Online monitoring of exhaled breath markers in Mouse Intensive Care Unit via infrared spectroscopy combined with luminescence sensors and Bayesian methodology

Dissertation zur Erlangung des Doktorgades
Dr. rer. nat der Fakultät Naturwissenschaften
der Universität Ulm

Vorgelegt von:

Dipl.-Chem. Felicia Seichter
aus Reutlingen

2020
Amtierender Dekan:
Prof. Dr. Thorsten Bernhardt

Gutachter:
Prof. Dr. Boris Mizaiikoff
Prof. Dr. Dr. Peter Radermacher

Tag der Promotion:
8.12.2020
Für meine Großmutter

“The ones who love us never really leave us”

(Harry Potter and the Prisoner of Askaban)
5.3 Paper II: Response-surface fits and calibration transfer for the correction of the oxygen effect in the quantification of carbon dioxide via FTIR spectroscopy ........................................... 62

5.4 Paper III: Online monitoring of carbon dioxide and oxygen in exhaled mouse breath via substrate-integrated hollow waveguide - Fourier transform infrared - luminescence spectroscopy ................... 75

5.5 Paper IV: Strategies for 13C enrichment calculation in Fourier-transform Infrared CO2 spectra containing spectral overlapping and nonlinear abundance-amount relations utilizing response surface . 87

5.6 Paper V: Metabolic monitoring using online analysis of 13C enriched carbon dioxide in exhaled mouse breath via iHWG-FTIR LS spectroscopy and Bayesian Sampling ........................................ 101

Acknowledgements .................................................. 117

Curriculum Vitae .................................................... 118

Publications ......................................................... 120

Eidesstattliche Erklärung ........................................... 123
Abstract

The research presented in this cumulative thesis focuses on the development of analyzer systems capable of monitoring relevant physiological parameters in exhaled breath in a mouse intensive care unit (MICU) and the development of corresponding data analysis strategies. The thesis is based on five peer-reviewed journal articles resulting from the interdisciplinary research collaboration of the Institute of Analytical and Bioanalytical Chemistry (IABC) at Ulm University and the Institute of Anesthesiologic Pathophysiology and Method Development (APV) at Ulm University Medical Center.

Within the framework of this thesis, two major goals were achieved. Firstly, an infrared spectroscopic sensing system based on a modified Fourier transform infrared (FTIR) spectrometer using substrate-integrated hollow waveguide (iHWG) technology pioneered at IABC in combination with luminescence based oxygen sensors (LS) was developed and integrated into the MICU at the APV for continuously monitoring exhaled mouse breath during medical experiments. The main aim was the seamless integration into existing medical instrumentation and medical routine. Secondly - and even more important – was the development of a data analysis procedure for the established iHWG-FTIR-LS system for oxygen, carbon dioxide and $^{13}$C enrichment. For that purpose, a complex calibration and data mining strategy was developed using non-linear calibration algorithms for the oxygen sensors via Hierarchical Models and Lagrange Multipliers enabling calibration transfer. Furthermore, an algorithm for humidity compensation was introduced next to the implementation of a response surface calibration strategy compensating for oxygen effects on the CO$_2$ calibration. In addition to the calibration transfer algorithm, direct calibration of $^{13}$C enrichment in terms of mole fraction using a novel curve resolution algorithm termed RABBIT-MCR was introduced, and a non-linear response surface calibration via a Hierarchical Model approach was achieved. All data mining methods were based on Bayesian sampling and modeling ensuring that results from one calibration are incorporated into later data evaluation routines (e.g., oxygen concentration into the CO$_2$ algorithm etc.), essentially building upon each other. This resulted in a unified data analysis procedure for the derived physiological parameters (i.e., RQ, $f_{fat}$, $f_{carb}$, $f_{prot}$, TEE) in biologically relevant concentration ranges.
The key milestone of this thesis was certainly the development of this fully integrated data analysis strategy, giving access to relevant physiological parameters (RQ, $f_{fat}$, $f_{carb}$, $f_{prot}$, TEE) at a temporal resolution not available to date, owing to the applied Bayesian methodology. This data analysis routine is actually independent of the experimental setup developed in the framework of this thesis and may thus be adapted and extended to alternative analyzer systems offering the same core data (i.e., oxygen, carbon dioxide concentration, $^{13}$C enrichment, $^{13}$C glucose plasma enrichment, etc.). Another key novelty of this work is the development of two new chemometric algorithms: (i) calibration transfer using Hierarchical Models and Lagrange Multipliers, which is necessary for daily recording of a calibration curve using only few calibration samples, and (ii) a new curve resolution algorithm termed RABBIT-MCR, for the deconvolution of overlapping spectral signals into their major contributions. The data algorithms and calibration methods were tailored to and tested for the experimental setup developed in the present thesis, however the underlying principles apply to any analytical system analyzing (breath) gas or other vapor phase matrices.

In summary, this thesis shows the development of an analysis system capable of monitoring relevant physiological parameters (i.e., oxygen concentration, carbon dioxide concentration, $^{13}$C enrichment, RQ, $f_{fat}$, $f_{carb}$, $f_{prot}$, TEE) in the exhaled breath of mice in a mouse intensive care unit, and in particular, the development of advanced data calibration and quantification routines using innovative chemometric tools including Bayesian statistics and non-linear calibration methods for automated daily use in clinical settings.
Zusammenfassung

Die wissenschaftliche Forschung, die in dieser kumulativen Doktorarbeit präsentiert wird, fokussiert sich auf der Entwicklung eines Analysesystems, das in der Lage ist, wichtige physiologische Parameter in einer Mäuse-Intensivstation zu überwachen, sowie auf die Entwicklung grundlegender Datenanalysestrategien. Die Arbeit basiert auf fünf referierten wissenschaftlichen Veröffentlichungen, die das Ergebnis einer interdisziplinären Forschungskooperation zwischen dem Institut für Analytische und Bioanalytische Chemie der Universität Ulm (IABC), und dem Institut für Anästhesiologische Pathophysiologie und Verfahrensentwicklung des Universitätsklinikum Ulm (APV) darstellen.

Zusammenfassung

Modellierung, wobei die Ergebnisse jeder Kalibrierung in die späteren Kalibrierungen integriert werden (d.h., die Sauerstoffkonzentration ergänzt später den CO2 Kalibrieralgorithmus, usw.), wodurch diese aufeinander aufbauen. Dies resultiert letztendlich in einem Datenauswerteverfahren, das physiologische Parameter wie RQ, \( f_{fat} \), \( f_{carb} \), \( f_{prot} \) und TEE in biologisch sinnvollen Konzentrationen ableiten kann.

Zentraler Meilenstein dieser Arbeit im Speziellen ist sicherlich die Entwicklung der Datenauswertestrategie, die dank der Bayes-Methode Zugang zu den obengenannten physiologischen Parametern (RQ, \( f_{fat} \), \( f_{carb} \), \( f_{prot} \) und TEE) in biologisch sinnvollen Konzentrationsbereichen mit einer bislang nicht verfügbaren Zeitauflösung bietet. Der Datenanalysealgorithmus ist tatsächlich unabhängig von dem in dieser Arbeit entwickelten experimentellen Messsystem und kann auf andere Analysesysteme, die die gleichen Kerndaten (d.h., Sauerstoff- und Kohlendioxidkonzentration, \(^{13}\)C Anreicherung im Atemgas und \(^{13}\)C Plasma-Glucose Anreicherung) abgreifen, entsprechend erweitert und angepasst werden. Eine weitere Innovation dieser Forschungsarbeit ist die Entwicklung von zwei neuartigen chemometrischen Algorithmen: (i) Kalibrationstransfer mit hierarchischen Modellen und Lagrange Multiplikatoren, sodass für die täglichen Aufnahme von Kalibrierkurven wenigen Kalibrierproben zwingend nötig sind; (ii) die Einführung eines neuartigen Kurvenauflösungs-Algorithmus - RABBIT-MCR - der für die Zerlegung von überlappenden spektralen Signalen in die wesentlichen molekularen Beiträge eingesetzt wurde. Die Datenalgorithmen und Kalibriermethoden sind zwar für das in dieser Arbeit entwickelte Messsystem maßgeschneidert und getestet worden, aber die zugrunde liegenden Prinzipien können für jedes (Atem-)Gas Analysesystem eingesetzt und adaptiert werden.

Zusammengefasst zeigt diese Doktorarbeit die Entwicklung eines Analysesystems, welches in der Lage ist, relevante physiologische Parameter (d.h., Sauerstoff- und Kohlendioxidkonzentration, \(^{13}\)C Anreicherung im Atemgas, RQ, \( f_{fat} \), \( f_{carb} \), \( f_{prot} \), TEE) in der Mäuseintensivstation kontinuierlich zu überwachen, und insbesondere die Entwicklung der zugrunde liegenden Kalibrierungs- und Quantifizierungsalgorithmen unter Nutzung innovativer chemometrischer Werkzeuge wie der Bayes-Statistik und nichtlinearen Kalibriermethoden, wodurch eine vollständige Automatisierung des Messablaufs im täglichen Routinegebrauch im klinischen Umfeld gegeben ist.
Zusammenfassung

Zentraler Meilenstein dieser Arbeit im Speziellen ist die Entwicklung einer Datenauswertestrategie, die dank der Bayes-Methodologie Zugang zu den obenge nannten biologischen Parametern (RQ, \( f_{fat} \), \( f_{carb} \), \( f_{prot} \) und TEE) in biologisch sinnvollen Konzentrationen in einer vorher nie verfügbaren Zeitauflösung bietet. Der Datenanalyse-Algorithmus ist tatsächlich unabhängig von dem in dieser Arbeit verwendeten Messsystem und kann auf alle mögliche Analysesysteme, die die gleichen Kerndaten (Sauerstoff- und Kohlendioxidkonzentration, \(^{13}\text{C} \) Anreicherung im Atemgas und \(^{13}\text{C} \) Plasma-Glucose Anreicherung) bieten, erweitert und angepasst werden. Eine weitere Neuerung dieser Arbeit ist die Entwicklung von zwei neuartigen Chemometrie-Algorithmen: der erste der Kalibrationstransfer mit Hierarchischen Modellen und Lagrange Multiplikatoren, der für die Aufnahme von täglichen Kalibrierkurven mit nur wenigen Kalibrierproben zwingend nötig ist, und der zweite der neue Kurvenauflösungs-Algorithmus RABBIT-MCR, der für die Zerlegung von überlappenden Spektren-Signalen in ihre Bestandteile verwendet wird. Die Datenalgorithmen und Kalibriermethoden sind zwar für das in dieser Arbeit verwendete Messsystem zugeschnitten, aber die zugrunde liegenden Prinzipien können für jedes (Atem-)Gas-System verwendet und adaptiert werden.
List of Papers

The presented thesis is based on the following papers, which are already published in internationally peer-reviewed journals.


### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>APV</td>
<td>Institute of Anesthesiologic Pathophysiology and Method Development</td>
</tr>
<tr>
<td>EE</td>
<td>Energy expenditure</td>
</tr>
<tr>
<td>$f_{\text{carb}}$</td>
<td>contribution of carbohydrate oxidation on the carbon dioxide production</td>
</tr>
<tr>
<td>$f_{\text{fat}}$</td>
<td>contribution of fat oxidation on the carbon dioxide production</td>
</tr>
<tr>
<td>$f_{\text{prot}}$</td>
<td>contribution of protein oxidation on the carbon dioxide production</td>
</tr>
<tr>
<td>FTIR</td>
<td>Fourier Transform Infrared</td>
</tr>
<tr>
<td>GC-MS</td>
<td>Gas Chromatography Mass Spectrometry</td>
</tr>
<tr>
<td>HM</td>
<td>Hierarchical Models</td>
</tr>
<tr>
<td>HMC</td>
<td>Hamiltonian Monte Carlo</td>
</tr>
<tr>
<td>HWG</td>
<td>Hollow Wave Guide</td>
</tr>
<tr>
<td>IABC</td>
<td>Institute of Analytical and Bioanalytical Chemistry</td>
</tr>
<tr>
<td>ICL</td>
<td>Interband Cascade Laser</td>
</tr>
<tr>
<td>iHWG</td>
<td>Substrate - Integrated Hollow Wave Guide</td>
</tr>
<tr>
<td>IR</td>
<td>Infrared</td>
</tr>
<tr>
<td>LM</td>
<td>Lagrange Multiplier</td>
</tr>
<tr>
<td>LMM</td>
<td>Linear mixed models</td>
</tr>
<tr>
<td>LS</td>
<td>Luminescence Spectroscopy</td>
</tr>
<tr>
<td>MCMC</td>
<td>Monte Carlo Markov Chain</td>
</tr>
<tr>
<td>MCR</td>
<td>Multivariate Curve Resolution</td>
</tr>
<tr>
<td>MCR-ALS</td>
<td>Multivariate Curve Resolution Alternating Least Squares</td>
</tr>
<tr>
<td>MF</td>
<td>Mole fraction</td>
</tr>
<tr>
<td>MICU</td>
<td>Mouse Intensive Care Unit</td>
</tr>
<tr>
<td>NDIRS</td>
<td>Non-Dispersive Infrared Spectrometry</td>
</tr>
<tr>
<td>NUTS</td>
<td>No U-Term Sampler</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial Least Squares Algorithm</td>
</tr>
<tr>
<td>RABBIT MCR</td>
<td>Rotation and Angle-Bending Bayesian induced Transformation MCR</td>
</tr>
<tr>
<td>RSM</td>
<td>response surface methodology</td>
</tr>
</tbody>
</table>
### Zusammenfassung

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ</td>
<td>Respiratory quotient</td>
</tr>
<tr>
<td>TEE</td>
<td>Total energy expenditure</td>
</tr>
<tr>
<td>TTR</td>
<td>Tracer-to-Tracee ratio</td>
</tr>
<tr>
<td>UBT</td>
<td>Urea breath test</td>
</tr>
</tbody>
</table>
1 Aims and Overview of the Thesis

The aim of this thesis was to develop an analysis system capable of monitoring important biological parameters in exhaled breath at the Mouse Intensive Care Unit (MICU) at a high time resolution of one data point per minute, if possible, on-line. In order to reach this goal, it was not only important to build an analyzer system that can be physically integrated into the ventilation system of the mouse, but even more crucial to develop a data analysis strategy that generates data of medical relevance by not only correcting physical effects like oxygen or humidity influence and tracing the error propagation for exact error bands, but by using meta data gained from other analytical methods also used during the medical trial, e.g., GC-MS, other medical parameters like the respiratory quotient (RQ), or even the contributions of fat, carbohydrate and protein oxidation on the carbon dioxide production ($f_{fat}, f_{carb}, f_{prot}$) or total energy expenditure (TEE).

Two important technologies were necessary for the success of this work: first is an analytical measurement system capable of recording physiological parameters in mouse breath, which is a challenge due to the low breath volumes available in mice and the need to fully integrate it into an already existing intensive care setup without disturbing the medical operation taking place. First steps towards this were already done at the institute and the setup presented in this thesis is derived from these works [1–3]. But the second and even more important technology is the Monte-Carlo Markov Chain (MCMC) driven Bayesian Sampling methodology as basis of the designed data analysis algorithm. Here, it was implemented using the software/programming language Stan. The main crucial point and advancement of this thesis is that the data algorithm itself is not strictly bound to the instrumental setup used and developed here. The experimental findings of this work concerning interferences and non-linearity corrections as well as the entanglement of simple parameters like oxygen, carbon dioxide or $^{13}$C enrichment concentration with derived physiological parameters can be adapted to any analysis system that
1 Aims and Overview of the Thesis

offers the main important parameters (inhaled and exhaled oxygen concentration, exhaled carbon dioxide concentration, \(^{13}\text{C}\) enrichment in exhaled CO\(_2\) and plasma glucose) at sufficient resolution and accuracy in the MICU setup, like, e.g., the ICL setup developed later on at IABC by Tütüncü et al [4–6]. The modular composition of the data analysis algorithm as well as the possibilities of the Bayesian methodology in setting up models using probability distributions offers easy access to modify and adapt the data algorithm even further in order to implement other instrumental setups or incorporate physical or physiological/medical corrections for improving analytical accuracy. Besides using Bayesian statistics for error propagation through the algorithm and gaining statistical ensured error bounds, in the course of this thesis, it was also necessary to develop completely new chemometric methods (a new and extremely efficient calibration transfer algorithm based on Hierarchical models and Lagrange Multipliers and a new Multivariate Curve Resolution algorithm termed RABBIT-MCR) and combine it with already existing approaches (e.g., response surface methodology [7, 8]). The results were published in five peer-reviewed papers, as referred to in the list of papers. They can be separated into the two main aims of the thesis:

1. Development of a data analysis procedure for the iHWG-FTIR-LS system for oxygen, carbon dioxide and \(^{13}\text{C}\) enrichment: implementation of a non-linear calibration algorithm for the oxygen sensors using Hierarchical models and Lagrange Multipliers for calibration transfer and including a humidity compensation, implementation of a response surface calibration compensating oxygen effects on the CO\(_2\) calibration, also including the calibration transfer algorithm mentioned before, and lastly, direct calibration of \(^{13}\text{C}\) enrichment in terms of mole fraction using a new curve resolution algorithm called RABBIT-MCR, non linear response surface calibration with a Hierarchical model approach. All are based on Bayesian Sampling and Modeling and results from one calibration are incorporated into later algorithms (e.g., oxygen concentration in the CO\(_2\) algorithm and so on) and build on each other. The algorithms are presented in the journal articles ‘Nonlinear Calibration Transfer Based on Hierarchical Bayesian Models and Lagrange Multipliers: Error Bounds of Estimates via Monte Carlo – Markov Chain Sampling’, shown in 5.2; ‘Response-surface fits and calibration transfer for the correction of the oxygen effect in
the quantification of carbon dioxide via FTIR spectroscopy\(^1\), shown in 5.3 and ‘Strategies for 13C enrichment calculation in Fourier-transform Infrared CO\(_2\) spectra containing spectral overlapping and nonlinear abundance-amount relations utilizing response surface\(^1\), shown in 5.5.

2. Integration of the data analysis procedure and the measurement system into the MICU setup as well as data analysis procedure for derived physiological parameters (RQ, \(f_{fat}\), \(f_{carb}\), \(f_{prot}\), TEE) in biologically relevant concentration ranges. The results were published in the journal articles ‘Online monitoring of carbon dioxide and oxygen in exhaled mouse breath via substrate-integrated hollow waveguide - Fourier transform infrared - luminescence spectroscopy \(^1\), shown in 5.4, and ‘Metabolic monitoring using online analysis of \(^{13}\)C enriched carbon dioxide in exhaled mouse breath via iHWG-FTIR LS spectroscopy and Bayesian Sampling\(^1\), shown in 5.6.
2 Introduction

2.1 Motivation

Mice are one of the leading mammalian model organisms for basic genetic research and for investigating human diseases due to the fact that they are nearly genetically identical to humans, with many biochemical and physiological pathways being conserved [9, 10].

Next to ethical concerns, the ability to modify genes in laboratory mice for medical trials as well as the access to fluid and tissue samples not possible in human patients offer an important opportunity for an insight in the metabolic pathways while investigating the biological processes of critical diseases and evaluating new therapy ways.

Coming from human research, it was shown that next to the usual invasive collection of fluids like blood or urine or tissue samples, monitoring of exhaled breath can offer additional insight in the metabolic process without influencing the underlying biological condition. While many commercial systems for the breath metabolic monitoring in human patients or larger animals exist, there are few existing systems that offer the same metabolic parameters for mice and rodents. The small size of a mouse and its even smaller breath volume are a particular challenge for many of the existing techniques due to the available sample volume of few milliliters. An even more challenging task is the on-line monitoring of breath during ventilation in a surgical environment like an Intensive Care Unit.

Sepsis and septic shock are among the leading causes of death in general and particular in hospitals with a high mortality rate even after diagnosis [11–13]. This and medical research into hemorrhagic shock in combination with a thorax trauma using mouse and porcine models in intensive and trauma care are one of the main research topics of the cooperating Institute of Anesthesiologic Pathophysiology
and Method Development at Ulm University Medical Center and part of a DFG Sonderforschungsbereich (SFB 1149) and the DFG supported Graduate College PULMOSENS.

The Mouse Intensive Care Unit (MICU) at the Institute of Anesthesiologic Pathophysiology and Method Development is one of the units investigating metabolic processes and therapeutic methods for intensive care using murine and porcine models [14–40].

The goal of this thesis was to develop an analysis system capable of monitoring important biological parameters in exhaled breath at the Mouse Intensive Care Unit (MICU), in particular the underlying data calibration and quantification algorithm using new chemometric tools like Bayesian statistics and non linear calibration methods and automate it for daily use in an medical environment.

2.2 State of the Art

The scope of this chapter is to present background information on the analytical and chemometric techniques used in this thesis as well as a short overview on the medical motivation on breath analysis in mice models. In addition, the state of the art in this research field in respect to the obtained data will be presented.

2.2.1 Metabolic monitoring and breath analysis in mice

Mouse models are among the most important model organisms in the study of human diseases and metabolic functions [41]. The mouse has a very strong genetic similarity with the human genome [42] and shares many features of the human metabolism [41]. One advantage of using murine models in medical research is the ability to manipulate and knock out genes in the mice genome for better understanding of certain metabolic and disease pathways and the influence of selected genes on these conditions [41]. Animal models also offer access to invasive procedures like blood, organ or tissue samples that are not available in correspondent human trials due to ethical restrictions.

Exhaled breath analysis is a rapidly growing field of research in medicine. Since
the first pioneer work of Linus Pauling in 1971 [43], over 800 volatile and semi-volatile components have been detected in breath [44]. Via the aveoli, biomarkers in blood or metabolism products are transferred into exhaled breath and therefore it is possible to detect diseases or get an insight into the metabolic processes of the body by monitoring the presence or concentration of exhaled markers. Breath consists of a mixture of nitrogen, oxygen, carbon dioxide, water and other volatile constituents in decreasing concentration [44]. The field of breath analysis can be separated into two main topics of use cases: one the detection of endogenously produced biomarkers that indicate underlying metabolic states and medical conditions and the other the detection of artificially induced isotope tracer enrichments in breath via dosage of labeled tracers to a patient, e.g., the administration of $^{13}$C urea in the urea breath test (UBT) for the detection of *Helicobacter pylori* [45].

The measurement of Energy Expenditure (EE) is a significantly important parameter in medicine and medical research. Nearly all metabolic measurements are done using respirometry, i.e. the measurement of the oxygen consumption rate $vO_2$ and carbon dioxide excretion rate $vCO_2$, from which the respiratory quotient $RQ = vCO_2 / vO_2$ and the EE can be calculated [46].

While many commercial instruments for the monitoring of vital parameters in human patients are available (capnography, pulse oximeters, blood gas analyzers) [47, 48], few such devices are available for small animal trials like mice or rats. For measurement of EE in rodents, two techniques are common: direct and indirect calorimetry ones.

Direct calorimetry measures heat loss using four different techniques: isothermal direct calorimetry, Heat-sink direct calorimetry, direct convection calorimetry and direct differential calorimetry. In all, the animal is enclosed in a chamber and through different techniques, the radiated heat of the animal is measured [49].

In indirect calorimetry, as mentioned above, the EE is measured using flow-through respirometry [46, 50]. The animal sits in a (large enough) chamber and the input and output stream of oxygen and carbon dioxide is measured. Both open-circuit (air enters the chamber and $CO_2$ may be absorbed or not) and close circuit systems (the $CO_2$ in the chamber is absorbed and only oxygen enters the chamber) are possible [50]. One limitation of these measurements is the dead volume of the chamber and its resulting time delay (from several minutes up to a few hours). The time delay is proportional to the volume of the chamber as well as the flow rate.
of the system. In awake animals, the chamber needs to be a certain size to avoid stress responses, meaning a downsizing of the chamber volume is not possible. Increasing the flow rate can also decrease the time delay, yet the resulting dilution necessitates a significantly improved accuracy of the gas analyzers [46].

For all techniques, the housing of the animal in a metabolic chamber is necessary; which is not viable for the surgical environment in the MICU.

An important information for interpreting the RQ and EE is compartmentalizing the whole body CO$_2$ production into the contributions of protein, carbohydrate and fat oxidation ($f_{fat}$, $f_{carb}$, $f_{prot}$) as principal energy sources. Commonly, this is done using theoretical relation functions and the cumulative urinary nitrogen excretion for assessing the fraction of protein oxidation [51–53]. Due to limitations where parallel metabolic processes like gluconeogenesis or lipid biosynthesis lead to overestimation through elevated RQ values or problems in the accurate and complete detection of urinary nitrogen excretion, this method has been superseded by $^{13}$C tracer study protocols suggested by Wolfe et al [53] for detecting carbohydrate oxidation. By infusing $^{13}$C tracer labeled glucose in a constant protocol and measuring plasma tracer enrichment as well as breath CO$_2$ enrichment values, it is possible to assess the fractional contribution of glucose oxidation and through this of fat and protein oxidation as well [53].

Next to detecting the oxygen consumption rate $vO_2$ and carbon dioxide excretion rate $vCO_2$ in breath, tracing $^{13}$C enrichment values in breath is therefore an important keystone for monitoring metabolic parameters.

### 2.2.2 Breath gas analysis using infrared spectroscopy

Due to its molecular and in particular isotopic specificity, infrared (IR) spectroscopy is inherently eligible for gas and especially breath gas analysis. Except for mono-elementary diatomic molecules like, e.g., oxygen or nitrogen, all molecules show an IR absorption and unlike liquid probes, where the strong absorption of a solvent like water can be problematic, in gas phase the interaction between analytes and molecules is strongly reduced. Out of many possible IR implementation methods, non-dispersive techniques, laser based absorption and FTIR (Fourier transform) spectroscopy are the most common.
For laser and non-dispersive based techniques, commercial instruments are already available for determining the $^{13}$C enrichment in breath, arising from $^{13}$C breath tests from humans, e.g., like the UBT; and $CO_2$ and oxygen are tracked by capnography.

FTIR is a less commonly technique used. Kindness and Marr used FTIR for the determination of $^{13}$C/$^{12}$C enriched material [54] while Esler et al reached high precision of 0.1 $\%$ by combining FTIR spectroscopy with a classical least squares (CLS) algorithm [55]. Zanasi et al detected $^{12}$CO$_2$ and $^{13}$CO$_2$ trapped within a polymer using FTIR [56] and Mohn et al used FTIR to detect the $^{13}$C/$^{12}$C ratio in the atmospheric CO$_2$ [57, 58]. Hofstetter et al built a CO$_2$ gas sensor measuring $^{12}$CO$_2$ and $^{13}$CO$_2$ using a quantum cascade detector and a FTIR spectrometer [59].

FTIR spectrometers are usually less sensitive than comparable laser assisted or non-dispersive techniques, yet one advantage of using a more broadband spectrometer and light source is that simultaneous detection of several breath analytes, e.g., isoprene, acetone, carbon monoxide, nitric oxide, or even humidity by surveying water absorption lines is possible.

Yet, in the special case of mouse breath, not many commercial techniques of all kinds of analytical occurrence are usable due to the low sample volumes available. Key point is the need for a very small gas sample cell that is yet sensitive enough for the applications examined. Core technology for this and prevalent in the Mizaikoff research group is the hollow waveguide method. The usual long-pass gas cells like the White or Harriot cell [60–62] need too much of a sample volume to be of use in mouse breath measurements where only few mL are available. A hollow waveguide is a hollow light guide that propagates the light thorough while simultaneous serving as a small-volume gas cell. Early prototypes were used in the off-line breath analysis of mouse exhaled breath during earlier studies in the Mizaikoff group [1–3].

The substrate-integrated hollow waveguide (iHWG) is a technology pioneered by the Mizaikoff group under the auspices of a project with the U.S. Department of Energy by Lawrence Livermore National Laboratory (LLNL) [63, 64], implementing a robust and easy to fabricate waveguide that has already been used in several applications: for chemometric studies in the detection of isobutylene, methane,
CO₂, butane, and cyclopropane gas [65], process analysis [66, 67], environmental monitoring [68–72], catalysis research [73], quantum cascade and interband cascade laser sensors [4, 74, 75], fiber coupled sensors [76], chemical activity research [77] and breath analysis [3, 5, 6, 78–83] among others.

In 2013, Fortes et al presented a first prototype of the experimental setup used in this thesis, which involved a commercial FTIR spectrometer with a small footprint (Bruker ALPHA), a iHWG gas cell and oxygen flow cell sensors, but still in an open setup [3, 78]. This setup was further developed during the course of the present thesis.

After the end of the experimental work in this thesis, even further applications have been developed in the Mizaikoff research group: (i) a breath analyzer based on an interband cascade laser, dual-channel iHWG and luminescence oxygen sensor as successor to the FTIR-iHWG system presented in this thesis [4–6] and (ii) the combination of the FTIR-iHWG system with a differential ion mobility spectrometer (DMS) [82, 83].

### 2.2.3 Oxygen sensors

There are four major methods for determining oxygen [84]: the classical Winkler titration [85], electrochemical sensors, pressure based ones and optical methods. Out of those, electrochemical and optical sensors are most commonly used for medical applications, e.g., determination of oxygen partial pressure pO₂ in breath.

While there are many amperometric and potentiometric sensors developed for high-temperature applications, e.g. the Lambda probe in cars [86], for ambient conditions, most are based on the Clark electrode [87]. It consists of a platinum cathode and a reference silver-silver chloride anode enclosed in an oxygen permeable membrane. Oxygen is reduced at the cathode to hydroxide [88, 89]. The main disadvantage is the consumption of oxygen and interferences of other gases like ozone, chloride, nitrogen oxides or hydrogen sulfide [84, 90].

The most popular sensing system for oxygen in medical applications are optical fiber sensors based on luminescence quenching (phosphorescence or fluorescence). A luminescent dye is exited by incoming radiation and emits radiation at longer wavelengths due to long exited state life times. This emission is quenched by dynamic
collision with molecular oxygen in its triplet state. The reduction in luminescence intensity as well as its life time is proportional to the oxygen concentration [84, 88–90].

Core technology of the sensor system is the embedding of a luminescent dye in a host polymer. Kautsky and Hirsch developed the first system in 1931 based on tryptafalvin adsorbed on a silica gel support [91]. Bergman described the first fluorescent sensor system with a UV light source and a fluoranthrene layer on porous glas [92]. Many other research groups [93–107] further developed the systems using fiber optics [101], different luminescence principles [100], different readout techniques [107] and, in particular, different dye systems [95, 106, 108]. Nowadays, the most commonly used fluorescence indicators are ruthenium based complexes and metalloporphyrin complexes [89, 90] but many other indicators are possible [84, 109].

Mathematically, the quenching can be modeled by the Stern-Vollmer equation [90]:

\[
\frac{F}{F_0} = \frac{\tau}{\tau_0} = 1 + K_{SV}[O_2]
\]

with \(F_0\) the incident fluorescence intensity, and \(F\) the fluorescence intensity after quenching, \(\tau\) respectively \(\tau_0\) the luminescence decay times of a sample with or without oxygen, \(K_{SV}\) the Stern-Vollmer constant and \([O_2]\) the concentration of oxygen in the sample. This suggests an ideal model with a linear relationship between \(\frac{F}{F_0} / \frac{\tau}{\tau_0}\) and \([O_2]\) in the Stern-Vollmer plot. In reality, there is often a non-linearity for higher oxygen concentrations in certain dye-polymer systems due to physical effects [84, 88–90] and non-linear models like an extended two-site model [110, 111] or other theoretical fitting procedures [93, 112–120] have been suggested.

To complete the sensor, next to a physical combination of excitation light source and detector, a read out system is necessary. Numerous methods have been implemented [84], but two of them are most common: sensing based on measurement of luminescence intensity and luminescence decay time. The former is dependent in the concentration of the luminophore, photo-bleaching and optoelectrical properties of light source and detector and cannot compensate ambient light levels or background luminescence [84, 89]. For this reason, the more robust decay based detection is usually preferred. Here, two techniques dominate: frequency domain and time domain fluorometry. In frequency domain, the excitation light is mo-
dulated, resulting in an also modulated but phase-shifted luminescence intensity [84, 89]. The life time can now calculated by the phase shift \( \phi \) and modulation frequency \( f \) via

\[
\tau = \frac{\tan \phi}{2\pi f}
\]  

(2.2)

In the time domain or rapid life time detection, the excitation light source is switched on for a short time and then turned off again. The detector is now switched on at two different time points on the decay curve to measure the photon count of the luminescence. The life time is calculated by

\[
\tau = \frac{t_2 - t_1}{\ln \frac{A_1}{A_2}}
\]  

(2.3)

where \( A_1 \) and \( A_2 \) the photon counts at two different times \( t_1 \) and \( t_2 \).

Most commercially available oxygen sensors are based on the optical quenching technique.

While there are other possible techniques for sensing oxygen (radioisotopic techniques, electron parametric resonance [90], direct spectroscopic sensing in UV [121], VIS [122, 123] and NIR [124, 125], absorption-based probes using a chromogenic reaction with oxygen or a colorshift [84]), they are either expensive or use irreversible reactions and therefore are used only in quite limited applications [84, 90].

### 2.2.4 Bayesian analysis and chemometric background

Gelman et al. [126, 127] defined Bayesian Interference as the “process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations.” [126]. Core of it is the Bayes’ Theorem:

\[
P(\theta | y) = \frac{P(\theta, y)}{P(y)} = \frac{P(\theta)P(y|\theta)}{P(y)}
\]

(2.4)

where the posterior density \( P(\theta | y) \) of parameter \( \theta \) and data \( y \) is a function of the prior distribution \( P(\theta) \) and the sampling (data) distribution \( P(y|\theta) \).
To sum up, the key point of Bayesian interference is to develop model $P(\theta, y)$ and calculate $P(\theta|y)$ in an appropriate way [126].

This computation usually takes place using Monte Carlo Markov Chain (MCMC) methods. They work by drawing series of correlated samples that converge towards the desired target distribution [128]. There are several software packages that offer implementation of this sampling: among them BUGS [129], Jags [130] and Stan [131, 132], the latter one was used in this work.

Stan, named after mathematician Stanislaw Ulam, one of the developers of the Monte Carlo method, is an open source C++ software that computes the log-posterior density using the No-U-Turn-Sampler (NUTS) [132], a Hamiltonian Monte Carlo (HMC) variant that combines the greater efficiency in the random walk behavior of the HMC with an automated step-size tuning that is necessary for HMC [128]. In Stan, it is possible to formulate predictive Bayesian models using a simple modeling language without having to adapt or touch the underlying MCMC sampling mechanism.

Hierarchical Bayesian models (HM), also called multilevel linear models [127] or linear mixed models (LLM) [133], are statistical models that combine a written hierarchical (multi level) form with Bayesian estimation [134]. According to Gelman [127], in a hierarchical model, the model parameters (regression coefficients) have a probability model and it is possible to estimate this from data using a second level model with its own parameters (hyperparameters). The difference between multi-level and classical regression models is that variation between groups (schools) is modeled [127], e.g., if the general shape of a calibration curve is known (first level model), the day-to-day variance of the group of daily calibration curves can be modeled (second level model).

Rao [135] first presented this in describing growth curves with random coefficients and estimating the variabilities of them. Rosenberg [136] later extended this to linear regression models whose parameters are estimated using empirical Bayesian estimators. Harville [137] showed that it is possible to assess coefficients for an individual curve and their covariance if measurements of the individual curve as well as group mean and group covariance values are available. This results in the approach called ‘random coefficients’, as presented by Laird [138] as an EM (Estimation-Maximization) algorithm to estimate both coefficients for individual
curves as well as their group distribution, and later summarized by Longford [139].
Hierarchical models combine the random coefficients approach with the Bayesian
interference mentioned earlier.

Vogt et al [8] used this approach for the calibration of an NDIR sensor of $^{13}$C offset
values that are dependent on CO$_2$ and O$_2$ concentration and were calibrated using
a nonlinear response surface. This key paper was the basis and trigger of all data
calibration algorithms developed later in this thesis.

Box et Wilson [140, 141] first introduced the response surface methodology (RSM).
It comprises of a school of statistical techniques for building empirical models in
which a desired response (output) is related to several predictors (input) in a (multi-
dimensional function (response curve or surface) [142]. The advantage of this
is that a (nonlinear) relationship between response and its influencing values can
be found and optimized empirically, even if the exact theoretical relationship is
not known. These relationships are often assumed as polynomials and one goal of
RSM is to find the optimum response surface. This can be done using, e.g., regular
polynomial optimization, ridge regression, Lasso or regularization approaches
[143].

Another, less common approach that shows promise for the combination of the
response surface approach with Hierarchical models and other external conditions
(calibration samples) is Lagrange Multiplier Optimization [144, 145], which is an
approach using constrained optimization. Local maxima or minima of a function
are searched while being constrained to external conditions (i.e., other equations
have to be satisfied by the chosen parameters). Geometrically, as a visualization,
this means finding the extrema spot in an surface that lies on an curve running
through the surface (the constraint). Lagrange multipliers introduce an extra va-
riable (the multiplier) to solve this constrained optimization problem [145].

When designing a calibration algorithm for a routine measurement setup, next to
finding and estimating the optimum calibration function, it is often desirable to
minimize concurrent calibration efforts, in particular when running several sen-
sors at the same time. At the same time, drift of the instrument makes frequent
recalibration necessary. One solution for this are calibration transfer approaches.
In general, at start, an extensive calibration of the system in question is recorded
and later transferred using a chosen algorithm. A detailed review and overview
of most existing techniques can be found in Noord [146] as well as in Feudale et al [147]. To sum them up, most calibration transfer algorithms were developed for the calibration transfer between two instruments/sensors and it is difficult to find methods for transferring a) calibration of univariate signal responses (e.g. a chemical sensor like an oxygen sensor), b) temporal calibration transfer on the same instrument, though this can be easier realized than the transfer proposed between instruments [146, 147], and c) most important, the transfer of nonlinear or multidimensional calibration relationships (response surfaces).

The last puzzle piece of the chemometric background of this thesis is the resolution of overlapping spectral signals, e.g., the $^{13}\text{C}/^{12}\text{CO}_2$ IR signal, into their underlying contributing constituents. One common approach for this is multivariate curve resolution. A set of multicomponent response data is decomposed into a concentration matrix and an spectral matrix of the pure contained analytes, often by using constraints (non-linearity, closure, kinetic modeling...) during the optimization algorithm. This strategy is implemented in many different algorithms, among them MCR-ALS (Multivariate Curve Resolution Alternating Least Squares) [148–151], Resolving Factor Analysis (RFA) [152] and Gentle [153]. The first has been implemented in many iterations and applications [148, 154–158] and has become a quite popular chemometric technique.

MCR-ALS determines concentration values of different chemical constituents and their unknown spectral contributions in parallel. The unique decomposition can be sometimes difficult when both the spectra and the concentration contribution of the underlying analytes are less known or unknown. To resolve this, constraints in terms of concentration or spectral response are often implemented. One trade-off of MCR is that the exact number of contributing analytes has to be known or at least determined via other methods. MCR-ALS is also a linear method, i.e. it assumes a linear relationship between spectral response and concentration (analog the Beer-Lambert law). Non-linear relationships in MCR are not part of the common available MCR implementations and rarely found in literature.
3 Results and discussion

The goal of this research work was to develop an analysis system capable of monitoring important physiological parameters in breath in the Mouse Intensive Care Unit. In order to reach this, it was not only important to build a measurement system that can be physically integrated into the ventilation system of the mouse, but even more crucial to develop a subsequent data analysis algorithm that generates data of medical relevance by not only correcting physical effects like oxygen or humidity influence and tracing the error propagation for exact error bands, but by using data gained from other analytical methods also used during the medical trial, e.g. GC-MS, other medical parameters like the respiratory quotient RQ or more important the contributions of fat, carbohydrate and protein oxidation on the carbon dioxide production or total energy expenditure (TEE).

Chief technology in this work is Monte-Carlo Markov Chain driven Bayesian Sampling using the software Stan [132, 159, 160]. By defining probability distributions for experimental parameters and sampling over 5000 + iterations in one simple run, it is possible to gain good statistics and therefore error propagation and even gain access to physiological parameters that are not readily available from analytical system itself or even when reference values at this time stamp are missing.

Analytical basis for the later calculation and the heart of the on-line monitoring system in the MICU is a modified FTIR spectrometer (Bruker ALPHA, Bruker Optik GmbH, Ettlingen, Germany) using the iHWG technology driven by this work group [63, 64] in combination with commercial LS flow-cell oxygen sensors (FireStingO2, Pyro Science GmbH, Aachen, Germany). Earlier iterations of this setup were presented by Fortes et al [3].

The five papers comprising the presented thesis can be divided into two subdivisions. The first one containing Papers I, II and IV consists of the description of the new data analysis algorithms developed that are necessary for the calibration
of the iHWG-FTIR-LS system in order to gain the necessary analytical output of carbon dioxide (CO₂), oxygen (O₂) and, if applicable, ¹³C enriched carbon dioxide (¹³CO₂) while reaching the necessary analytical precision requirements necessary despite interfering physical effects. The second subdivision containing Papers III and V deals with the implementation of these calibration algorithms as well as the physical analytical system into the MICU. Going even one step further, it was shown that it is now possible to calculate derived physiological meta data like the RQ (Paper III + V) or the contributions of fat, carbohydrate and protein oxidation as well as EE (Paper V) in a resolution of one sample per minute, even if necessary corresponding reference values, e.g., ¹³C plasma glucose, are only available once per hour.

The on-line monitoring system in the MICU consists of two important fundamental analytical techniques. FTIR spectroscopy is used to measure the two CO₂ isotopes occurring while oxygen is measured by luminescence flow cell sensors. The advantage of using a broad wavelength commercial spectrometer over more specialized and precise instruments like NDIRS or laser spectroscopy is the possibility of detecting other physiological relevant breath molecules like isoprene [79–81] or carbon monoxide, or even possible interferences in the measurement (water for humidity in the oxygen determination) without needing to greatly change the experimental setup, e.g. light source, spectrometer or measurement cell. This is in particular possible because the concentration range expected for CO₂ and ¹³CO₂ in mouse breath measurements in question lies in the few percent range, which does not demand the high precision usually needed for applications using NDIRS or laser spectroscopy.

For better analytical precision, both components have to be calibrated daily. This is physically and time constrained by the fact that the measurement system is highly integrated into the surgical environment of the ICU, meaning that the calibration effort needs to be as time and sample efficient as possible and has to be practicable in its implementation by the medical personnel operating the MICU without large training effort or time effort during trial.

Oxygen, like all mono-elementary diatomic molecules, has no IR absorption and therefore needs another analytical technique for detection. In this research thesis, commercial LS sensors were used for oxygen monitoring. They work by detecting the quenching of a fluorescence signal stimulated in a special fluorescence dye via
oxygen, which is a strong fluorescence quencher. This quenching signal can be related to the oxygen concentration contained in the sample. The commercial sensors already come with an integrated calibration offered by the manufacturer that can be re-calibrated using up to two calibration points (0 and 20 % oxygen). The main reason why it was necessary to develop a new calibration algorithm for the oxygen sensors despite having an already existing calibration, was that the manufacturer factory calibration was optimized for simplicity and robustness in a high precision concentration range of 0 - 50 % oxygen, which unfortunately was not sufficient for the analytical precision demands and expected physiological range in ICU (0 - 100 % oxygen). Another reason for the advanced calibration transfer algorithm developed for the sensor was that the sensors showed a considerably large day-to-day variation which made a daily calibration necessary in order to reach the required precision and which also could not be resolved using the offered recalibration by the manufacturer. In fact, the manufacturer recalibration sometimes made the accuracy of the oxygen value even worse. Later on, it was even identified that the oxygen signal is dependent on the humidity concentration sample, which is a critical issue in breath, since breath samples compared to usually dry calibration samples contain a certain amount of humidity (Paper III).

The core of the calibration algorithm of the oxygen sensor is driven by the need of a non-linear relationship between the raw signal of the sensor and the desired oxygen concentration (Paper I). At the same time, a full calibration set spanning an entire calibration curve is not possible due to the constraints purported by the medical environment mentioned above. The solution is a non-linear calibration relationship using a rational function with four coefficients and combining it with new chemometric techniques like a calibration transfer algorithm based on Lagrange Multipliers, and a calibration model called Hierarchical Models developed by Gelman et al [127]. The latter uses the knowledge that calibration curves recorded at similar conditions can be seen as a family of curves and their mean coefficients as well as Day-to-Day variance can be determined. A daily calibration curve is seen as a sample of this distribution. The calibration transfer on the other hand uses calibration samples in order to transfer an entire calibration between sensors or in this case different days. Combining now the method of Hierarchical Models and Lagrange Multiplier Optimization, the topical calibration coefficients and therefore actual calibration curve can be achieved using only two calibration
samples. This also becomes important later on since calibration samples are also necessary to calibrate the CO₂ determination. The Bayesian Sampling included into the Stan program containing the data analysis offers realistic concentration values and in particular error boundaries through modeling of sampling errors as well as simulation of a more complex error model in dependence to the oxygen content.

The calibration procedure developed in Paper I was the first important step for the more complex calibration procedure for CO₂ in the FTIR system (Paper II). The standard method for correlating FTIR signals to concentration is the univariate one-wavelength approach of the Beer-Lambert law or more advanced a PLS approach [1]. This is only correct if the correlation between signal and concentration is in the linear range as defined by the Beer-Lambert law and no other spectral interferents at the wavelength range of the signal are present. Both conditions are not applicable for situation of the FTIR system in the MICU. The physiological CO₂ concentrations expected during medical trials are in the nonlinear signal range of the FTIR spectrum. Earlier research [1] has also shown that different oxygen concentrations with the same CO₂ concentration result in vastly different FTIR signals. This was eminently substantiated during this research. A causality between oxygen concentration and FTIR CO₂ signal loss was shown, though the exact physical reason or explanation for the nonlinearity and signal loss could not be identified using the methods and equipment available. For an exact root cause analysis, even more extensive and expansive research would have been necessary. Since this was not the goal of this thesis, it was deemed sufficient to correct for the signal loss using an empirical function. This however means that the usual calibration equation correlating CO₂ concentration and spectral signal had to be extended to include an oxygen term, resulting in an three-dimensional response surface calibration.

Multivariate data analysis algorithms like Principal Component Analysis (PCA) are often helpful in reducing large spectral data sets to just a few values. In the case of the CO₂ FTIR calibration, the spectral peak could be reduced to a single principal component and its score. Several wavelength ranges of the FTIR spectra, several preprocessing methods as well as different calibration relationships for response surface models including the PCA Score, CO₂ and oxygen concentration of calibration sets were tested and the optimum combination determined. As in the
case of the oxygen sensor, the FTIR needs an daily calibration due to fluctuations. The Lagrange Multiplier-Hierarchical Model calibration transfer approach introduced in Paper I was now extended for the multidimensional response surface, which could be successfully achieved again using two calibration samples. Here the true strength of the new calibration strategy is demonstrated: a measurement system with two different analytical techniques and complex underlying non-linear calibration relationship could be prepared for daily calibration using only three calibration samples (one of each calibration sample of each calibration system is identical with the other). While it might have been possible to record an entire calibration curve for the oxygen sensor each day, the same is physically impossible for the FTIR system since one calibration set measurement takes 5 hours or more. A calibration transfer algorithm is therefore a must.

These two calibration algorithms can now be combined for the on-line breath marker monitoring in the MICU of non-isotope tracer enriched mice (Paper III). Multiple days of medical mouse trials [161] were monitored with the analysis system integrated into the respiratory system of the MICU and analyzed using the calibration approaches of the first two studies. Via the exhaled CO/he signal loss using an empirical function. This however means that the usual calibration equation correlating CO and oxygen concentration in the breath of the intubated mice as well as the known inhaled oxygen concentration, it is possible to calculate the respiratory quotient RQ, an important metabolic parameter. This study also shows why it was necessary to develop such a novel and complex calibration concept for the measurement system. While it might have been possible to solve some of the individual problems arising in the studies mentioned above using already existing solutions, the constraints and requirements of the system in the MICU made an general solution involving several calibration approaches and algorithms interlocking in each other necessary to achieve the solution presented in this study, capable of generating all the physiological parameters desired in one large calibration data program. Without it and the underlying Bayesian sampling integrated into the software, it would not have been possible to solve the problem of humidity dependence of the oxygen measurement. This interference was only detected in the real breath samples and not in the dry calibration (where humidity was 0 %) since in the earlier study (Paper I) the possibility of non-dry samples was not considered and therefore no investigations into a humidity effect of the
3 Results and discussion

Oxygen signal was performed. Although a real time humidity measurement using a humidity sensor integrated into the setup, which might have even further improved the accuracy of the oxygen measurements, was not possible since the mouse trials were already finished at this time, it was possible to build a humidity error model using few off-line measurements and the Bayesian approach, modeling the influence of the humidity variation and effect on the error propagation through all parameters. This was only possible because the entire signal and calibration chain and its measurement error propagation was modeled using Bayesian data analysis and statistics in a extensive interlocked calibration model. This way, measurement interferences on one sensor signal and their error propagation could be modeled and traced back through the entire calibration chain for all resulting parameters, resulting in valid and realistic values with statistically accurate error estimations.

With these results, it was possible to gain time-resolved curves of biologically significant parameters in breath like exhaled CO₂, exhaled oxygen, oxygen intake and RQ with a per-minute resolution. Tentative metabolic interpretation of these curves correlating so-called Recruitment Maneuvers, medical therapy strategies during mechanical ventilation, to dips and spikes in the oxygen and CO₂ concentration, was even possible.

For medical trials using ¹³C isotope labeling, an additional strategy for the calibration of the ¹³C tracer ratio in CO₂ is necessary. There are two possible strategies for this calibration. The first is to consider the ¹³CO₂ as its own independent signal and build a calibration algorithm determining the absolute ¹³CO₂ concentration, only in the second step calculating the ¹³C isotope ratio [1]. There are disadvantages to this approach: an entire additional calibration for ¹³CO₂ is necessary with the same potential oxygen dependence as for ¹²CO₂. This means not only finding a spectral range without overlapping interference from the ¹²CO₂, but also an additional source of error in the isotope ratio and an extra step in the ¹³C isotope tracer ratio determination. Additionally, the preparation of pure ¹³CO₂ samples is expensive, time-consuming and prone to errors, since unlike CO₂ samples, which can be prepared using a gas mixing pump, the ¹³CO₂ have to be prepared by hand since pure ¹³CO₂ gas is very expensive. Even then, it is impossible to prepare totally pure ¹³CO₂ samples since even 99.9 % isotope-pure ¹³CO₂ contains traces of ¹²CO₂.

The second strategy is using the fact that it is desired to determine only the ra-
3 Results and discussion

tio of $^{13}$CO$_2$ against $^{12}$CO$_2$ and therefore it is not necessary to know the exact $^{13}$CO$_2$ concentration. E.g., in GC-MS analysis, the $^{12}$CO$_2$ signal is taken as internal standard against the $^{13}$CO$_2$ signal, making it possible to directly calibrate against the $^{13}$C tracer-to-tracee ratio TTR. This approach was chosen in the study detailed in Paper IV. The same observations occurring in the CO$_2$ determination (nonlinear signal relationship, oxygen effect) are assumed to be applicable for the $^{13}$CO$_2$ quantification. It seemed therefore expedient to try adapting the nonlinear calibration approaches of the first two studies (Paper I and II) to this new challenge.

It is also supposed that a daily calibration model using a calibration transfer would also improve calibration accuracy and precision. However, due to the time effort necessary for manufacturing the $^{13}$CO$_2$ calibration samples (a complete calibration set takes 3 - 5 days), it was not possible to build a model for the day-to-day variability, which is a necessity for the HM-LM calibration transfer approach established in this thesis.

A first step for finding a response surface calibration that can correct for interferences orthogonal to the signal relationship in question is reducing the multivariate signal, e.g., a spectrum, to a single response value. In the case of TTR determination, due to the presence of $^{12}$CO$_2$ in the signal, an additional step normalizing the $^{13}$CO$_2$ signal to the $^{12}$CO$_2$ signal is important, i.e., generating a signal quotient that correlates to the TTR. There are two different approaches for this step: one are classical normalization approaches (dividing the spectrum by a norm representing the $^{12}$CO$_2$ signal), the other spectral decomposition algorithms based on PCA decomposition (MCR approaches). Two classical (normalization to area under curve or peak maximum) and two spectral decomposition approaches (MCR-ALS and RABBIT-MCR) were tested using two data sets - one collected for this thesis, the other already measured during my diploma thesis [162]. The RABBIT-MCR algorithm (Rotation and Angle Bending Bayesian induced Transformation) was specially developed for this study. It operates in a PCA subspace generated by coordinate transformation and is related to MCR techniques. The aim of using MCR techniques in this study is to generate two loadings out of the spectra that uniquely belong to the $^{12}$CO$_2$ and $^{13}$CO$_2$ signal respectively with no overlap. Reaching this, the corresponding scores and their quotient can be directly correlated to the TTR. The spectral decomposition methods performed better than the clas-
Results and discussion

Classical approaches, since the latter had difficulties completely eliminating the $^{12}\text{CO}_2$ signal.

The normalized scores can now be used in finding the best response surface equation. For this, different relationships between normalized signal and TTR, oxygen and carbon dioxide concentration were tested. The study showed a possible correlation of the signal to total $\text{CO}_2$ concentrations, making the latter correction term necessary. In fact, the calculations showed later on that the oxygen influence is somehow canceled out while forming the TTR quotient, removing the oxygen correction term, while a $\text{CO}_2$ interaction term is necessary.

The tracer mole fraction (MF) is related to TTR and both can be converted into each other. The study also showed that spanning the response surface for MF gives better results than using TTR directly.

Again, in this study, it was actually not important which MCR algorithm is used for spectral decomposition (it is recommended to pick one that offers the bests results for a particular problem), but the approach for empirically correcting interferents on the signal using a response surface can be adapted universally to other spectral challenges. The Bayesian sampling in the background offers modeling of measurement variation and definition of special complex error relationships, resulting in valid and realistic values with statistically accurate error estimations.

This second-to-last study (Paper IV) was the last puzzle piece necessary for monitoring breath markers in the MICU when using $^{13}\text{C}$ labeling and the metabolic monitoring approach developed in the final study of this thesis (Paper V). It not only presents a completed data analysis algorithm for the iHWG-FTIR-LS system used in this study, going even one step further by using meta data like physiological knowledge and reference values gained by other systems in the MICU, e.g., GC-MS measurements of blood plasma, it is a) possible to generate biologically relevant values, e.g., for RQ and b) gain access to important metabolic parameters never available before, like, e.g., the contributions of fat, carbohydrate and protein oxidation on the carbon dioxide production or TEE.

Combining the data analysis approaches for oxygen, $\text{CO}_2$ and TTR, it was now possible to determine the exhaled $\text{CO}_2$ ($^{12}\text{CO}_2$ and $^{13}\text{C}_2$ fraction as well as total $\text{CO}_2$), exhaled oxygen, oxygen intake, TTR, respectively MF, of the $^{13}\text{C}$ tracer and RQ in breath during murine experiments investigating the reaction of psy-
chologically stressed mice to thorax trauma and hemorrhagic shock in the MICU. One issue that also occurred earlier on (Paper III) was that for some mouse experiments, RQ values lower than 0.7 were reached. This is medically highly unlikely since total fat oxidation produces a RQ of 0.7 and physiologically meaningful RQ values lie between 1 and 0.7 [53]. The true source of these too low values for RQ are difficult to ascertain. Errors in the CO₂ and oxygen measurement and determination are suggested since medical attempts to explain are hard to substantiate.

Using physiological relationships between RQ and the relative contributions of fat, carbohydrate and protein oxidation on the carbon dioxide production ($f_{fat}$, $f_{carb}$ and $f_{prot}$) as well as the $^{13}$C glucose tracer concentration in blood plasma, a Bayesian model estimating physiological meaningful values for RQ, $f_{fat}$, $f_{carb}$ and $f_{prot}$ as well as TEE was deployed by efficient sampling as well as priors, which is one of the strengths of Bayesian MCMC sampling. Sampling was also helpful when estimating $^{13}$C glucose tracer plasma concentrations since GC-MS measurements only take place up to six times during an experiment (maximum 3 times for glucose tracer plasma concentrations) have to be guessed and all other time points have to be estimated. Bayesian methodology is inherently helpful for this by offering to define probability distributions estimating unknown parameters. Even if the exact biological value is not known at a certain time point, it is possible to guess at a correlated concentration using directed NUTS sampling and gain statistic estimates and error bounds, as was shown before with the humidity correction for the first mouse experiments (Paper III).
4 Conclusions

In order to develop an analyzer system capable of measuring important analytical parameters like physiological relevant components in breath, two components are necessary: (i) a physical measurement system capable of detecting the parameters of interest at sufficient analytical accuracy and precision; (ii) equally important is an appropriate calibration and data analysis algorithm that ideally not only offers access to the parameters directly measured by the system but, using meta data gained from other systems or physiological and medical knowledge as well as new powerful data analysis techniques like Bayesian Sampling and Statistics. They can open the door to other very metabolically relevant parameters like the RQ or even the contribution of different oxidation channels to the whole body carbon dioxide production ($f_{fat}$, $f_{carb}$, $f_{prot}$), or TEE, leading to a better understanding of metabolic processes modeled in medical trials like mouse phenotyping and metabolic monitoring. While it might not be possible to offer all of these parameters on-line in real time due to the calculating time and power necessary, or perhaps the access to reference values is not available on-line (e.g., GC-MS derivate analysis of blood or urine samples), the complex off-line data analysis offers a new powerful tool for better interpreting and evaluating metabolic processes in an medical trial system like the MICU and gaining unprecedented insight in the metabolic system at a per-minute resolution.

This was made possible by the development of two key technologies in this research:

1. Further development and physical integration of an iHWG-FTIR-LS system using a modified FTIR spectrometer using the iHWG technology in combination with commercial LS flow-cell oxygen sensors into the existing medical setup of the MICU at the Institute of Anesthesiologic Pathophysiology and Method Development.
2. Development of a data analysis procedure for the iHWG-FTIR-LS system for oxygen, carbon dioxide and $^{13}$C enrichment: implementation of an non-linear calibration algorithm for the oxygen sensors using Hierarchical Models and Lagrange Multipliers for calibration transfer and including a humidity compensation, implementation of a response surface calibration compensating oxygen effects on the CO$_2$ calibration, also including the calibration transfer algorithm mentioned before, and lastly, direct calibration of $^{13}$C enrichment in terms of mole fraction using a new curve resolution algorithm called RABBIT-MCR, non-linear response surface calibration with a Hierarchical Model approach. All are based on Bayesian sampling and modeling, and results from one calibration are incorporated into later algorithms and built on each other.

3. And as last step: Integration of the data analysis procedure and the measurement system into the MICU setup as well as data analysis procedure for derived physiological parameters (RQ, $f_{fat}$, $f_{carb}$, $f_{prot}$, TEE) in biologically valid concentrations.

In the end, it was possible to accompany over 70 mouse trials at the MICU (part of them with $^{13}$C enrichment experimental protocols) using the new analysis setup and to evaluate the resulting breath data, showing that the developed system and data analysis algorithm procedure can be integrated into routine work at the MICU.

Key milestone of this thesis in particular is the development of a data analysis strategy giving access to derived physiological parameters (RQ, $f_{fat}$, $f_{carb}$, $f_{prot}$, TEE) in biologically valid concentrations in a resolution not available before thanks to the Bayesian methodology approach. This analysis algorithm is in fact independent of the experimental setup as used in this work and can be adapted and extended to all analytical systems offering the same core data (oxygen, carbon dioxide concentration and $^{13}$C enrichment, $^{13}$C glucose plasma enrichment). Due to the modular composition of the data analysis procedure, it is easy to implement further corrections and compensations of the experimental system used, if necessary, into the existing algorithm (see the oxygen or humidity correction).

Another novelty in this work is the development of two new chemometric algorithms: Calibration transfer using Hierarchical models and Lagrange Multipliers,
which is necessary for the recording of a daily calibration curve using only few calibration samples, and a new curve resolution algorithm, RABBIT-MCR, for the deconvolution of overlapping spectral signals into their components.

The data algorithms and calibration methods are tailored to the experimental system used in this thesis, but the underlying principles can be applied and adapted for any system analyzing (breath) gas.

To conclude, this thesis shows the development of an analysis system capable of monitoring important physiological parameters (oxygen concentration, carbon dioxide concentration, $^{13}$C enrichment, RQ, $f_{\text{fat}}$, $f_{\text{carb}}$, $f_{\text{prot}}$, TEE) in exhaled breath at the Mouse Intensive Care Unit (MICU). It demonstrates in particular the underlying data calibration and quantification algorithm using new chemometric tools like Bayesian statistics and non linear calibration methods, which is fully automated for daily use in an clinical environment.
Literatur


Literatur


Literatur

[16] Clair Hartmann u. a. „In-Depth Characterization of the Effects of Cigarette Smoke Exposure on the Acute Trauma Response and Hemorrhage in Mice“. In: SHOCK (Feb. 2018), S. 1. ISSN: 1073-2322. DOI: 10.1097/SHK.0000000000001115.


[23] Michael Gröger u. a. „Effects of Pretreatment Hypothermia During Resuscitated Porcine Hemorrhagic Shock“. In: Critical Care Medicine 41.7 (Juli 2013), e105–e117. ISSN: 0090-3493. DOI: 10.1097/CCM.0b013e31827c0b1f.

Literatur


[26] Katja Baumgart u.a. „Cardiac and metabolic effects of hypothermia and inhaled hydrogen sulfide in anesthetized and ventilated mice“. In: *Critical Care Medicine* 38.2 (Feb. 2010), S. 588–595. ISSN: 0090-3493. DOI: 10.1097/CCM.0b013e3181b9ed2e.

[27] Florian Simon u.a. „Comparison of cardiac, hepatic, and renal effects of arginine vasopressin and noradrenaline during porcine fecal peritonitis: a randomized controlled trial“. In: *Critical Care* 13.4 (2009), R113. ISSN: 1364-8535. DOI: 10.1186/cc7959.


[29] Vladislava Simkova u.a. „The effect of superoxide dismutase overexpression on hepatic gluconeogenesis and whole-body glucose oxidation during resuscitated normotensive murine septic shock“. In: *Shock* 30.5 (Nov. 2008), S. 578–584. ISSN: 1540-0514. DOI: 10.1097/SHK.0b013e31816a6e0f.

[30] Eberhard Barth u.a. „Effects of ventilation with 100% oxygen during early hyperdynamic porcine fecal peritonitis“. In: *Critical Care Medicine* 36.2 (Feb. 2008), S. 495–503. ISSN: 0090-3493. DOI: 10.1097/01.CCM.0B013E318161FC45.


[34] G Albuszies u.a. „Effect of increased cardiac output on hepatic and intestinal microcirculatory blood flow, oxygenation, and metabolism in hyperdynamic murine septic shock“. In: *Critical care medicine* 33.10 (2005), S. 2332–2338.

[35] Balazs Hauser u.a. „Ethyl pyruvate improves systemic and hepatosplanchnic hemodynamics and prevents lipid peroxidation in a porcine model of resuscitated hyperdynamic endotoxemia“. In: *Critical Care Medicine* 33.9 (Sep. 2005), S. 2034–2042. ISSN: 0090-3493. DOI: 10.1097/01.CCM.0000178177.03979.CE.


[40] B Santak u.a. „In vivo quantification of endotoxin-induced nitric oxide production in pigs from Na15NO3-infusion.“ In: *British journal of pharmacology* 122.8 (Dec. 1997), S. 1605–10. ISSN: 0007-1188. DOI: 10.1038/sj.bjp.0701553.


Literatur


[56] Riccardo Zanasi u.a. „Determination of 13 C/ 12 C Carbon Isotope Ratio“. In: Analytical Chemistry 78.9 (Mai 2006), S. 3080–3083. ISSN: 0003-2700. DOI: 10.1021/ac0517038.

[57] Joachim Mohn u.a. „Continuous field measurements of d13 C–CO2 and trace gases by FTIR spectroscopy“. In: Isotopes in Environmental and Health Studies 44.3 (2008), S. 241–251. ISSN: 1025-6016. DOI: 10.1080/10256010802309731.


[95] Otto S Wolfbeis, Marc J.P. Leiner und Hermann E. Posch. „A new sensing material for optical oxygen measurement, with the indicator embedded in an aqueous phase“. In: Mikrochimica Acta 90.5-6 (1986), S. 359–366. ISSN: 00263672. DOI: 10.1007/BF01199278.


[107] Max E. Lippitsch u.a. „Fibre-optic oxygen sensor with the fluorescence decay time as the information carrier“. In: *Analytica Chimica Acta* 205.C (1988), S. 1–6. ISSN: 00032670. DOI: 10.1016/S0003-2670(00)82310-7.

38


[121] Peter B. Arnouldse u.a. „Instrumentation for the Breath-by-Breath Determination of Oxygen and Carbon Dioxide Based on Nondispersive Absorption Measurements“. In: Analytical Chemistry 64.2 (1992), S. 200–204. ISSN: 15206882. DOI: 10.1021/ac00026a020.


[135] C. Radhakrishna Rao. „The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves“.

[136] Barr Rosenberg. „Linear regression with randomly dispersed parameters“.
In: *Biometrika* 60.1 (1973), S. 65–72.


[147] Robert N. Feudale u. a. „Transfer of multivariate calibration models: A review“. In: *Chemometrics and Intelligent Laboratory Systems* 64.2 (Nov. 2002), S. 181–192. ISSN: 01697439. DOI: 10.1016/S0169-7439(02)00085-0.


5 Papers

5.1 Declaration of Authorship and Permissions

5.1.1 Paper I

FS and JV co-wrote the paper. FS did all experiments and data evaluation. JV contributed the idea of the calibration transfer algorithm. BM and PR proof-read the paper and gave financial and organizational support.

Reprinted with permission from Elsevier according to the right of personal use. The work has been published first in the journal Analytica Chimica Acta.

5.1.2 Paper II

FS and JV co-wrote the paper. FS did all experiments and data evaluation. JV and FS developed the data algorithm routine together. JV contributed the idea of the response surface. BM and PR proof-read the paper and gave financial and organizational support.

Reprinted with permission from Elsevier according to the right of personal use. The work has been published first in the journal Analytica Chimica Acta.

5.1.3 Paper III

FS wrote the paper, did all calibration experiments and data analysis and was involved in the routine collection of the mouse data. ET and LTH were involved in the routine collection of the mouse data and proof-read the paper. JV assisted in writing the paper and developing the data analysis. PR and JV contributed the
medical interpretation of the data. SK and MG did the mouse experiments and proof-read the paper. UW did the GC-MS validation and proof-read the paper. PR and BM proof-read the paper and gave financial and organizational support.

Reprinted with permission from IOP Publishing. All rights reserved.

5.1.4 Paper IV

FS wrote the paper, did all experiments and data analysis. JV gave the idea of the RABBIT algorithm and JV and FS co-developed the data analysis algorithm. UW did the GC-MS measurements. PR and BM proof-read the paper and gave financial and organizational support.

Reprinted with permission from Elsevier according to the right of personal use. The work has been published first in the journal Analytica Chimica Acta.

5.1.5 Paper V

FS wrote the paper, did all calibration experiments, all data analysis and was involved in the routine data collection of the mouse data. ET and LTH were involved in the routine collection of the mouse data. JV assisted in writing the paper and developing the data analysis algorithm and developed the metabolic monitoring derived data analysis routine. UW did the GC-MS measurements. SK and MG did the mouse experiments. PR and JV contributed medical interpretation. BM and PR proof-read the paper and gave financial and organizational support.

This article was published under the terms of the Creative Commons Attribution 4.0 licence (CC BY 4.0), https://creativecommons.org/licenses/by/4.0.
5.2 Paper I: Nonlinear Calibration Transfer Based on Hierarchical Bayesian Models and Lagrange Multipliers: Error Bounds of Estimates via Monte Carlo – Markov Chain Sampling
Nonlinear calibration transfer based on hierarchical Bayesian models and Lagrange Multipliers: Error bounds of estimates via Monte Carlo – Markov Chain sampling

Felicia Seichter a, Josef Vogt b, *, Peter Radermacher b, Boris Mizaikoff a, **

a Institute of Analytical and Bioanalytical Chemistry, Ulm University, 89081, Ulm, Germany
b Institute of Anesthesiologic Pathophysiology and Method Development, Ulm University Medical Center, 89081, Ulm, Germany

ARTICLE INFO

Article history:
Received 1 March 2016
Received in revised form 9 September 2016
Accepted 9 November 2016
Available online 16 November 2016

Keywords:
Calibration transfer
Lagrange Multipliers
Fluorescence
Hierarchical models
Oxygen sensor
Bayesian statistics

ABSTRACT

The calibration of analytical systems is time-consuming and the effort for daily calibration routines should therefore be minimized, while maintaining the analytical accuracy and precision. The ‘calibration transfer’ approach proposes to combine calibration data already recorded with actual calibrations measurements. However, this strategy was developed for the multivariate, linear analysis of spectroscopic data, and thus, cannot be applied to sensors with a single response channel and/or a non-linear relationship between signal and desired analytical concentration. To fill this gap for a non-linear calibration equation, we assume that the coefficients for the equation, collected over several calibration runs, are normally distributed. Considering that coefficients of an actual calibration are a sample of this distribution, only a few standards are needed for a complete calibration data set. The resulting calibration transfer approach is demonstrated for a fluorescence oxygen sensor and implemented as a hierarchical Bayesian model, combined with a Lagrange Multipliers technique and Monte-Carlo Markov-Chain sampling. The latter provides realistic estimates for coefficients and prediction together with accurate error bounds by simulating known measurement errors and system fluctuations. Performance criteria for validation and optimal selection of a reduced set of calibration samples were developed and lead to a setup which maintains the analytical performance of a full calibration. Strategies for a rapid
determination of problems occurring in a daily calibration routine, are proposed, thereby opening the possibility of correcting the problem just in time.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

One key step when setting up an analysis system or developing a quantification technique for a particular analyte is designing an appropriate calibration strategy. For routine use, time efficiency and economic viability should be considered next to the analytical performance, as a substantial effort is required for determining an appropriate calibration model/function and maintaining it. Hence, one may be inclined to reuse a once established calibration model as long as possible in the day-to-day routine, but this may conflict with the demands on accuracy and precision. We therefore herein consider a technology for measuring a scalar signal that is supposed to be converted to a scalar analyte property, such as chemical concentration, using a calibration curve. One can define a theoretical calibration equation based on the physical measurement process; however, the physical properties could be affected by sensor imperfections, which are usually not captured by such a theoretical equation. If these imperfections cannot be eliminated by simple means and add to a nonlinear relationship between measurable signal and analyte quantity, one may resort to empirical functions. Empirical functions are defined via signal values at certain data or sampling points and allow for a smooth interpolation for intermediate values of the independent variable. Polynomial functions can cover any smooth curvature, as long as the polynomial order is high enough, but cannot efficiently approximate functions with one inflection point connecting two asymptotic lines, i.e. a response as shown for example in Fig. 1. For a good fit to reference grid points, a higher order polynomial (n > 10) is necessary. But when interpolating between the grid points, the polynomial tends to oscillate around the expected curve. These effects can be ameliorated by regularization [1], yet, series of well-defined calibration samples are needed for an accurate approximation [2]. A rational function could better approximate the above mentioned curve profile with only few coefficients. However, for a statistically reliable determination and for an acceptable error performance, the number of calibration samples should at least exceed the number of calibration coefficients. This may require an effort one cannot spare in routine calibration scenarios. We consider the case where the general shape of the calibration curve does not distinctively change from one calibration to the next or may be comparable between different but identically constructed instruments. Nevertheless, it is assumed that a previously established calibration cannot be used at present because the difference between the actual sensor response and a previously measured one exceeds the actual machine precision in some range of the analyte concentration, thus leading to a systematic calibration error. As a historic calibration may not be used directly, it would be desirable to exploit at least the fact that all calibration curves share a similar appearance. This previously gathered information should therefore be smartly combined with the information obtained from just one or two actual calibration samples yielding a useful complete calibration data set with minimal additional effort. Calibration transfer is a suitable approach to solve this problem [3] for a linear multivariate quantification. Corresponding strategies were primarily developed in the field of near infrared spectroscopy. The different strategies comprise a way to remove the fluctuation of sensor status/difference between sensors etc. via multivariate transformation (i.e., standardization). Typically, complete actual spectra or some local features, collected for a certain reference condition, are transformed to match spectra found during the initial calibration, and then the calibration is applied to the transformed spectra [3,4]. In addition, multivariate methods/preprocessing steps can be explored that are more robust against variance effects. The relative merits of such calibration transfer strategies are discussed in many publications; Noord as well as Feudale et al. give an excellent overview on the topic [3,4]. While most publication refer to multivariate linear systems, there is no straightforward way to adapt or modify these calibration transfer schemes to a system that (i) only provides a univariate signal (i.e. like a chemical sensor) and/or (ii) is subject to a nonlinear response function. To the best of our knowledge, no corresponding strategies can be found in literature, and are thus the scope of this contribution. As a solution, we propose an approach that is based on the assumption that calibration curves, collected at comparable conditions, form a family of curves whose similarity can be captured via appropriate statistics. Statistical concepts characterizing a group of lines collected under comparable conditions have evolved over time. In 1965, Rao [5] considered growth curves with random coefficients and provided equations to estimate the variability of the coefficients. Rosenberg [6] extended this approach in 1973 for curves collected at different subjects and described the inter-subject variability of curves with a multivariate normal distribution for the curve coefficients. Harville [7] developed formulas to assess coefficients for an individual curve and their covariance, if measurements for the individual curve next to group mean values and the group covariance were available. Laird [8] provided an efficient EM algorithm to estimate both the coefficients for individual curves and their group distribution. This

![Fig. 1. Raw value $\Delta \phi$ [$^\circ$] vs. oxygen concentration [%]. Custom calibration using an empirical function (see Eqn. (3)), fitted to a set of 11 calibration points spread over the range of 0–100% at 10% intervals. Blue solid line: fit, dotted red line: 95% confidence interval, black dots: experimentally obtained raw values (phase shift). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image-url)
class of approaches was termed ‘random coefficients’ model or regression, and has been summarized in a seminal textbook by Longford [9]. Herein, an improved approach to estimate the group-variance was presented. Examples how to adapt the random coefficient approach to multivariate problems were provided and applied to curves with any shape that can be covered by polynomial functions in the independent variable. Also discussed is the gain in precision of the determined polynomial coefficients, if a curve is treated as member of group of different curves, compared to the precision obtained if only actual measurements were used to define the polynomial. This gain in precision paves the way for a coefficient based calibration transfer, as proposed in our present study. The random coefficient approach has been applied to a scalar NDIR (non-dispersive infrared) sensor signal that relies on two independent variables [2]. The corresponding response surface was described by a polynomial in the two independent variables. A principal component analysis on the variance of the polynomial coefficients, collected over several calibrations, allowed isolating a subspace comprising the potential values of the coefficients, driven by day-to-day variability. The dimension of this subspace is much smaller than the number of unknown coefficients. Thus, confining the coefficients to this subspace dramatically reduced the number of samples necessary for an actual calibration. Polynomials appear to be the method of choice since they allow the retracing of almost any shape of calibration curve. However, as discussed before, a ‘regularization’ or smoothing of the polynomial is necessary. The optimal tuning for such a smoothing is critical and we are not aware of any applications linked with random coefficient models. As we will demonstrate later, only a few coefficients are necessary to approximate curve such as described above with a rational function that, unfortunately, is nonlinear in the coefficients. However, solely for calibration curves that are linear in their coefficients, such as polynomials, one can easily derive formulas for the distribution of values predicted by the calibration curve, i.e. concentration values for a given signal measurement. Such a distribution considers the propagation of the uncertainty in the parameters defining the calibration curve and in the measurements to predicted concentrations. Corresponding formulas are difficult to define for calibration equations that are inherently nonlinear. Approximate solutions for nonlinear equations have been provided in Refs. [10,11], however, precise estimates for confidence ranges have to be assessed via bootstrapping or resampling methods [12], which in turn require an iteration over different sets of curves, usually providing only approximate statistical features. Hence, we turn to the Bayesian equivalent to the random coefficient approach, the ‘hierarchical models’ [13–17]. According to Gelman [13,14], Bayesian statistics is based on the explicit use of probabilities for quantifying uncertainty in inferences based on a statistical analysis. The first step in a Bayesian data analysis is setting up a full probability model or definition of a joint probability distribution for all observable and unobservable quantities in a problem. Applied to a calibration problem, this means defining a probability in which specific signals can be observed for samples with a given analyte composition. For such a definition, one needs a function, the calibration equation, which defines a functional relation between analyte in the sample and the signal, observed as response of the sensors. The shape of calibration curves is characterized by a set of coefficients $\beta$, in the simplest case slope and intercept. In Bayesian methods, these coefficients are treated as stochastic instead of being an unknown deterministic quantity like in the more classical ‘frequentist’ framework [18]. As a stochastic quantity, the coefficients have a distribution, denoted as $P(\beta)$. For a full probability model, one also has to consider a measurement error — here we refer to it as $\sigma$. Let $D$ denote the observed data values, in the present case the sensor response. One important distribution for the full probability model is the likelihood that $D$ can be found for given set of calibration coefficients, and for given measurement error — it is denoted as $P(D|\beta,\sigma)$. In a traditional approach, this likelihood would be maximized to obtain an estimate for the coefficients $\beta$, which is usually solved by regression. In a Bayesian context however, the full probability model is defined as the joint probability distribution $P(D|\beta,\sigma)P(\beta)$. The next step in a Bayesian analysis is termed conditioning on data or estimating the conditional probability distribution of the calibration coefficients given the observed data. According to the Bayes’ Theorem [14,15,19], this conditional probability is proportional to the just mentioned product, or:

$$P(D|\beta,\sigma) \propto P(D|\beta,\sigma)P(\beta)$$

The distribution on the left side of Eqn. (1), $P(D|\beta,\sigma)$, is the posterior distribution. By obtaining a posterior distribution, interval estimates can be constructed and uncertainty information about estimates provided, since in Bayesian analysis, these intervals represent the most probable region of model parameters based on all available information [18]. In Eqn. (1), the distribution $P(\beta)$ acts as a prior for the calibration coefficients. Priors reflects any prior knowledge of the coefficients and uncertainty achieved from earlier data. Most the time, so-called vague flat priors (with a uniform distribution) are used such that the posterior distribution is mainly affected by the data. As an advantage of Bayesian analysis, information based on pre-existing knowledge can be implemented using priors [20]. The key feature of the hierarchical approach is that the prior distribution $P(\beta)$ is treated as an unknown, which, in the present case, has to be derived from data of the different calibration runs. It reflects the day to day distribution of the coefficients and allows implementing the concept that an actual calibration data set belongs to a group of independently performed calibrations with well defined group mean values for their calibration coefficients and case-to-case variance. With Bayesian analysis, a descriptive probability model to observe data given some parameters can be linked with a likelihood function that reflects our prior knowledge about the involved parameters. It allows calculating a posterior distribution for predicted quantities and with it the uncertainty in prediction. This posterior distribution reflects what we should believe about the value of the prediction, given the experimental data [20]. In a similar logic, a predictive distribution can be defined for estimating the analyte concentrations, which is based on new measurements and the determined coefficients. The corresponding distribution for prediction is based on a prediction equation, derived by converting the calibration equation and the coefficient distribution given in Eqn. (1) serves as a prior. The product of distributions used in the right side of Eqn. (1) or that for the predictive distribution might be difficult to convert into an explicit or analytically given distribution of coefficients for nonlinear calibration equations, not to mention the predictive distribution. At first glance, this mirrors the situation described for the random coefficients models. However, as a big difference, such an analytical solution is not required. Most statistical Bayesian software packages simply ‘draw a valid sample’ from such a distribution product and reconstruct the resulting ‘posterior’ distribution from a large set of samples. Corresponding software packages (e.g. ‘Winbugs’ [21]) are freely available with tutorial examples of hierarchical regression models, such as the ‘orange tree’ example for nonlinear hierarchical models [10] (Winbugs, example volumes II). Carlin [22, chapter 17] discusses implementations of a linear hierarchical model and Gelman et al. [13] consider hierarchical models applied on functions that are nonlinear in the unknown coefficients. Given their versatility, Bayesian statistics are therefore slowly making their way into chemometrics [18] [23,24]. In this study, we have developed a Bayesian calibration transfer model implemented via
the software package 'Stan' [25]. 'Stan' allows defining a statistical 'model' for almost any aspect of Bayesian statistics. Based on such a model it derives the posterior probability distribution of the model parameters from a prior probability distribution and a likelihood function. It is possible to define different priors and probability distributions for every parameter to be determined. A particular feature in Stan is that it is very effective at converging from diffuse random initialization, though user-specified values can also be implemented [26,27]. Stan utilizes a compact and concise scripting language such that the algorithm developed for the specific situation is readily shared and reproduced among different users. Moreover, error propagation from measurements or from the uncertainty in the case-to-case variability towards the prediction for curve values can be easily obtained. This feature will be used to define performance criteria for validation, for optimization of the minimal set of calibration samples and to test the applicability of the approach. Furthermore, we will explore whether statistical properties of a group of sensors can be used for the efficient calibration of a new member (i.e., a new sensor) of the same product family. The resulting implementation fully exploits the potential for stability, accuracy and precision provided by the sensor technology, obtaining a useful calibration with minimal measurement effort, yet accurate and realistic error bounds.

1.1. Example application: nonlinear calibration of an oxygen sensor

As practical implementation, the calibration transfer strategy was applied for the determination of the \(O_2\) concentration using the luminescence (LS) quenching effect of a fluorophore immobilized in a commercially available, fiber-coupled, flow-through cell (FireStingO2, Pyro Science GmbH, Aachen, Germany). Oxygen quenches the fluorescence of an exited dye immobilized in a sensor matrix, resulting in an optical of signal that is reduced in fluorescence intensity, fluorescence lifetime, and shift of the phase angle in dependence on the oxygen concentration. Noise and signal fluctuations of the sensor increase for concentrations above 50%, as less and less photons arrive at the detector and random detector noise gains prevalence. Therefore the manufacturer confines the high precision range to 0–50% oxygen [28] and provides a factory calibration that favors simplicity and robustness rather than precision and accuracy. Fluorescence quenching is described by the Stern-Volmer equation [29] or, for a more realistic scenario, two quenching dyes can be considered [30] using two different constants \(K_{S1}\) and \(K_{S2}\) and relative contributions \(f_1\) and \(f_2\) to the combined quenching:

\[
\frac{F}{F_0} = \frac{f_1}{1 + f_1 K_{S1}[O_2]} + \frac{f_2}{1 + f_2 K_{S2}[O_2]} \tag{2}
\]

Here, \(F_0\) is the incident fluorescence intensity, and \(F\) the fluorescence intensity after quenching. Experimentally determined quenching values over the entire range of 0–100% oxygen reveal that the two dye version of Eqn. (2) is necessary to describe the quenching response. Unfortunately, this equation cannot be converted to a function that is linear in the unknown coefficients, and hence, in the present case, has been replaced by the following empirical form:

\[
\Delta \phi_1 = \frac{c_1 x_t^2 + c_2 x_t + c_3}{x_t^2 + c_4} + \epsilon_t \quad \text{or} \quad \Delta \phi_1 = F(x_t, c_t) + \epsilon_t \tag{3}
\]

Here, \(x_t\) pertains to the independent variable, i.e., the oxygen concentration \([O_2]\), \(\Delta \phi\) to the phase shift [31], the raw unit for the measured signal and \(\epsilon_t\) refers to a randomly measured error, which will be considered in Eqn. (1). As Fig. 1 indicates, this equation accurately reflects the signal-to-concentration relationship. Eqn. (3) defines the shape of the calibration curve via the coefficient set \(c_1, \ldots, c_4\), which is denoted as \(c_t\), and hence, the actual state of the sensor is captured by \(c_t\). Actual \(c_t\) values are determined by the calibration, and can thus be applied in a prediction equation. The latter can be obtained by solving Eqn. (3) for the independent variable. The resulting prediction is denoted as \(y_t\), for discrimination against the measurement, and is formally described as:

\[
y_t = G(\Delta \phi_1, c_t) \tag{4}
\]

The calibration transfer will therefore be applied to determine the coefficients of \(c_t\).

Fig. 1 shows a decreasing signal at higher oxygen concentrations, which is associated with an increased measurement error in this range. The typical error distribution of an oxygen sensor for different oxygen concentration (here for several different days and sensors) is shown in Fig. 2 — the dependency on the oxygen concentration is clearly visible. This error is a physical phenomena caused by the naturally occurring intensity loss and resulting higher signal fluctuation, and a heteroscedastic measurement error. Heteroscedasticity introduces a certain bias if not considered adequately. To avoid this one requirement of most classical chemometric techniques is a homoscedastic error distribution. One of the advantages of Stan is that each error included in the model in question can be characterized via an individual, even nonlinear, function, i.e. allowing the integration of heteroscedastic errors. As displayed in Fig. 3, the error can be represented by a function:

\[
s_t = a_0 + a_1 |O_2| \tag{5}
\]

It reflects a full recreation of the error function in the model. The Bayesian analysis obtained with the MCMC sampling of Stan allows to assess the propagation of measurement errors to the final error bounds for the oxygen concentration prediction. The latter should be very accurate and realistic.

2. Material and methods

2.1. Experimental

All gas calibration samples were prepared from pure technical grade nitrogen and pure medical grade oxygen (both MTI IndustryGase AG, Neu-Ulm, Germany) using a gas mixing pump (DIGAMIX 2 M 301, H. Wösthoff Messtechnik GmbH, Bochum,
The oxygen sensors were calibrated using 11 mixtures of oxygen and nitrogen spanning 0–100% oxygen. Data evaluation was done using MATLAB 8.2 (R2013b, The Mathworks Inc, Naticks/ MA, USA) and MatlabStan 2.7.0.0 [32]/Stan 2.8.0 [25]. Six different sensors, named Sensor 1 to 6 (S1–S6) were studied over the period of one year. S1, S2, S5, and S6 represent an older generation of oxygen sensors, while S3 and S4 were purchased later. Functional properties of the sensors were the same however. All sensor sets were measured in 3–46 replicates or calibration days, thus resulting in 122 different data sets. Sensor S3 and S4 are supposed to be used under conditions where a calibration transfer is required for efficient and accurate calibration, and the calibrations for the other sensors were therefore considered the pool of previous measurements necessary to assess the day-to-day or sensor-to-sensor variability.

2.2. Numerical methods

The calibration transfer method is based on calibration Eqn. (3), and the data set described above. Following major steps have to be executed, as described later in detail:

1. As the key step of the calibration transfer: define an approach, called ‘Lagrange Multipliers’, to combine mean values and the case-to-case covariance of calibration coefficients gained from a sufficient, large enough calibration set pool with a minimal number of actual measurements to generate an actual calibration curve.

2. Define a hierarchical model, as outlined in Table 4, to use collected calibration sets to assess mean values and variance/covariance of coefficients across different sensors or different calibration sets of one specific sensor. Use the multivariate distribution of coefficients and the Lagrange Multipliers to determine actual calibration coefficients. Sample the corresponding predictions and generate parallel predictions from conventionally generated calibration curves.

3. Develop performance criteria for the calibration transfer and use them to determine the best combination of calibration samples for a calibration transfer by examining quality criteria.

4. Use these performance criteria to validate the calibration transfer model, and to evaluate its predictive capabilities via internal and external test samples.


2.2.1. Lagrange Multiplier approach

For the sake of simplicity, we assume that only one measurement with the index \(\text{t}\) is made to characterize the actual calibration. The independent variable of the calibration (i.e., concentration) shall be called \(x_t\), while the dependent variable/sensor response shall be called \(\Delta \phi_t\). We denote the set of actual calibration coefficients with \(c_t\). This vector \(c_t\) should satisfy Eqn. (3) for a given independent variable \(x_t\) and for a measured response \(\Delta \phi_t\). The key requirement is that this \(c_t\) should be a sample of a normal distribution around the mean coefficients \(\bar{c}\) and inter-day variance \(\Sigma\). The \(c_t\) values that satisfy both criteria could be determined using a Lagrange Multiplier approach [33–36]. Here, the coefficients have to satisfy an equation that is linear in the unknown coefficients, but should be applied for the nonlinear calibration Eqn. (3). Hence, this equation is rearranged to become linear in the calibration coefficient set \(\mathbf{c}\). This can be achieved for rational functions in the dependent variable as follows:

\[
\begin{align*}
\Delta \phi_t (c_4 + c_5 x_t) &= c_1 x_t^2 + c_2 x_t + c_3 \\
\Delta \phi_t x_t &= (x_t^2, x_t, 1, -\Delta \phi_t) \\
&= (x_t^2, x_t, 1, -\Delta \phi_t) (c_1, c_2, c_3, c_4)
\end{align*}
\]

(6)

The left side of the last row of Eqn. (6) corresponds to a transformed response variable, which is noted as \(R = \mathbf{B}^T \mathbf{c}_t\). For the response variable \(R\), one obtains the following linear equations:

\[
R = \mathbf{B}^T \mathbf{c}_t; \quad \text{with} \quad \mathbf{B} = (x_t^2, x_t, 1, -\Delta \phi_t)
\]

(7)

From the requirement that \(c_t\) should also be as close as possible to the mean coefficient set \(\bar{c}\), one can define the Lagrange-Norm

\[
Q = \frac{1}{2} (c_t - \bar{c})^T \Sigma^{-1} (c_t - \bar{c}) + \lambda \left( \mathbf{B}^T \mathbf{c}_t - R \right)
\]

(8)

Here, the first part pertains to the assumption that \(c_t\) is a sample of a normal distribution around \(\bar{c}\) and \(\Sigma\), and the second part reflect the restraint that \(\Delta \phi_t\) lies on the calibration curve. The Lagrange-Norm (8) is then minimized against \(\alpha\) and \(\lambda\). Minimizing by \(\lambda\) gives

\[
\frac{\partial Q}{\partial \lambda} = 0, \quad \text{which gives} \quad \mathbf{B}^T \mathbf{c}_t = R
\]

(9)

Minimizing (8) by \(\mathbf{c}_t\) results in

\[
\frac{\partial Q}{\partial \mathbf{c}_t} = 0, \quad \text{which yields} \quad \Sigma^{-1} (c_t - \bar{c}) + \lambda B = 0
\]

(10)

or:

\[
c_t = \bar{c} - \lambda \Sigma B
\]

Insertion in (9) results in

\[
\begin{align*}
\mathbf{B}^T (\mathbf{c} - \lambda \Sigma B) &= R \\
\mathbf{B}^T \Sigma B &= R - \mathbf{B}^T \bar{c} \\
\lambda &= \left( \mathbf{B}^T \Sigma B \right)^{-1} \mathbf{B}^T \bar{c} - \left( \mathbf{B}^T \Sigma B \right)^{-1} R
\end{align*}
\]

(11)

This yields the factor \(\lambda\), which can be inserted into Eqn. (10) to obtain the desired \(\mathbf{c}_t\). If more calibration samples are used, \(\lambda\) and \(\mathbf{R}\) are vectors and \(\mathbf{B}\) is a matrix. In the following, we subsume the equations above to a 'Lagrangian function':

\[
\mathbf{c}_{\text{LGF}} = \text{LGF}(\mathbf{c}_t, \Sigma, \Delta \phi_{\text{limited}})
\]

(12)

2.2.1.1. Hierarchical model. ‘Stan’ is an open-source Bayesian programming language that allows defining a statistical model with a structure that is outlined in Fig. 3 and exemplified for the case of a nonlinear calibration based on Eqn. (3). The data set \(\mathcal{D}\) is defined in the Data Block, at the top of the figure. Unknown parameters, which have to be determined from data, are defined in the parameter block—in the present example the coefficients \(\mathbf{c}\) and parameters to describe a signal dependent measurement error. The software samples these parameters from a prior distribution, which can be specified in the model block. Otherwise a default, non-informative prior is used. In our case, we choose flat priors in a specified range (uniform distribution), derived from expected values gained from earlier calibration and measurements (i.e., the default initialization in Stan) as initialization values. In the model block, the actual, underlying statistical model is defined, and reflects the probability \(P(\mathcal{D} | \mathbf{c})\), as defined in Eqn. (1). First, a measurement error for each calibration level is defined. Then the latter is used to assess the probability that a measurement can be linked to the calibration curve for each replicate. This results in a combined probability of finding the data given the parameters to consider. Note that \(P(\mathcal{D} | \mathbf{c})\), as used in Eqn. (1), is realized in a series of normal distributions \(N(\mathbf{y}_x, \sigma_i)\) for individual measurements, where the mean value is calculated from \(F(\mathbf{x}_c, \alpha)\), which is a nonlinear function in the coefficients \(\mathbf{c}\). A heteroscedastic measurement error \(\sigma_i\) defines the variance of this distribution. The lines in the Model Block demonstrate a key feature of Stan: Instead of using an analytical expression for a complex conditional probability, a sampling statement is defined, which may be composed of different elements. During a simulation run, ‘Stan’ picks or samples a set of parameters, calculates the probabilities for this sample and checks whether the combined probability is large enough to be consistent with the measured data and the proposed distributions. If they are consistent, the parameter set is accepted as a sample of the underlying, unknown parameter distribution; otherwise, it is rejected. Strategies for rejection/acceptance, and for picking up an appropriate parameter set, follow the principles of Monte-Carlo Markov-Chain (MCMC) sampling [22,37,38]. Typically, a sequence of 2000 or more parameter sets is sampled, and MCMC-update and rejection criteria are designed to ensure that each set is a sample of the underlying distribution. During sampling for each valid set of parameters, the generated quantities block is called. Here, the actual set of parameters can be used for prediction: values on the calibration curve at different levels can be assessed for each set of parameters. O2 concentration values can be predicted based on measured signals for different levels and replicates. About 50 replicate actual measurements were used for each level, which vary only in their current measurement error range. The resulting predictions depend and fluctuate with the actual parameter set and the measurement error and are collected in a sampling chain comparable to the sampled parameters. The distributions of parameters and derived values, such as predicted concentrations, in the sampled chains reflect the underlying ‘true’ distributions. Therefore, these sequences can be used to reconstruct the underlying distributions and can be used to calculate statistical parameters including average, standard deviation and 5% and 95% quantiles. In the present case, it would be very difficult to devise an analytical expression for the distribution of the predicted \(O_2\) concentration or \(y_{ij}\) values for the nonlinear prediction equation, given the uncertainty in the average day-to-day coefficient values and given the measurement error. The simple sampling statement in this block provides adequate results. A complete ‘Stan’ script to execute this example can be found in the Appendix as well as the
supplementary material. ‘Stan’ uses a very efficient MCMC variant, i.e., the Hamiltonian Monte Carlo/No-U-Turn Sampler [39] and can be set up on all major operation systems [25]. Interfaces to most major statistical software packages are available, including, at the moment R, MATLAB [32], Python, Julia and Stata. Since the underlying ‘Stan’ programming script remains the same for each platform and interface, an effortless transfer between the different software platforms and therefore, rapid and facile exchange of finished scripts is enabled. A full description and manual of the ‘Stan’ programming language can be found in the supporting information of the software [25–27,40].

The case-to-case variability of the calibration coefficients is described by a mean coefficient set \( \mathbf{c} \) and covariance matrix \( \Sigma \), which form a multivariate normal distribution and should be determined from several measured calibration sets. The determination scheme outlined in Fig. 5 is extended by expanding the data set to cover several calibration sets, and \( \mathbf{c} \) and \( \Sigma \) are added to the parameter list (Fig. 4). This means that on the upper level of the sampling, the software draws a sample of \( \mathbf{c} \) and \( \Sigma \) from a default, non informative prior distribution. These values are used to define the probability \( P(c) \), which is used in the model block as prior distribution. The model block defines statements to sample from hierarchical product \( P(\mathbf{c}|P(c)) \) used in Eqn. (1). For sampling, an additional higher-level loop is put around the two loops for replicate measurements and calibration levels. The higher-level loop pertains to different calibration sets. Here an actual coefficient set \( c_i \) is sampled from \( P(c) \) and these coefficients are used in the inner loop to sample from the probabilities to find specific measurement values. This sampling is repeated for each of the different calibration sets. In the inner loop, we also use a limited set of measurements for each calibration set, which is denoted \( \Delta \mathbf{c}_\text{limited} \) to determine coefficients based on the Lagrange-Multiplier approach using Eqn. (12). The sampling across different levels yields the parameters for the case-to-case variability of the coefficients, and the other parameters for the inner loops. As the uncertainties in the coefficients due to measurement errors are covered in the inner loop, they do not affect the estimates for the case-to-case variability. In the generated quantities block of the expanded script, the Lagrange-derived coefficients and the coefficients from the hierarchical model are both used to obtain predicted concentrations, which are denoted with \( \mathbf{z} \) for the Lagrange-Model. The chain of sampled predictions is then used to derive a quality index for the full model, and for the reduced Lagrange-model.

### 2.2.2. Quality index for a full calibration

One important parameter to assess the quality of a calibration is the root mean square error of prediction (RMSEP). It is defined for each calibration level as:

\[
\text{RMSEP}_{\text{level}} = \sqrt{\frac{\sum (y_i - \bar{x}_i)^2}{n}}
\]

where \( y_i \) is the concentration predicted from the calibration and \( \bar{x}_i \) the real concentration of sample \( i \). The lower the RMSEP, the better the predictions of the calibration model fit the expected values. The average RMSEP (RMSEP) over all runs, days and levels is tentatively a useful indicator for the overall calibration performance of a model. However, herein random measurement errors and prediction errors increase with increasing oxygen concentration levels, and a plain average of RMSEPs over the different levels can be dominated by the errors at higher levels. Optimizing such an average RMSEP can result in minimization of the error at high oxygen concentration levels. For an objective quality criterion, the errors at each level have to be weighted with a factor. To develop such a factor, one has to consider that the RMSEP of the Lagrange fit cannot be better than the RMSEP of the single or hierarchical fit, since the underlying calibration samples determine the general quality of the fit. The aim of the calibration transfer is to maintain possibly close the predictive capabilities of a full calibration while reducing the calibration effort. To quantify this with a scalar value, a relative error variable \( \Delta \text{RMSEP} \), is defined for each level \( i \) using the RMSEP of the Lagrange fit (RMSEP\(_i\)) and the RMSEP of the hierarchical fit (RMSEP\(_{\text{IM}}\)):

\[
\Delta \text{RMSEP}_i = \frac{\text{RMSEP}_{i\text{IM}} - \text{RMSEP}_{i}}{\text{RMSEP}_{i\text{IM}}}
\]

The sum of the \( \Delta \text{RMSEP} \) of one Lagrange fit over all levels is now considered as a measure of success of the calibration transfer over the entire calibration range, and is defined as:

\[
\Sigma \Delta \text{RMSEP} = \sum_{i} (\Delta \text{RMSEP}_i)
\]

\( \Sigma \Delta \text{RMSEP} \) is then calculated for each calibration run. A low \( \Sigma \Delta \text{RMSEP} \) is an indicator that the reduced calibration transfer fit comes close to the predictive abilities of a full calibration, i.e., that the calibration transfer was successful. If calibration sample values of a day deviate from the usual expected accuracy, resulting in a ‘bad’ calibration, the values for \( \text{RMSEP}_{i\text{IM}} \) and \( \text{RMSEP}_{i} \) will be equally worse and therefore cancel each other out. Hence, \( \Sigma \Delta \text{RMSEP} \) cannot detect such a situation. In order to account for this problem, \( \Sigma \Delta \text{RMSEP}^2 \) is defined. Instead of using separate \( \text{RMSEP}_{i\text{IM}} \) calculated for each calibration day as weighting factor, the average \( \text{RMSEP}_{i\text{IM}} \) of all fits in the calibration model is used:

\[
\Sigma \Delta \text{RMSEP}^2 = \sum_{i} (\Delta \text{RMSEP}_i)^2
\]

This value allows detecting calibrations that provide less accuracy than expected compared to the usual calibrations.

### 2.2.3. Quality index for an actual calibration transfer

For a typical calibration based on multiple samples, problems with an individual sample, either during calibration sample preparation or caused by unusual sensor fluctuation, may be identified, if the corresponding measurement points are off the calibration curve. If the number of measurements is smaller than the number of calibration coefficients, and if the coefficients for the calibration curve are determined by Lagrange Multipliers, then the measured points always lie on the calibration curve and the lack of fit criterion fails. For the same reason, the criterion based on \( \Sigma \Delta \text{RMSEP} \) cannot be used. Erroneous measurements might drive the determined location of \( c_i \) away from the mean value \( \mathbf{c} \) and thus, may have the same effect as a sensor drifting from its previously determined mean value. In both cases, the key assumption of the present calibration transfer scheme is challenged, i.e., that the coefficient values to be determined (\( c_i \)) are part of a normal distribution around the mean values \( \mathbf{c} \) and covariance matrix \( \Sigma \). With increasing distance from the mean value, probability decreases that \( c_i \) is a sample of the multi-normal distribution around \( \mathbf{c} \) and \( \Sigma \). Therefore...
probability is defined as:

\[ P(\Delta \phi_{\text{pred}} | \Delta \phi_{\text{real}}) \sim \text{Normal}(\Delta \phi_{\text{real}}, \sigma \Delta \phi) \]  

(19)

Values of \( P(\Delta \phi_{\text{pred}} | \Delta \phi_{\text{real}}) \) of an actual calibration that have changed compared to the values gained during model building indicate that the control point lies off the calibration curve, indicating problems with the calibration transfer. In addition, Eqn. (17) can also be calculated for the reduced calibration set used with a calibration transfer. Thus, three criteria for the applicability of the calibration transfer are available: \( P(\Delta \phi_{\text{pred}} | \Delta \phi_{\text{real}}) \), \( P(\Delta \phi_{\text{pred}} | \Delta \phi_{\text{real}}) \) and \( \Delta \text{RMSEP}_i^2 \), calculated for the reduced sample set. While building the calibration model, values for these criteria can be assessed from the data that were used to build the calibration model. Corresponding values can be collected over the MCMC runs performed with model building. This provides some reference values and distributions for the quality criteria describing proper calibration transfer. For comparison, MCMC sampling distributions can be collected for the criteria calculated from two or three samples measured for the actual calibration transfer. If the new distribution of a criterion differs extensively from those of the calibration model, then the performance of an actual calibration transfer may be compromised. This helps with quickly deciding during routine
Table 1
Comparison of possible combinations of training sets and calibration set points. The row in bold face font indicates the best choice for the training model, which was applied subsequent calculations.

<table>
<thead>
<tr>
<th>O₂ conc. [%]</th>
<th>RMSEP</th>
<th>SD</th>
<th>RMSEP</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Set 1</td>
<td>Set 2</td>
<td>Set 3</td>
<td>Set 1</td>
</tr>
<tr>
<td>0</td>
<td>1.76</td>
<td>2.11</td>
<td>44.00</td>
<td>41.24</td>
</tr>
<tr>
<td>20</td>
<td>44.75</td>
<td>1.66</td>
<td>46.20</td>
<td>2148</td>
</tr>
<tr>
<td>30</td>
<td>1.32</td>
<td>1.76</td>
<td>0.68</td>
<td>77.38</td>
</tr>
<tr>
<td>100</td>
<td>1.24</td>
<td>1.20</td>
<td>6.38</td>
<td>100.67</td>
</tr>
<tr>
<td>0, 100</td>
<td>1.17</td>
<td>1.01</td>
<td>0.87</td>
<td>28.00</td>
</tr>
<tr>
<td>0, 20</td>
<td>1.31</td>
<td>0.89</td>
<td>0.65</td>
<td>15.14</td>
</tr>
<tr>
<td>0, 30</td>
<td>1.29</td>
<td>0.72</td>
<td>0.61</td>
<td>13.82</td>
</tr>
<tr>
<td>20, 30</td>
<td>0.73</td>
<td>0.90</td>
<td>0.74</td>
<td>66.13</td>
</tr>
<tr>
<td>30, 100</td>
<td>1.52</td>
<td>0.76</td>
<td>0.70</td>
<td>69.72</td>
</tr>
<tr>
<td>0, 20, 30</td>
<td>0.72</td>
<td>0.87</td>
<td>0.75</td>
<td>5.93</td>
</tr>
<tr>
<td>0, 0, 30</td>
<td>0.93</td>
<td>0.73</td>
<td>0.73</td>
<td>15.37</td>
</tr>
<tr>
<td>20, 30, 100</td>
<td>0.77</td>
<td>0.78</td>
<td>0.66</td>
<td>63.14</td>
</tr>
<tr>
<td>0, 20, 30, 100</td>
<td>0.76</td>
<td>0.70</td>
<td>0.63</td>
<td>7.02</td>
</tr>
</tbody>
</table>

Table 2
Repeatability of MCMC sampling.

<table>
<thead>
<tr>
<th></th>
<th>RMSEP</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.6070</td>
<td>4.4881</td>
</tr>
<tr>
<td>Std.dev.</td>
<td>0.0006</td>
<td>0.0223</td>
</tr>
</tbody>
</table>

Table 3
External model validation using random subsets cross validation.

<table>
<thead>
<tr>
<th>#Training set/#Test set</th>
<th>RMSEP</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 1/1A</td>
<td>1.12</td>
<td>1.52</td>
</tr>
<tr>
<td>Std.dev.</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>Mean 2/2A</td>
<td>0.85</td>
<td>0.87</td>
</tr>
<tr>
<td>Std.dev.</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>Mean 3/3</td>
<td>0.61</td>
<td>0.64</td>
</tr>
<tr>
<td>Std.dev.</td>
<td>0.03</td>
<td>0.05</td>
</tr>
<tr>
<td>Mean 3/2A</td>
<td>0.60</td>
<td>0.65</td>
</tr>
<tr>
<td>Std.dev.</td>
<td>0.03</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 4
External validation of sensor drift over time and transfer across different sensors.

<table>
<thead>
<tr>
<th>#Training</th>
<th>#Test</th>
<th>RMSEP</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor drift over time</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2A</td>
<td>0.96</td>
<td>0.75</td>
</tr>
<tr>
<td>Std.dev.</td>
<td></td>
<td>0.64</td>
<td>0.53</td>
</tr>
<tr>
<td>Transfer across different sensors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1-S4 (1)</td>
<td>55/56(1B)</td>
<td>1.28</td>
<td>1.47</td>
</tr>
<tr>
<td>S3 (2)</td>
<td>54/2B</td>
<td>0.70</td>
<td>0.59</td>
</tr>
<tr>
<td>Std.dev.</td>
<td></td>
<td>0.50</td>
<td>1.31</td>
</tr>
<tr>
<td>S3/S4 (3)</td>
<td>51/52/55/56</td>
<td>0.61</td>
<td>1.87</td>
</tr>
<tr>
<td>Std.dev.</td>
<td></td>
<td>0.61</td>
<td>3.97</td>
</tr>
</tbody>
</table>
Fig. 6. Normal Calibration: a) $P(c_t|c)$ vs. $P(Df_{\text{pred}}|Df_{\text{real}})$ for full calibration sets, b) $P(c_t|c)$ vs. $\Sigma\Delta\text{RMSEP}_2$ for full calibration sets, c) $P(c_t|c)$ vs. $P(Df_{\text{pred}}|Df_{\text{real}})$ for reduced calibration sets, d) $P(c_t|c)$ vs. $\Sigma\Delta\text{RMSEP}_2$ for reduced calibration sets.

Fig. 7. Set 1 modified: first column: $P(c_t|c)$ vs. $P(Df_{\text{pred}}|Df_{\text{real}})$ for reduced calibration sets, second column: $P(c_t|c)$ vs. $\Sigma\Delta\text{RMSEP}_2$ for reduced calibration sets. First row: wrong test gas measured (wrong $\Delta\phi$ value), second row: too fast start of measurement (large standard deviation), third row: both cases combined (worst case scenario).
work whether a calibration is having problems and if it is possible to correct it in real-time.

2.2.4. Validation of the calibration transfer

Three different strategies for validating the calibration transfer were performed, based on the training and test sets shown in Fig. 5. For each validation strategy, a specific training set was validated against two different data sets, i.e., one test set consisted of selected calibration runs removed from the full set and kept aside for validation. Two versions were explored: (i) a test set of calibrations, collected after a certain time, was validated against a training set collected before to assess sensor drift, and (ii) training and test sets were randomly selected from the same set of calibrations, for external validation. In the later strategy, the calibration runs collected for one sensor were validated against calibration runs collected for another sensor, to explore a calibration transfer from known sensors to unknown sensors. In detail, the corresponding validations steps are:

1. Validation Set 1: consisting of all sets of S1, S2, S3, S4. As test set 1A: selected sets removed from the full set of calibration runs. As test set 1B: set consisting of all calibration runs of S5 and S6
2. Validation Set 2: set consisting of all runs of S3. As test set 2A: selected sets removed from the full set of calibration runs. As test set 2B: set consisting of all calibration runs of S4
3. Validation Set 3: set consisting of all calibration runs of S3 and S4. As test set 3: selected sets removed from the full set of calibration runs

In a full calibration, relying only on actual calibration samples, 11 different levels or samples were used. From these 11 calibration levels, a limited set is combined with statistical information on the coefficients for a calibration curve to define an actual calibration. Four calibration levels were readily available: 0%, 20%, 30% and 100% oxygen. For brevity, in the present study only these four levels have been considered for the calibration transfer.

3. Results and discussion

3.1. Determination of optimum calibration model and set points

To define a calibration model, one has to select the training set used to define the case-to-case variability and select the test set for later validation. Possible combinations are outlined in Fig. 5. However, at this point of time, no test sets were removed yet, and all data sets are in the training set. Before validation, one has to determine the best number and combination of calibration set points used for the Lagrange Multiplier calculation. The 15 possible combinations of using up to four set points from four possible calibration levels were investigated. Both \( \text{RMSEP} \) and \( \Sigma \text{RMSEP} \) were averaged over all levels and days and were used as comparison points. Here, the criteria for “best” was a compromise between the minimum number of set points to minimize calibration efforts, while still maintaining the error levels acceptable.

From all possible models, the version using training set 3 consisting of S3 and S4 and two calibration set points (i.e., 0 and 30% oxygen), clearly shows the best performance. The combination of these two set points establishes the best compromise between calibration stability and availability of calibration gas. Concurrently, Table 1 shows that, as previously mentioned, the \( \text{RMSEP} \) is vulnerable to larger error variances of higher oxygen levels, which limits its applicability as indicator for the performance of the calibration transfer. In contrast, \( \Sigma \text{RMSEP} \) indicates reliably that the calibration transfer was successful, since only the deviation from the error expected for the given data is considered. This parameter is therefore more sensitive to a particular deficient calibration transfer vs. \( \text{RMSEP} \), in which extreme values may cancel each other out. To show the reproducibility of the MCMC sampling, the optimum model was repeated five times. The results are summarized in Table 2 and demonstrate excellent reproducibility of the MCMC sampling.

3.2. Model validation

The calibration transfer model was validated to explore the performance of the transfer in the case of new (i.e., “unknown”) calibrations, but also to evaluate its robustness, if different calibration sets are removed. 25% of the calibration samples used before were randomly removed before the model building process and were used as test sets. This external validation was repeated 10 times. Three different calibration set models were compared using the optimum number and combination of calibration set points shown in Table 1. Table 3 indicates that the values for \( \text{RMSEP} \) and \( \Sigma \text{RMSEP} \) of the training and test sets of the three models are quite stable and comparable when randomly extracting sets from calibration runs for validation. Comparing the averages over the 10 repetitions of each validation run with the obtained values summarized in Table 1, both calibration and validation are quite close and no significant deterioration of the predictive abilities can be detected, which is confirmed by the associated standard deviations. Only the model based on calibration set 2 (only S3) shows larger standard deviations and deviation from the full set model. It appears that the calibration robustness of this model is reduced vs. the other models, if calibration runs are removed.

3.3. Calibration of sensor drift

For the calibration transfer, it is assumed that coefficients randomly fluctuate around mean values. A systematic drift of the coefficients over time would certainly challenge the applicability of the transfer model. It is therefore important to determine whether the calibration transfer is stable over a period of a few months without requiring recalibration. As a test, several calibration sets of S3 recorded during a later period were removed, and used as a validation set. When testing them against calibration model 2 and 3, the results shown in the upper part of Table 4 were achieved. The \( \Sigma \text{RMSEP} \) value is quite stable for calibration and validation, and comparable to the values achieved for the random validation. Consequently, the sensor drift for this example case appears negligible.

3.4. Calibration transfer for new unknown sensors

As derived from real-world analytical scenarios, it is frequently desirable to apply an established calibration transfer model to systems/sensors that have not been included in the previous calibration process; in the best case, these sensors may be used without a full calibration routine executed for the new sensor. The lower part of Table 4 presents several combinations for calibration transfer. While calibration model 2, validated with test model 2B, shows promising results and nearly no deterioration in calibration performance, the other models reveal increased deviations. This indicates that in some instances, there may be some loss in calibration accuracy, if an unknown sensor is calibrated via the transfer routine without inclusion in the calibration model. Hence, whether the calibration transfer is successful for an unknown and non-calibrated sensor needs to be assessed for each individual case.
3.5. Quality control of a calibration

We have proposed to use \(P(\text{ct}\mid \varphi)\) and \(P(\Delta \varphi_{\text{pred}}\mid \Delta \varphi_{\text{real}})\) as defined in Eqs. (17)–(19) as criterion parameters for the quality of a calibration transfer. Plots of \(P(\text{ct}\mid \varphi)\) vs. \(P(\Delta \varphi_{\text{pred}}\mid \Delta \varphi_{\text{real}})\) and \(P(\text{ct}\mid \varphi)\) vs. \(\Delta \text{RMSEP}\) for both full calibration and reduced Lagrange calibration transfer set are shown in Fig. 6. They include the results of all 2000 MCMC chains calculated for 35 different calibration runs and show the distributions of the quality criteria. For the reduced calibration set, the two calibration samples at 0% and 30% O\(_2\) concentration, which were used to assess the coefficient, plus one control sample at 20% O\(_2\) were applied. The lower panels in Fig. 6, pertaining to the reduced set, show a pattern comparable to the upper panels, which reflect the fit quality obtained with the 11 measurements of the full calibration set. This indicates that the three measurements of the reduced set were in fact sufficient to assess the fit quality of an actual calibration, and similar enough to serve as a ‘rapid warining parameter’ (i.e., cross-check) for failed calibrations. The values for the quality criteria are around 2 with two exceptions. A closer look at the corresponding individual calibration curves indicates that they were still acceptable in terms of calibration performance, despite the fact that they do not have the high quality of the other calibration data sets. With the quality criteria, it is desired that two different failures should be detected. Firstly, an accidental measurement of the wrong test gas, resulting in a wrong \(\Delta \varphi\) value and secondly, an accidental calibration measurement before reaching equilibrium of the test gas in the instrument after switching from one concentration level to the next, thus causing a larger than usual standard deviation. To explore the sensitivity of the quality criteria against possible failures, the raw calibration data of one calibration run was artificially altered to simulate first, a case of a deficient measurement of one calibration sample, by increasing all \(\Delta \varphi\) values at 0% O\(_2\) by several units from the expected value of 50. Secondly, the case of a measurement prior reaching equilibrium was simulated by lowering the first two \(\Delta \varphi\) values at 0% O\(_2\) by several units, thus resulting in a much higher standard deviation than usually expected. The third investigated case was a combination of both. All cases herein are modeled after the most common experimentally observed calibration errors for such sensors and thus serve as excellent model-cases for problems occurring during real-world routine calibration situations. Evidently, such problems can be recognized and remedied during routine calibration protocols. The results of these effects are shown in the three panels in Fig. 7.

Here, a distinct difference of the altered set (blue) compared to the others is immediately evident. The parameter used as quality control criteria show a significantly worsened parameter value and larger spread compared to those achieved by a not disturbed calibration data set. Recalling the definition of probabilities, it shows that the \(\varphi\) values calculated from the ‘wrong’ calibration set are considerably more instable and vary significantly more compared to a correctly calibrated data set, which in turn impacts on the predictive performance of that calibration.

The probabilities \(P(\text{ct}\mid \varphi)\) and \(P(\Delta \varphi_{\text{pred}}\mid \Delta \varphi_{\text{real}})\), the calibration error representation \(\Delta \text{RMSEP}\), as well as their graphical representations appear to be excellent indicators for rapid recognition of problems during a daily calibration transfer routine. Therefore, these results allow for a rapid and immediate correction of potential problems. Only in the case of persisting problems, a full recalibration of the sensor in question has to be considered.

4. Conclusion and outlook

We have developed a Bayesian calibration transfer algorithm using Lagrange Multipliers and a hierarchical model that is implemented within a multi-platform software package. Using Monte-Carlo Markov-Chain sampling, an accurate statistic estimation of error bounds concerning any aspect of the statistical model pertaining to calibration transfer is ensured. This includes estimates for the coefficients of a nonlinear calibration equation, determination of the raw measurement error as a function of the signal intensity, estimates for the variability of the calibration coefficients across different calibration runs, and – derived from these parameters, quantitative estimates for the prediction error associated with the calibration transfer. This provides the basis for defining quality criteria for a full calibration and for a calibration transfer, which can be used to detect problems with daily calibration routine for correcting problems just in time. With these tools, the calibration transfer approach developed herein could be optimized to exploit the potential for stability, accuracy and precision provided by the applied sensor technology for obtaining valid calibration models with minimal experimental effort. A key concept of the present approach is to represent a nonlinear response curve by a few coefficients and perform statistics on these coefficients. Generally, this concept may be applied also to more complex, e.g., spectroscopic data. With a principal component analysis (PCA) performed on optical spectra, the actual shape of the spectrum may be represented by a few ‘scores’. The latter may be nonlinear functions dependent on the composition of the sample, which may be completely characterized by a few coefficients. Using such a conversion, the calibration transfer concept developed herein may be applied on spectra with score values that are nonlinear, e.g., in the concentration of the sample components. While the present study only the implementation of the calibration transfer algorithm for univariate oxygen sensors is shown, this calibration strategy was also developed for the multivariate quantification of carbon dioxide via an infrared gas sensor combining a Fourier transform infrared spectrometer (FTIR) with a highly miniaturized substrate-integrated hollow waveguide (iHWG) gas cell [41–46] for correcting the influence of oxygen on the carbon dioxide signal. As both sensor systems – fluorescence for O\(_2\) and infrared for CO\(_2\) – are calibrated using the same calibration gases (i.e., mixtures of O\(_2\) and CO\(_2\) in nitrogen) and within the same procedure, the prerequisites for applying the developed calibration transfer protocols are immediately evident for minimizing the calibration efforts in daily routine work. This combined sensors system is currently operated during routine studies at the Mouse Intensive Care Unit (MICU) of the Institute of Anesthesiologic Pathophysiology and Method Development at Ulm University, and serves as an on-line monitoring system for quantifying oxygen and carbon dioxide in exhaled breath of a ventilated mice [47–49]. As will be shown during future studies, the combination of the calibration transfer algorithms discussed herein with such higher-order sensing concepts will therefore facilitate the integration of these sensing technologies into the ventilation system of the MICU, in particular since the daily calibration efforts will be minimized.

Acknowledgement

The authors would like to thank the Federal State of Baden-Württemberg (Landesgraduiertenförderungsgesetz, grant number 1212-LGFG-E) for partial financial support. This work was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract DE-AC52-07NA27344. This project was partially funded under LLNL sub-contract Nos. B598643 and B60301. The Boehringer Ingelheim Ulm University BioCenter (BIU, grant number D.5008) is acknowledged for financial support.
Appendix

functions(
    real prediction(vector coeffs, real O2)
    // this is the calibration function
    { return((coeffs[1] * O2 + coeffs[2] * D2 + coeffs[3]) / (D2 + coeffs[4])); }
    real O2prediction(vector coeffs, real dphi)
    // the inverse calibration, converts measurements to concentration
    { real a; real b; real c;
      real squareroot;
      a<coeffs[1];
      b<coeffs[2]-dphi;
      c<coeffs[3]-dphi*coeffs[4];
      squareroot=--sqrt(b*b-4*a*c);
      return((-b-squareroot)/(2*a)*100);
    }
    real RMSEP(vector y, vector yhat, int N)
    // root mean squared error of prediction
    { real summed;
      summed<-squared_distance(y,yhat);
      return(sqrt(summed/N)); }
)

data
    { int<lower=0> levels; //number of calibration levels/data points
      vector[levels] O2;
      vector[levels] dphi;
    }

    transformed data
    { vector[levels] O2_trans;
      O2_trans<-O2/100;
    }

    parameters
    { vector[4] Coefs;
      real<lower=0> sigma; // the measurement error is also determined
    }

    model
    { real pred;
      for (n in 1:levels)
      { pred<-prediction(Coefs, O2_trans[n]);
        dphi[n]~normal(pred,sigma);
      }
    }

    generated quantities
    { vector[levels] O2_prediction;
      real RMSEP_calibration;
      for (n in 1:levels)
      { O2_prediction[n]<-O2prediction(Coefs,dphi[n]);
        RMSEP_calibration<--RMSEP(O2,O2_prediction,levels); }
Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.aca.2016.11.025.

References

5.3 Paper II: Response-surface fits and calibration transfer for the correction of the oxygen effect in the quantification of carbon dioxide via FTIR spectroscopy
Response-surface fits and calibration transfer for the correction of the oxygen effect in the quantification of carbon dioxide via FTIR spectroscopy

Felicia Seichter a, Josef Vogt b, *, Peter Radermacher b, Boris Mizaikoff a, **

a Institute of Analytical and Bioanalytical Chemistry, Ulm University, 89081 Ulm, Germany
b Institute of Anesthesiologic Pathophysiology and Method Development, Ulm University Medical Center, 89081 Ulm, Germany

HIGHLIGHTS
• The background gas matrix can significantly interfere with an FTIR peak of carbon dioxide.
• A response surface including data reduction via PCA successfully corrects these effects.
• Calibration efforts are greatly reduced by a newly developed calibration transfer.
• Optimum model parameters are determined and cross-validated for the FTIR quantification.

ABSTRACT
During routine Fourier-Transform Infrared Spectroscopy (FTIR) based quantification of carbon dioxide in breath, it is necessary to account for a non-linear signal response to the analyte concentration and disturbance factors arising from the gas background matrix. These factors as well as day-to-day fluctuation should be corrected via calibration. We present a novel strategy to combine the information of previous calibrations with a minimal number of actual calibration measurements to obtain a precise calibration.

After decomposition of the FTIR spectra via principal component analysis (PCA) into scores (corresponding to intensity) and loadings (corresponding to spectral curves), an empirical response surface fit equation between scores, analyte concentration and disturbance factors is established. The fit equation can be characterized via the coefficients determined by calibration. Out of a pool of coefficients gained from several calibrations, a multivariate inter-day distribution is generated. By requiring the coefficient set of the actual calibration to be a sample of the multivariate inter-day distribution, the number of necessary routine calibration samples is reduced to two. The corresponding coefficients are determined using the Lagrange Multipliers approach and the inter-day variability of coefficients is estimated using Bayesian statistics and Hierarchical models. The best calibration parameters in terms of calibration equation, wavelength region, preprocessing options and choice of routine calibration samples were determined; optimized for minimal number of calibration samples.

© 2017 Elsevier B.V. All rights reserved.
1. Introduction

Carbon dioxide is not only a component in air, sea water, product in combustion or an important greenhouse gas, it also plays an important part in many metabolic processes in living organisms.

As an important biological building block, biomarker and metabolite, it is, e.g., involved in photosynthesis, alcohol fermentation, the tricarboxylic acid/citric cycle, pentose phosphate pathway, fatty acid catabolism, acetyl-CoA pathway as well as many more [1–3]. Within all these processes, CO₂ is generated and hence, whole body CO₂ production can be used as a marker of metabolic activity. In medicine, the partial pressure of carbon dioxide in arterial (pCO₂) in blood samples is determined on routine basis via blood gas analyzers, and the ratio of pCO₂ to expired CO₂ concentration serves as an essential marker for pulmonary functioning during acute treatment of patients in the hospital [4]. This may serve as a show case that a valuable diagnostic parameter can be derived when carbon dioxide concentration in exhaled breath is combined with other related metabolic parameters, opening the path to non-invasive breath analysis.

There are many ways of detecting and quantifying carbon dioxide in gas, including the popular infrared (IR) spectroscopy, since carbon dioxide is a strong IR absorber. When detecting in a varying gas background matrix, e.g., in the breath of artificially resired patients, which can range from 20 to 80% oxygen, the system needs to be either unaffected from such background effects or the resulting effect needs to be corrected, whether by experimental modification or via data analysis strategies.

Fig. 1 shows two FTIR spectra of the same CO₂ concentration in a background matrix of different O₂ concentrations (20 and 80% respectively). A loss of IR signal is clearly visible.

There are several possible causes for non-linearity or signal loss (not caused by concentration changes) in IR absorption spectra. On one side, there are effects caused by the FTIR instrument (detector and source non-linearity, Fourier transform/apodization, instrument setup, etc.) [5–18], while on the other, physical effects like pressure broadening, collision broadening or a phenomenon called optical collision effect are possible [19–45]. The effect of oxygen or other background gases on an IR signal have been described in literature [19–45], while Griffiths et al. describe the possible influences of the FTIR instrument and instrumentation [46].

However, it is not the aim of this work to explore the exact cause of either the non-linearity or the signal loss. Since physical system modifications to minimize these effects are not possible and we refrain from physically motivated calibration procedures, it is sufficient for the scope of this work that effects like line broadening or that of apodization were sufficiently reproducible from one calibration day to next to allow a predictable effect of oxygen concentration and dose response relation for carbon dioxide when the system configuration or the measurement procedure is left unchanged. We therefore aim to develop a data analysis procedure to remove the influence of background signal changes suitable for routine work and regardless of underlying cause.

For an application of a FTIR spectrometer in the monitoring of carbon dioxide in exhaled breath, it is desired that both effects are handled and corrected simultaneously. The process of calibration transfer has been generally described as the transfer of an extensive calibration recorded on one instrument (primary, master or parent instrument) to a second instrument (secondary, slave or child instrument) using less samples than for a full calibration via chemometric methods. Further information on calibration strategies can be found in the reviews of Feudale [47] and de Noord [48]. Published methods usually consider the transfer between two physically different instruments, but the concept can be translated to the calibration transfer of one fluctuating instrument over time (i.e., two different temporal states).

A calibration transfer pertaining to different states of the same machine can easier than a transfer between instruments because detectors or measurement principles etc. stay the same. The following factors however, can complicate a transfer between different states:

- a small, but yet distinct spectrometer drift between days
- during recording: occurrence of an IR-inactive interferent (change in background gas matrix, i.e. oxygen concentration) whose concentration is not known beforehand and varies over time due to outside influences (breathing pattern, general status, etc.) and nonlinear interference on IR-signal
- no clearly defined “master” system without interference exists: there always occurs a small influence on IR signal
- non-linear relationship between spectrometer response and prediction value (non-linear absorption range)
- simultaneous calibration of two different systems (FTIR and oxygen sensors) with the same reference samples at the same time

While many calibration transfer methods exist for removing spectrometer fluctuations and other signal perturbations, to the best of our knowledge, no transfer method exists that is capable of establishing a calibration using only a few samples to correct most of the challenges listed above. Moreover, to the best of our knowledge, no effort has been made to develop a procedure that handles all these issues and leads to a fast and streamlined data analysis and calibration system whose maintenance can be performed by non-expert personnel.

Earlier, we developed a novel calibration strategy with a Bayesian calibration transfer algorithm using Lagrange Multipliers and a hierarchical model that is implemented within a multi-platform software package using Monte-Carlo Markov-Chain sampling for the calibration transfer of a commercial oxygen sensor with a non-linear sensor response [49]. We adapt and extend this
approach for the requirements of the FTIR system. We present the correction of the "oxygen effect" on the carbon dioxide IR signal (and possibly all other IR signals) and simultaneous correction of system drifts using a PCA scores based response surface fit and a calibration transfer approach based on hierarchical models and Lagrange Multipliers.

2. Material and methods

2.1. Experimental procedure

The experimental design for calibrating the IR sensor consisted of a randomized set of 25 gas mixtures spanning the biologically expected concentration of 1–5% CO\textsubscript{2} and 20–80% oxygen. The combination and run order of the samples was built using a full factorial design of experiment, DOE, (PLS Toolbox 7.9.3, Experiment Designer, Eigenvector Research, Inc., Manson/WA, USA). One mixture (1% CO\textsubscript{2}/20% oxygen) was repeated 5 times interspersed in the set to monitor changes in the system and mixing pump. The set was measured on 14 days, resulting in 2700 data sets. The FTIR spectra were collected from 4000 to 400 cm\textsuperscript{-1} (averaging 20 background scans (nitrogen) and 20 sample scans, apodization function Blackmann-Harris 3-Term, 5 to 10 repeats per mixture). Data evaluation was done using MATLAB 8.6 (R2015b, The Mathworks Inc, Naticks/MA, USA), PLS Toolbox 7.9.3[Eigenvector Research, Inc., Manson/WA, USA] and MatlabStan 2.7.0.0 [50]/Stan 2.9.0 [51]. All gas calibration samples were manufactured from pure technical grade nitrogen, pure technical grade carbon dioxide and pure medical grade oxygen (all MTI Industriegase, Neu-Ulm, Germany) using a gas mixing pump (DIGAMIX 2M 301, H. Wösthoff Messtechnik GmbH, Bochum, Germany).

2.2. Optical setup

The system to be calibrated is based on a Bruker ALPHA FTIR Spectrometer with a custom made gas cell and modified sampling chamber (Bruker Optik GmbH, Ettlingen, Germany) [52,53]. A 7.5 cm straight channel substrate-integrated hollow waveguide (iHWG) serves gas cell as well as wave guide [54]. Radiation from the spectrometer is focused into the iHWG and out into the internal DTGS detector of the ALPHA via two 1° gold-coated off-axis parabolic mirrors (OAPM) with an effective focal length of 2”[Janos Technology Inc, Keene/NH, USA]. The entire sample chamber is flushed with nitrogen during measurements. A FireStingO2 oxygen flow cell sensor (Pyro Science GmbH, Aachen, Germany) is integrated into the gas outlet. Between measurements and during background calibration, the gas cell was flushed with nitrogen. Fig. 2 shows the experimental setup.

2.3. Data analysis procedure

CO\textsubscript{2} has three viable spectral regions in the mid-infrared (MIR), as shown in Fig. 3, and we focus on the most intensive, the asymmetric stretch band \(\nu_2\) in region I. Due to the configuration of our setup, the \(\nu_3\) band (region II) is half cut off by the barium fluoride windows of the gas cell, while the overtone bands around 2700 cm\textsuperscript{-1} (region III) show high signal-to-noise ratios due to low detector sensitivity and low light source intensity.

The first step of the procedure after preprocessing of the IR spectra (baseline correction using a weighted least squares algorithm and optional mean centering) is data reduction of the multivariate signal of the IR spectrometer via Principal Component Analysis (PCA). Decomposition shows that nearly 99% of the spectral variance can be explained using just one principle component (PC) with negligible residuals. Fig. 4 plots the PC scores against CO\textsubscript{2} and O\textsubscript{2} concentration for different calibration runs. The response surface is almost planar and reproducible for different calibration days. Earlier research [55] has shown that it is possible to describe and correct a reproducible influence on the quantification of an analyte signal via a response.
surface. Accordingly, to capture the first order approximated planar response surface and potential minor deflections from it, we use a 25 sample calibration protocol as depicted in Fig. 4 and try to describe the measured surface with a polynomial in the variables for the CO₂ and O₂ concentrations. The polynomial may contain product terms like [CO₂][O₂] to cover potential nonlinearities. Fitting the measured score values to such a polynomial yields an equation that can be restructured to express the unknown CO₂ concentration as a function of the measured score and the interfering O₂ concentration. Fitting to the measurements of a single calibration day or run, the coefficient of the resulting polynomial should be adapted to the current state of the equipment and thereby, the inter-day variability of the coefficients should reflect the inter-day variability of the measurement system. This approach faces two major challenges: The effort for 25 sample measurements is too large for a routine calibration and the optimal structure of the fitting polynomial is unknown.

To minimize the number of measurements needed for a single calibration we focus on the coefficients that define the shape of the response surface. We assume that the day-to-day variability of the measurement instrument be can be described by a multivariate normal day-to-day distribution of the coefficients. In a first step, this variability can be assessed using Bayesian hierarchical models. The key concept of our approach is that we treat the coefficients of a new, to be defined, calibration as a sample of this day-to-day distribution. With this constraint, only a few actual calibration samples are necessary for an exact determination of the calibration coefficients. The corresponding step, the second of our calibration transfer, uses a Lagrange Multiplier approach. With the additional information from previous calibrations, only two defined calibration samples are sufficient to obtain a calibration accuracy that is comparable to a full calibration. For validation, a third sample can be included to the calibration. The full methodology using the example of a calibration of a fiber-optic luminescence based oxygen sensor has already been described [49]. We adapted the procedure to calculate the coefficients of the polynomial response surface fit using two defined calibration samples. The full mathematical derivation and how we treat measurement errors in the CO₂
calibration are presented in the Appendix.

Several fitting approaches were explored:

a) single fit: an individual fit to determine the coefficients from 25 actual calibration samples;
b) hierarchical fit/HM: using 25 actual calibration samples under the constraint that the actual calibration coefficients are a sample of the multivariate day-to-day distribution of the calibration coefficients.
c) Lagrange fit/LG: determination of actual calibration coefficients from two actual calibration samples with the same restraints for the actual coefficients as for b).

We avoid any a-priori assumption for the selection of the optimal polynomial structure except that only polynomial terms up to the second order like \([\text{CO}_2][\text{O}_2]\) or \([\text{O}_2][\text{O}_2]\) will be used. We test all feasible combination of polynomial elements (compare equations (1)–(7)) and will use the variant that both gives a good predictive precision when used with the calibration transfer and that requires a minimal number of polynomial elements.

Predictive precision is assessed with several parameters: RMSEP, which covers the (averaged) prediction error for the 25 calibration samples of the single calibration run, \(\text{RMSEP}\), which pertains to the average RMSEP for 14 different calibration runs (averaging over all calibration samples and runs), and \(\Sigma \Delta \text{RMSEP}\), which describes the difference between \(\text{RMSEP}\) for a full hierarchical fit and \(\text{RMSEP}\) calculated from two calibration samples and the LG approach - it measures the average loss in prediction quality when using 2 instead of 25 actual calibration samples.

The Watanabe-Akaike information criterion (WAIC) [56] can provide some information about the optimal polynomial structure of the calibration equation. It has been widely implemented in Bayesian Statistics and Stan models [57–60]. Its test value decreases with the goodness of fit and increases with the number of parameters used for fitting. In the present case, the number of parameters increases with the number of polynomial elements. In general, a lower WAIC is preferred. The definition of WAIC can be found in section Appendix A in the Appendix.

3. Results

One of the first exploratory steps when dealing with multivariate data is data reduction via Principal Component Analysis (PCA). As mentioned above, decomposition via PCA shows that a single PC captures over 99% of the spectral variance and the resulting spectral residuals are negligible, reflecting only spectral noise. This implies that the shape of the CO\(_2\)-FTIR spectrum, as collected under the present conditions, is not affected by changes in the O\(_2\) and CO\(_2\) concentrations in the sample producing the spectrum. Only the height or intensity of the spectrum is altered and this intensity change is reflected in changes of the PCA score value. Therefore, one should be able to establish a mathematical or functional relationship between observed score values and O\(_2\) and CO\(_2\) concentration values.

3.1. Nonlinear effects of O\(_2\) and CO\(_2\) on the IR-signal

Fig. 4 implies that the PC score of a spectrum is proportional to the underlying CO\(_2\) concentration and affected by the oxygen content in the sample. The oxygen effect is demonstrated in Fig. 5. Here, the scores of spectra with exactly the same CO\(_2\) concentration but oxygen concentrations ranging from 20% to 80% show an exemplary signal loss with increasing oxygen concentration. To assess how this signal loss, when ignored, affects predictions for the CO\(_2\) concentrations, we developed a quadratic prediction equation from score values obtained with different CO\(_2\) concentration values at fixed 20% oxygen. We then used this prediction equation to assess CO\(_2\) concentration values from the scores shown in Fig. 5. The resulting CO\(_2\) predictions demonstrate a sensitivity against oxygen, as indicated in Fig. 5. This exemplary calculation for one carbon dioxide concentration and one exemplary calibration set shows that the influence of oxygen on the quantitative determination of carbon dioxide cannot be ignored and must be accounted for in the calibration process. Details about the predictive approach used here can be found in the supplementary material together with further values for the oxygen effect.

3.2. Response surface calibration

To describe the response surface indicated in Fig. 7 several possible fit equation were tested.

Fig. 5. Influence of oxygen concentration on the CO\(_2\) quantification. Empty black squares: averaged score of PC1 of 4% carbon dioxide (The real carbon dioxide concentration set at the mixing pump is 3.89% for all samples.) with different oxygen concentrations fo one example day. Filled blue triangles: predicted carbon dioxide concentration at different oxygen levels assuming an oxygen concentration of 20L. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
\[ S = c_1 + c_2[CO_2] + c_3[CO_2]^2 + c_4[O_2] + c_5[O_2]^2 + c_6[O_2][CO_2] \]

(1)

\[ S = c_1 + c_2[CO_2] + c_3[CO_2]^2 + c_4[O_2] + c_5[O_2]^2 \]

(2)

\[ S = c_1 + c_2[CO_2] + c_3[CO_2]^2 + c_4[O_2] + c_5[O_2]^2 \]

(3)

\[ S = c_1 + c_2[CO_2] + c_3[CO_2]^2 + c_4[O_2] + c_5[O_2]^2 \]

(4)

\[ S = c_1 + c_2[CO_2] + c_3[CO_2]^2 + c_4[O_2][CO_2] \]

(5)

\[ S = c_1 + c_2[CO_2] + c_3[CO_2]^2 + c_4[O_2] \]

(6)

\[ S = c_1 + c_2[CO_2] + c_3[CO_2]^2 + c_4[O_2]^2 \]

(7)

S is the score of PC1 and c₁ to c₆ are the calibration coefficients.

3.3. Comparison of model performance

To explore whether in routine work the calibration transfer can replace a full calibration approach we compared average RMSEP of the standard full calibration fit (henceforth called single fit) with a fit using a hierarchical model (hierarchical fit/HM) and with a fit obtained with our calibration transfer approach (Lagrange/LG fit).

As can be seen in Table 3, the average values of the Lagrange calibration transfer fit are quite close to those of the single and hierarchical fit, meaning that the calibration performance of the calibration transfer approach is sufficient for routine work. It is therefore possible to reduce the calibration effort during daily calibration without losing calibration power.

3.4. Determination of optimum calibration model and set points

Several parameters affect the performance of the calibration transfer using the Lagrange Multiplier approach and should be optimized. The first one is the number and composition of calibration samples (calibration set points) that are later used to build the calibration. Another possibility of improving results are different preprocessing methods performed in advance to PCA. A third possibility is the choice of spectral region to analyze, which may pertain to the selection of absorption bands used or to the removal of spectral regions with instability due to spectral saturation. The quality of fit for an optimization trial is assessed with the averaged RMSEP over all days and calibration points which reflects the general calibration accuracy of the model and with \( \Sigma \Delta \text{RMSEP} \), which reflects the loss in precision induced by the reduced set of actual calibration points.

3.4.1. Optimal set points

An important influence on the calibration performance of the calibration transfer is the choice of calibration set points. Due to the time constraints of the routine application the calibration is intended for, no more than two set point/gas samples are practically possible. Earlier research has shown that one set point is usually not sufficient in terms of error performance and calibration efficiency is much improved using two set points, which is also correct in our case. In order to determine the model with the optimum calibration performance, all combinations of the 25 calibration points using the Lagrange calibration transfer algorithm with two set points were tested. The results are shown in Table D in the Appendix. Fig. 8 plots the change of the prediction error against the average prediction error for the different combinations.

There is a cluster of combinations that show nearly identical calibration performance and no clear optimum combination exists. From this cluster, set point 2 (1% CO₂, 30% O₂) and set point 21 (5% CO₂, 20% O₂) were intuitively chosen for all later calculations because they span the concentration region that will most often occur during the routine work.

3.4.2. Preprocessing and spectral window

Preprocessing methods are one of the most common steps before starting multivariate data analysis for further improving results. Since the IR spectra recorded are already very clean with very low noise levels and low background effects, only a baseline correction using a weighted least squares algorithm and mean centering were tested. Another possibility for stabilizing and improving performance is the choice of the spectral window. As indicated in Fig. 1 both ¹²C0₂ and ¹³C0₂ contribute the observed CO₂ spectrum with a slight overlap. It is possible that later on in routine work samples with an enriched ¹³C0₂ concentration might...
be analysed. The enriched $^{13}$CO$_2$ content may change the shape of CO$_2$ spectrum. Hence, two different models were prepared: The first includes the entire $v_2$ region while the second considers only the spectral range where no $^{13}$CO$_2$ bands overlap.

A possible complication with CO$_2$ spectra generated in the concentration range of 1%–5% as in the present setup is that for high concentration samples the signal for lines becomes unstable because they are close to saturation caused by a nearly total absorption, particular in the range of the peak maximum. We therefore tested whether removing the instable area improved performance.

The hierarchical model is derived from 25 times 14 measurements of calibration samples. For such a large number of measurements, the recording of a few specific samples or entire calibration day might be faulty due to undetected problems or operator error. We found one measurement day which could be considered a clear outlier because one of the two calibration set point measurements necessary for the calibration transfer was faulty. While the set was used for general building of the hierarchical model, it was removed from all latter evaluations since no meaningful calibration transfer is possible with a missing calibration set point.

The following twelve possible models for preprocessing methods and spectral windows, shown in Table 4, were investigated.

The best results could be reached by combined baseline correction and mean centering while excluding the largest peak tip. Comparing the best results for the $v_3$ peak to the overtone region, though the former is closer to saturation, still better results than for region III could be reached. In the presence of $^{13}$CO$_2$, the optimum is achieved using no preprocessing.

### 3.5. Model validation

A random subset - based cross validation is applied to ensure that the preprocessing/spectral window combination selected in section 3.4.2 remains the best choice to calibrate data sets that were not used for model building. This was done by randomly removing

---

**Table 1**

Comparison of possible combinations for surface fit equation. The row in bold face font indicates the best choice for the fit equation, which was applied to subsequent calculations.

<table>
<thead>
<tr>
<th>Equation Nr.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSEP</td>
<td>0.029</td>
<td>0.024</td>
<td>0.102</td>
<td>0.023</td>
<td><strong>0.022</strong></td>
<td>0.104</td>
<td>0.074</td>
</tr>
</tbody>
</table>

**Table 2**

Average regression coefficients of response surface fit with standard deviation and 95% confidence limits.

<table>
<thead>
<tr>
<th></th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$c_3$</th>
<th>$c_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-5.300</td>
<td>2.335</td>
<td>-0.060</td>
<td>-0.00204</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.004</td>
<td>0.009</td>
<td>0.0011</td>
<td>0.00003</td>
</tr>
<tr>
<td>5% quantile</td>
<td>-5.306</td>
<td>2.320</td>
<td>-0.0639</td>
<td>-0.00209</td>
</tr>
<tr>
<td>95% quantile</td>
<td>-5.294</td>
<td>2.350</td>
<td>-0.0561</td>
<td>-0.00199</td>
</tr>
</tbody>
</table>

**Table 3**

Comparison of achieved mean RMSEP of different calibration approaches.

<table>
<thead>
<tr>
<th>Calibration approach</th>
<th>RMSEP</th>
<th>Std.dev.</th>
<th>5% quantile</th>
<th>95% quantile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single fit</td>
<td>0.022</td>
<td>0.002</td>
<td>0.019</td>
<td>0.026</td>
</tr>
<tr>
<td>HM fit</td>
<td>0.022</td>
<td>0.002</td>
<td>0.019</td>
<td>0.026</td>
</tr>
<tr>
<td>LG fit</td>
<td>0.027</td>
<td>0.009</td>
<td>0.016</td>
<td>0.044</td>
</tr>
</tbody>
</table>
15% (in our case 2 of the 14) of the sets in the calibration pool and designating them as validation sets. The remaining sets were used to build the model as shown above and the resulting model was used to predict the validation set data. The whole procedure was repeated 20 times for each possible model. Table 5 shows the corresponding results.

In the present case, the full calibration pool contains, with limited redundancy, all variations necessary for building a stable hierarchical model. During cross validation, two data sets were removed from the full pool which sometimes contained data information that was critical for model building. In consequence the iterative algorithm failed to converge to an optimal solution as judged by the number of iteration cycles and very poor prediction errors. These runs were not used for evaluation. Yet, for later routine work, the full calibration pool will be used, where no such instabilities ever occurred, and these happenings can be safely ignored. Table 4 shows that the difference of the error parameters between known calibration and "unknown" validation set is negligible. Again, Model M6 generated the best results.

3.6. Validation of sensor drift

Two types of drift in an analytical system need to be corrected via calibration methods: the first is the inter-day drift of a sensor system. This drift is supposed to be corrected by a calibration, in particular with our calibration transfer model using the inter-day covariance matrix. The second is a sensor drift during the calibration day. Since in most cases, it can only be corrected by several recalibrations during the day, which is a distinct disadvantage during routine work and often not possible, intra-day variance is to be avoided as far as possible with practical modification of the system in question. To explore the intra-day variance of FTIR system as well as the variability of gas mixing pump system, quality measurements of control samples with a fixed sample composition (1% CO2 in 20% O2 and nitrogen) were included in the measurements of each calibration set. The control samples were repeated after measurements of every fifth calibration set point. Changes in their values should reflect the drift of the entire system during the day in the calibration transfer model. In order to show the general intra-day variance of the system, they were completely removed from the calibration sets and predicted after model building. Table 6 shows the average CO2 concentration as well as standard deviation of the five quality control samples for all 14 days, predicted representatively by the hierarchical model, and their overall values. The results indicate that no significant drift takes place during the day and between days and most fluctuation are within the error margin of the system and evaluation method.

4. Conclusion and outlook

Summing up, the approach combines three different evaluation strategies: a) extraction scores or intensities from multivariate spectra by a principle components analysis, b) calibration based on a nonlinear relation or response surface between scores and gas concentrations, and c) evaluation of statistical properties of the coefficients of the response surfaces as a basis for a calibration transfer. We established a response surface that reflects a minimal representation of O2 and CO2 interactions that affect FTIR spectra. The coefficients quantifying the interactions can be determined with high precision. The minimal representation of the nonlinear interaction allows incorporating the response surface into routine work and implementation of a fast daily calibration routine.

Table 4
Comparison of possible preprocessing methods and spectral regions as well as analytical performance. The row in bold face font indicates the best choice, which was applied to subsequent calculations.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Preprocessing method</th>
<th>Spectral region [cm⁻¹]</th>
<th>RMSEP</th>
<th>ΣΔRMSEP</th>
<th>ΣΔRMSEP²</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>none</td>
<td>2100–2500</td>
<td>0.0427</td>
<td>14.77</td>
<td>13.94</td>
</tr>
<tr>
<td>M2</td>
<td>baseline</td>
<td>2100–2500</td>
<td>0.0416</td>
<td>13.29</td>
<td>12.88</td>
</tr>
<tr>
<td>M3</td>
<td>baseline + mean centering</td>
<td>2100–2500</td>
<td>0.0424</td>
<td>13.33</td>
<td>12.86</td>
</tr>
<tr>
<td>M4</td>
<td>none</td>
<td>2100–2500, 2356–2366 excluded</td>
<td>0.0278</td>
<td>14.95</td>
<td>13.80</td>
</tr>
<tr>
<td>M5</td>
<td>baseline</td>
<td>2100–2500, 2356–2366 excluded</td>
<td>0.0264</td>
<td>12.86</td>
<td>12.55</td>
</tr>
<tr>
<td>M6</td>
<td>baseline + mean centering</td>
<td>2100–2500, 2349–2366 excluded</td>
<td>0.0263</td>
<td>12.68</td>
<td>12.52</td>
</tr>
<tr>
<td>M7</td>
<td>none</td>
<td>2349–2400</td>
<td>0.0661</td>
<td>13.90</td>
<td>13.50</td>
</tr>
<tr>
<td>M8</td>
<td>baseline</td>
<td>2349–2400</td>
<td>0.0720</td>
<td>17.20</td>
<td>16.04</td>
</tr>
<tr>
<td>M9</td>
<td>baseline + mean centering</td>
<td>2349–2400</td>
<td>0.0721</td>
<td>17.00</td>
<td>15.61</td>
</tr>
<tr>
<td>M10</td>
<td>none</td>
<td>2349–2400, 2356–2366 excluded</td>
<td>0.0420</td>
<td>14.97</td>
<td>14.32</td>
</tr>
<tr>
<td>M11</td>
<td>baseline</td>
<td>2349–2400, 2356–2366 excluded</td>
<td>0.0471</td>
<td>16.93</td>
<td>15.90</td>
</tr>
<tr>
<td>M12</td>
<td>baseline + mean centering</td>
<td>2349–2400, 2356–2366 excluded</td>
<td>0.0474</td>
<td>16.83</td>
<td>15.81</td>
</tr>
<tr>
<td>M13</td>
<td>baseline + mean centering</td>
<td>3438–3800</td>
<td>0.0546</td>
<td>15.90</td>
<td>14.70</td>
</tr>
</tbody>
</table>

Table 5
Results of random subset cross validation: overall RMSEP, ΣΔRMSEP and ΣΔRMSEP² for calibration and validation sets for 20 repeats - average ± standard deviation.

<table>
<thead>
<tr>
<th>Model name</th>
<th>Calibration</th>
<th>Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSEP</td>
<td>ΣΔRMSEP</td>
</tr>
<tr>
<td>M1</td>
<td>0.0467 (± 0.0033)</td>
<td>13.73 (± 2.09)</td>
</tr>
<tr>
<td>M2</td>
<td>0.0432 (± 0.0027)</td>
<td>12.16 (± 1.21)</td>
</tr>
<tr>
<td>M3</td>
<td>0.0515 (± 0.0023)</td>
<td>14.98 (± 1.56)</td>
</tr>
<tr>
<td>M4</td>
<td>0.0493 (± 0.0063)</td>
<td>14.34 (± 1.40)</td>
</tr>
<tr>
<td>M5</td>
<td>0.0454 (± 0.0040)</td>
<td>12.96 (± 1.36)</td>
</tr>
<tr>
<td>M6</td>
<td>0.0455 (± 0.0075)</td>
<td>13.09 (± 1.58)</td>
</tr>
<tr>
<td>M7</td>
<td>0.0311 (± 0.0204)</td>
<td>13.97 (± 2.35)</td>
</tr>
<tr>
<td>M8</td>
<td>0.0352 (± 0.0130)</td>
<td>13.21 (± 1.74)</td>
</tr>
<tr>
<td>M9</td>
<td>0.0379 (± 0.0167)</td>
<td>13.21 (± 1.25)</td>
</tr>
<tr>
<td>M10</td>
<td>0.0694 (± 0.0030)</td>
<td>13.63 (± 1.78)</td>
</tr>
<tr>
<td>M11</td>
<td>0.0717 (± 0.0068)</td>
<td>14.18 (± 1.16)</td>
</tr>
<tr>
<td>M12</td>
<td>0.0715 (± 0.0077)</td>
<td>14.26 (± 1.58)</td>
</tr>
</tbody>
</table>
applied a Bayesian calibration transfer algorithm using Lagrange Multipliers and a hierarchical model that is implemented within a multi-platform software package. Optimum model conditions in terms of choice of fit equations, calibration set points, spectral region and preprocessing steps were determined and the resulting models validated via random-subset cross validation. While the process was optimized for the requirement of the later routine application in question (carbon dioxide and oxygen quantification in exhaled mouse breath), it can be adapted to other possible FTIR analytes and gas background matrices that showed similar effects. The routine calibration procedure and data analysis algorithm will be implemented in an on-line combined iHWG FTIR system for the simultaneous quantification of carbon dioxide and oxygen in exhaled breath of a ventilated mouse and be operated during routine studies at the Mouse Intensive Care Unit (MICU) of the Institute of Anesthesiologic Pathophysiology and Method Development at Ulm University.

Acknowledgements

The authors would like to thank the Federal State of Baden-Württemberg (Landesgraduiertenförderungsgesetz, grant number: 1212-LGF-E) for partial financial support. This work was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract DE-AC52-07NA27344. This project was partially funded under LLNL by Lawrence Livermore National Laboratory (LLNL) under Contract formed in part under the auspices of the U.S. Department of Energy.

Appendix A. Mathematical definitions

The root mean square error of prediction (RMSEP) is defined as following:

$$\text{RMSEP}_j = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (y_{ji} - x_j)^2}$$  \hspace{1cm} (A.1)

where $j$ pertains to one of 25 different calibration samples with $x_j$ as real concentration, $y_{ji}$ are predictions pertaining to this sample, as derived from the calibration function using $i$ ($i = 1$ to $m$) different score values. The quality criteria $\Delta \text{RMSEP}$ is calculated as following:

$$\Delta \text{RMSEP}_j = \frac{\text{RMSEP}_{j,G} - \text{RMSEP}_{j,HM}}{\text{RMSEP}_{j,HM}} \hspace{1cm} (A.2)$$

$$\Sigma \Delta \text{RMSEP} = \sum_{j} (\text{abs}(\Delta \text{RMSEP}_j)) \hspace{1cm} (A.3)$$

First a relative error variable $\Delta \text{RMSEP}_j$ is defined for each level $j$ using the RMSEP of the Lagrange fit ($\text{RMSEP}_{j,G}$) and the RMSEP of the hierarchical fit ($\text{RMSEP}_{j,HM}$). Then $\Delta \text{RMSEP}_j$ of one Lagrange fit is averaged over all levels, $n$ being the number of levels.

$$\Sigma \Delta \text{RMSEP}_2 = \frac{\text{RMSEP}_{j,G} - \text{RMSEP}_{j,HM}}{\text{RMSEP}_{j,HM}} \hspace{1cm} (A.4)$$

$$\Sigma \Delta \text{RMSEP}_2 = \sum_{j} (\text{abs}(\Delta \text{RMSEP}_j)) \hspace{1cm} (A.5)$$

The Watanabe-Akaike information criterion (WAIC) is based on the following equation[57]

$$e_{\text{d}p_{\text{WAIC}}} = ld_{\text{d}p} - p_{\text{WAIC}} \hspace{1cm} (A.6)$$

Further explanation on the different terms of the WAIC can be found in Refs. [56-60] The expected log pointwise predictive density for a new dataset ($d_{\text{d}p}$) is calculated here as

$$ld_{\text{d}p} = \sum_{i=1}^{n} \log \left( \frac{1}{S} \sum_{s=1}^{S} p(y_{i}, \theta_{s}) \right) \hspace{1cm} (A.7)$$

considering data $y_1, y_2, \ldots$ independently modeled using parameters $\theta$ and by sampling the posterior probability $p(\theta)$ from $\theta_s$ with $s = 1 \ldots S$. The estimated effective number of parameters $p_{\text{WAIC}}$ is defined as

$$p_{\text{WAIC}} = \sum_{i=1}^{n} \frac{1}{S} \left( \log p(y_{i}, \theta_{s}) \right) \hspace{1cm} (A.8)$$
where $V^S_{-1}$ represents the sample variance, $V^S_{-1}a_0 = \frac{1}{\lambda} (a_0 - \overline{y})^2$.

In order to make WAIC comparable to other information criteria like the Akaike information criterion (AIC) or Bayesian information criterion (BIC), WAIC is usually presented on the deviance scale:

$$\text{WAIC} = -2\ell_{dp_{WAIC}}$$ (A.9)

### Appendix B. Mathematical derivation of calibration transfer method

For the sake of simplicity, we assume that only one measurement with the index $t$ is made to characterize the actual calibration. The independent variables of the calibration (i.e., the concentration of carbon dioxide) shall be called $x_t$, the known variable $y_t$ the oxygen concentration, while the dependent variable/PCA score shall be called $S_t$. We denote the set of actual calibration coefficients with $\mathbf{c}_t$. This vector $\mathbf{c}_t$ should satisfy equation (B.1) for given independent variable $x_t$, known variable $y_t$ and for a measured response $S_t$. The key requirement is that this $\mathbf{c}_t$ should be a sample of a normal distribution around the mean coefficients $\overline{T}$ and inter-day variance $\Sigma$. The $\mathbf{c}_t$ values that satisfy both criteria could be determined using a Lagrange Multiplier approach. Here, the coefficients have to satisfy an equation that is linear in the unknown coefficients. The underlying calibration equation is the following

$$S_t = \mathbf{c}_t^T \overline{T} + \varepsilon_t$$ (B.1)

The left side of the last row of equation (B.1) corresponds to our response variable, which is noted in the following as $R_t = S_t$. For the response variable $R_t$, one obtains the following linear equations:

$$R = \mathbf{B}^T \mathbf{c}_t; \quad \text{with } \mathbf{B} = \begin{pmatrix} 1, x_t, x_tA^2, x_t y_t \end{pmatrix}$$ (B.2)

From the requirement that $\mathbf{c}_t$ should also be as close as possible to the mean coefficient set $\overline{T}$, one can define the Lagrange-Norm

$$Q = \frac{1}{2} (c_t - \overline{T})^T \Sigma^{-1} (c_t - \overline{T}) + \lambda \mathbf{B}^T \mathbf{c}_t - R$$ (B.3)

Here, the first part pertains to the assumption that $\mathbf{c}_t$ is a sample of a normal distribution around $\overline{\mathbf{c}}$ and $\Sigma$, and the second part reflect the restraint that $\phi_t$ lies on the calibration curve. The Lagrange-Norm (B.3) is then minimized against $\mathbf{c}_t$ and $\lambda$. Minimizing by $\lambda$ gives

$$\frac{\delta Q}{\delta \lambda} = 0 , \quad \text{which gives } \mathbf{B}^T \mathbf{c}_t = R$$ (B.4)

Minimizing (B.3) by $\mathbf{c}_t$ results in

$$\frac{\delta Q}{\delta \mathbf{c}_t} = 0 , \quad \text{which yields } \Sigma^{-1} (c_t - \overline{T}) + \lambda \mathbf{B} = 0$$ or: $c_t = \overline{T} - \lambda \mathbf{B}$

(B.5)

Insertion in (B.4) results in

$$\mathbf{B}^T (\overline{T} - \lambda \mathbf{B}) = R$$

$$\mathbf{B}^T \Sigma B = R - \mathbf{B}^T \overline{T}$$

$$\lambda = \left( \mathbf{B}^T \Sigma B \right)^{-1} \mathbf{B}^T (\overline{T} - \mathbf{B}^T \Sigma B)^{-1} R$$ (B.6)

This yields the factor $\lambda$, which can be inserted into Equation (B.5) to obtain the desired $\mathbf{c}_t$.

### Appendix C. Error propagation in the calibration fit

The error in the carbon dioxide quantification is covered by the sampling statement of the Stan statistical modeling language

$$S_t \sim \text{normal} (\overline{S}, \sigma_S)$$ (C.1)

It indicates that the measured score value $S_t$ is sampled from a normal distribution with the predicted score $\overline{S}$ as mean value and $\sigma_S$ as standard deviation and simulates the random fluctuations of the instrument. For the determination of the hierarchical model multiple runs were analysed such that $\sigma_S$ can be determined as a function of the intensity of the score values. It reflects the increase of measurement errors with scores values and thereby with CO2 concentration values. The following formula is used:

$$\sigma_S = \frac{\sigma_{e1}}{\text{signal}_{max} - S}$$ (C.2)

where $\sigma_{e1}$ and $\text{signal}_{max}$ are determined as unknown parameters in the model and $\overline{S}$ is the current predicted score value. For the determination of $\text{signal}_{max}$ a normally distributed prior was used, that was estimated from replicate measurements of score values of calibration samples. This $\sigma_S$ pertains to measurement errors that are averaged over all calibration days. To assess the predictive power of the calibration transfer/lagrange approach score measurements of two samples are used to define a calibration curve. The predictive power depends on their measurement error. Actual values for $\sigma_S$ are estimated from 5 replicate score measurements. For further information how to implement errors in the Stan code see Ref. [49].

The error in the carbon dioxide quantification is modeled using two parts. The first is defined in the sampling statement of the Stan statistical modeling language

$$S_t \sim \text{normal} (\overline{S}, \sigma_S)$$ (C.3)

It indicates that the measured score value $S_t$ is sampled from a normal distribution with the predicted score $\overline{S}$ as mean value and $\sigma_S$ as standard deviation. This statement simulates the random fluctuation of the instrument and therefore the measurement error for a sample on the day to be calibrated.

The second part pertains to an estimate for $\sigma_S$. It approximates the fact that the measurement error increases with scores values and thereby with CO2 concentrations values, or:

$$\sigma_S = \frac{\sigma_{e1}}{\text{signal}_{max} - S}$$ (C.4)

where $\sigma_{e1}$ and $\text{signal}_{max}$ are determined as unknown parameters in the model and $\overline{S}$ is the current predicted score value. For the determination of $\text{signal}_{max}$ a normally distributed prior was used, that was estimated from replicate measurements of score values of calibration samples. For further information how to implement errors in the Stan code see Ref. [49].

### Appendix D. Supplementary table: Comparison of possible set points
Table D7

<table>
<thead>
<tr>
<th>Set Number</th>
<th>CO2/(%)</th>
<th>O2/(%)</th>
<th>SD</th>
<th>RMSEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>20</td>
<td>1</td>
<td>0.15</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>30</td>
<td>2</td>
<td>0.26</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>40</td>
<td>3</td>
<td>0.37</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>60</td>
<td>4</td>
<td>0.51</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>80</td>
<td>5</td>
<td>0.69</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>20</td>
<td>6</td>
<td>0.12</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>30</td>
<td>7</td>
<td>0.22</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>40</td>
<td>8</td>
<td>0.32</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>60</td>
<td>9</td>
<td>0.42</td>
</tr>
<tr>
<td>10</td>
<td>2</td>
<td>80</td>
<td>10</td>
<td>0.52</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
<td>20</td>
<td>11</td>
<td>0.13</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>30</td>
<td>12</td>
<td>0.23</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>40</td>
<td>13</td>
<td>0.33</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>60</td>
<td>14</td>
<td>0.43</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
<td>80</td>
<td>15</td>
<td>0.53</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>20</td>
<td>16</td>
<td>0.14</td>
</tr>
<tr>
<td>17</td>
<td>4</td>
<td>30</td>
<td>17</td>
<td>0.24</td>
</tr>
<tr>
<td>18</td>
<td>4</td>
<td>40</td>
<td>18</td>
<td>0.34</td>
</tr>
<tr>
<td>19</td>
<td>4</td>
<td>60</td>
<td>19</td>
<td>0.44</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>80</td>
<td>20</td>
<td>0.54</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>20</td>
<td>21</td>
<td>0.15</td>
</tr>
<tr>
<td>22</td>
<td>5</td>
<td>30</td>
<td>22</td>
<td>0.25</td>
</tr>
<tr>
<td>23</td>
<td>5</td>
<td>40</td>
<td>23</td>
<td>0.35</td>
</tr>
<tr>
<td>24</td>
<td>5</td>
<td>60</td>
<td>24</td>
<td>0.45</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>80</td>
<td>25</td>
<td>0.55</td>
</tr>
</tbody>
</table>

References

5.4 Paper III: Online monitoring of carbon dioxide and oxygen in exhaled mouse breath via substrate-integrated hollow waveguide - Fourier transform infrared - luminescence spectroscopy
Online monitoring of carbon dioxide and oxygen in exhaled mouse breath via substrate-integrated hollow waveguide Fourier-transform infrared-luminescence spectroscopy

Felicia Seichter¹, Erhan Tütüncü¹, Leila Tamina Hagemann¹, Josef Vogt², Ulrich Wachter³, Michael Gröger³, Sandra Kress⁴, Peter Radermacher⁵ and Boris Mizaikoff⁶

¹ Institute of Analytical and Bioanalytical Chemistry, Ulm University, D-89081 Ulm, Germany
² Institute of Anesthesiologic Pathophysiology and Method Development, Ulm University Medical Center, Helmholtzstrasse 8/1, D-89081 Ulm, Germany

E-mail: boris.mizaikoff@uni-ulm.de

Keywords: exhaled breath analysis, mouse, oxygen, carbon dioxide, FTIR, substrate-integrated hollow waveguide, infrared and luminescence online monitoring

Abstract
Exhaled breath offers monitoring bio markers, as well as diagnosing diseases and therapeutic interventions. In addition, vital functions may be non-invasively monitored online. Animal models are frequently used in research for determining novel therapeutic approaches and/or for investigating biological pathways. The exhaled carbon dioxide concentration, exhaled and inhaled oxygen concentration, and the subsequent respiratory quotient (RQ) offer insight into metabolic activity. One may adapt breath sampling systems and equipment designed for human applications to large animal studies. However, such adaptations are usually impossible for small animals due to their minuscule breath volume. Here, we present a system for the online monitoring of exhaled breath in a 'mouse intensive care unit' (MICU) based on a modified Fourier-transform infrared spectrometer equipped with a substrate-integrated hollow waveguide gas cell, and a luminescence-based oxygen flow-through sensor integrated into the respiratory equipment of the MICU. Thereby, per-minute resolution of O₂ consumption and CO₂ production was obtained, and the 95% confidence range of the determined RQ was ±0.04 or approximately ±5% of the nominal value. Changes in the RQ value caused by intervention in either the metabolic or respiratory system may therefore reliably be detected.

1. Introduction
Apart from the determination of biomarker patterns serving as an indication for certain conditions and diseases, the development of online analytical systems capable of quantifying or at least monitoring analytes in exhaled breath and relevant constituents in real time is among the major aims of breath analysis. Online analysis promises to simplify sample collection as well as avoid compromising samples [1]. As a non-invasive method and for online measurements, breath analysis is among the favored strategies versus, e.g. invasive blood sampling, which only provides a snapshot of the patient’s condition. Carbon dioxide and oxygen are not only among the most abundant components in breath, but their concentration in exhaled breath also offers an important diagnostic into the metabolic status of a patient. Carbon dioxide is produced and released by the oxidation of carbohydrates, fats and proteins. These oxidative processes also consume oxygen. Each of these processes consumes a specific amount of oxygen and produces a specific amount of CO₂, which also generates a specific amount of heat. These contributions can be derived from the literature, which are either based on stoichiometry or directly measured, such as heat production after combustion. From these established physical principles, linear additive relations of oxygen consumption (VO₂ in mg/min) and CO₂ release (VCO₂ in mg/min), and associated oxidative loss of carbohydrates, loss of fat and loss of protein (all losses in mg/min) as independent variables can be established. These equations can
be rearranged to provide the so-called ‘respiratory quotient’ (RQ), i.e. the ratio of VCO₂ over VO₂ as a function of substrate utilization:

\[
RQ = \frac{VCO_2}{VO_2} = \frac{a_2 f_{\text{carb}} + b_2 f_{\text{fat}} + c_2 f_{\text{prot}}}{a_3 f_{\text{carb}} + b_3 f_{\text{fat}} + c_3 f_{\text{prot}}}
\]

\[= f_{\text{carb}} + 0.7 f_{\text{fat}} + 0.81 f_{\text{prot}}, \]

(1)

where \( f_{\text{carb}} \) gives the relative contribution of the carbohydrate oxidation to VO₂. Other terms such as \( f_{\text{fat}} \) and \( f_{\text{prot}} \) are defined accordingly. Observed values for the RQ are rather insensitive to changes in the protein oxidation. Hence, if protein oxidation is replaced by a feasible mean constant value, the resulting error may be neglected [2]. Thus, feasible estimates for carbohydrate and fat oxidation may be obtained from RQ values alone. These metabolic parameters may then be used to exemplarily assess a shift from glucose oxidation to fat oxidation, which provides a first and important information about the overall metabolic conditions [3, 4].

There is a growing demand for breath diagnostics in animals either for veterinary diagnostics [5] or for surveillance and research in pre-clinical animal models. In intensive care research, it is often necessary to consider the entire organism and its metabolic connections. This emphasizes the utility of mouse models simulating clinical intensive care conditions, which can be performed at a reasonably large number of replicates, and allow postmortem exploration of the involved and affected organs [6–10].

For studies involving larger animals, one may adapt breath sampling systems and equipment designed for human medicine. The key factors for analyzing the respiratory gas exchange rates (i.e. VO₂ and VCO₂) are calculated from the product of the actual flow rates and concentration values for exactly the same points of time, which in turn requires precise alignment of the time profile of the concentration values and gas flow. In humans, flow sensors and concentration sensors can be placed in a parallel arrangement within the main gas stream [11], which still causes only minimal flow resistance due to the rather large diameter of the gas tubes of the respiratory system, despite splitting the gas stream. However, a significant flow resistance will arise, if this main stream arrangement is scaled down for a gas line with a diameter of 1–3 mm, which requires positioning of the sensors one after another, and extensive corrections for the time offset during the measurements is required. During mechanical ventilation, mice require respiratory rates of approx. 150 cycles per minute, and a minute ventilation of 20 ml min⁻¹. Retracing a single breath cycle with 20 sampling points would therefore require a time resolution of one measurement every 20 msec. During that short time period, the mean gas release is approx. 7 μl. Such minuscule breath volumes do not meet the rather large sample volumes demanded by conventional analyzer systems designed for clinical applications. While capnography monitors adapted for use with rodents could handle the small volumes, they fail to provide O₂ measurements. Balances for O₂ uptake and CO₂ release for mice are typically assessed via so-called metabolic chambers [12]. However, due to the dead volume, the data obtained by these methods only provides coarse time averages of the net uptake and release [13]. Moreover, for obvious reasons, the use of a metabolic chamber is not possible during mechanical ventilation.

With the current state-of-the-art, the authors are not aware of any approach that allows online breath gas monitoring with the small breath volumes available during mouse studies, which moreover take advantage of reasonably priced and ready-to-implement optical sensing strategies. As a compromise between breath-cycle resolved monitoring and coarse time averages, the present study focuses on a time scale resolution of 1 min, which provides insight into the metabolic or physiologic response to a variety of deliberate experimental challenges. Consequently, we present a novel optical sensing system for the routine online monitoring of exhaled breath in the mouse intensive care unit (MICU). This analyzer is based on a Fourier-transform infrared (FTIR) spectrometer equipped with a substrate-integrated hollow waveguide (iHWG) pioneered by the Mizaikoff team [14, 15], which simultaneously serves as a miniaturized gas cell and IR photon conduit, and a luminescence-based flow-through oxygen sensor integrated into the respiratory equipment of the MICU. This sensing system offers an overall data sampling rate of one data point per minute. Previously optimized and now automated calibration algorithms and multivariate correction/data evaluation routines including statistical sampling of measurement error, offer rapid, reliable and comfortable access to real-time data in routine operation with accurate and realistic error bounds at the required levels. The quality of measurements obtained with the developed system, and its potential to capture dynamic changes in the gas exchange are explored using exemplary mouse data.

2. Material and methods

2.1. Experimental setup

The developed breath analyzer system comprises a custom-modified IR spectrometer for determining CO₂ via iHWGs, and flow-cell-based luminescence sensors for oxygen measurements.

The IR component is based on a Bruker ALPHA FTIR Spectrometer (Bruker Optik GmbH, Ettlingen, Germany) equipped with a custom-made iHWG gas cell developed by the Mizaikoff team [16, 17]. A 75 mm straight-channel iHWG with a channel cross-section of 2 mm² was simultaneously used as a miniaturized gas cell, and as an IR photon waveguide [14]. Using iHWGs is beneficial due to the minute required
gas volume, their high volumetric optical efficiency, the rapid sampling rate, the possibility of temperature control and their mechanical stability [14, 17–26]. Radiation from the IR spectrometer is focused into the iHWG, and propagated onto the internal DTGS detector of the ALPHA via two gold-coated off-axis parabolic mirrors (OAPMs; diam: 1") with an effective focal length of 2" (Janos Technology Inc., Keene, NH, USA). A flow-through oxygen sensor (FireStingO2, Pyro Science GmbH, Aachen, Germany) is integrated into the gas outlet of the iHWG. Between measurements and during background collection, the iHWG is purged with synthetic air. A FlexiVent Module 1 respiratory system (Scireq, Montreal, Canada) is used for mechanical ventilation. A gas reservoir of approximately 100 ml is inserted between the expiratory branch of the respiratory system and the optical sensors, which serves as a damping element to smooth fluctuations in the gas concentration during a respiratory cycle. The developed breath analysis system was integrated into the existing respiratory system of the MICU unit, as shown in figure 1.

2.2. Calibration design

The experimental design for calibrating the IR sensor consisted of a randomized set of 25 gas mixtures across the biologically expected concentration range of 1%–5% CO\textsubscript{2}, and 20%–80% oxygen, as can occur in ventilated patients. The combination and executed order of the samples was established using a full factorial design-of-experiment (PLS Toolbox 7.9.3, Experiment Designer, Eigenvector Research, Inc., Manson, WA, USA). One mixture (1% CO\textsubscript{2}/20% oxygen) was repeated five times interspersed in the set to monitor changes of the system and of the mixing pump. This set was measured over 14 d, thereby resulting in 2700 data sets. The IR spectra were collected from 4000–400 cm\textsuperscript{−1} (averaging 20 background scans (nitrogen) and 20 sample scans; apodization function: Blackmann–Harris three-term, five or ten repeats per mixture). The oxygen sensors were calibrated using 11 mixtures of oxygen and nitrogen spanning 0%–100% oxygen, repeated over 19 d, resulting in 385 calibration data sets. The calibration is performed in dry gas (0% RH/relative humidity), and at a flow of 200 ml min\textsuperscript{−1}. While the flow effects are negligible, the oxygen sensor signal depends on the sample humidity. Expired air has a temperature of approx. 34 °C and RH of 95%. With the transfer through the tubing, the air cools down to ambient temperature, and the humidity reaches saturation at 100% with some even condensing at the inner tube wall.

Figure 3 shows the change in δϕ signal of the oxygen sensor resulting from the RH at concentration levels of 16% O\textsubscript{2}. The humidity effect, when expressed as a change in the δϕ signal, increases from zero to about 0.59 at saturation. The data for this correction were gained by an experiment where a mixture of 16% O\textsubscript{2} was enriched in humidity by a humidity generator (Owlstone Humidity Generator OHG–4, Owlstone Inc., Cambridge, UK) from close to 0% to 100% RH and the signal measured by the same oxygen sensor as the one integrated into the setup. The humidity was measured using a capacitive humidity sensor (SF52, Mitchell Instruments, Ely, UK), where the dielectric strength of hygroscopic polymer is altered by the water vapor content.

The calibration of the sensor was performed using dry gas mixtures, since the accuracy of reference gases
with a set humidity is questionable using gas bottles. It was assumed that the expired air should be saturated, and ideally the difference in humidity between the breath gas and calibration sample should be 100%. With the transfer through the tubing, some humidity may be picked up by the calibration samples, and in turn some humidity may be lost for the respiration samples. Both processes reduce the difference between the calibration and breath gas. While routine measurements of the RH for each sampling point are not available, a difference of 95% between the calibration gas and breath samples was assumed, accounting for minor changes. According to figure 3, this translates to a change in $\delta \varphi$ of $0.59 \pm 0.039$. For data evaluation, each measurement of $\delta \varphi$ was corrected for this offset, and an uncertainty of $\pm 0.039$ was added next to the measurement error to each measurement, if the error propagation from the raw measurements to $O_2$ production and RQ values was estimated as outlined in figure 2.

Data evaluation was performed using MATLAB 8.6 (R2015b, The Mathworks Inc., Naticks, MA, USA),
PLS Toolbox 7.9.3 (Eigenvector Research, Inc., Manson, WA, USA) and MatlabStan 2.7.0.0 [27]/Stan 2.11.0 [28].

All gas calibration samples were prepared from pure technical grade nitrogen, pure technical grade carbon dioxide and pure medical grade oxygen (all MTI Industriegase, Neu-Ulm, Germany) using a gas mixing pump (DIGAMIX 2 M 301, H. Wösthoff Messtechnik GmbH, Bochum, Germany). For routine calibrations, three certified test gases were used: 5% CO₂ and 20% oxygen in nitrogen (Cal 1), 0.99% CO₂ and 29.9% oxygen in nitrogen (Cal 2), and 2.49% CO₂ in nitrogen (Cal 3) (all MTI Industriegase, Neu-Ulm, Germany).

During mouse experiments, one CO₂ IR spectrum every minute, and one O₂ data point every second was obtained. The 60 oxygen data points were averaged for noise reduction, thus resulting in one final time-synchronized data point for oxygen and carbon dioxide every minute during online mouse breath monitoring. Prior to starting routine measurements, an IR background was sampled using ambient air. After around half-way through an average mouse experiment (i.e. after approximately 3 h), the sensor measurements were stopped for a short period, enabling us to record a new IR background. For daily calibration, five IR spectra and at least 100 oxygen data points of the two/three calibration gases were recorded at the end of the procedure and averaged. The system reaches a dynamic range for calibration (i.e. ignoring humidity) of 1%–5% for carbon dioxide, and 10%–90% for oxygen. The system was humidity calibrated for 16% oxygen, which is the usual oxygen content in the exhaled breath of mice using air as the inspiratory gas mixture.

2.3. Data analysis procedure

The data analysis routine was based on a calibration transfer algorithm previously developed and optimized by the authors for the present optical sensing system [29, 30]. This data analysis strategy is based on hierarchical Bayesian models and Lagrange Multiplier optimization. Monte-Carlo Markov Chain sampling provides realistic estimates for coefficients and prediction together with accurate error bounds by simulating known measurement errors and system fluctuations. Using only three calibration gases, both the FTIR system and oxygen sensors are simultaneously calibrated for daily routine use. Fully automated Matlab scripts and wrappers offer comfortable and rapid data evaluation for routine studies without the requirement of expert users.

Figure 2 provides an outline of the data evaluation model including its error propagation scheme. The day-to-day variance (i.e. represented by the calibration sets used to build the hierarchical model), knowledge gained on the fit equation, fit coefficients gained by the hierarchical model and actual calibration samples analyzed to represent the current sensor status were used to build the actual calibration model by using a calibration transfer algorithm based on Lagrange Multiplier optimization. Combined with the sensor signal, concentrations of the breath markers are calculated. The corresponding uncertainty estimation is derived from the uncertainty in calibration coefficients as well as from modeling of the current random measurement error from the actual calibration samples and sensor response. Readers interested in the full error propagation procedure are referred to the supplementary material.

The calibration of the oxygen sensor is based on an empirical rational function as follows:

\[
\delta \varphi = \frac{\delta c_1 [O_2] + \delta c_2 [O_2]}{O_2} + \delta c_4,
\]

with \( \delta \varphi \) representing the raw data recorded by the oxygen sensor. This value is corrected for an offset in \( \delta \varphi \) expected for a difference of 95% RH between the calibration samples and breath samples. \( [O_2] \) refers to the oxygen concentration in vol%, and \( c_i \) to the current calibration coefficients. The uncertainty in the humidity values is considered as a propagation of errors from the measurements and other sources, as depicted in figure 2. Here, random \( \delta \varphi \) offset values in the range of 0.59 ± 0.039 were generated, whereby the ±0.039 variability covers both an uncertainty in humidity difference and calibration slope.

The IR data evaluation uses the main IR absorption peak of CO₂, and first applies a data reduction routine via principal component analysis (PCA) to one score, and then corrects for other interferents (i.e. oxygen) via the corresponding response surface. The calibration equation is therefore written as follows:

\[
S = c_1 + c_2 [CO_2] + c_3 [CO_2]^2 + c_4 [O_2][CO_2],
\]

with the \( S \) score of principal component 1, \([CO_2]\) carbon dioxide concentration in vol%, \([O_2]\) oxygen concentration in vol%, and \( c_i \) the current calibration coefficients. More detailed information can be found in our associated previous studies [29, 30]. Next to the exhaled oxygen and carbon dioxide concentration directly obtained from the two sensor systems, the difference between inhaled and exhaled oxygen, i.e. \( \Delta O_2 \) describes the amount of consumed oxygen, and is another relevant metabolic parameter. Since in the current MICU setup, the mouse is ventilated using compressed air, the inhaled oxygen concentration is determined by analyzing the oxygen content in compressed air. The entire calibration procedure is visually summarized in figure 4. By using the large number of sampling iterations gained from Stan software, it is possible to calculate the average standard deviation and 95% confidence interval values, thereby offering robust and realistic error boundaries for all calculated values.

3. Animal experiments

The study was approved by the federal authorities for animal research of the Regierungspräsidium Tübingen.
The study was conducted during the year 2016, and was recently published [31], where anesthesia, surgical instrumentation and the experimental design have been discussed in detail. Animals were mechanically ventilated with a small animal ventilator (FlexiVent, Scireq, MO, Canada) using a pressure-controlled mode. The initial ventilator settings were \( F_{O2} = 0.21 \), respiratory rate 150 min\(^{-1}\), tidal volume 4–6 ml kg\(^{-1}\),

(approved animal experimentation number: 1190, 24-09-2014), Baden-Württemberg, Germany, and performed in adherence to the National Institutes of Health Guidelines on the Use of Laboratory Animals and the European Union 'Directive 2010/63 EU' on the protection of animals used for scientific purposes.

Figure 4. Visual flow chart of the calibration and data analysis procedure. Black arrows represent data passing onto the next processing step. Values in the light green boxes (\( O_2 \) in, \( O_2 \) ex, \( CO_2 \) ex, \( \Delta O_2 \), RQ) are the parameters gained by the data analysis procedure during a measurement cycle. FTIR: Fourier-transform infrared spectrometer, \( \delta \phi \): raw value of oxygen sensor, PCA: principal component analysis, Cal Set: calibration sets used for building the model, HM: hierarchical model, CT: calibration transfer, LM: Lagrange Multiplier optimization approach, Cal samples: actual calibration samples analyzed (i.e. daily calibration), \( O_2 \) in: inhaled oxygen concentration, \( O_2 \) ex: exhaled oxygen concentration, \( CO_2 \) ex: exhaled carbon dioxide concentration, \( \Delta O_2 \): difference between inhaled and exhaled oxygen, RQ: respiratory quotient.

J. Breath Res. 12 (2018) 036018 F Seichter et al
inspiratory/expiratory time ratio 1:2, and positive end-expiratory pressure (PEEP) 3 cmH₂O. The respiratory rate was modified to maintain an arterial partial pressure of CO₂ (pCO₂) between 30–40 mmHg, and the PEEP was varied according to the arterial pO₂ (if pO₂/FiO₂ > 300 mmHg: PEEP = 3 cmH₂O; pO₂/FiO₂ < 300 mmHg: PEEP = 5 cmH₂O; pO₂/FiO₂ < 200 mmHg: PEEP = 8 cmH₂O). Recruitment maneuver (RMs, 5 s hold at 18 cmH₂O) were repeated every 30 min to avoid atelectasis formation due to the anesthesia and supine positioning. Each hour, the RM was followed by a compliance measurement, where the lung was inflated to a volume of 1 ml and the thoracic pressure was measured. Thereafter, a second RM was performed. General anesthesia was sustained by continuous intravenous administration of ketamine, midazolam and fentanyl to reach deep sedation. The experiment started with a thorax trauma, followed by surgical instrumentation. Hemorrhagic shock was performed about 1 h later by removing 30 μl · g⁻¹ of blood and by titrating the mean arterial pressure (MAP) to 35 mmHg via further removal or retransfusion of blood. Fluid administration was temporarily stopped. At the start of the resuscitation phase, shed blood was retransfused, together with the administration of hydroxyethyl starch 6% and noradrenaline (NA) titrated to maintain MAP ≥ 50 mmHg. At the end of the experiment, the animals were exsanguinated via blood withdrawal from the arterial catheter under deep anesthesia.

4. Results and discussion

The data presented are those of one animal (wild type C57BL/6J strain) of the study group, undergoing a combined blunt chest trauma and hemorrhagic shock.

4.1. Physiology and metabolism

The data analysis provides inhaled and exhaled oxygen concentration, their difference (ΔO₂), and the exhaled carbon dioxide concentration, which together allow estimation of the respiratory quotient in exhaled mouse breath. The experiment started around 8:00 and ended approximately 7 h later, delivering one data point per minute on the current metabolic status of the mouse. Figure 5 shows the determined respiratory concentration data. After thorax trauma, arterial and central venous catheters were implemented for monitoring and drug administration. Then, the mouse was connected to the ventilation system, while breath gas recording was initiated a while beforehand, in order not to hinder the medical procedures. This connection is reflected in a first jump of the measured O₂ consumption, as shown in figure 5(c). Continuous recording was maintained until the mouse died, which coincided with a sharp drop of O₂ consumption. The first connection was followed by a ragged time course for both O₂ consumption and CO₂ production. The ragged time course was due to RMs [32, 33], which started 100 min after beginning and were repeated at an interval of 30 min. RMs were used in mice during mechanical ventilation with low tidal volumes, in order to re-expand at electatic lung regions, which had collapsed due to supine positioning, anesthesia effects and reduced diaphragmatic muscle tone, and thereby restore and/or maintain static thoraco-pulmonary compliance [34]. RMs are characterized by a transient increase of the airway pressure during a couple of seconds, which will re-open non- or poorly-aerated alveoli. As a consequence, respiratory system mechanics and gas exchange are improved. Besides the intermittent peaks, the time course of the RQ data shows a significant drop to values close to 0.7 towards the end of the protocol. The down drift of the RQ values lasts for 3–4 h. For such a long period, the CO₂ output by respiration has to match the CO₂ production by oxidation. Any increase in tissue/blood CO₂ content may subsequently induce hemodynamic variations, because CO₂ is well established as a potent systemic vasodilator as well as a pulmonary vasconstrictor. The down drift indicates an increasing preference for fat oxidation with increasing duration of the experiment, and thus rules out respiratory effects of hypoventilation, which act on a shorter timescale. However, acute variations in the gas exchange are evident in response to the intermittent RMs. Figure 6 takes a closer look at the O₂ uptake data during an RM. This demonstrates a sharp drop in the O₂ uptake in response to an RM that lasts about 5 min, followed by a steady increase towards the level that was observed just before the maneuver. The reduced uptake in the first phase of the maneuver may result from two phenomena. First, an enforced ‘breath holding’ that last for a few seconds. The observed response may last longer as it is stretched out with the damping reservoir that is inserted in the expiratory branch of the measurement system. The second process is linked to high intrathoracic pressures that were induced with RMs, which may have a detrimental effect on cardiac output [35], and may thereby reduce pulmonary perfusion and impair the gas exchange. The rebound in pulmonary perfusion may explain the recovery of the O₂ uptake that is seen about 10 min after an RM. Figure 6(b) indicates that this pattern is reproducible for repetitive RMs. The expiratory CO₂ release follows the same pattern, except that transient increase is evident in the first response to an RM. This can be explained by a CO₂ accumulation during the ‘breath holding’ phase, and a transiently increased washout of CO₂ directly after an RM, most likely due to the RM-induced transitory increase in alveolar ventilation.

After the first RM, approx. 2 h after beginning a hemorrhagic shock was induced, which lasted for 1 h. The reduced blood volume during shock causes a reduced organ perfusion and thereby reduced oxygen supply. An oxygen supply/demand mismatch may
develop, which is considered to be a key determinant of hemorrhagic damage [36, 37], and should be reflected in a transient drop of the O\textsubscript{2} uptake. Figure 6 shows a smoothed uptake curve that was tentatively corrected for the perturbations caused by RMs. The first response to the hemorrhagic shock is superimposed by effects of an RM performed five minutes earlier, which complicates such a correction. Nevertheless, the corrected uptake curve shows a drop to minimal O\textsubscript{2} uptake values during hemorrhagia and a subsequent compensatory increase. The data shown so far give a hint about the information that can be obtained with continuous monitoring of respiratory O\textsubscript{2} and CO\textsubscript{2} data.

Figure 5. Exemplary results of one mouse experiment: (a) exhaled oxygen; (b) difference between inhaled and exhaled oxygen; (c) exhaled carbon dioxide; (d) RQ. Thicker solid line: mean, thinner solid line: 95% confidence interval. Red: no correction, blue: with humidity correction.

Figure 6. Physiologic interpretation of the results. (a) Time course of the O\textsubscript{2} uptake and a tentative correction for the effect of intermittent RMs to expose the impact of a hemorrhagic shock. Black solid line: measured difference between the O\textsubscript{2} concentrations of inspired and expired air. Red dotted line: tentative interpolated function. Green area indicates the time frame of a hemorrhagic shock. (b) CO\textsubscript{2} production and O\textsubscript{2} consumption during several RMs.
4.2. Precision and accuracy in RQ determinations

Figure 5 demonstrates how RQ values are derived from O₂ and CO₂ concentration measurements, and the effect of humidity on these estimates. According to figure 3, an increase in humidity from zero to 100% increases the δφ signal by 0.6, which translates to an approximate decrease of 16% to 15.3% in the O₂ concentration. In consequence, the breath gas measurements are underestimated, and correcting this shift increases the average O₂ in mouse breath concentration measurement by approximately 0.7%. This is reflected in figure 5, which shows the results with and without correction side by side. As expected, there is no significant influence on the CO₂ quantification. However, the exhaled oxygen values are raised by the oxygen correction. Subsequently, ΔO₂ is lowered, and RQ is raised. This is actually a desired result, since in other mice data the RQ was sometimes below the biological limit of 0.6, which indicates that it was underestimated. Actual RH values cannot be determined during routine experiments. Therefore, it was assumed that the difference in RH for breath gas measurements and calibration gas measurements fluctuates in the range of 95% ± 3%. These fluctuations should translate into variations of ± 0.039 in the δφ signal, and ±0.059 in the O₂ concentrations, correspondingly, the RQ estimates fluctuate likewise.

The uncertainty in the breath sample humidity is captured in a sequence of (Bayesian MCMC) sampling runs, which were generated to assess the error propagation. For each sample run, simulated random fluctuations of humidity were used for correction, and combined with simulated measurement errors and the resulting ‘noisy’ data were used to assess the RQ values.

There are two sources of fluctuations that affect the estimated RQ values. One source relates to per-second or per-minute fluctuations in the sensor signals for O₂ and CO₂, and are called ‘short-term fluctuations’. The other source is uncertainties in the coefficients for the calibration curves of O₂ and CO₂. These values fluctuate from day to day, yet were constant during any given measurement day. These uncertainties are referred to as ‘long-term fluctuations’.

Based on the sampling strategy, RH is assigned to the long-term fluctuations. If one is primarily interested in changes of the RQ values with time or in response to an experimental intervention such as the hemorrhagic shock, then long-term fluctuations may safely be ignored, and the fixed values for the mean calibration coefficients and humidity corrections can be used. To isolate the effect of long-term versus short-term fluctuations, simulations were performed where different fluctuations were turned off. Figure 7 shows the resulting confidence ranges for a time section of the protocol. If only short-term fluctuations are considered, then the uncertainty in RQ determinations is reduced by 50%. Ignoring the effect of the RH reduces the uncertainty to approx. 2/3 of the initial starting values. If short-term fluctuations were ignored, but all long-term variations considered, barely any improvement is evident. This can be explained by the fact that the variances of the fluctuation add up, however, not their standard deviation (SD). In other words, the overlay of one source with a fluctuation (i.e. an SD) of 0.2 with a variable that has an SD of 1.0 results in an overlaid SD of 1.02. These data indicate that the precision (i.e. approx. twice the SD) in absolute values for RQ estimates is approx. ±0.04 or less than 5% of the nominal value. Schadewaldt et al [37] explored the precision in RQ estimates of commercial instruments including the former ‘Deltatrac’ gold-standard. They found precision values, which were about twice the values determined in the present study with the system developed herein. Our data indicate that imprecision in long-term factors evidently plays a major role. If one is primarily interested in changes in respiratory values observed in response to a challenge, fluctuations may be ignored, which almost doubled the precision. Our study furthermore demonstrates that these responses cover metabolic changes caused by a hemorrhagic shock, and serve as indicators of inflicted damage [35, 36]. The extent of a drop in O₂-uptake following an RM reflects the vulnerability of the circulatory system and cardiac function [34]. Hence, the approach developed in the course of this study is a viable strategy to assess the O₂/CO₂ gas exchange, and its dynamic alterations. Thus, a valuable tool for characterizing the underlying metabolic and physiologic conditions has been established and will be further evolved for widespread routine usage.

5. Conclusions and outlook

In this study, an innovative optical sensor system is presented combining infrared and luminescent sensing principles for the online monitoring of relevant metabolic parameters (i.e. exhaled oxygen and carbon dioxide, difference between inhaled and exhaled oxygen and the RQ) in the exhaled breath of ventilated mice at a temporal resolution of one data point per minute. Thereby, the immediate metabolic status of the mouse during medical trials may be continuously monitored. It was shown that continuous monitoring may reliably capture actual values of the RQ with a 95% confidence interval of ±0.04. Moreover, an unexpected sensitivity of RQ values against RMs was revealed for the first time via the developed sensing strategy, which may be indicative for pulmonary or circulatory problems and will be investigated in more detail during future studies.

The developed optical sensing system comprises a modified FTIR spectrometer combined with an iHWG and a luminescence-based flow-through oxygen sensor integrated into the respiratory system of a MICU. Optimized and automated calibration routines along with statistical sampling offer a rapid,
reliable and comfortable data evaluation strategy providing accurate and realistic error boundaries resulting from average, SD and 95% confidence interval values at each data point.

The developed system is nowadays in daily use at the MICU located at the Institute of Anesthesiologic Pathophysiology and Method Development at Ulm University Medical Center, continuously proving the reliability of the developed system in routine exhaled breath analysis scenarios.

Acknowledgments

The authors gratefully acknowledge the continuous support of the team operating the MICU at the Institute of Anesthesiologic Pathophysiology and Method Development at the Ulm University Medical Center for assistance during the mouse breath studies. The authors would also like to thank the Federal State of Baden-Württemberg (Landesgraduiertenförderungsgesetz, grant number: 1212-LGFG-E) for partial financial support to FS. The development of the sensing technology was performed in part under the auspices of the US Department of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract DE-AC52-07NA27344. This project was partially funded under LLNL sub-contract Nos. B598643 and B60301. The Boehringer Ingelheim Ulm University BioCenter (BIU, grant number: D.5008) is acknowledged for financial support. The animal studies were in part supported by an SFB-Grant of the German Research Foundation (DFG, grant number 1149). Last but not least, the BMBF project APOSEMA and MSCA RISE project TROPSENSE (funded by the Horizon 2020 Framework Programme of the European Union under grant agreement number 645758) are acknowledged for partial financial support.

The authors declare no conflict of interest and that the funding agencies had no influence on the work presented and on the publication of this paper.

ORCID iDs

Boris Mizaikoff https://orcid.org/0000-0002-5583-7962

References

2008 The effect of superoxide dismutase overexpression on hepatic gluconeogenesis and whole-body glucose oxidation during resuscitated normotensive murine septic shock Intensive Care Med. 35 344–9


[9] Baumgart K et al 2010 Cardiac and metabolic effects of hypothermia and inhaled hydrogen sulfide in anesthetized and ventilated mice Crit. Care Med. 38 588–95


[19] Sieger M and Mizaikoff B 2016 Optimizing the design of GaAs/AlGaAs thin-film waveguides for integrated mid-infrared sensors Photonics Res. 4 100


[31] Hartmann C et al 2018 In-depth characterization of the effects of cigarette smoke exposure on the acute trauma response and hemorrhage in mice Shock (https://doi.org/10.1097/SHK.0000000000011115)


5.5 Paper IV: Strategies for 13C enrichment calculation in Fourier-transform Infrared CO2 spectra containing spectral overlapping and nonlinear abundance-amount relations utilizing response surface
Strategies for $^{13}$C enrichment calculation in Fourier-transform infrared CO$_2$ spectra containing spectral overlapping and nonlinear abundance-amount relations utilizing response surface fits

Felicia Seichter $^a$, Josef Albert Vogt $^b$,*, Ulrich Wachter $^b$, Peter Radermacher $^b$, Boris Mizaikoff $^a$,**

$^a$ Institute of Analytical and Bioanalytical Chemistry, Ulm University, 89081, Ulm, Germany
$^b$ Institute of Anesthesiologic Pathophysiology and Method Development, Ulm University Medical Center, 89081, Ulm, Germany

HIGHLIGHTS

- PCA analysis of IR spectra of $^{13}$CO$_2$/12CO$_2$ mixtures gives 2 relevant components.
- The PCA results can be transformed to obtain non-negative loadings and scores.
- The transformed loadings match the IR spectra expected for $^{13}$CO$_2$/12CO$_2$.
- The transformed scores are non-linear in the $^{13}$CO$_2$, 12CO$_2$ and O$_2$ sample content.
- The non-linearity can be described using a response surface fit.

ABSTRACT

The metabolism can be explored via $^{13}$C labeling of biological active substances and subsequent quantification of $^{13}$C enrichment in the exhaled carbon dioxide in breath. The resulting tracer enrichment values can be determined by Fourier-transform Infrared Spectroscopy (FTIR), since different CO$_2$ isotopologues result in distinct absorption lines in the spectrum. The corresponding determination poses two challenges: first, FTIR absorbance can contain a nonlinear relationship between analyte amount and spectral signal and second, the spectral peaks for the different isotopologues overlap. The overlap precludes a separate calibration to assess the isotopologue concentration values and with it a determination of enrichments from concentration values. We propose here, first, a data reduction step like Principal Component Analysis (PCA) to convert the spectral information into one score pertaining to the $^{13}$CO$_2$ enrichment. In a second step, a calibration function between score and enrichment values was established. The enrichment score can be derived by normalizing a subset of the spectrum by some measure for the 12CO$_2$ sample content. Alternatively, the overlapping spectra were decomposed into two isotopologue spectra and the intensity of the separated spectra was used to form an enrichment score. For spectral separation, either Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) was used or a novel decomposition strategy developed for this paper called Rotation and Angle-Bending Bayesian induced Transformation - Multivariate Curve Resolution (RABBITeMCR) that operates in a Principal Component Analysis (PCA) subspace and is derived from MCR. We compared $^{13}$C enrichment estimates from FTIR CO$_2$ spectra using different normalization variants with the two spectral separation models. In conclusion, the two spectral separation variants performed nearly equal, but better than any normalization variant.

© 2019 Published by Elsevier B.V.

1. Introduction

$^{13}$C labeling plays an important role in the investigation of metabolic processes. By labeling biologically active substances like...
amino acids or carbohydrates with the \(^{13}\)C isotope, it is possible to trace their metabolism in the body by detecting their excretion as \(^{13}\)CO\(_2\) in breath and therefore gain an insight in metabolic processes in the body. In particular, one of the first medical breath tests was the \(^{13}\)C urea breath test [1]. The \(^{13}\)C enrichment in carbon containing samples like, e.g., the \(^{13}\)CO\(_2\) content in \(^{13}\)C breath tests, can be quantified as the tracer-to-tracer ratio (TTR) or the tracer mole fraction (MF) which are defined as:

\[
TTR = \frac{[^{13}\text{CO}_2]}{[^{12}\text{CO}_2]} \quad \text{MF} = \frac{[^{13}\text{CO}_2]}{[^{12}\text{CO}_2] + [^{13}\text{CO}_2]}
\]

with \([^{13}\text{CO}_2]\) the \(^{13}\)CO\(_2\) concentration and \([^{12}\text{CO}_2]\) the \(^{12}\)CO\(_2\) concentration. TTR can be transformed into MF and vice versa using the following equation:

\[
MF = \frac{TTR}{1 + TTR}
\]

A key number in metabolic studies is the expiratory \(^{13}\)CO\(_2\) release, which is usually assessed from the product of ‘whole body \(^{13}\)CO\(_2\) production’ times the \(^{13}\)CO\(_2\) enrichment in breath gas. Whole body \(^{13}\)CO\(_2\) production can be estimated from capnography or respiratory gas exchange in the frame of indirect calorimetry [2]. Consequently, only the \(^{13}\)CO\(_2\) isotope ratio is determined and the analytical challenge is reduced to assess \(^{13}\)CO\(_2\) to \(^{12}\)CO\(_2\) ratios from \(^{13}\)CO\(_2\) and \(^{12}\)CO\(_2\) derived signals. There are many techniques for quantifying the \(^{13}\)C enrichment, among them Fourier-transform Infrared Spectroscopy (FTIR). \(^{13}\)CO\(_2\) isotopologues show different rotational absorption lines and vibration peaks in the IR spectrum since the transitions are mass-dependent. Fig. 1 shows the theoretical lines for \(^{13}\)CO\(_2\) and \(^{12}\)CO\(_2\) of a segment of the CO\(_2\) IR spectrum [3], which we used for determination.

The spectra of \(^{12}\)CO\(_2\) and \(^{13}\)CO\(_2\) overlap in the wave number range between 2275 and 2325 cm\(^{-1}\). Moreover, the ratio of \(^{13}\)CO\(_2\) over \(^{12}\)CO\(_2\) in biological samples is usually below 10\% and in the overlapping range, the measured abundance lines are dominated by \(^{12}\)CO\(_2\) contributions. There are different options to quantify the TTR or MF values from a FTIR spectrum. First, one could select a spectral region for \(^{13}\)CO\(_2\), which contains only \(^{13}\)CO\(_2\) lines, i.e. at the right border of the spectrum shown in Fig. 1 and a corresponding segment of the spectrum for \(^{12}\)CO\(_2\), i.e., at the left border of the spectrum and assess the ratios of the corresponding areas. Such a strategy captures only a fraction of the spectral information and one would loose the quality control ‘built-in’ in the multivariate analysis of full spectra: unknown interferents may be detected if the measured spectrum cannot be explained by an overlay of the \(^{12}\)CO\(_2\) and \(^{13}\)CO\(_2\) spectra. Because of these considerations, the complete spectrum as shown in Fig. 1 should be analyzed. This in turn requires separating the overlapping spectra of the different compounds. Most of the established spectra decomposition approaches have in common that a large measured set of mixed multicomponent response data is expressed in a matrix \(D\), where a column of the matrix pertains to a measured analyte, and an individual element of the row gives the signal observed by a spectral line. This data matrix is then decomposed into the product of two matrices, a concentration matrix \(C\), where a column pertains to a sample and a row contains the concentration of the constituents of the sample. The second matrix \(S\) contains line intensities or spectra of the analyte in question. Elements of \(S\) and \(C\) reflect concentration values or line intensities and should therefore be non-negative. Based on the ‘non-negativity’ constraint, a decomposition of the data matrix is possible which in turn can be solved for the concentration matrix \(C\) Multi Curve Resolution Alternating Least Squares (MCR-ALS) [4–7], Resolving Factor Analysis (RFA) [8] and Gentle [9] are based on this concept with the first having found many applications in current research [4,10–14].

MCR-ALS is designed to determine concentration values of different analytes and their unknown spectral distribution in parallel. With unknown spectra and limited information about sample concentrations a unique decomposition is difficult. Hence, the different MCR-ALS variants use additional information about reference samples with known concentration, restraints concerning the peak shape of spectra or the time course of concentration values (uni-modality) or precursor -product relations for concentrations obtained during process monitoring (closure). These approaches require in most cases that observed line intensities are a linear function of the analyte concentration. However, our previous studies [2,15] demonstrated that for \(^{12}\)CO\(_2\) concentrations below 3–5\% the measured shape of IR-spectrum does not change significantly. Yet, we could detect a nonlinear relation between the overall height or intensity of the lines and the \(^{12}\)CO\(_2\) concentration in sample. Such a height-related non-linearity cannot be corrected with preprocessing and due to this nonlinear relationship concentration estimates gained by MCR-ALS cannot be reliably converted to the ratio of two isotopologue or TTR enrichments, which was our primary point of interest. Moreover, restraints pertaining to unimodality or closures, pivotal features of MCR-ALS, cannot be used in the present case.

As a solution, we adopted a two step approach: first, we derived a marker or score for the isotope ratio, in which a part of the spectrum, extracted for the \(^{13}\)C intensity, was normalized by some measure for the \(^{13}\)C content. At this stage, a spectral decomposition of the line intensities for a given optical setup may come into effect and we strove for an excellent separation of spectral peaks. For the second step, we expected a nonlinear relation between this ratio marker and the true isotope ratio of the sample. This relation was assessed using a calibration or regression of scores against isotope ratio. The proposed two step approach with signal separation followed by calibration required more computation effort than techniques that combine the decomposition and quantification step but the gain in calibration accuracy and precision is assumed to be significant for problems that necessitate non-linear calibration relationships and response surfaces. In this paper, we compared a \(^{13}\)C
enrichment calculation from FTIR CO2 spectra using normalization to separation either using standard MCR-ALS as a reference or on a newly developed decomposition algorithm. It is called Rotation and Angle-Bending Bayesian induced Transformation - Multivariate Curve Resolution (RABBIT - MCR), was derived from MCR and operates on the subspace of a Principal Component Analysis (PCA). We compared 13C enrichment estimates from FTIR CO2 spectra using different normalization variants with the two spectral separation models.

2. Material and methods

2.1. Experimental methods

Two different data sets were used for testing the data evaluation strategies: the first one was the main one to be calibrated and is in use for the on-line measurement of mouse breath in an intensive care unit. The second, first published in Ref. [16], was mainly used to prove the effectiveness of the spectra decomposition scheme. All gas calibration samples were mixed in two steps: A base mixture of pure technical grade nitrogen, pure technical grade carbon dioxide, and pure medical grade oxygen (all MTI Industriegase, Neu-Ulm, Germany) was generated using a gas mixing pump (DAGIMIX 2 M 301, H. Wösthoff Messtechnik GmbH, Bochum, Germany). In a second step, these mixtures were spiked with 99.9% 13CO2 (Isotec, Sigma-Aldrich, St. Louis/MO, USA) using a manual turbulent mixing method with 50 ml Luer-lock syringes (B. Braun Melsungen AG, Melsungen, Germany). Data evaluation was done using MATLAB 8.6 (R2015b, The Mathworks Inc, Naticks/MA, USA), PLS Toolbox 7.9.3 (Eigenvector Research, Inc., Manson/WA, USA) and MatlabStan 2.7.0.0 [17]/Stan 2.14.0 [18]. MCR-ALS was done using the MCR ALS Toolbox [19,20]. Oxygen and 12CO2 concentrations were calculated using procedures described in earlier works [2,15,21].

2.1.1. Substrate integrated HWG FTIR setup for online breath analysis (data set 1/D1)

The main system used to generate data set 1 is based on a Bruker ALPHA FTIR Spectrometer with a custom made gas cell and modified sampling chamber (Bruker Optik GmbH, Ettlingen, Germany) [22,23]. A 7.5 cm straight channel substrate-integrated hollow waveguide (iHWG) serves as gas cell as well as wave guide [24]. Radiation from the spectrometer is focused into the iHWG and out into the internal DTGS detector of the ALPHA via two 1” gold-coated off-axis parabolic mirrors (OAPM) with an effective focal length of 2” (Janos Technology Inc, Keene/NH, USA). The entire sample chamber is flushed with nitrogen during measurements. A Fire-StingO2 oxygen flow cell sensor (Pyro Science GmbH, Aachen, Germany) is integrated into the gas outlet. Between measurements and during background calibration, the gas cell was flushed with nitrogen. The exact setup has already been described elsewhere [2,15,25]. The design for calibrating the IR sensor consisted of a set of 25 gas mixtures spanning the biologically expected concentration of 1–5% CO2 and 20–80% oxygen. The combination and run order of the samples was randomized using a full factorial DOE (PLS Toolbox 7.9.3, Experiment Designer, Eigenvector Research, Inc., Manson/WA, USA). Each gas mixture was then spiked with different 13CO2 volumina resulting in 13C enrichments of 0, 10, 20 and 30% TTR. The FTIR spectra were collected from 4000 to 400 cm−1 (averaging 20 background scans (nitrogen) and 20 sample scans, apodization function Blackmann-Harris 3-Term, 5 repeats per mixture). The data set consisted of 1578 measurements in total, 1164 of which the exact TTR concentration was known and that were used for calibration, 34 which where further validated using GC-MS (validation set 1/V1) and 209 samples of which the exact TTR concentration was known (calibration/reference samples during other mouse studies) and which were also used for validation (validation set 2/V2).

2.1.2. Conventional HWG FTIR setup for offline breath analysis (data set 2/D2)

The second setup (data set 2) consisted of a FTIR-HWG setup comprising a FTIR spectrometer (IRcube, Bruker Optics, Ettlingen, Germany) and a gas sensing module using a custom HWG (2.5 cm length, outer diameter 4.1 mm, inner diameter 2.0 mm) embedded into a customized gas cell assembly sealed with ZnSe windows. Radiation emanating after absorption within the gas cell was directly coupled to a liquid nitrogen cooled mercury-cadmium-telluride (MCT) detector (FTIR-16-1.0, InfraRed Associates, Stuart/FL, USA). The entire sensor system was encased in a flexible plastic housing and kept under pure nitrogen (MTI IndustrieGase AG, Neu-Ulm, Germany) for providing stable measurement conditions. The setup has been described in Seichter et al. [16]. Calibration samples were prepared as described above. Calibration samples included pure 12CO2 and 13CO2 samples as well as TTR enriched samples (mixtures of 12CO2 and 13CO2) in the expected concentrations in mouse breath. The data set consisted of 928 measurements in total, 695 with exact, known TTR concentrations that were used for calibration. The FTIR spectra were collected from 4000 to 400 cm−1 (averaging 200 nitrogen background- and 200 sample scans, apodization function Blackmann-Harris 3-Term, 5 repeats per mixture). The set also included several breath samples of mice. Two successive breath samples (1 mL each) were collected during several mouse experiments every hour during a total measurement period of 5 h via 1 mL plastic syringes from anesthetized and ventilated mice. From the entire sample volume, 0.5 mL was used for IR analysis; the remaining sample was used for collecting GC/MS validation data, which was later used as reference.

2.2. Data analysis procedure

The most IR intensive asymmetric stretch band ν2 of CO2 from 2200 to 2500 cm−1 was used. All spectra were baseline-corrected using a Weighted-Least-Squares Algorithm (PLS Toolbox, Eigenvector Research, Inc., Manson/WA, USA). Decomposition via PCA showed that nearly all of the spectral variance (99%) can be explained by only two principal components (PC). We aimed to transform these loadings to obtain a score pertaining to the 13CO2 concentration normalized by the 12CO2 concentration that can be regressed against TTR or MF. The first option, a normalization approach is explained in section 2.2.1. Here, a subset of the spectrum was normalized to the 12CO2 peak, to minimize the impact of 13CO2 on the principle components. The second option was a spectra decomposition/separation that provided loadings and scores unique to the two underlying peaks belonging to the two isotopologues and is expanded in section 2.2.2.

2.2.1. Normalization

Consider two spectral lines with abundances, that are linear in one of the isotopologues, or \( x_1 = k_1 [^{13}\text{CO}_2] \) and \( x_2 = k_2 [^{12}\text{CO}_2] \). Their signal ratio is:

\[
x_1/k_1 = 1^{13}\text{CO}_2 / 1^{12}\text{CO}_2 \quad \text{or} \quad \frac{[^{13}\text{CO}_2]}{[^{12}\text{CO}_2]} = \text{TTR} = \frac{k_2}{k_1} \frac{x_1}{x_2}
\]  

(3)

Such a relation also holds if the numerator or denominator contain a sum of signals pertaining to the appropriate isotopologue. More generally, any sum of signals in the \(^{13}\text{CO}_2\) range normalized over any measure for the \(^{12}\text{CO}_2\) signal intensity provides a score or index that is proportional to the TTR value. Hence, for
normalization, we divided $^{13}$CO$_2$ absorption values of the spectrum by a norm $n_x$, which was derived from spectral areas covering only $^{12}$CO$_2$ absorption. We compared the 1-Norm and Inf-Norm normalization algorithms included in the PLS Toolbox. The 1-Norm reflects the area under a peak segment:

$$n_x = \sum_{i=1}^{N} |x_i|$$

(4)

and for the inf-Norm a fixed (usually maximum) peak value is used. Earlier works [15] have shown that the removal of an unstable peak area (e.g., the maximum of the $^{12}$CO$_2$ peak due to values close to saturation) can improve results. Hence, we combined different definitions of $n_x$ with an optional removal of certain peak areas and we tested whether better results could be achieved when including the whole spectral area or only the area containing the $^{13}$CO$_2$ peak in the PCA. The different combinations of $n_x$ definitions, window selection for PCA and exclusion of saturable peak areas resulted in 26 different normalization models as shown in Table 1.

### 2.2.2. Spectral decomposition

Successful decomposition delivers loadings resembling the pure isotope spectra together with the corresponding scores, which in the present case refer either to the amount of $^{12}$CO$_2$ or $^{13}$CO$_2$ in a sample. Hence, the latter could be used to assess enrichment (saturation) can improve results. Hence, we combined different normalization models as shown in Table 1.

2.2.3. MCR-ALS

Like all MCR models, MCR-ALS [19] tries to decompose a data matrix $D$ into components comparable to eqn (6). It uses an iterative 'alternating least squares' approach to obtain the desired non-negativity. The iteration requires reasonable starting values and therefore, some of the MCR methods (among them MCR-ALS) start with a noise reduced decomposition as defined in (7). Non-negative data matrices for concentration profiles and spectra are determined by alternatively keeping one matrix constant under

$$D = SL^T + E$$

(6)

Removing the error matrix (residuals) $E$ leads to the relationship

$$\hat{D} = SL^T$$

(7)

The equation above serves as a data cleansing step. But the "score" matrix $S$ and the loading matrix $L$ usually contain negative elements, and different approaches were established to enforce non-negativity.

### Table 1

<table>
<thead>
<tr>
<th>Model name</th>
<th>Spectral region [cm$^{-1}$]</th>
<th>remove peak maximum?</th>
<th>normalization method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional normalization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>2315–2201</td>
<td>yes</td>
<td>area 2398-2360 cm$^{-1}$</td>
</tr>
<tr>
<td>M2</td>
<td>2500–2099</td>
<td>yes</td>
<td>area 2398-2360 cm$^{-1}$</td>
</tr>
<tr>
<td>M3</td>
<td>2315–2201</td>
<td>yes</td>
<td>area 2400-2349 cm$^{-1}$</td>
</tr>
<tr>
<td>M4</td>
<td>2500–2099</td>
<td>yes</td>
<td>area 2400-2349 cm$^{-1}$</td>
</tr>
<tr>
<td>M5</td>
<td>2315–2201</td>
<td>yes</td>
<td>area 2400-2349 cm$^{-1}$</td>
</tr>
<tr>
<td>M6</td>
<td>2500–2099</td>
<td>yes</td>
<td>area 2400-2349 cm$^{-1}$</td>
</tr>
<tr>
<td>M7</td>
<td>2315–2201</td>
<td>yes</td>
<td>area 2500-2398 cm$^{-1}$</td>
</tr>
<tr>
<td>M8</td>
<td>2500–2099</td>
<td>yes</td>
<td>area 2500-2398 cm$^{-1}$</td>
</tr>
<tr>
<td>M9</td>
<td>2315–2201</td>
<td>yes</td>
<td>area 2500-2398 cm$^{-1}$</td>
</tr>
<tr>
<td>M10</td>
<td>2500–2099</td>
<td>yes</td>
<td>area 2500-2398 cm$^{-1}$</td>
</tr>
<tr>
<td>M11</td>
<td>2315–2201</td>
<td>yes</td>
<td>area 2500-2398 cm$^{-1}$</td>
</tr>
<tr>
<td>M12</td>
<td>2500–2099</td>
<td>yes</td>
<td>area 2500-2398 cm$^{-1}$</td>
</tr>
<tr>
<td>M13</td>
<td>2315–2201</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M14</td>
<td>2500–2099</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M15</td>
<td>2315–2201</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M16</td>
<td>2500–2099</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M17</td>
<td>2315–2201</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M18</td>
<td>2500–2099</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M19</td>
<td>2315–2201</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M20</td>
<td>2500–2099</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M21</td>
<td>2315–2201</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M22</td>
<td>2500–2099</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M23</td>
<td>2315–2201</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M24</td>
<td>2500–2099</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M25</td>
<td>2500–2099</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>M26</td>
<td>2315–2201</td>
<td>no</td>
<td>area 2500-2099 cm$^{-1}$</td>
</tr>
<tr>
<td>Normalization via spectra decomposition</td>
<td></td>
<td></td>
<td>RABBIT-MCR (median scores)</td>
</tr>
<tr>
<td>M27</td>
<td>2500–2099</td>
<td>yes</td>
<td>RABBIT-MCR (median scores)</td>
</tr>
<tr>
<td>M28</td>
<td>2500–2099</td>
<td>yes</td>
<td>RABBIT-MCR (mean scores)</td>
</tr>
<tr>
<td>M29</td>
<td>2500–2099</td>
<td>yes</td>
<td>MCR-ALS</td>
</tr>
</tbody>
</table>
non-negativity restraints while determining the elements of the other matrix with a linear regression, where the residuals between the PCA reproduced data and the products between $S$ and $L^T$ are minimized. At this step additional restraints like uni-modality for spectra or the time profile of concentration can be defined, which constitute the power of these approaches. However, as we are interested in ratios of concentrations and not in their absolute values, most of the restraints provided by MCR-ALS cannot be used and we simply focus on the power of spectral separation provided by MCR-ALS, using only reference values and non-negativity.

2.2.4. RABBIT-MCR

A first step in many MCR variants is a decomposition based in a PCA as outlined in equation (7) after all interferences in the spectrum not belonging to the absorption are removed via pre-processing steps based on baseline correction and detrending. The obtained scores and loading matrices $S$ and $L$ still contain negative values. Their occurrences can be minimized with an appropriate transformation, based on a transformation matrix $T$. As $T^{-1}ST$ gives the identity matrix, it can be inserted between the two factors on right of equation (7) to obtain

$$D = S T^{-1} T L^T = S T^{-1} \left( L T^T \right)^T$$ (8)

Thereby, the two factors are the transformed scores and loadings of the form:

$$S_T = ST^{-1}; \quad L_T = LT^T$$ (9)

RFA [8], Gentle [9] and the new RABBIT-MCR approach use such a transformation strategy. The transformation matrix $T$ is determined from the requirement that the transformed loadings closer resemble the expected resolved spectra of the data they are derived from or result in chemically meaningful spectra. RABBIT-MCR builds the transformation from a rotation of the coordinate system of the PCA space followed by a bending of angles between the axes of the new coordinates:

$$T = RB$$ (10)

where $R$ is a rotation matrix and $B$ angle bending matrix. The determination of the transformation matrix is the most crucial step of the success of the RABBIT-MCR. In order to limit the number of possible solutions and find a spectral meaningful one, two constraints analogous to those possible in MCR are used.

The first is non-negativity in scores and loadings since it is assumed that the spectra and concentrations they represent are non-negative. The second is compactness, which confines the range of one spectrum on a minimal number of lines, or the other way round, the number of spectral lines for one component with an absorbance of zero (i.e. no absorption) should maximal. The “compacting of spectral peaks” constraint suppresses small remnant peaks of one isotope peak in areas where the other isotope absorbs. RABBIT-MCR is implemented as a Bayesian Model which allows to assess how the quality of decomposition translates into the precision of the calibration derived in the second step. The full mathematical derivation of RABBIT-MCR can be found in the appendix Appendix A. It is implemented in the statistical language Stan [27] and the corresponding code used in this paper is included in the supplementary material. The analysis provides a distribution of scores but not their optimal value. The obtained scores can be presented as the mean or the median value of the distribution of estimated scores. Hence, we use as ‘normalization’ model the mean and median variant of RABBIT-MCR (model M27 and M28 in Table 1) and compare with model M29 for MCR-ALS.

2.2.5. Response surface fit

Earlier, we showed that the determination of carbon dioxide via FTIR depended on the oxygen concentration of the sample [15]. Moreover, we observed that the $^{12}$CO$_2$ lines approach saturation for $^{12}$CO$_2$ concentrations above 3–5%. It is therefore in the realm of possibilities that the $^{13}$C determination could depend on the $^{12}$CO$_2$ and/or the oxygen concentration. We tested several different calibration fits to determine the optimum fit and find out whether the above mentioned molecules interfere with the quantification and have to be included in the calibration model. The following fit equations were tested:

$$S = c_1 + c_2 TTR + c_3 TTR^2$$ (C1)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [O_2]$$ (C2)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [O_2] TTR$$ (C3)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [O_2] TTR + c_5 [O_2]$$ (C4)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [CO_2]$$ (C5)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [CO_2] TTR$$ (C6)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [CO_2] + c_5 [CO_2]^2$$ (C7)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [CO_2] + c_5 [CO_2] TTR$$ (C8)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [CO_2] + c_5 [CO_2]^2 + c_6 [CO_2] TTR$$ (C9)

$$S = c_1 + c_2 TTR + c_3 TTR^2 + c_4 [CO_2] + c_5 [O_2]$$ (C10)

S refers to a normed score as defined in equation. The same equations can be established for MF by interchanging it for TTR. We tested whether better or more stable calibration models could be compiled using MF instead of TTR.

2.3. Model validation (data set 1)

The models and their calibration performance were validated using two external validation sets that were not included during spectral decomposition/calibration model building. The first contained 34 samples measured in the IR and measured and validated simultaneously using GC-MS (V1) [28]. The second (V2) contained 209 reference calibration samples from a mouse calibration with three reference concentrations: two different reference calibration gases with natural $^{13}$C enrichment and one calibration with artificial higher enrichment at around 11% TTR, all with a known $^{12}$CO$_2$, O$_2$ and $^{13}$C TTR concentration. These two validation sets were used to assess the calibration performance of the models and for choosing the optimum model for FTIR TTR quantification in mouse breath.

2.4. Model validation (data set 2)

The second data set was validated using an external validation set of 163 mouse breath samples that were measured simultaneously with GC-MS. For 54 of those samples, TTR values were calculated using a different calibration procedure, as presented in an earlier work [16]. The TTR (MF) results gained by our new
procedure were compared to those GC-MS and other FTIR values.

3. Results

3.1. Normalization and spectral decomposition

In a first step we derived a measure for the tracer/tracer ratio from overlapping spectra. We tested three different options.

3.1.1. Normalization

Fig. 2 shows an example of the normalization approach. After normalization using a quantity for the $^{12}$CO$_2$ content, the resulting spectra had nearly the same intensity for the $^{12}$CO$_2$ part of the spectrum and the $^{13}$CO$_2$ part varied with the $^{13}$CO$_2$ content of the sample. This variability could be assessed with one single PCA component with an intensity or score that was proportional to the $^{13}$CO$_2$ content. Fig. 2 demonstrates that the profile of the $^{13}$CO$_2$ component failed to reflect the expected profile as shown in Fig. 1 especially the overlapping range was overestimated. It appeared to be very difficult to completely eliminate the $^{12}$CO$_2$ influence and we observed that most normalized spectra still contained artifacts in the $^{12}$CO$_2$ peak area.

3.1.2. Spectral decomposition via RABBIT-MCR

In RABBIT-MCR, first scores and loadings were calculated using a PCA decomposition. Fig. 3 shows the corresponding loadings and demonstrates the subsequent steps. The first two loadings (PC1 and PC2) contained over 99% of the spectral information, but they could not unequivocally be assigned to a corresponding spectral peak. In the next step, the PCA subspace was rotated and Fig. 3 shows the loadings after this transformation. Although the loadings were now separated, they still contained negative elements and artifacts in areas where no spectral information should be. To improve the separation further, the axis of the rotation matrix was bent by an angle till spectral separation took place. To find the optimal angle, constraints (non-negativity of scores and loadings as well as compactness of peaks) were applied. The resulting loadings in Fig. 3 showed a good separation; they resembled closely the expected isotope-pure spectra as shown in Fig. 1, which is based on the absorption lines of $^{12}$CO$_2$ and $^{13}$CO$_2$ according to the HITRAN data base [3].

The bottom row of Fig. 3 pertains to data set 2, obtained with a different optical setup. Again, the PCA alone provided an incomplete separation, but both loadings spanned nearly all of the spectral variance. The overlap was almost resolved by a rotational transformation and completely dealt with by the subsequent angle-bending step.

3.1.3. Spectral decomposition via MCR-ALS

Data set 1 and Data set 2: MCR-ALS using non-negativity constraints on spectra and concentrations on the calibration data set resulted in the resolved spectra shown in Fig. 3, last column, for data set 1 and data set 2. Unlike the loadings gained via RABBIT-MCR, the loadings belonging to the $^{12}$CO$_2$ isotopologue show a small peak in the spectral range belonging to the $^{13}$CO$_2$ peak in both cases. This small peak does not exist in HITRAN spectra shown in Fig. 1. It might be caused by the natural $^{13}$CO$_2$ labeling of CO$_2$, which always accounts for about 1% and hence all calibration samples contained at least the natural $^{13}$CO$_2$ labeling.

3.1.4. Quality of decomposition

The success of separation as well as spectral artifacts could be assessed from spectral data which were reconstructed from scores and resolved loadings, and then compared with corresponding

![Fig. 2. Normalization: In this example, spectra were normalized over the maximal, non distorted signal of the $^{12}$CO$_2$ peak around 2370 cm$^{-1}$ inset: loadings obtained after a PCA performed on signals in the non overlapping $^{13}$CO$_2$ range.](image-url)
Fig. 3. Results of spectra separation. First three columns: RABBIT-MCR: original PCA loadings, transformed loadings, updated loadings, last column: MCR-ALS. Upper row: data set 1, lower row: data set 2.

Fig. 4. Residuals after reconstruction via RABBIT-MCR and MCR-ALS. First row: original spectra, second row: reconstructed Spectra (RABBIT-MCR), third row: residuals (RABBIT-MCR), fourth row: reconstructed spectra (MCR-ALS), fifth row: residuals (MCR-ALS). First column: D1 complete data, second column: D1 test data, third column: D2 calibration data, fourth column: D2 complete data.
original spectra to obtain the residuals. Fig. 4 compares the original spectra with reconstructed spectra and shows the residuals obtained from the RABBIT-MCR and MCR-ALS procedures on data set 1 and 2. Both methods show similar residuals with no clear data artifacts for data set 1. In data set 2, the residuals of RABBIT-MCR and MCR-ALS are quite similar, however some samples have an extremely high carbon dioxide concentrations (larger 5% CO2), meaning that some spectral lines reach the saturation limit of the experimental FTIR setup. It resulted in deformed peak shapes that differ greatly from the usual peak shape which conflicts with our central assumption that the observed spectra can be explained by an overlay of two components with constant spectra. The spectrometer software has difficulties with signal values for an almost complete absorbance leading to grossly over- or underestimated values, that cannot be explained with a smooth transition towards spectra deformed by saturation. It prevented a reliable reconstruction for these samples with our approach and some other non-linear decomposition approaches [29] as well.

The residuals for the second optical setup (third column, third row) show such systematic deviations, however only the 13CO2 part of the spectra was inflicted and peak deviations accounted for about 5% of the spectral intensity. A similar ratio of deviation to observed signal can be seen for the first optical setup, but here the deviation are more randomly distributed over the entire spectrum. This data indicated that spectral deformation accounted for a part of less than 5% of observed signal intensity.

### 3.2. Calibration- and predictive quality

The normalization models provided a single score for TTR and the spectral decomposition approach yielded two different scores for the 12CO2 and 13CO2 content which can be converted with equation (2.2.5) to TTR scores. The latter were then utilized for testing the different calibration equations. Some of these equations try to correct a potential oxygen interference. For Data Set 1, the oxygen and 13CO2 values used in those fits were determined using the procedure described earlier [2,15]. Oxygen concentration values were not measured for Data Set 2 and, in consequence, the oxygen effect on 12CO2 measurements could not be corrected and of the proposed fit equations only Eq. (C1) was tested, which relies on just the TTR or MF measurements. Different combinations of normalization models as shown in Table 1 and fit equations were tested and Table S1 (Supplemental material) shows the resulting RMSEP of the V1 and V2 validation sets as a measure of their performance. Fig. 5 shows a graphical presentation of Table S1, zooming in on the best results. Models below the red line perform better than the RABBIT-MCR model. This is the case for normalization models based on equation (C8) and (C9). However, across all normalization and separation models the best results can be reached using the MF approach and fit equation (C6), which can be seen in the lower left corner of Fig. 5.

The RABBIT-MCR has the lowest RMSEP for validation set V1 and the performance of MCR-ALS is quite close. Thus, the incomplete, remnant separation observed for MCR-ALS does not impair the usability of corresponding normed scores. We assumed that normalization/separation models with a RMSEP error (average across equation (C6) and (C9)) above 0.3 were based on a less than optimal normalization strategy and dismissed them as unsuitable for quantification. The remaining ‘suitable’ models included all MCR variants and all normalization models that used the same non-overlapping 13CO2 range for the PCA on normalized spectra as depicted in Fig. 2 (M3, M5, M9, M13, M15 and M21). The best normalization model, M5, normalized over the largest possible non-overlapping range for 12CO2. This strategy reached a performance close to that obtained with the MCR approaches. For all ‘suitable’ normalization/separation models, the RMSEP for MF was about 67% of the TTR values. This might be due to the fact that the denominator for the TTR definition is only the tracer score whereas for MF, the denominator is the sum of the tracer and tracee scores. This sum might be more robust against random errors as compared to the single term used in the TTR definition. It was also shown [30] that for tracer/tracee mixtures with overlapping signals in a spectrum where each component signal is a linear function the component concentration, the linearity is preserved for a calibration based on MF, whereas non-linearities arise for TTR based calibrations. These additional non-linearities may cause the difference.

### 3.3. Optimum model (data set 1)

The optimal calibration equations were (C6), (C8) and (C9). They all contained mixed terms between TTR and the 12CO2 concentration and no term pertaining to the O2 concentration, despite our earlier observation that carbon dioxide quantification [15] is oxygen-dependent. For a cautious explanation, we assume that the oxygen effect can be approximated with a multiplicative factor, that is comparable for both isotopologues such that this factor is canceled out when the tracer to tracee quotient is formed. Earlier research has shown that the absorption slope is different for 12CO2 and 13CO2 calibration fits [16,31] and we observed a different downward drift of these calibration lines for larger enrichment values. This different bending could explain why the TTR score, which reflects a quotient between 12CO2 and 13CO2 signals, changes with higher total carbon dioxide concentration and is not constant, as presuming when using calibration equations that only include TTR as variable. For the ‘suitable’ normalization/separation models, the RMSEP1 obtained with calibration equation (C1) is in average three times larger that the average RMSEP1 obtained with the optimal calibration equation (C6). This further underscores the importance of the TTR/CO2 interaction.

Fig. 6 shows the response surface and the calibration data obtained using Fit - eqn. (C6). The scatter around the calculated surface shown in Fig. 6 suggests that the measurement error increased.
with increasing CO2 concentration values, probably because the signal to noise ratio increased for lower absorbance values.

Fig. 7 shows the real vs. predicted MF concentration of the validation set 1 and 2. A good agreement could be seen for validation set 1 (Fig. 7). A larger deviation could be seen for the 11% TTR samples of validation set 1. This was not necessarily caused by weakness of the calibration fit. The higher-enrichment samples were manufactured by hand and unlike the samples of V1, the concentration not validated by GC-MS. V2 was also used to test the temporal stability of the calibration fit. The calibration samples were measured half a year before the samples of validation set 2 and the set of samples making up the validation set 2 were measured on separate days distributed over 7 months. Fig. 8 shows the averaged prediction of the reference samples of all calibration days of the V2.

The figure shows that no clear drift in the predictions is visible. The first reference sample is based on natural occurring $^{13}$CO$_2$, which is in the range of 1.09%–1.12% and most fluctuations were statistical. The second reference samples were prepared manually by gas mixing, with a target MF of 11%. The fluctuations shown here are caused by errors in sample preparation and by random measurement errors. It can be assumed that the calibration model is stable over a long time and a daily re-calibration is not necessary since daily system fluctuations are eliminated by the calibration strategy.

3.4. Optimum model (data set 2)

Unlike D1, D2 was only used to explore to what extend the spectra decomposition algorithms can be applied to another optical setup. The preceding sections have shown that both MCR-ALS and RABBIT-MCR can successfully separate the isotope spectral peaks. For validation, 163 samples of mouse breath validated via GC-MS were used. Tables 2 and 3 show the RMSEP calculated for the different models and TTR/MF fit tested. Two RMSEP were calculated for Table 2: one with the full data set and one with the 52 samples for which values of an additional FTIR calibration published in one of our earlier works [16] existed. As mentioned before, only Eq. (C1) could be used to build a TTR/MF calibration model. Section 3.3 showed that the RMSEP will drop to approximately one third if sample CO2 concentration values are considered for the response surface. Thus it can be inferred that RMSEP around 2, as shown Table 2 could be reduced to values around 0.7 for an optimal response surface equation. Table 2 also contains the RMSEP values of a previous FTIR calibration [16] for comparison. This calibration first used a partial least square analysis to assess apparent concentrations of the two CO2 isotopologues, converted the concentrations to TTR values and then used a linear calibration between expected and determined TTR values. It thereby straightened out some of the nonlinearities that were considered at present with eqn. (C6). This PLS-based approach reached an RMSEP value that is half of the value obtained with the MCR-approaches using only TTR values but it could not reach the optimal value expected for eqn (C6). This may hint at the importance of considering the nonlinearities for an calibration design.

4. Conclusion

In this paper, we compared the results for the $^{13}$C enrichment calculation in FTIR CO2 spectra using classic approaches of area or maximum normalization to a standard MCR-ALS approach as well
as a newly developed algorithm. In our approach, spectra decomposition was the first and most important step to obtain enrichment scores. Our approach accomplishes good results in cases where non-linearities do not affect the shape of the spectral components, e.g., like a saturation that changes the spectral profile. After this, the data size of spectral values across hundreds of lines could be reduced down to a single score pertaining to enrichment. For spectra decomposition, we present a novel strategy based on the rotation and angle bending of a Principal Component Analysis (PCA) subspace and derived from MCR, called Rotation and Angle-Bending Bayesian induced Transformation - Multivariate Curve Resolution (RABBIT - MCR). It used constraints that are optimized for our application: the resolving of the two overlapping carbon dioxide isotopes in their FTIR peaks. It is expected that for other applications (other analytes or analytical methods), the constraints will have to be adapted to the physical knowledge of the method and species in question. We showed that a MCR-ALS decomposition gives the same results when the additional information of provided by samples with known concentration values was used. We replaced this information with the restraint of "compactness". This resembles to certain extend the MCR-ALS variant proposed by Windig et al. They proposed that a minimal angle-bending between the columns of the loading matrix, when used as additional restraint next to non-negativity could improve the contrast or difference in between spectral lines. It is likely to work in the present case, but one cannot achieve more than the complete separation of lines. Next to complete separation the RABBIT-MCR, as a Bayesian approach provides statistical features like confidence ranges for all estimates, which are difficult to obtain with the MCR-ALS variants. The enrichment scores obtained by spectra decomposition were used in a second step for calibration. There is a nonlinear relationship between score and enrichment values that can be characterized and calibrated by response-surface fitting. We showed that it is possible to calculate TTR/MF ratios and therefore $^{13}$C enrichment ratios from FTIR spectra using RABBIT-MCR and a response surface calibration with very good results and errors without having to calibrate/quantify the exact $^{13}$CO$_2$ concentration. For the present case, we cannot rule out minor spectral deformations, but with their present intensity, they cannot challenge the validity of our concept. The requirement that non-linearities are confined on the score to enrichment relations is the only constraint we put on potential non-linearities. These non-linearities may arise from any optical and experimental setup imperfections, molecule interactions or minor saturation effects. Our approach is not designed for pronounced saturation effects. This situation is better covered by the Nonlinear-MCR ALS approach [29]. While earlier studies [2,15,21] have shown that these results can be even further improved by mapping the daily variance of the system using calibration transfer algorithms, the calibration effort needed for one calibration set (the recording of which took over 5 days of work due to the time needed for preparing and measuring the samples) makes tracing the daily fluctuations very difficult since it is very difficult to next to impossible to complete a calibration set during one day - either an extensive reduction of the number of calibration samples in the set via Design-of-Experiment (possibly losing calibration accuracy due to not mapping the entire calibration range) or by pre-preparing calibration gases (which would be very expensive due to the price of pure $^{13}$CO$_2$ and might not be stable over time) might be a possibility but will take a large effort and planning to undertake. The analysis algorithm present here can be used to calculate $^{13}$C enrichment ratios in mouse studies in the analysis setup established in the mouse intensive care unit (MICU) of the Institute of Anesthesiologic Pathophysiology and Method Development, Ulm University Medical Center [2] without even needing more calibration effort than the one already implemented for oxygen and carbon dioxide quantification and furthering the list of analysis values gained by the system in place.

![Graph](image_url)

(a) reference sample 1 (1% CO$_2$ in 30% oxygen, natural enrichment)

(b) reference sample 3 (1% CO$_2$ in 30% oxygen, 11% TTR).

Fig. 8. Temporal stability of the calibration model: averaged predicted MF concentration collected over different measurement days (validation set 2). Data point: averaged value over 5 samples, error bar: standard deviation of the mean. Black line: Expected MF concentration.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>RMSEP (TTR) of validation set 3 (data set 2).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>RMSEP (54 samples)</td>
</tr>
<tr>
<td>MCR-ALS</td>
<td>2.0084</td>
</tr>
<tr>
<td>RABBIT-MCR (mean loading)</td>
<td>6.0221</td>
</tr>
<tr>
<td>RABBIT-MCR (median loading)</td>
<td>2.0041</td>
</tr>
<tr>
<td>FTIR (PLS)</td>
<td>1.0026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3</th>
<th>RMSEP (MF) of validation set 3 (data set 2).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>RMSEP (163 samples)</td>
</tr>
<tr>
<td>MCR-ALS</td>
<td>1.7431</td>
</tr>
<tr>
<td>RABBIT-MCR (mean loading)</td>
<td>4.0222</td>
</tr>
<tr>
<td>RABBIT-MCR (median loading)</td>
<td>1.7934</td>
</tr>
</tbody>
</table>
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

F. Seichter: Methodology, Conceptualization, Data curation, Formal analysis, Software, Validation, Writing - original draft.
Josef Albert Vogt: Methodology, Conceptualization, Validation, Writing - original draft, Writing - review & editing. Ulrich Wachter: Investigation. Peter Radermacher: Project administration, Supervision. Boris Mizakoff: Funding acquisition, Project administration, Resources, Supervision.

Acknowledgement

The authors would like to thank the Federal State of Baden-Württemberg (Landesgraduiertenförderungsgesetz, grant number: 1212-LFGF-E) for partial financial support to FS. The development of the sensing technology was performed in part under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract DE-AC52-07NA27344. This project was partially funded under LLNL sub-contract Nos. B598643 and B60301. The Boehringer Ingelheim Ulm University BioCenter (BIU, grant number: D.5008) is acknowledged for financial support. Last but not least, the partial financial support by the BMBF project APOSEMA, the MSCA RISE project TROPSENSE (funded by the Horizon 2020 Framework Programme of the European Union under grant agreement n° 645758) and the DFG within the SFB “CRC1149 TRAUMA” and the Research Training Group PULMOSENS (“GRK2203”) are acknowledged. The authors acknowledge the support of the team operating the mouse intensive care unit (MICU) at the Klinik für Anästhesiologie, Sektion Anästhesiologische Pathophysiologie und Verfahrensentwicklung (now: Institute of Anesthesiologic Pathophysiology and Method Development, Ulm University Medical Center), and especially Sandra Kress for assistance during the mouse breath studies.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aca.2019.10.038.

Appendix A. RABBIT-MCR

The notation used in this paper is consistent with traditional matrix notation: matrices are represented by boldface, upper case characters (D), vectors (as column vectors) by boldface lower case characters (a), and scalars by italic lower case characters (n). The transpose of vectors and matrices is denoted by superscript T (S^T). For every invertible transformation R_x holds R_x^{-1}R_x = I. With the constraint that R_x is orthogonal, i.e. for a vector l_i of this matrix holds l_i^*l_i = 1. Thereby the loadings are normed. Keeping this normalization for the transformed loadings results in:

$D = S \times L^T$ (A.1)

This is a data reducing step (several hundred spectral elements reduced to a few, e.g. 2) as well as a data clean up step (removal of spectral noise and random errors). L is orthogonal, meaning

$L^T \times L = I$ (A.2)

with I the identity matrix. If Eq. (A.1) is right multiplied with L an equation for the determination of S is gained.

$D \times L = S \times L^T \times L = S$ (A.3)

Estimating the scores from A.3, D can be approximated in terms of an optimal linear regression.

First transformation: Rotation of scores and loadings

Since the single value decomposition of the PCA is only guided by maximum variance, the resulting loadings in PC space are usually a superimposition of several spectra of the underlying pure components and can have negative values. Accordingly, the scores are an overlap of concentration values an can be negative. For an transformation of the scores and loadings into chemically meaningful pure spectra and concentrations according to the Multivariate Curve Resolution principle the first constraint is therefore that a loading vector can be unequivocally assigned to a spectral component. This equals a transformation which achieves that all elements are larger/equal zero. The transformed loadings matrix $L_R$ is a product of the loading matrix L and a transformation matrix R

$L_R = L \times R$ (A.4)

In principle, every transformation matrix is possible that results in similar results as Eq. (A.1).

$D = S \times L^T = S_R \times L_R^T$ (A.5)

For every invertible transformation $R_x$ holds $R_x^{-1}R_x = I$. Substituted in Eq. (A.5), this results in:

$S \times L^T = S \times I = L^* \times S_R \times L^T$ $=$ $S \times R_1^{-1}R_x \times L^T$ (A.6)

Therefore a transformation as following

$L_{R_x} = L \times R_{x^T}$; $S_{R_x} = S \times R_x^{-1}$; (A.7)

in scores and loadings that decompose the spectral matrix D according to (A.1). The original decomposition L is orthogonal, i.e. for a vector $l_i$ of this matrix holds $l_i^*l_i = 1$. Thereby the loadings are normed. Keeping this normalization for the transformed loadings results in:

$L_{R_x}^* \times L_{R_x} = I$ or $R_xR_x^T = I$ (A.8)

From this follows that the transformation matrix R has to be orthonormal, or $R^T = R^R = L$. With the constraint that R is orthonormal, Eq. (A.6) is transformed to:

$S \times L^T = S \times I \times L^T = S \times R^{-1}R \times L^T$ or

$L_R = L \times R; S_R = S \times R$. (A.9)

A rotation matrix is a special form of an orthonormal matrix can

Decomposition into PCA space

As seen in Eq. (6), a spectral matrix can be decomposed as following:

$D = S \times L^T$ (A.1)

This is a data reducing step (several hundred spectral elements reduced to a few, e.g. 2) as well as a data clean up step (removal of spectral noise and random errors). L is orthogonal, meaning

$L^T \times L = I$ (A.2)

with I the identity matrix. If Eq. (A.1) is right multiplied with L an equation for the determination of S is gained.

$D \times L = S \times L^T \times L = S$ (A.3)

Estimating the scores from A.3, D can be approximated in terms of an optimal linear regression.

First transformation: Rotation of scores and loadings

Since the single value decomposition of the PCA is only guided by maximum variance, the resulting loadings in PC space are usually a superimposition of several spectra of the underlying pure components and can have negative values. Accordingly, the scores are an overlap of concentration values an can be negative. For an transformation of the scores and loadings into chemically meaningful pure spectra and concentrations according to the Multivariate Curve Resolution principle the first constraint is therefore that a loading vector can be unequivocally assigned to a spectral component. This equals a transformation which achieves that all elements are larger/equal zero. The transformed loadings matrix $L_R$ is a product of the loading matrix L and a transformation matrix R

$L_R = L \times R$ (A.4)

In principle, every transformation matrix is possible that results in similar results as Eq. (A.1).

$D = S \times L^T = S_R \times L_R^T$ (A.5)

For every invertible transformation $R_x$ holds $R_x^{-1}R_x = I$. Substituted in Eq. (A.5), this results in:

$S \times L^T = S \times I = L^* \times S_R \times L^T$ $=$ $S \times R_1^{-1}R_x \times L^T$ (A.6)

Therefore a transformation as following

$L_{R_x} = L \times R_{x^T}$; $S_{R_x} = S \times R_x^{-1}$; (A.7)

in scores and loadings that decompose the spectral matrix D according to (A.1). The original decomposition L is orthogonal, i.e. for a vector $l_i$ of this matrix holds $l_i^*l_i = 1$. Thereby the loadings are normed. Keeping this normalization for the transformed loadings results in:

$L_{R_x}^* \times L_{R_x} = I$ or $R_xR_x^T = I$ (A.8)

From this follows that the transformation matrix R has to be orthonormal, or $R^T = R^R = L$. With the constraint that R is orthonormal, Eq. (A.6) is transformed to:

$S \times L^T = S \times I \times L^T = S \times R^{-1}R \times L^T$ or

$L_R = L \times R; S_R = S \times R$. (A.9)

A rotation matrix is a special form of an orthonormal matrix can
be easily constructed. For the case of \( n_c = 2 \), a one-parameter representation results

\[
\mathbf{R} = \begin{pmatrix}
\cos(\theta) & -\sin(\theta) \\
\sin(\theta) & \cos(\theta)
\end{pmatrix}
\]  

(A.10)

with \( \theta \) the rotation angle from the original coordinate system.

**Second transformation: Angle bending**

The first transformation step \( \mathbf{R} \) is a rotation of the orthogonal coordinate axes. In most cases, this rotation alone is not enough to achieve positive scores as well as positive loadings. Therefore, the second transformation step is an angle bending where the angle between the coordinate axes is “bent” from 90° to, e.g., 100 or 80°. Let \( z \) be the angle of deviation from 90°. The angle bending \( \mathbf{B} \) is defined as the following transformation

\[
\mathbf{B} = \begin{pmatrix}
\cos(\alpha) & \sin(\alpha) \\
\sin(\alpha) & \cos(\alpha)
\end{pmatrix}
\]  

(A.11)

For \( \alpha = 45° \), both axes fall on a line, therefore, \( \alpha \) is restricted to 45° to 45°. Since the angle bending transformation is not orthonormal, the loadings matrix loses its orthonormal attribute.

**Combined transformation**

The overall transformation \( \mathbf{T} \) is a connection of rotation and angle bending transformation, i.e.:

\[
\mathbf{T} = \mathbf{R} \times \mathbf{B}
\]  

(A.12)

\( \mathbf{T} \) has the following two properties:

- It needs only two parameters: \( \theta \) and \( \alpha \)
- The length of the transformed vector is equal to the length of the original vector

The transformed scores and loadings result from:

\[
\mathbf{S} \times \mathbf{L}_T = \mathbf{S} \times \mathbf{1} \times \mathbf{T} = \mathbf{S} \times \mathbf{TT}^{-1} \times \mathbf{L}_T
\]

or

\[
\mathbf{L}_T = \mathbf{L} \times \left( \mathbf{T}^{-1} \right)^T \mathbf{S}_T = \mathbf{S} \times \mathbf{T}.
\]  

(A.13)

If \( \mathbf{L}_T \) is known, \( \mathbf{S}_T \) can also be calculated using the following equation:

\[
\mathbf{S}_T = \mathbf{D} \left( \mathbf{L}_T^\top \right)^+.
\]  

(A.14)

Here, the pseudoinverse \( (\mathbf{L}_T^\top)^+ \) is used, since \( \mathbf{L}_T \) is not a square matrix. In order to restrict all possible solutions to only those chemically meaningful, two constraints are utilized: one the non-negativity of scores and loadings and the “compactness”. Both are explained in detail in the following sections.

**Determination of parameters \( \alpha \) and \( \theta \)**

\( \theta \) and \( \alpha \) are determined in the context of Bayesian Monte Carlo Markov Chain sampling. A probability \( P(\mathbf{T}) \) is defined that a suitable transformation is found. In our case, \( P(D|\mathbf{T}) \) is a product of the probability that the non-negativity constraint is satisfied \( P(T|\mathbf{NN}) \) and the probability that the “compactness” constraint is satisfied, \( P(\mathbf{T}|C) \).

\[
P(\mathbf{T}) = P(T|\mathbf{NN})P(\mathbf{T}|C)
\]  

(A.15)

In the following sections, the two constraints and their probabilities are defined and their maximum value determined.

**Non-negativity**

Since a transformed loading \( \mathbf{L}_T \) is supposed to correspond to a spectrum, its single elements need to have positive values since absorption spectra are always positive by definition. The scores \( \mathbf{S}_T \) correspond to concentrations and are therefore also supposed to be positive. There are lines in every spectrum that are ideally supposed to be zero but become negative due to random occurrences. In an ideal case, the number of non-negative elements is zero, but can be more than zero with declining probability \( P(T|\mathbf{NN}) \). For a measure of total-negativity it is necessary to record the magnitude of negativity of an element. The total-negativity \( p_T \) (penalty sum) is implemented as follows:

- Start: \( p_T = 0 \)
- Loop over all vectors \( \mathbf{l}_i \) of \( \mathbf{L}_T \)
  - Loop over all elements \( l_{ij} \) of vector \( \mathbf{l}_i \)
  - If \( l_{ij} < 0 \) then \( p_T = p_T + l_{ij} \)
  - Return the penalty sum \( p_T \) as result.
- Repeat procedure for elements \( s_{ij} \) of \( \mathbf{S}_T \) and return penalty sum.

If \( p_T = 0 \), \( P(T|\mathbf{NN}) \) is the largest since the non-negativity constraint is perfectly satisfied. The larger \( p_T \) gets, the lower the probability \( P(T|\mathbf{NN}) \) that the constraint is satisfied. In the Bayesian notation of the Stan software this is implemented with a half-sided Cauchy probability distribution for the penalty sum round zero with scale or width parameter \( \sigma_1 \):

\[
f(p_T; 0, \sigma_1) = \frac{1}{\pi \sigma_1} \left[ \frac{\sigma_1^2}{p_T^2 + \sigma_1^2} \right]
\]  

(A.16)

In order to weight the penalty sums for scores and loadings, the elements \( l_{ij} \) and \( s_{ij} \) are divided by a scale parameter \( \text{scaleScores} \) and \( \text{scaleLoadings} \). These two parameters as well as \( \sigma_1 \) (which determines how strongly deviations from the zero value are weighted) are the tuning parameters for a successful transformation.

**Compactness**

The second constraint deals with spectral overlap of two components. It can happen that the \( \mathbf{L}_T \) of a component are completely positive but include elements of another components (spectral residuals). This is dealt with by the compactness-constraint. It demands that a spectrum includes as few spectral lines/elements as possible. Unlike the unimodality constraint common in MCR, the compactness can deal with absorption bands that have several peak maxima, as occurring in IR spectroscopy. It should only be used in cases where the approximate peak shape of the pure components is known to avoid misleading solutions. It is assumed that it can be roughly approximated over how many lines/elements the desired spectrum stretches. Therefore it is possible to guessimate how many elements or lines are supposed to be zero. Let \( N_{\text{max}} \) be a generous ‘a priori’ estimate of this number. The compactness-constraint is now satisfied with a probability \( P(T|\mathbf{C}) \) of 1 if the real number of ‘zero-elements’ \( s_{\text{zero}} \) is smaller than \( N_{\text{max}} \). The compactness-constraint decreases rapidly if the number of ‘zero-elements’ \( s_{\text{zero}} \) is less than \( N_{\text{max}} \). Again, a sum analog to Appendix A.5.1 is defined. This time, the sum stands for the number of ‘zero-elements’ \( s_{\text{zero}} \) and is supposed to be as large as possible.
\[ s_{\text{zero}} = 0 \]

Loop over all vectors \( l_i \) of \( L \):

- Loop over all elements \( l_{ij} \) of vector \( l_i \):

\[ s_{\text{zero}} = s_{\text{zero}} + \text{weight}(l_{ij}) \]

\( \text{weight}(x, a_1) = \exp\left(-\frac{x}{a_1}\right) \)

The probability \( P(T|C) \) is defined as a sigmoid function, implemented as a logit-parameterized version of the Bernoulli distribution:

\[ P(T|C) = \frac{1}{1+\exp(-\text{weight}(s_{\text{zero}}-\sigma_\text{1}))} \]

It is supposed to be close to 1.

There are three tuning parameters for the compactness constraint: \( \sigma_1 \) (weighting factor for values close to zero), \( \sigma_0 \) (tuning the slope of the Bernoulli Logit function) and \( \sigma_C \) (number of ’zero-elements’ for which \( P(T|C) = 0.5 \)). \( \sigma_C \) may serve as a prior estimate for \( \sigma_C \).

By tuning the six parameters, \( \text{scaleForScores} \), \( \text{scaleForLoadings} \), \( \sigma_1 \) (for non-negativity constraint) and \( \sigma_0 \) and \( \sigma_C \) (for compactness constraints), the transformation parameters \( \eta \) and \( \theta \) can be determined for an optimum transformation resulting in chemically meaningful scores and loadings. The mean (or sometimes median) transformed loadings and scores of all Monte Carlo – Markov Chain samplings runs are used later on for calculation of normed scores and TTR regression.

References

5.6 Paper V: Metabolic monitoring using online analysis of 13C enriched carbon dioxide in exhaled mouse breath via iHWG-FTIR LS spectroscopy and Bayesian Sampling
Metabolic monitoring via on-line analysis of $^{13}$C-enriched carbon dioxide in exhaled mouse breath using substrate-integrated hollow waveguide infrared spectroscopy and luminescence sensing combined with Bayesian sampling

Felicia Seichter 1, Josef Vogt 1, Erhan Tütüncü 2, Leila Tamina Hagemann 1, Ulrich Wachter 1, Michael Gröger 1, Sandra Kress 1, Peter Radermacher 1 and Boris Mizaikoff 2

1 Institute of Anesthesiologic Pathophysiology and Method Development, Ulm University Medical Center, Helmholtzstrasse 8/1, 89081, Ulm, Germany
2 Institute of Analytical and Bioanalytical Chemistry, Ulm University, Albert-Einstein-Allee 11, 89081, Ulm, Germany
E-mail: boris.mizaikoff@uni-ulm.de

Keywords: exhaled breath analysis, mouse model, FTIR, on-line monitoring, substrate-integrated hollow waveguide, luminescence sensing, metabolic monitoring

Abstract

In studies that target specific functions or organs, the response is often overlaid by indirect effects of the intervention on global metabolism. The metabolic side of these interactions can be assessed based on total energy expenditure (TEE) and the contributions of the principal energy sources, carbohydrates, proteins and fat to whole body CO$_2$ production. These parameters can be identified from indirect calorimetry using respiratory oxygen intake and CO$_2$ dioxide production data that are combined with the response of the $^{13}$CO$_2$ release in the expired air and the glucose tracer enrichment in plasma following a $^{13}$C glucose stable isotope infusion. This concept is applied to a mouse protocol involving anesthesia, mechanical respiration, a disease model, like hemorrhage and therapeutic intervention. It faces challenges caused by a small sample size for both breath and plasma as well as changes in metabolic parameters caused by disease and intervention. Key parameters are derived from multiple measurements, all afflicted with errors that may accumulate leading to unrealistic values. To cope with these challenges, a sensitive on-line breath analysis system based on substrate-integrated hollow waveguide infrared spectroscopy and luminescence (iHWG-IR-LS) was used to monitor gas exchange values. A Bayesian statistical model is developed that uses established equations for indirect calorimetry to predict values for respiratory gas exchange and tracer data that are consistent with the corresponding measurements and also provides statistical error bands for these parameters. With this new methodology, it was possible to estimate important metabolic parameters (respiratory quotient (RQ), relative contribution of carbohydrate, protein and fat oxidation $f_{\text{carb}}$, $f_{\text{fat}}$ and $f_{\text{prot}}$, total energy expenditure TEE) in a resolution never available before for a minimal invasive protocol of mice under anesthesia.

1. Introduction

After trauma and/or during sepsis the stress response comprising release of glucagon, catecholamines and glucocorticoids as well as the increased cytokine formation changes in the hormonal system under conditions of trauma or sepsis stimulate glucose turnover [1], but simultaneously induce a shift from carbohydrate to lipid oxidation [2]. Such a shift to preferential lipid oxidation is unfavorable and may have undesired side effects: such a switch in fuel utilization is also associated with a lower ‘yield’ of the mitochondrial respiration [3]. FFA accumulation may cause ‘lipotoxicity’, characterized by organ ‘steatosis’ in the liver, kidney, and heart [4], ultimately leading to hyperinflammation and organ damage [5–7]. Sepsis- and trauma-related hemodynamic compromise requires catecholamine therapy, which...
may, however, further worsen the metabolic stress and impair immune function [8]. Detailed understanding of the metabolic status as well as its response to therapeutic interventions, especially with respect to the balance between lipid and carbohydrate oxidation, is desirable, in order to allow estimating whether a pathophysiological condition and/or a therapeutic intervention results in enhanced or eventually attenuated metabolic stress.

Indirect calorimetry is a non-invasive tool to assess the whole body metabolic rates [9]. Key numbers of indirect calorimetry are O$_2$ uptake (VO$_2$) and the CO$_2$ release (VCO$_2$), which allows quantifying the respiratory quotient (RQ), i.e. the ratio of VCO$_2$ over VO$_2$. The RQ value is a function of the relative contribution of carbohydrate, protein and fat oxidation ($f_{\text{carb}}$, $f_{\text{fat}}$ and $f_{\text{prot}}$). Indirect calorimetry benefits from stoichiometric considerations applied to the complete oxidation of the principal energy sources like carbohydrate, protein or fat. For each group, it can assess the amount of oxygen needed and the amount of CO$_2$ and energy that is produced with complete oxidation. As stoichiometry holds regardless of the individual steps or pathways of degradation, indirect calorimetry provides reliable estimates as long as the catabolic processes are not overlaid by synthetic processes. Simultaneous gluconeogenesis or lipid biosynthesis increase the RQ and lead to biased estimates [10]. The different relative contributions can be unraveled using an independent determination of whole body CO$_2$ production caused by protein breakdown.

The latter can be approximated from the cumulative urinary nitrogen excretion. A framework of theoretical calculations to assess the rate of carbohydrate, fat and protein oxidation from RQ values and nitrogen excretion has been developed [10–12]. Using these concepts, the link between caloric intake, fat loss and muscle wasting can be explored [13] as well as how to avoid it with specific feeding regimes [14]. However, the combined use of indirect calorimetry and nitrogen excretion data is limited: An accurate determination of the cumulative nitrogen excretion is complicated: urinary nitrogen excretion consists of loss as urea and other nitrogen containing molecules like ammonia and creatinine, and a parallel determination of the three constituents is difficult. Moreover, a long observation period is required for an accurate determination. In critically ill patients, a considerable fraction of the nitrogen may not be excreted but can be retained in expanding fluid spaces in different forms, like, e.g. urea, ammoniac or amino acids, which can be further aggravated by increasing plasma concentrations, eventually caused by impaired kidney function [11].

Lengthy protocols linked with nitrogen loss have largely been superseded with $^{13}$C glucose tracer studies to assess carbohydrate oxidation. Wolfe [12] proposed an extension of the gas exchange protocol, where a $^{13}$C labeled glucose tracer is given using constant intravenous infusion with the measurement of plasma tracer enrichment values and respiratory $^{13}$CO$_2$ enrichment. Under these conditions, the fractional contribution of the carbohydrate oxidation can be assessed, which in turn allows breaking down the different contributions of carbohydrate, fat and protein oxidation to the whole body gas exchange.

Principles of indirect calorimetry, extended with tracer protocols should also hold for mice models. The use of indirect calorimetry to assess the energy metabolism of mice is well established [15], but so far, requires the use of an ‘open flow’ metabolic chamber, which can house more animals and requires that the gas input to and output from the chamber are converted to individual gas exchange values [16, 17]. The usual metabolic chamber setup can only be applied on awake animals.

We have developed a protocol to assess the gas exchange of a mouse under anesthesia [18–21] and we also have performed a series of $^{13}$C tracer studies to assess the carbohydrate metabolism in a mouse model of critical conditions like sepsis or hemorrhagic shock [22–25]. Thus, the groundwork is laid for the implementation of a mouse protocol that combines the gas exchange and tracer protocols to obtain a separate determination of the carbohydrate, protein and fat oxidation in a mouse model that mimics a mouse intensive care unit (MICU). We arrive at minute-resolution values of TEE, RQ, $f_{\text{carb}}$, $f_{\text{fat}}$ and $f_{\text{prot}}$. These values are derived from breath parameters like oxygen intake, unlabeled CO$_2$ and $^{13}$CO$_2$ production using a non-invasive, on-line breath iHWG-IR-LS spectroscopy system. The only additional invasive element is the collection of a few plasma samples for $^{13}$C tracer enrichment measurements.

This approach integrates a variety of different data, like O$_2$ concentration in the inspired air or the plasma enrichment of a glucose tracer. Measurement errors in the different input data may propagate to unfeasible metabolic result.

To avoid these effects, the entire data set was analyzed with a Bayesian statistical model. This model pushes respiratory data in the frame of their measurement error to values that provide RQ values that are consistent with RQ values, derived from the sum of $f_{\text{carb}}$, $f_{\text{fat}}$ and $f_{\text{prot}}$ values that were weighted by their stoichiometric gas exchange factors. In addition the model enforces that the sum to the relative contributions equals one that each value stays in the range of zero to one, and that $f_{\text{carb}}$ comes close to tracer derived values. This Bayesian model allows deducing statistical error bands from Monte Carlo Markov Chain (MC-MC) sampling runs [26].

We tested the applicability of the analysis algorithms in preliminary study using published data [27] of a mural experiment that explored the impact of chronic stress on the robustness against a hemorrhagic shock. The corresponding data were expanded.
with on-line gas exchange measurements. Based on these data we try to identify prerequisites for optimal performance as well as weak points of this concept and what measures should be taken to avoid them. This should lead to a novel methodology to use respiratory data, e.g. for RQ, O₂ and CO₂ concentrations to calculate important metabolic parameters in a resolution never available before, like, e.g. the contributions of fat, carbohydrate and protein oxidation on CO₂ production or TEE.

2. Material and methods

2.1. Experimental setup

The breath analyzer system for the MICU is based on a custom-modified IR spectrometer (Bruker ALPHA FTIR Spectrometer, Bruker Optik GmbH, Ettlingen, Germany) for determining CO₂ via substrate-integrated hollow waveguides (iHWGs), and flow-cell-based luminescence sensors for oxygen measurements (FireStingO₂, Pyro Science GmbH, Aachen, Germany). The complete setup integrated into the respiratory system of the MICU has already been described elsewhere [11] and is shown in figure 1 as schematic and photograph. Radiation from the FTIR spectrometer was focused through a 7.5 cm straight channel iHWG used simultaneously as miniaturized gas cell (volume: 3 ml) and as wave guide and re-focused on the detector using two gold-coated off-axis parabolic mirrors (Janos Technology Inc, Keene/NH, USA). Between measurements and during background zero calibration, the iHWG was purged with synthetic air. The flow cell oxygen sensor was integrated into the gas outlet of the iHWG.

The measuring system was calibrated using the procedures described in [9–12], further explained in the section 2.3.

A zero calibration (background measurement) of the FTIR system was taken at the beginning and in the middle of the medical trial (around 10:30–11:00) by disconnecting the setup from the respiratory system and flushing it with the synthetic air used for ventilation, thereby compensating the atmospheric concentration of carbon dioxide contained in the synthetic air.

2.2. Animal experiments

The study had been approved by the federal authorities for animal research of the Regierungspräsidium Tübingen (approved animal experimentation number: 1190, 24.09.2014), Baden-Württemberg, Germany, and the Animal Care Committee of the University of Ulm, Baden-Württemberg, Germany, and performed in adherence with the National Institutes of Health Guidelines on the Use of Laboratory Animals and the European Union ’Directive 2010/63 EU’ on the protection of animals used for scientific purposes. Details have been published previously [27], and here we complement this data with online measurements for the CO₂ production and O₂ uptake, which were performed in the same protocol. We refer later to recruitment maneuvers [28] (RM, 5s hold at 18 cmH₂O), which were repeated every 30 min to avoid atelectasis formation due to the anesthesia and supine positioning. Each hour, the RM was followed by a compliance measurement, where the lung was inflated to a volume of 1 ml and the thoracic pressure was measured. Thereafter, a second RM was performed.

The experiment started with a thorax trauma, followed by a surgical instrumentation and a hemorrhagic shock induced by removing 30 µl/g of blood and by titrating mean arterial pressure (MAP) to 35 mmHg via further removal or re-transfusion of blood. Fluid administration was temporarily stopped. At the start of the resuscitation phase, shed blood was re-transfused, together with the administration of hydroxyethyl starch and noradrenaline (NA) titrated to maintain MAP ≥50 mmHg.

Immediately after re-transfusion, a uniformly labeled, non-radioactive ¹³C₆ glucose (obtained from Campro Scientific, Berlin Germany) was infused first with a constant priming rate of 0.47 mg/g*h ¹³C₆ glucose to fill up the body pools for 0.5 h followed by constant infusion of 0.186 mg/g*h ¹³C₆ glucose. Blood samples were taken from the tip of tail vein in the last four consecutive hours of the protocol. The ¹³C tracer enrichment in plasma glucose was determined using an Agilent 6890 gas chromatograph connected with an Agilent 5973 mass spectrometer (Agilent technologies, Santa Clara USA). To determine the isotopic enrichment, the trifluoroacetate derivate [27, 29] was used and the mass traces of m/z 319 for unlabeled glucose, and m/z 325 for the glucose tracer were measured in single ion monitored (SIM) mode with electron impact (EI) ionization.

2.3. Data analysis procedure

Data evaluation was performed using MATLAB 8.6 (R2015b, The MathWorks Inc, Naticks/MA, USA), PLS Toolbox 7.9.3 (Eigenvector Research, Inc, Manson/WA, USA), R 3.5.1 [30] and RStan 2.17.3 [31].

One IR spectrum every minute and one O₂ data point every second was obtained. The 60 oxygen data points were averaged for noise reduction, thus resulting in one final time-synchronized data point for oxygen concentration measurement and one IR spectrum every minute during on-line mouse breath monitoring.

The primary data analysis routine was based on the calibration transfer and data analysis algorithms developed in earlier works [18–21]. Monte-Carlo Markov Chain (MCMC) sampling provides realistic estimates for coefficients and prediction together with accurate error bounds by simulating known measurement errors and system fluctuations.

Oxygen concentration values were gained using hierarchical Bayesian model and Lagrange Multiplier
optimization for calibration transfer of a nonlinear calibration function based on only two daily calibration samples [18]. As proposed in [20], a humidity correction was implemented. From the Fourier transformed IR spectrum, total CO\textsubscript{2} concentration were calculated using the same Lagrange Multiplier optimization for calibration transfer and a nonlinear response surface [20]. Total CO\textsubscript{2} concentration was separated into \textsuperscript{12}CO\textsubscript{2}, \textsuperscript{13}CO\textsubscript{2} and \textsuperscript{13}C enrichment values (mole fraction respectively Tracer-to-Tracee Ratio, TTR) using another nonlinear response surface calibration with a Rotation and Angle-Bending Bayesian induced Transformation—Multivariate Curve Resolution (RABBIT—MCR) deconvolution step [21]. Based on these measurements for the O\textsubscript{2} and total CO\textsubscript{2} concentration, the RQ values could be calculated.

One of the observations in an earlier work [20] was that sometimes, RQ values reach biologically implausible low values of less than 0.6. A second,
novel approach presented in this work aimed to correct this by integrating the metabolic relationship of gas exchange as appropriate constraints into the Bayesian statistical model, resulting in biologically reasonable RQ values. Yet, the approach went even one step further: by including $^{13}$C glucose tracer concentrations and respiratory minute volume (RMV) collected via six reference measurements and interpolating them to all measuring points of the FTIR system, it was possible to estimate values for the fractional contribution of fat, protein and carbohydrate oxidation $f_{\text{fat}}$, $f_{\text{prot}}$ and $f_{\text{carb}}$ to total CO$_2$ production as well as the total energy expenditure (TEE).

3. Theory

In the following, $[O_2]_{\text{ex}}$ denotes the inspiratory O$_2$ gas concentration and $[O_2]_{\text{in}}$ and $[CO_2]_{\text{ex}}$ the expiratory O$_2$ and CO$_2$ gas concentrations. With $V_{\text{e}}$ as the respiratory minute volume, the gas concentration values were converted to the oxygen consumption, which equals $V_{\text{e}}\Delta O_2 = V_{\text{e}}([O_2]_{\text{in}} - [O_2]_{\text{ex}})$ and CO$_2$ production, i.e. $V_{\text{e}}[CO_2]$. The respiratory quotient (RQ) is defined as the ratio of oxygen consumption over CO$_2$ production, which gives

$$RQ = \frac{[CO_2]_{\text{ex}}}{\Delta O_2} \quad (1)$$

The following relations between RQ and $f_{\text{carb}}$, $f_{\text{prot}}$ and $f_{\text{fat}}$ were derived from the stoichiometric use of oxygen for the complete oxidation of the different metabolite classes and the resulting CO$_2$ production:

$$RQ = f_{\text{prot}} \cdot 0.83 + f_{\text{carb}} \cdot 1.0 + f_{\text{fat}} \cdot 0.7 \quad (2)$$

$$1 = f_{\text{prot}} + f_{\text{carb}} + f_{\text{fat}} \quad (3)$$

Note that equation (2) implies that the minimal value for RQ is 0.7 for $f_{\text{fat}} = 1$ and the maximum RQ is 1 for $f_{\text{carb}} = 1$. Equations (2) and 3 form two equations for three unknowns. With estimates for $f_{\text{prot}}$, i.e. derived from urinary nitrogen excretion, the relative contribution for carbohydrate and fat oxidation can be estimated from RQ.

Based on these results, one can also calculate the total energy expenditure (TEE) [32–34]. A simplified approximation for the total energy expenditure TEE can be derived using [12]

$$TEE = 16.49 \frac{VCO_2}{RQ} + 4.63 \cdot VCO_2 \quad (4)$$

Alternatively, estimates for $f_{\text{carb}}$ can be used for a complete determination. Combining equations (1) to (2) gives

$$f_{\text{prot}} \cdot 0.83 + f_{\text{carb}} \cdot 1.0 + f_{\text{fat}} \cdot 0.7 = \frac{VCO_2}{V O_2} \quad (5)$$

$$(f_{\text{prot}} \cdot 0.83 + f_{\text{carb}} \cdot 1.0 + f_{\text{fat}} \cdot 0.7) \cdot VO_2 = VCO_2 \quad (6)$$

From equation (3) follows

$$f_{\text{fat}} = 1 - f_{\text{prot}} - f_{\text{carb}} \quad (7)$$

By substituting $f_{\text{fat}}$, as defined in equation (7), in equation (2), $f_{\text{prot}}$ can be expressed as a function of $f_{\text{carb}}$.

$$RQ = f_{\text{prot}} \cdot 0.83 + f_{\text{carb}} + (1 - f_{\text{prot}} - f_{\text{carb}}) \cdot 0.7 \quad (8)$$

which gives:

$$f_{\text{prot}} = 7.69 \cdot RQ - 2.31 \cdot f_{\text{carb}} - 5.38 \quad (9)$$

$f_{\text{carb}}$ can be estimated from the plasma tracer enrichment and the exhalation of labeled $^{13}$CO$_2$ in breath [12] in a protocol involving a constant infusion of $^{13}$C labeled glucose:

$$f_{\text{carb}} = \frac{TTR_{\text{CO}_2, \text{breath}}}{TTR_{\text{PlasmaGlucose}}} \quad (10)$$

The ratio of $^{13}$C enrichment in the expired air to $^{13}$C plasma tracer is denoted as mean tracer ratio. This relation can be used in equation (10) to express $f_{\text{prot}}$ as a function of the measured RQ value and values obtained using a glucose $^{13}$C-tracer infusion regime.

3.1. Bayesian determination

The equations above demonstrate that $f_{\text{carb}}$, $f_{\text{prot}}$ and $f_{\text{fat}}$ can be estimated given the appropriate measurements. These calculations are based on multiple measurements like inspiratory and expiratory O$_2$ concentration and tracer enrichment values. Fluctuations in the range of the errors bounds for CO$_2$ and O$_2$ determinations may lead to error bounds of RQ values that are outside the physiological range of 0.7 to 1. Moreover, in equation (10), the derived RQ values are multiplied with a large factor, which may shift estimates for $f_{\text{prot}}$ towards values outside the permitted range of 0 to 1. On the other hand, it is conceivable that a minor shift of measured values, which stays in the range of the measurement error, may lead to feasible values of the fractional contributions.

Hence, we stipulate that estimated RQ values and estimates of fractional contributions of the different metabolite groups stay in a feasible range, which, in consequence, reduces the permitted range of error bounds for the gas exchange and tracer enrichment values. The method of choice for studying the interrelation between measurement errors, restraints for some intermediate values and the confidence range for the resulting estimates is the Bayesian methodology. We use the software and programming language Stan, a sampling variant [26] where measurement errors, restraints and functional relation between different variables are used to define a statistical model, including Bayesian priors. The corresponding
algorithm draws random samples from this model using a No-U-Term sampler, and from the collected samples, distributions of the different variables, e.g. in the present case, RQ values or predicted VO$_2$ and VCO$_2$, can be estimated. In following, we outline the principles used to define the appropriate statistical model. Starting point is the requirement:

$$f_{\text{fat}}, f_{\text{carb}}, f_{\text{prot}}: \text{each between } 0 \text{ and } 1, \text{ together with}$$

$$f_{\text{prot}} + f_{\text{carb}} + f_{\text{fat}} = 1 \quad (11)$$

The statistical modeling language allows defining a distribution that satisfies equation (12). Let $[f_{\text{carb}}, f_{\text{prot}}, f_{\text{fat}}]$ denote a random sample of this distribution. This random sample contains a value for $f_{\text{carb}}$ and one can assess the likelihood that the sample $f_{\text{carb}}$ value matches the corresponding measurements. In addition, based on equation (2), an RQ value can be calculated from the random contribution sample, which, due to the sampling definition and set prior, stays in the expected range of 0.7 to 1. We denote it with $\text{RQ}_{\text{sample}}$. Equation (1) can be rearranged to $V\text{CO}_2 = \text{RQ}_{\text{sample}} V\text{O}_2$ and hence, with a given likelihood, we can draw a sample value for $V\text{O}_2(\text{sampling})$ from its measurement distribution, and the product $V\text{O}_2(\text{sampling})\text{RQ}_{\text{sample}}$ gives a prediction for $V\text{CO}_2$ measurements. One can estimate the likelihood that this prediction stays in the range of the measurement error for $V\text{CO}_2$. These steps, taken together, allow estimating the likelihood that random samples of $[f_{\text{carb}}, f_{\text{prot}}, f_{\text{fat}}]$ can explain the measurements for O$_2$, CO$_2$ concentration values and $^{13}$C enrichment values in plasma and exhaled breath.

These steps are combined into one large Bayesian model approach using the programming language Stan. The statistical sampling provides estimates for metabolic parameters based on measurements and with it the statistical error propagation from the first measurement error up to the biological end values, e.g. TEE or $f_{\text{prot}}$.

This modeling approach requires confidence ranges or error bounds for all measurements. Figure 2 shows a graphical representation of the data analysis algorithm and how measurement parameters and results are connected to each other using previously described calibration transfer and response surface approaches for O$_2$, CO$_2$ concentration estimates [18–21], which also provide an estimate for their measurement error. One problem is the availability of $^{13}$C TTR values in plasma glucose. Compared to
the minute resolution of the FTIR-iHWG-LS system, only up to 3 plasma values per experiment are available. This means that an interpolation between the available reverence values to minute resolution is necessary. In a first step, a very simple interpolation was implemented. In the time range before the first plasma measurement, a sampling of normal distribution around the mean of all plasma measurements of the experiment and its standard deviation is used. After the first measurement, a linear interpolation between two measurement values takes place. This interpolation is very rudimentary and can result in steep jumps in the plasma TTR and resulting mean tracer ratio at the GC-MS measurement time points.

Figure 3 shows the interpolated values of the plasma TTR of all mice and therefore the interpolation process. The lines contain the interpolated (mean sampled) values, while the dots belong to the actual measurement points. Before the first GCMS-measurement point, the average from all resulting measurement points is assumed as a wide approximation as complement. The results before this point are not included in the later considerations, since the interpolation before this point is very approximate.

4. Results

The data of 12 mice experiments with a constant $^{13}$C glucose tracer infusion are presented and metabolic parameters were determined for the last two hours of the protocol. The following general metabolic conditions were found: During the last two hours glucose production increased from $2.9 \pm 1 \mu$mol g$^{-1}$ h$^{-1}$ to $3.7 \pm 0.8 \mu$mol g$^{-1}$ h$^{-1}$ and the oxidized fraction of the infusion glucose tracer decreased from $62 \pm 14\%$ to $55 \pm 8\%$, whereas the CO$_2$ production, calculated from the expiratory CO$_2$ concentration, measured by GC/MS, and the respiratory minute volume was constant at $26 \pm 2$ and $26 \pm 3 \mu$g l$^{-1}$ min$^{-1}$. Plasma glucose concentration was in the range of 1000 $\mu$g ml$^{-1}$ and plasma lactate was at 1.7 mmol l$^{-1}$.

During the $^{13}$C glucose tracer administration, the optical measurement system was implemented, resulting in a set of measurements for inhaled and exhaled oxygen concentration, exhaled CO$_2$ concentration values together with the plasma glucose $^{13}$C enrichment and the $^{13}$CO$_2$ in the expired air. For an analysis of these combined data, the sampling algorithm outlined above was used and 10 000 sampling runs were collected to derive statistical properties of all model parameters.

Figure 4 shows an overview of all $^{13}$C breath enrichment curves. The $^{13}$C enrichment is corrected for basal or pre-tracer infusion values. Spikes around 11:00 hr are caused by the fact that the measurement system had to be disconnected from the respiratory system once in the middle of each medical trial for recording a new background correction for the FTIR spectrometer. The tracer infusion rate was the same for all mice, it started by priming the various distribution pools with double infusion rate for one hour. Nevertheless, the mice show different exhaled TTR values and therefore different metabolism of the $^{13}$C glucose tracer. In particular, the rate of increase reach a plateau is different for the mice during the priming phase. In some mice, after priming a drop in TTR exhalation takes place. Mouse 10 in particular shows
Figure 4. $^{13}$C TTR in exhaled mouse breath during MICU experiment: solid lines: mean, dotted-dashed red lines: 95% confidence interval. Tracer infusion usually starts around 11:30–12:00 and runs for half an hour. Shown are the results of all 12 mouse experiments. Spikes around 11:00 are from a short interruption in the measurement during new zero calibration. The values are basal corrected from the values before tracer injection. The tracer infusion rate was the same for all mice, but different mice show different TTR exhalation and therefore different metabolism of the $^{13}$C glucose tracer.

Figure 5. Example results of one mouse experiment (mouse 1): solid lines: mean, dotted-dashed lines: 95% confidence interval. The lines in red are from the first, not metabolically corrected approach and the lines in black are from the second, metabolically corrected approach. Color code for the panel on the lower right: blue: $f_{fat}$, green: $f_{prot}$ and red: $f_{carb}$.
an extremely slow rise and never reaches the plateau before the end of the trial. Just comparing the different courses of the TTR curves reveals how different the mice and their metabolism react to the same trauma.

Figure 5 shows the time course of VO$_2$, VCO$_2$ and RQ of a typical experiment. A drop in VO$_2$, VCO$_2$ can be seen around 13:00 and 14:00 hrs, at the same time as recruitment maneuvers (RM) were performed, which lasts for 5 to 10 min and is followed by a rebound.

The red lines in figure 5 represent the VO$_2$, VCO$_2$ and RQ results of the first data analysis approach where RQ values were conventionally calculated based on the ratios of uncorrected VO$_2$ and VCO$_2$ measurements and where no restraints were imposed to confine the resulting RQ values to the range of 0.7 to 1. The black lines represent the second approach incorporating the biological restraints and where $^{13}$C glucose tracer values were used to assess the different contributions to total CO$_2$ production, as shown in the lower right panel of figure 5. While VO$_2$ and VCO$_2$ are still close to original (red) values with overlapping confidence intervals, the RQ values show now only biologically plausible values above 0.7. The correction is strongest for the time after 13:00, since, first, after that time point, the uncorrected RQ values touch the lower limit of 0.7 and, second, only after this time point, $^{13}$C glucose tracer values are available, as indicated in figure 5, and the time before this, only rudimentary assumptions are on hand. As mentioned before, the $^{13}$C glucose tracer measurements point are available only after this time point and only then, the second approach can unfold its full impact to derive an inherently consistent set values for $f_{fat}$, $f_{prot}$, $f_{carb}$ and RQ values.

Figure 6 presents the $f_{fat}$, $f_{prot}$ and $f_{carb}$ curves of several animals. While for most mice, $f_{fat}$, $f_{prot}$ and $f_{carb}$ can be well separated, the confidence intervals for mouse 4 and mouse 10 overlap, which was also the conclusion of the consideration of figure 6. Mouse 4 had problems with the O$_2$ and CO$_2$ calibration, resulting in large error bands that propagate to all derived measurements. This shows the advantage of the implemented Bayesian error propagation. The much larger confidence intervals and error bounds compared to other mice propagate through all calculations and indicate that the results of this experiment probably should not be included in later medical interpretations due to poor quality of the

---

**Figure 6.** Box-whiskers plot of all mouse experiments at time 13:30. Black: $f_{carb}$, red: $f_{prot}$. The whiskers correspond to the 95% confidence interval.

---

**Table 1.** Total Energy Expenditure for Mouse 1, over 6 subsequent hours.

<table>
<thead>
<tr>
<th>Hour</th>
<th>TEE in kJ/h</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.885</td>
</tr>
<tr>
<td>2</td>
<td>0.7965</td>
</tr>
<tr>
<td>3</td>
<td>0.747</td>
</tr>
<tr>
<td>4</td>
<td>0.7805</td>
</tr>
<tr>
<td>5</td>
<td>0.759</td>
</tr>
<tr>
<td>6</td>
<td>0.227</td>
</tr>
</tbody>
</table>
Figure 7. Fractional contribution of fat, protein and carbohydrate oxidation $f_{\text{fat}}$, $f_{\text{prot}}$ and $f_{\text{carb}}$ to total CO$_2$ production: solid lines: mean, dotted-dashed lines: 95% confidence interval. $f_{\text{fat}}$ (blue), $f_{\text{prot}}$ (green) and $f_{\text{carb}}$ (red). Shown are the results from all 12 mice experiments after the first plasma measurement. While for most mice, $f_{\text{fat}}$, $f_{\text{prot}}$ and $f_{\text{carb}}$ can be well separated, the confidence intervals for mouse 4 and mouse 10 overlap. In all mice, the contribution of fat oxidation dominates.

5. Discussion

We present a new method to assess minute-resolution values of TEE, RQ, $f_{\text{fat}}$, $f_{\text{carb}}$ and $f_{\text{prot}}$. These values are derived from breath parameters, i.e. O$_2$ uptake, CO$_2$ production and $^{13}$C tracer enrichment measured using an on-line breath iHWG-FTIR LS spectroscopy system, together with $^{13}$C plasma glucose enrichment values. The resulting measurements are analyzed with a Bayesian model that is based on established indirect calorimetry equations that consider the biological relationships between RQ and the relative contributions of fat, carbohydrate and protein oxidation on CO$_2$ production. These equations are combined with $^{13}$C tracer analysis. The combined model enforces self-consistent estimates for the all processes involved, i.e. RQ values that are consistent with the relative contribution to carbon dioxide production and measured gas exchange values.

Our major focus was on metabolic changes, which can be derived from breath test data combined with plasma tracer data. The key finding was a slight drop over time of the average RQ values from about 0.8 to 0.7, indicating a shift towards lipid oxidation. Simultaneously, glucose turnover was high, showed a tendency to increase, and glucose oxidation tended to decrease. This reflects glycolysis to lactate, lactate recycling to glucose, while mitochondrial ATP production is fueled by lipolysis. Rather than being caused by impaired O$_2$ availability, under these conditions, glycolysis is most likely reflecting aerobic glycolysis, e.g. as a result of adrenergic stimulation of the Na$^+$-K$^+$-ATPase system in skeletal muscle[36], such as it has been demonstrated in rodent hemorrhagic shock [37] and patients with sepsis [38].

The contribution of lipolysis to CO$_2$ productions exceeds that of carbohydrate and protein oxidation, which are both in the range of 15% to 25%, and basically reflects aerobic glycolysis and the inhibition of carbohydrate oxidation by lipolysis through the Randle Cycle[39], where fatty acid-derived acetyl-CoA inhibits the entry of pyruvate into the Krebs- or TCA-cycle. The variability of breath gas TTR values indicated that these processes are expressed to variable extends. In particular, the low breath and plasma
TTR values recorded for Mouse 10 (figure 5) cannot be traced back to this mechanism, as the RQ values \(\approx 0.8\) are too high for a pronounced shift towards lipolysis, hence indicating that other processes besides aerobic glycolysis like protein oxidation might be involved.

In contrast to the average over time, we observed of \(\text{VO}_2\) and \(\text{VCO}_2\) in the time-resolved data, which coincided with alveolar recruitment maneuvers performed every 30 min, which resulted in a transient increase of RQ, the extent of which varied between individual animals. ‘Recruitment maneuvers’ comprise an inspiratory hold at increased airway pressure to restore the otherwise cyclic (tidal) or even continuous alveolar collapse, which occurs as a result of general anesthesia, in particular in the supine position [40]. However, such a recruitment maneuver also aggravates any mechanical ventilation-induced reduction in venous return and redirects pulmonary blood flow to dependent lung regions, i.e. regions with lower ventilation/perfusion ratios and, thereby, potentially increases the dead space fraction [41]. Hence, both the measured \(\text{VO}_2\) and \(\text{VCO}_2\) signal decline, and the effect on RQ will depend on the relative importance of the respective increase in dead space ventilation and the perfusion of lung regions with low ventilation/perfusion ratios. The latter mainly determines \(\text{O}_2\) transfer, while \(\text{CO}_2\) elimination essentially depends on the former [42], but due to the different solubility and diffusibility of \(\text{O}_2\) and \(\text{CO}_2\), respectively, the extent of the transitory decline of the \(\text{VCO}_2\) and \(\text{VO}_2\) signals may vary according to the individual distribution of alveolar ventilation and pulmonary perfusion. Consequently, albeit both the measured \(\text{VO}_2\) and \(\text{VCO}_2\) signals transitorily decline, the calculated RQ as the quotient of these two may actually decrease or increase. No matter the direction of the RQ change, this effect will be particularly pronounced during conditions of reduced circulating blood volume (hypovolemia). In other words, respiratory cycle- and/or recruitment maneuver-dependent variations of the \(\text{VO}_2\), \(\text{VCO}_2\) and RQ values most likely reflect variations of pulmonary perfusion and/or the distribution of alveolar ventilation/perfusion ratios [42], beyond changes of overall metabolism and/or substrate utilization. These transitory changes in pulmonary gas exchange are stretched out by the damping reservoir, which is inserted in the expiratory branch of the measurement system, leading to measurable variations over a period of 10–20 min.

We used a mouse model in an intensive care unit setup to explore the hemodynamic and metabolic responses to a hemorrhagic shock. Thus, the question arises to which extend our results can be translated to the human situation. As far as metabolic data are concerned, one has to consider that murine metabolic rates are much faster than those in humans [43]. In our study, we targeted quasi-stationary conditions in plasma and tissue labeling, which was achieved about two hours after starting the tracer infusion. For human studies, this period requires at least to five or six hours. Taking into consideration this kinetic difference, our finding on aerobic glycolysis derived from the time-averaged response may in fact be transferred to human studies, especially since the underlying mechanisms are identical.

Our approach enforces RQ values over 0.7, which mirrors the assumption that all metabolites were completely oxidized to CO\(_2\), which in turn is excreted by respiration. This assumption is no longer valid when ‘semi-oxidized’ metabolites like acetone are expired. The latter is derived from the decarboxylation of ketone bodies, which are formed from acetyl-CoA under conditions of a high lipolytic rate. The formation of acetone from a 4-carbon segment of the alkane chain of a fatty acids requires 2 moles of \(\text{O}_2\) and produces 1 mole of \(\text{CO}_2\), yielding an RQ = 0.5. We collected a complete IR spectrum in the wave number range of 400 to 4000 of the breath gas and could not detect any signals for acetone; however this may change from species to species and depend on comorbidities like diabetes.

The hemorrhagic shock used in the present protocol causes tissue hypoxia, which leads to lactic acidosis production because of anaerobic metabolism. Due to the tissue hyperperfusion, both the measured breath gas \(\text{VO}_2\) and \(\text{VCO}_2\) decreases, but to a lesser extent for \(\text{VCO}_2\). Hence, during the period of hemorrhagic shock, the RQ rises, even to values >1.0. During, resuscitation with re-perfusion of shed blood, both \(\text{VO}_2\) and \(\text{VCO}_2\) increase beyond the initial values [44], which is referred to as the repayment of an \(\text{O}_2\) debt that had accumulated during the shock phase [45]. This ‘repayment’ decreases the RQ value, and, hence, may have contributed to the low RQ values seen in the last phase of our protocol.

Thus, monitoring the dynamic alterations of the gas exchange offers an insight into the multilayered conditions of the host, starting from energy metabolism, short term disturbances in oxidative metabolism and lung functionality. We previously successfully separated recruitment maneuver-related, short-term, alterations from long-term metabolic changes [21], which would also be possible in the present setup, if the tracer infusion phase is prolonged to cover more recruitment maneuvers.

There is a nonlinear relation between measured signals and \(\text{O}_2\) and \(\text{CO}_2\) concentration values, as sketched in figure 8. To minimize the calibration effort an actual calibration line is determined via ‘calibration transfer’ [18, 19] from two or three actual calibration points and previously collected information about the general shape of the calibration line, as indicated with the grey band in figure 8. The actual calibration points should be located on this confidence band to obtain a reliable actual calibration curve. Unfitting calibration points, like the red crosses in
figure 8 will be detected by our Bayesian approach and cause larger error bounds for the resulting actual calibration. These larger error bounds are propagated down to the final estimates. This propagation makes our setup susceptible to operating errors, like incomplete flushing of the system for calibration. This may explain the larger error ranges observed for mouse 4 and 10 in figures 6 and 7. In this way increased error bounds indicate potential problems and allow dismissing the afflicted results for some animals as not credible.

The number of measurement points for $^{13}$C glucose plasma enrichment values for an off-line blood analysis is restricted by the small murine blood volume. For compensation, we use a simplistic interpolation between these plasma measurement points. A smoother interpolation based on metabolic or kinetic models, which estimates the time course of $^{13}$C glucose plasma enrichment values based on a few measurement points would have been possible, yet, was beyond the scope of this study.

The presented methodology is not limited to the current setup, and may be extended to all analytical systems that offer the few key parameters in a sufficiently high time resolution. We assume that the analytical performance will even further improve with instruments with higher precision [46]. In addition, more $^{13}$C TTR measuring points of plasma glucose could be achieved by deploying an on-line sensor rather than off-line analysis of blood samples, which would, in turn, yield a higher time resolution and better approximations for $f_{\text{fat}}$, $f_{\text{carb}}$ and $f_{\text{prot}}$.

6. Conclusions

The aim of the paper was to present a novel way of gaining access to metabolic parameters like, e.g. the contributions of fat, carbohydrate and protein oxidation. This was successfully achieved and implemented in small animal studies. By combining the on-line, or at least, constant off-line measurement of VO$_2$ intake, VCO$_2$, and $^{13}$C tracer enrichment in breath and plasma glucose with Bayesian models, a new path to metabolic parameters usually only available via indirect calorimetry has been opened.

We present a new method to assess minute-resolution values of TEE, RQ, $f_{\text{fat}}$, $f_{\text{carb}}$ and $f_{\text{prot}}$. These values are derived from breath parameters like oxygen intake, carbon dioxide production and $^{13}$C tracer enrichment measured using an on-line breath iHWG-FTIR LS spectroscopy system as well as $^{13}$C plasma glucose enrichment values. The resulting measurements are analyzed with a Bayesian model that is based on established equations from indirect calorimetry that consider the biological relationships between RQ and the relative contributions of fat, carbohydrate and protein oxidation on the carbon dioxide production. These equations are combined with $^{13}$C tracer analysis. The combined model enforces self-consistent estimates for the all processes involved, i.e. RQ values that are consistent with the relative contribution to carbon dioxide production and measured gas exchange values. The Bayesian approach provides statistical error bands from MC-MC sampling runs.

The present study design and evaluation still has some limitations: A smoother interpolation between the plasma measurement points based on metabolic models is quite possible, but would go far beyond the scope of this first paper. More $^{13}$C TTR measuring points of plasma glucose would require an on-line sensor instead of blood drawings, due to the limited blood volume. However, the aim of the paper was to present a novel way of gaining access to important metabolic parameters like, e.g. the contributions of fat, carbohydrate and protein oxidation on the carbon dioxide production or TEE using the data analysis algorithms presented and this was successfully achieved.
By combining the on-line, or at least, constant off-line measurement of VO$_2$ intake, VCO$_2$, and $^{13}$C tracer enrichment in breath and plasma glucose with Bayesian models, a new path to metabolic parameters usually only available via indirect calorimetry has been opened. This methodology is not limited to the current setup but can extended to all analytical systems that offer the few key parameters just mentioned in a sufficiently high time resolution. It is expected that the analytical performance of the methodology will even further improve with instruments with even better precision than the setup used [28]. One current limitation is the availability of $^{13}$C glucose plasma enrichment values, which currently could be only gained by off-line blood analysis. If a better time resolution of these values can be achieved, either by implementing an biological model for estimating $^{13}$C glucose plasma enrichment values or, even better, deploying an on-line sensor for $^{13}$C glucose plasma measurements, even better approximations for $f_{fat}$, $f_{carb}$ and $f_{prot}$ could be achieved.

Nevertheless, with this new methodology, it was possible to achieve biologically relevant values, e.g. for RQ, oxygen and carbon dioxide in breath, and gain access to important metabolic parameters in a resolution never available before, like, e.g. the contributions of fat, carbohydrate and protein oxidation on the carbon dioxide production or TEE during murine experiments investigating the reaction of mice to thorax trauma and hemorrhagic shock.

**Acknowledgments**

The authors acknowledge partial support of this study by the project APOSEMA funded by the German BMBF within the M-Era.net program by the DFG under Grant Agreement No. GRK2203-PULMOSENS, and by the H2020 project VOGAS (Grant No. 824986). The Machine Shop at Ulm University is thanked for support during prototype development of the iHWGs.

**ORCID iD**

Boris Mizaikoff @ https://orcid.org/0000-0002-5583-7962

**References**

[17] Burnett C M L and Grobe J L 2014 Dietary effects on resting metabolic rate in C57BL/6 mice are differentially detected by indirect (O2/CO2 respirometry) and direct calorimetry Mol. Metabolism 3 460–4
[23] Baumgart K et al 2010 Cardiac and metabolic effects of hypothermia and inhaled hydrogen sulfide in anesthetized and ventilated mice Crit. Care Med. 38 588–93
Cystathionine-γ-lyase expression is associated with mitochondrial respiration during sepsis-induced acute kidney injury in swine with atherosclerosis. 


[32] Ratzlaff R, Nowak D, Gordillo D and Hata J S 2016 Mechanically ventilated, cardiothoracic surgical patients have significantly different energy requirements comparing indirect calorimetry and the penn state equations. JPEN J. Parenter Enteral Nutr. 40 959–65


[35] Speakman J R 2013 Measuring energy metabolism in the mouse - theoretical, practical, and analytical considerations. Front Physiol. 4 34


Acknowledgements

Special thanks to

1. Prof. Dr. Boris Mizaikoff
2. Prof. Dr. Peter Radermacher
3. Josef Vogt
4. Uli Wachter, Marina Fink, Sandra Kress and the people of the Department of Anesthesia
5. Tamina Hagemann, Erhan Tütüncü, Andreas Wilk, Katharina Wörle and all the others from the IABC
6. The people of the Wissenschaftliche Werkstatt Feinwerktechnik Uni Ulm
7. The DAAD and Landesgraduiertenförderung
8. Trauma SFB and Graduiertenkolleg PULMOSENS
9. My parents, family and friends
Curriculum Vitae
Curriculum Vitae

CV removed
Publications


5. **Felicia Seichter**, Josef A. Vogt, Peter Radermacher, and Boris Mizaikoff: "Response-surface fits and calibration transfer for the correction of the oxy-


Publications


Oral presentations

1. **Felicia Seichter**, Boris Mizaikoff: "Chemometric strategies in IR breath diagnostics". 2017, 3rd TROPSENSE Workshop, Ulm, Germany.


Poster presentations


Eidesstattliche Erklärung

Name: Felicia Seichter


Ludwigsburg, den ..............................................................

Felicia Seichter