but no clear systematic error behaviour is appearing

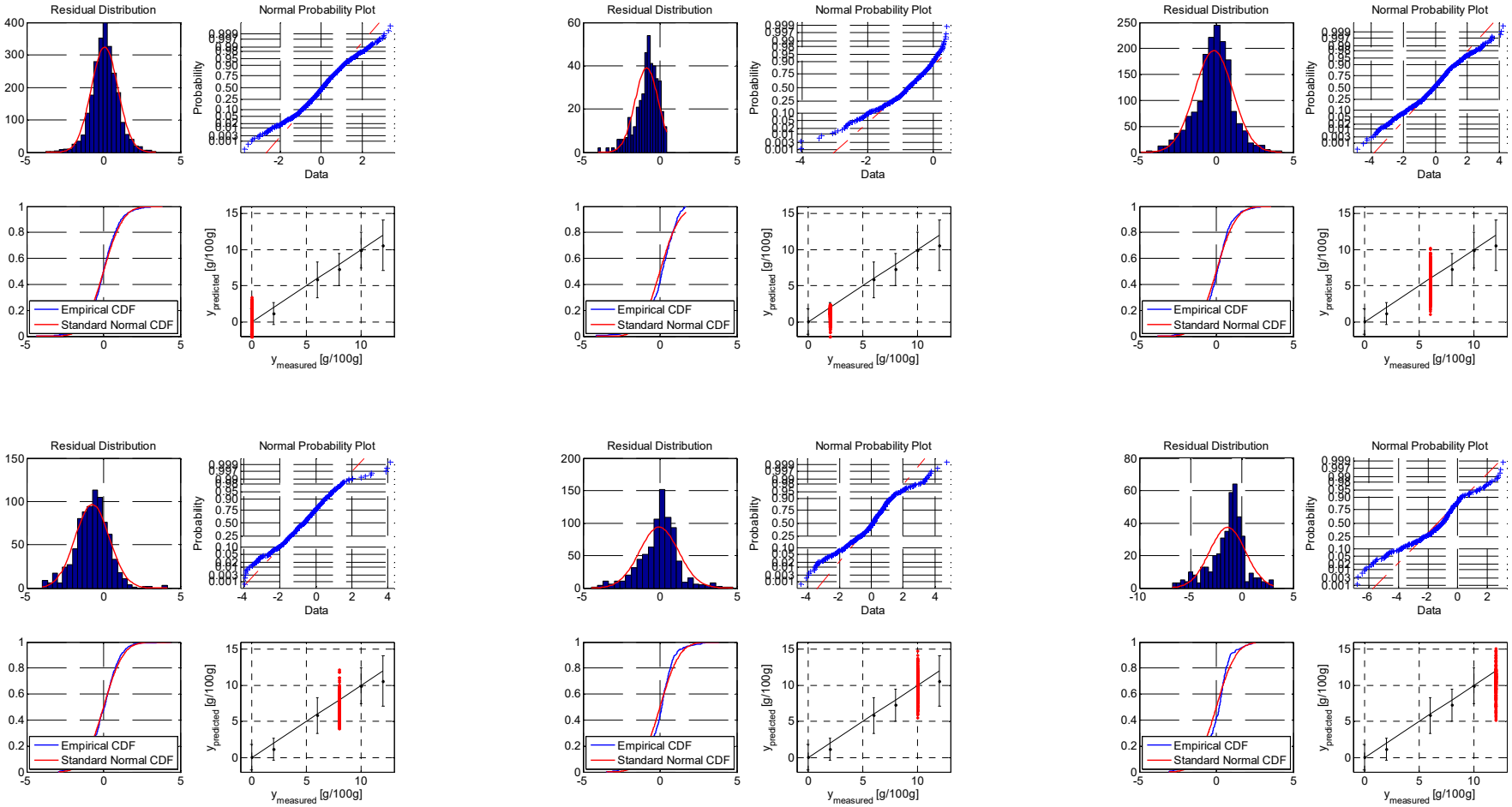


Figure A.7: resulting error statistics for individual concentration levels of SVR model solution; except for 2 g/100g (further investigations needed), no clear systematic error behaviour is visible

A.5 Additional results for Kernel PLS on full data set and EPO

The figures are supplementary to the results presented in the main document and present an addition to the final model error statistic (Figure A.8 for nine LV and Figure A.9 for 15 LV) and the single concentration level error statistics (Figure A.10 for nine LV and Figure A.11 for 15 LV).

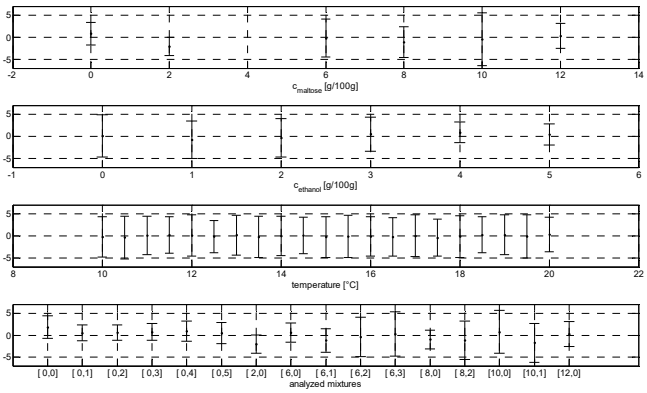


Figure A.8: resulting error and its standard deviation separated according to concentration level, temperature and the concentration combination in the samples for the model with nine LV; there are differences visible, but no clear systematic error behaviour is appearing

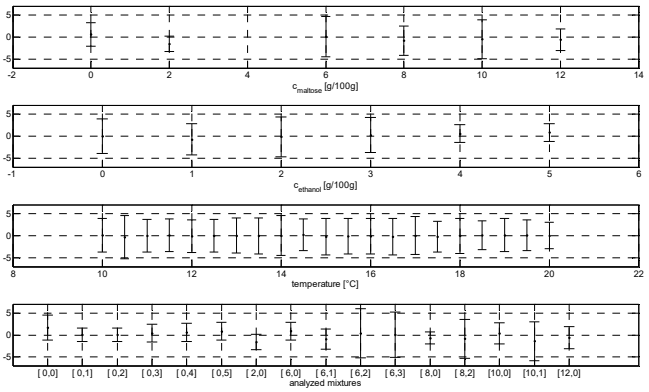


Figure A.9: resulting error and its standard deviation separated according to concentration level, temperature and the concentration combination in the samples for the model with 15 LV; there are differences visible, but no clear systematic error behaviour is appearing

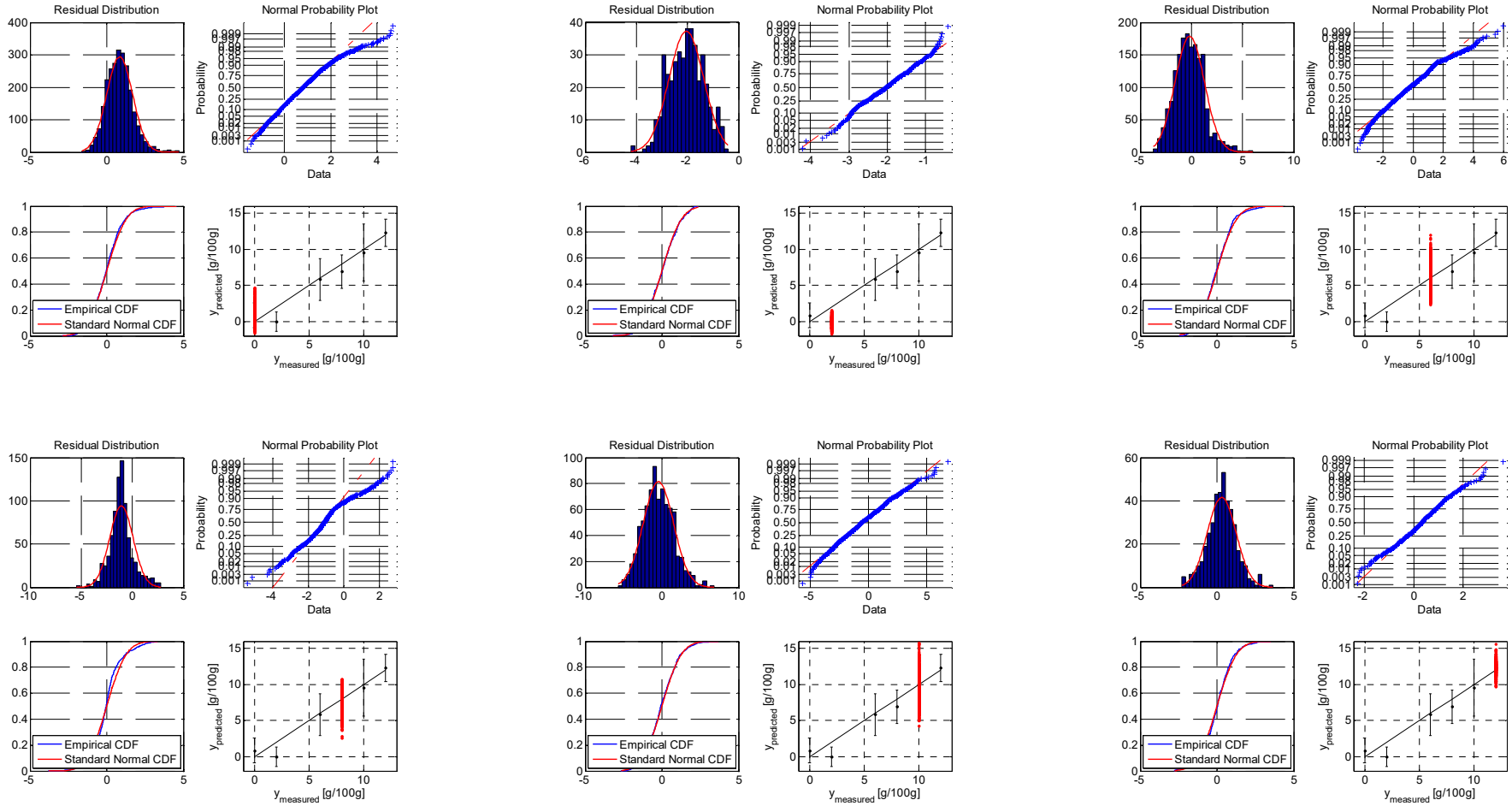


Figure A.10: resulting error statistics for individual concentration levels of kernel-PLS model solution with nine LV; except for 2 g/100g (further investigations needed), no clear systematic error behaviour is visible

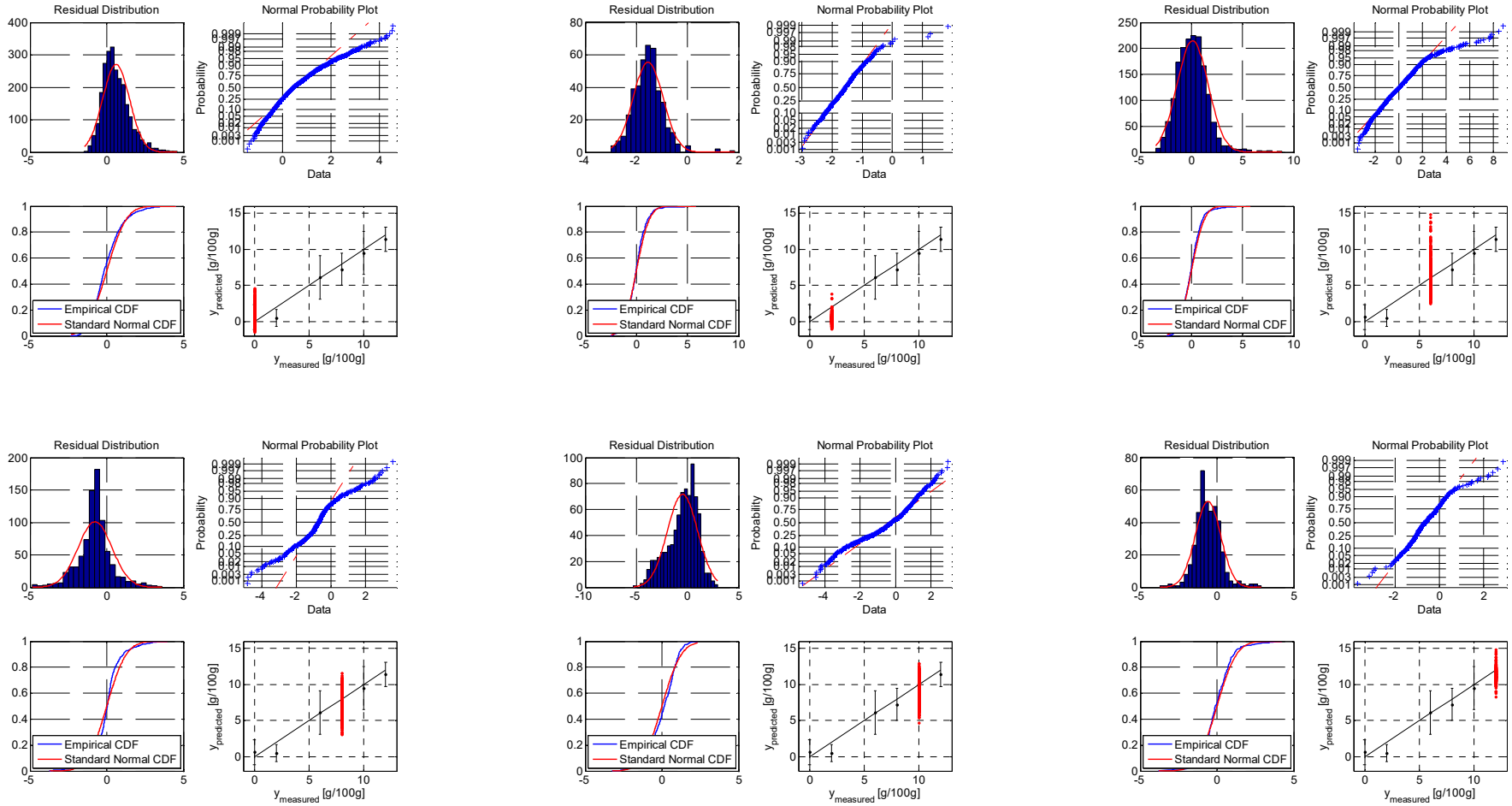


Figure A.11: resulting error statistics for individual concentration levels of kernel-PLS model solution with 15 LV; except for 2 g/100g (further investigations needed), no clear systematic error behaviour is visible

A.6 Kernel-PLS on full data set and polynomial extension

The figures are supplementary material to the results presented in the main document for the model solution created by polynomial extension of input data.

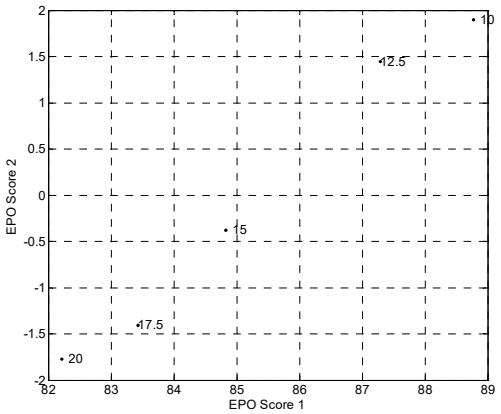


Figure A.12: Score plot of EPO investigation on the dataset of step size 2.5 °K and all samples with concentrations between zero and 14 g/100g maltose and zero to 5 g/100g ethanol; it is visible, that the first component is enough to clearly distinguish between the different temperatures.

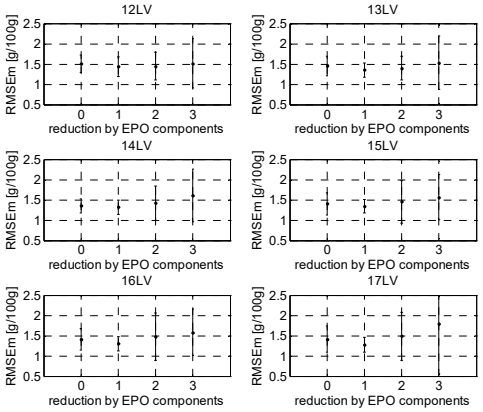


*Figure A.13: each model is compared individually in the pre-chosen area plotting residual error against number of used EPO components; error bars = three times standard deviation (3*σ); 14-15 LV and one EPO component seem to be suitable*

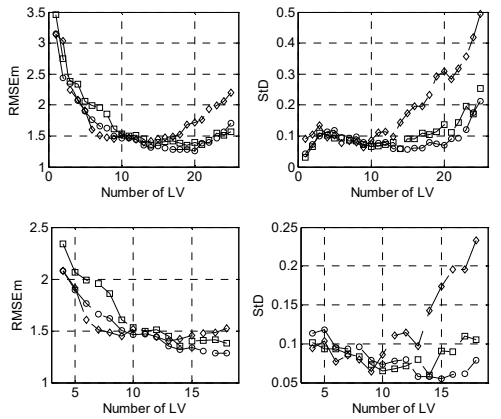


Figure A.14: comparison between mean prediction error (left side) and their standard deviation (right side) for original data set models (squares), data set reduced by one EPO component (circles) and reduced by two EPO components (diamonds); the top two figures show plot from one to 25, the figures at the bottom from three to 18 latent vectors; two EPO components are obviously too much for the presented case (error as well as standard deviation are mostly higher); although the differences between original and reduced data set by one EPO are quite low, there are better accuracies achieved 15 LV by the latter.

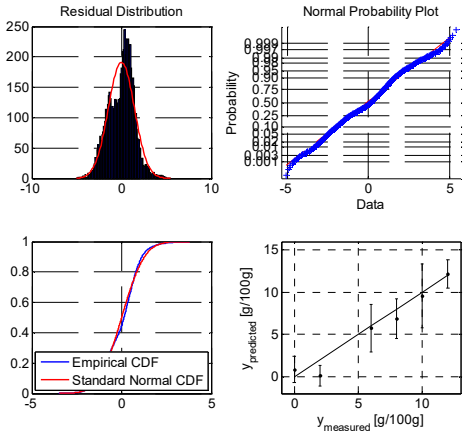


Figure A.15: comparison four graphical validation methods proving normal distribution of residual errors for kernel-PLS model solution on data set with polynomial extension; residual distribution (top left), linearized normal probability plot (top right), cumulative distribution function (down left) and parity plot with two times standard deviation as error bars (down right)

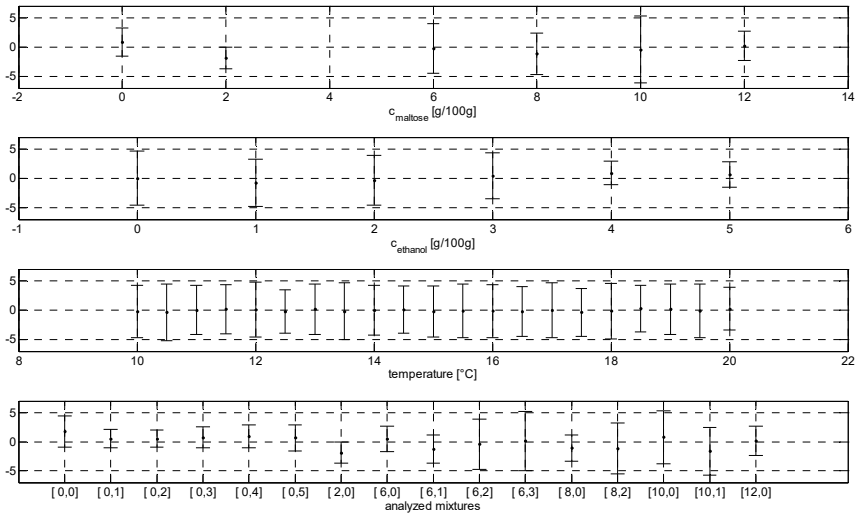


Figure A.16: resulting error and its standard deviation separated according to concentration level, temperature and the concentration combination in the samples; there are differences visible, but no clear systematic error behaviour is appearing

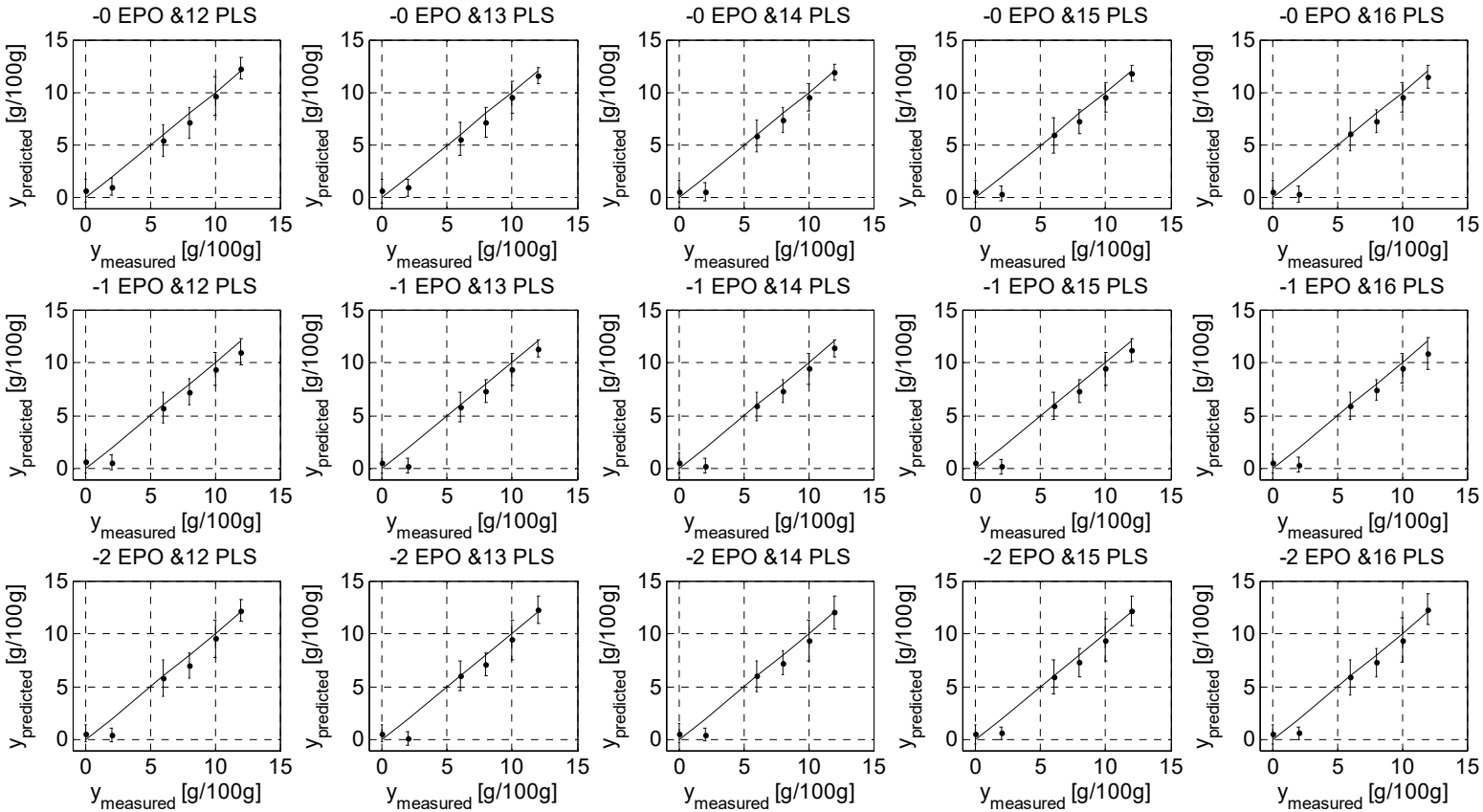


Figure A.17: visual inspection of parity plots (error bars of one times standard deviation) with rising number of EPO components (vertical) and rising number of latent vectors (horizontal); the differences are again only minor

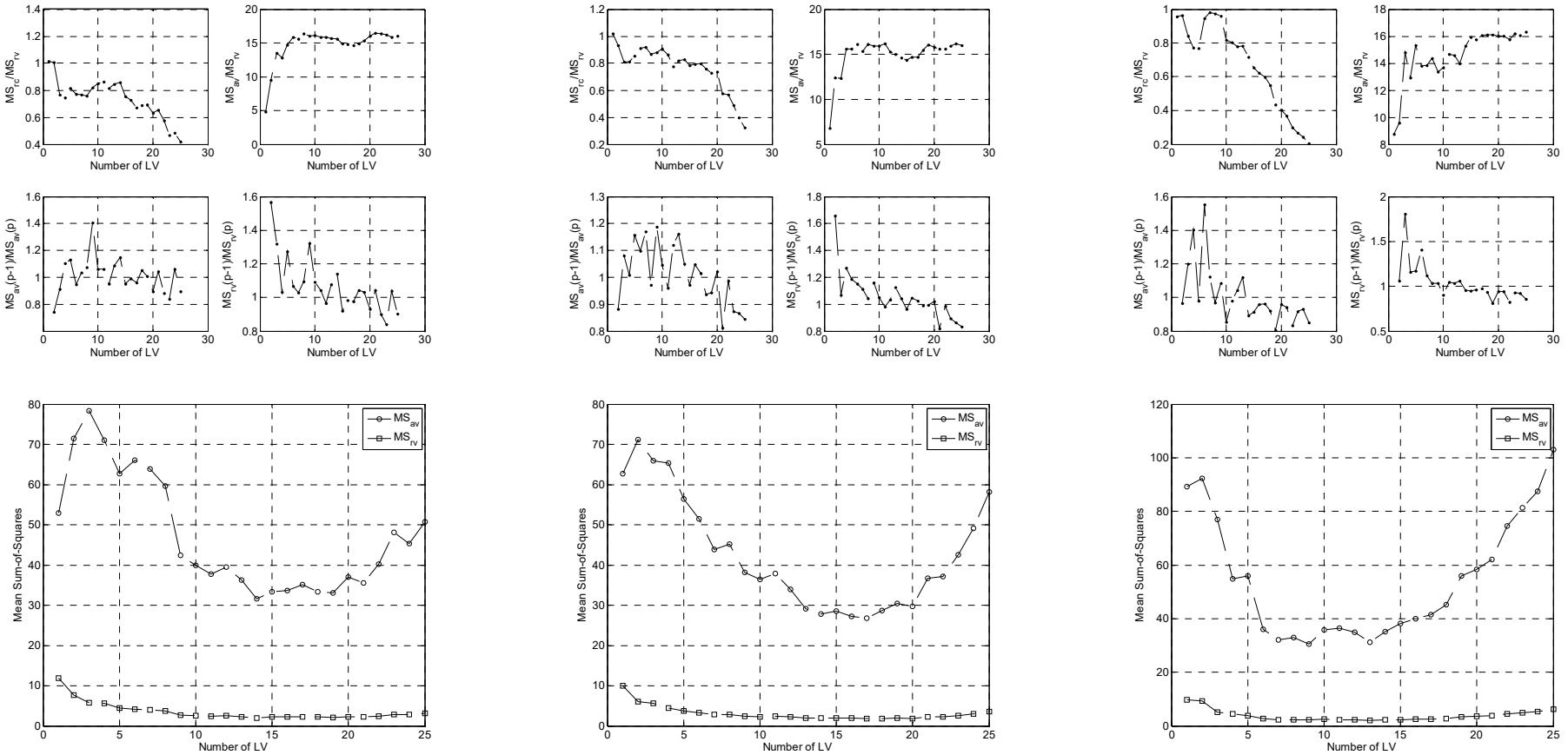


Figure A.18: plots of validity and precision as well as ratios of mean sum of squares in different variants using original data set (left), reduced by one (middle) and two EPO components (right). Figures on the top present ratios MS_{rc}/MS_{rv} (validity calibration/validity validation, top left, decreasing with rising number of components due to overfitting), MS_{av}/MS_{rv} (precision/validity, top right, convergence supports the assumption of no necessity to include more components or latent vectors), $MS_{av}(p\text{-}1)/MS_{av}(p)$ (precision, bottom left) and $MS_{rv}(p\text{-}1)/MS_{rv}(p)$ (validity, bottom right) – the latter both converge around one, which supports again the assumption of no necessity to include more components or latent vectors; figures at the bottom present precision and validity mean sum of squares as separate plot - dashed line with circles – MS_{av} (precision), dashed line with squares - MS_{rv} (validity) – a (local) minimum in both is preferable

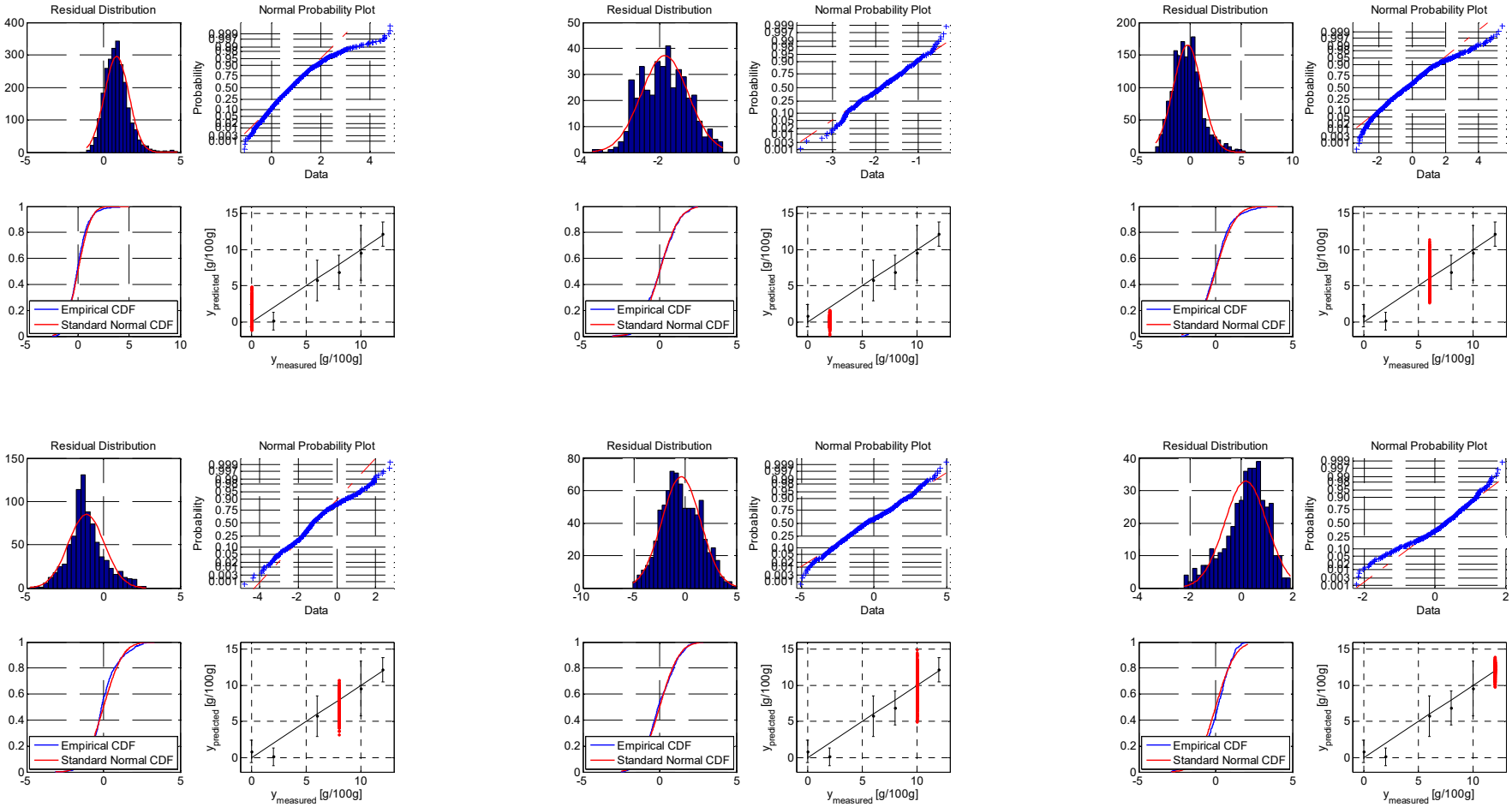


Figure A.19: resulting error statistics for individual concentration levels of kernel-PLS model solution with 15 LV; except for 2 g/100g (further investigations needed), no clear systematic error behaviour is visible

A.7 Kernel-PLS on binary mixture with maltose

The figures are supplementary material to the results presented in the main document for the model solution created on calibration data with binary mixtures of maltose only.

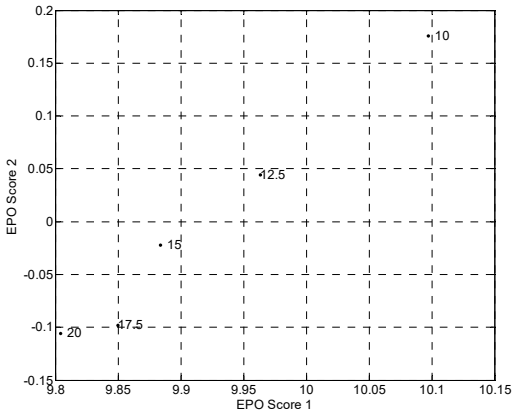


Figure A.20: Score plot of EPO investigation on the dataset of step size 2.5 °K and all samples with concentrations between zero and 14 g/100g maltose and zero to 5 g/100g ethanol; it is visible, that the first component is enough to clearly distinguish between the different temperatures.

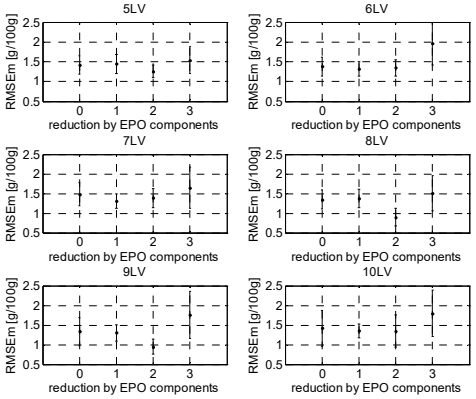


*Figure A.21: each model is compared individually in the pre-chosen area plotting residual error against number of used EPO components; error bars = three times standard deviation (3*σ); eight LV and two EPO component seem to be suitable*

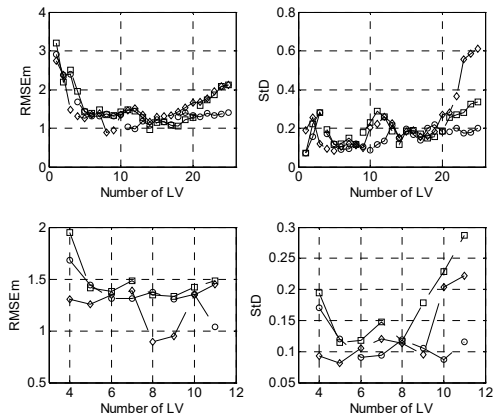


Figure A.22: comparison between mean prediction error (left side) and their standard deviation (right side) for original data set models (squares), data set reduced by one EPO component (circles) and reduced by two EPO components (diamonds); the top two figures show plot from one to 25, the figures at the bottom from three to 18 latent vectors; two EPO components are obviously too much for the presented case (error as well as standard deviation are mostly higher); although the differences between original and reduced data set by one EPO are quite low, there are better accuracies achieved around eight to 10 and 14/15 LV by the latter. Even though the error decreases after 10 LV again, eight to nine LV should be ideal considering the standard deviation, since the deviation rises for both data set afterwards.

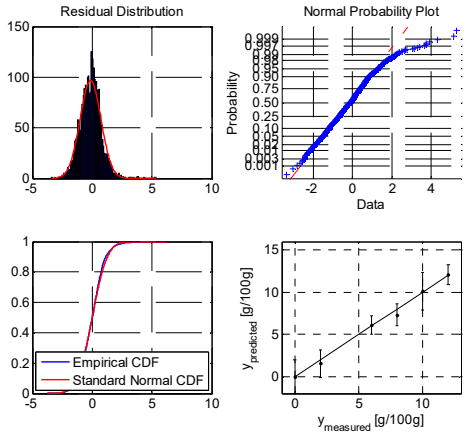


Figure A.23: comparison four graphical validation methods proving normal distribution of residual errors for kernel-PLS model solution on data set with binary maltose only; residual distribution (top left), linearized normal probability plot (top right), cumulative distribution function (down left) and parity plot with two times standard deviation as error bars (down right)

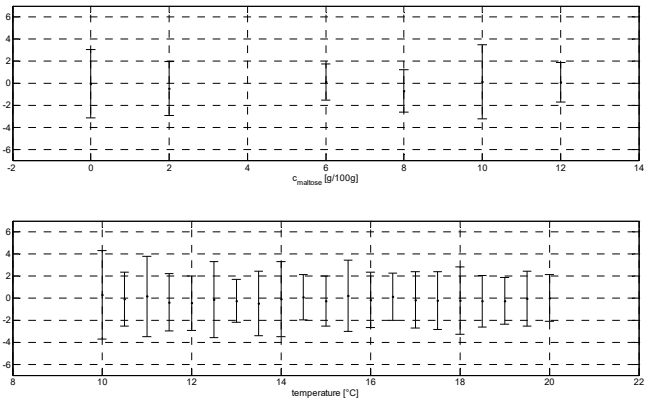


Figure A.24: resulting error and its standard deviation separated according to concentration level and temperature in the samples; there are differences visible, but no clear systematic error behaviour is appearing

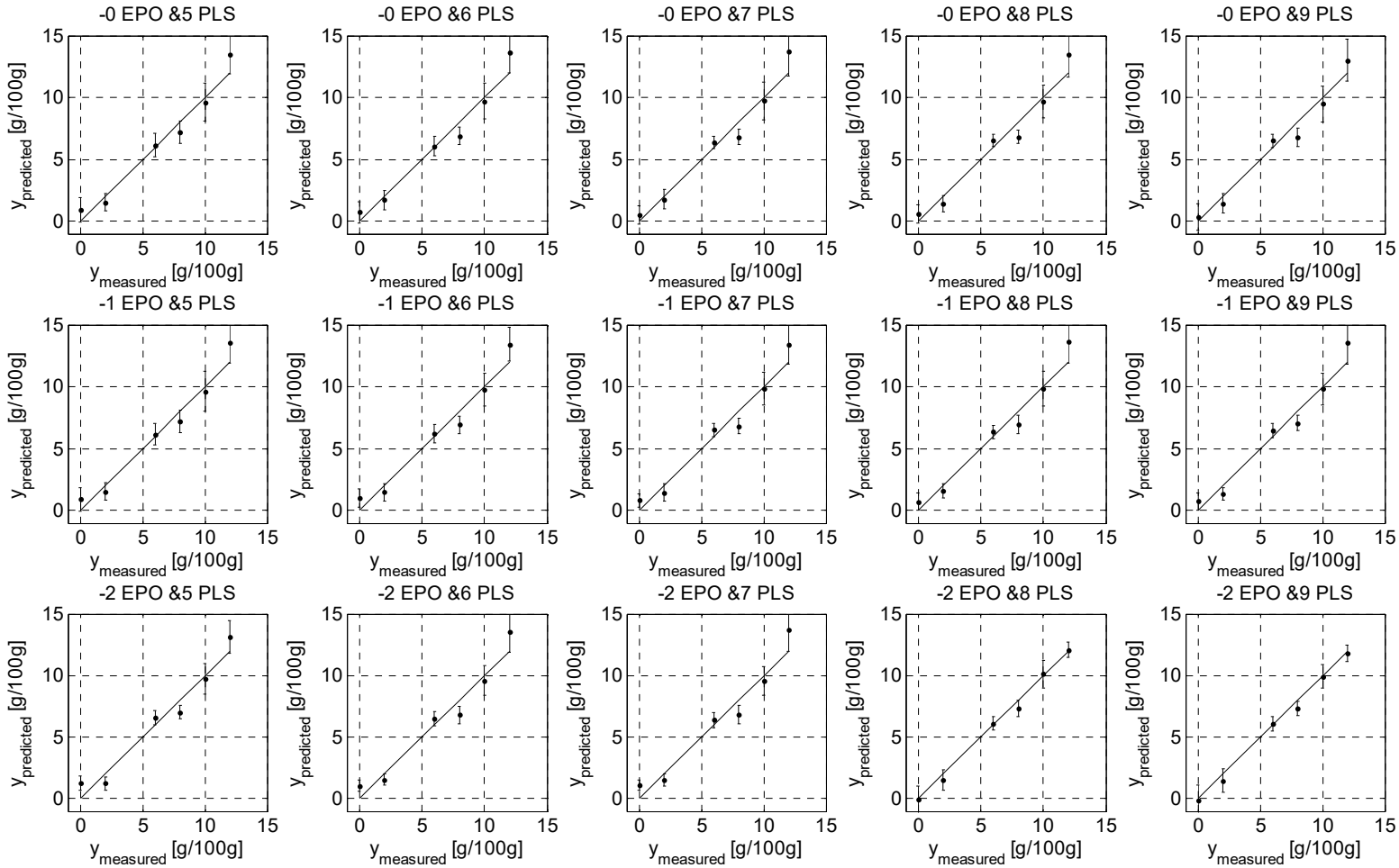


Figure A.25: visual inspection of parity plots (error bars of one times standard deviation) with rising number of EPO components (vertical) and rising number of latent vectors (horizontal)

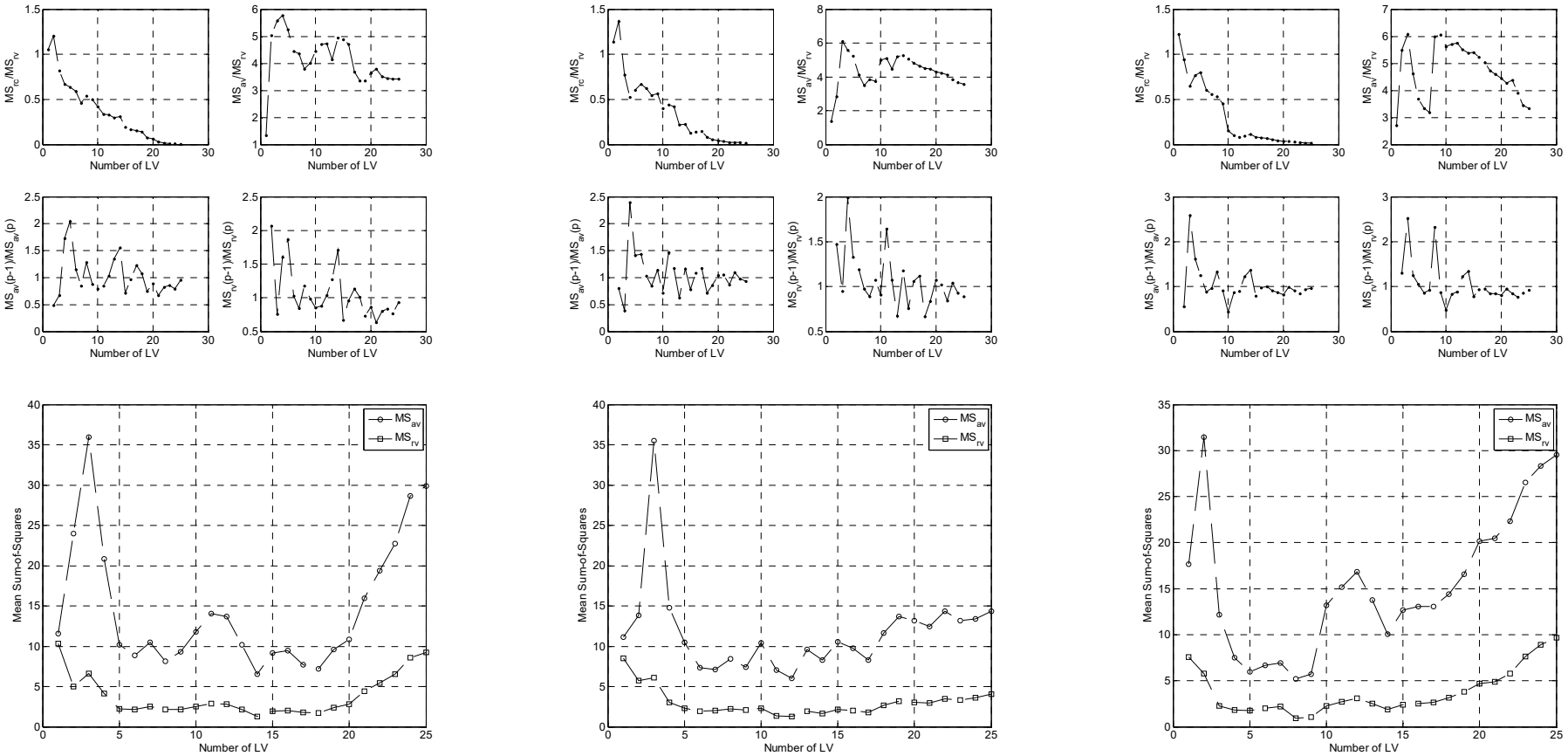


Figure A.26: plots of validity and precision as well as ratios of mean sum of squares in different variants using original data set (left), reduced by one (middle) and two EPO components (right). Figures on the top present ratios MS_{rc}/MS_{rv} (validity calibration/validity validation, top left, decreasing with rising number of components due to overfitting), MS_{av}/MS_{rv} (precision/validity, top right, convergence supports the assumption of no necessity to include more components or latent vectors), $MS_{av}(p\text{-}1)/MS_{av}(p)$ (precision, bottom left) and $MS_{rv}(p\text{-}1)/MS_{rv}(p)$ (validity, bottom right) – the latter both converge around one, which supports again the assumption of no necessity to include more components or latent vectors; figures at the bottom present precision and validity mean sum of squares as separate plot - dashed line with circles – MS_{av} (precision), dashed line with squares - MS_{rv} (validity) – a (local) minimum in both is preferable

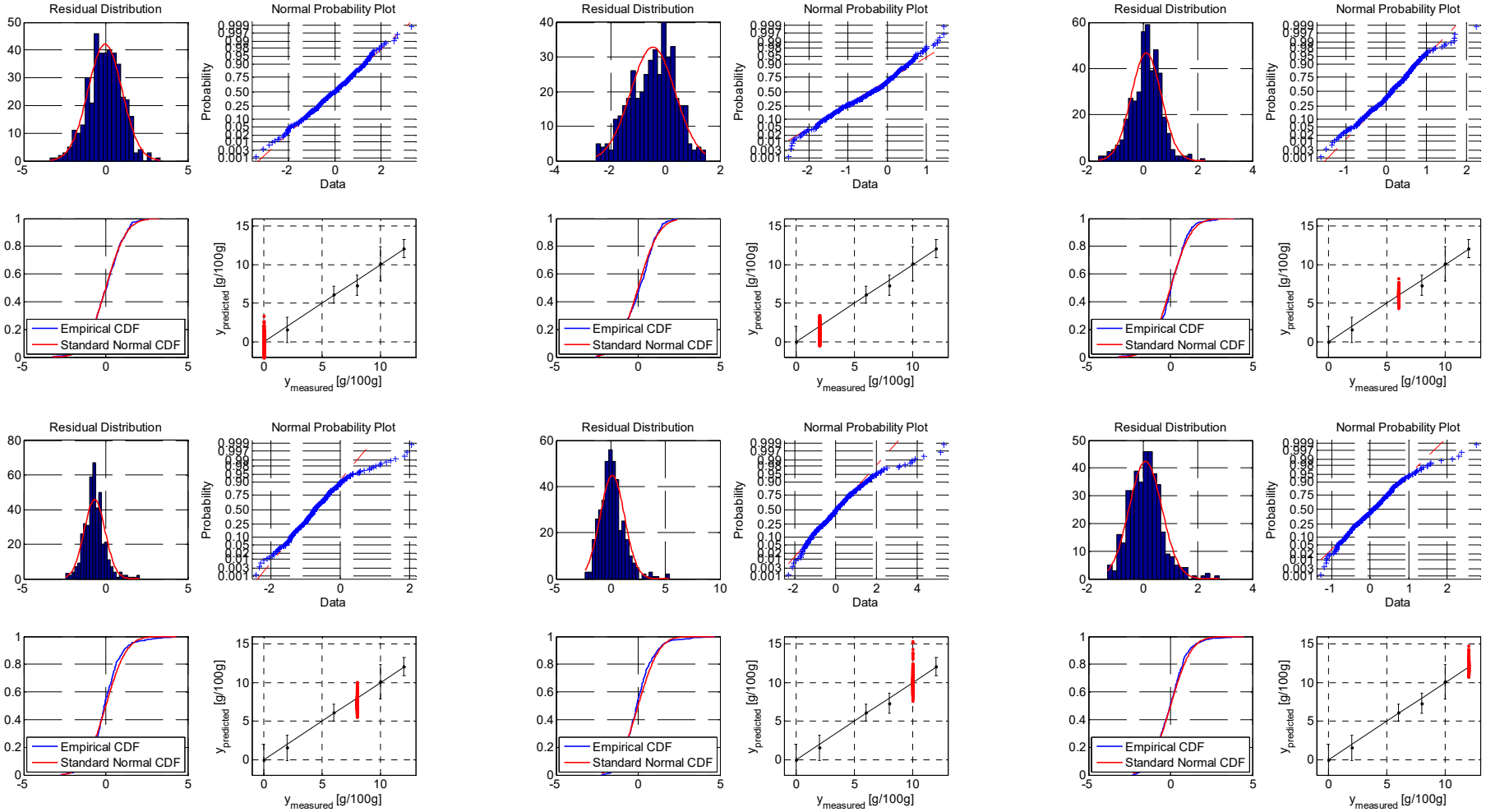


Figure A.27: resulting error statistics for individual concentration levels of kernel-PLS model solution; there are indications for systematic error behaviour visible (further investigations needed)